Pneumonia Detection using DL Model

Data Science Project

NCAI Healthcare Sector Training Program 2021

Prepared by

Alhanouf AlHunief, Nouf AlSaeed, Sdeem AlHadi

Supervised by

Aljowhara AlSahan, Hanan AlDosari, Layan AlAbdullatef, Reem AlRowaili, Sarah Enani, Suhailah AlAmri

Table of Contents

List of Tables

List of Figures

Abstract

- **Background**

  Pneumonia is an infection that can range in seriousness from mild to life-threatening. The exploration of x-ray images using Machine Learning (ML) and Deep Learning (DL) algorithms to detect some diseases, obtained a lot of researchers' attentions in the medical field, where it has shown promising results. The goal of this paper is to develop a deep learning model to detect pneumonia from chest x-ray images.

- **Methods**

  In this paper we propose a Convolutional Neural Network (CNN) models to detect pneumonic and non-pneumonic x-ray chest images; The dataset used in this work comprised of 5866 chest x-ray images, we used Keras framework specifically the sequential model to perform data exploration, data preprocessing, data modeling, and data evaluation.

- **Results**

  The results show that detecting pneumonia using chest x-ray based on the Convolutional Neural Network (CNN) can improve training speed and detection accuracy. Where the data evaluation showed a great result, and we achieved an accuracy of 85.9% respectively.

- **Conclusion**

  As a result of the magnificent outcomes accomplished in this work, we recommend that further studies be conducted to improve existing models using the Pre-trained Model, since they have proved their effectiveness and efficiency in achieving better and higher accuracy.

Keywords: Artificial Intelligence AI, deep learning DL, Pneumonia

1. Background

Pneumonia is a lung infection caused most usually by a virus or bacteria. In addition, it is the leading cause of death in children all over the world. Every year, approximately 1.4 million children die of pneumonia, accounting for 18% of all children under the age of five 1]. Every year, two billion people worldwide suffer from pneumonia [1]. In recent years, Artificial Intelligence AI approaches have been used to solve problems in a variety of medical diagnosis processes [2]. Making an accurate diagnosis and locating the source of Pneumonia and other problems in a timely manner remains a significant challenge for doctors in order to relieve patients' sufferings. Our goal is to improve the detection of pneumonia from lung x-ray images. So, we decided to use a sequential model form keras framework, using this method promises higher and better accuracy.

The remaining part of this work is organized as follows: Section 2 contains description of the dataset, preprocessing, and the model used. while Section 3 contains the conclusion and recommendation.

In this section, we went through related works, as we have founded a lot of similar researches via the use of search engines, we used google scholar and the Saudi Digital Library DSL and we picked some of them based on the method and models used in the research papers, we used different keywords to find them such as: pneumonia; chest X-ray; deep learning; image processing; chest x-ray; and we provided a quick overview of some key contributions from the current literature about pneumonia.

The Diagnosis of pneumonia using chest X-ray is one of the common methods used by medical experts, as this paper presented by (Wu, H. et al 3) proposed A novel hybrid system for detecting pneumonia from chest x-ray image; ACNN-RF which is an adaptive median filter Convolutional Neural Network (CNN) recognition model based on random forest (RF) in which they have followed three steps trained by a dataset that consists of 5863 images, the average recognition rate of detecting the pneumonia is up to 97% by the

proposed ACNN-RF which proved that the identification system is more effective than the previous traditional image identification system.

Another research paper about Deep Convolutional Neural Network (CNN) was proposed by (T. Rajasenbagam, et al 4) to detect pneumonia infection in the lung using chest x-ray images. the model was trained with 12,000 dataset images, where several techniques were used to the model such as, the content-based image retrieval technique to annotate the images in the dataset using metadata and further contents. The classification accuracy was 99.34% in the unseen chest x-ray images which showed that the classification model was greater than the other techniques.

The researchers (Khalid El Asnaoui et al. 5) have used automated detection and localization models of pneumonia on lung X-ray and computed tomography (CT) images. Then, they compared between the existing convolutional neural network (CNN) binary classification models which are (VGG16, VGG19, DenseNet201, Inception_ResNet_V2, Inception_V3, Resent50, MobileNet_V2, and Xception). The proposed work has been tested using X-ray and CT dataset which contain 5856 images (4273 pneumonia and 1583 normal). Interestingly, the obtained results showed that the Resent50, MobileNet_V2, and Inception_ResNet_V2 show highly satisfactory performance with rate of increase in training and validation accuracy (more than 96%). On the other hand, the results determined that CNN, Xception, VGG16, VGG19, Inception_V3, and DenseNet201 have low performance which was around 84% accuracy.

The work presented in (Tawsifur Rahman et al. 6) proposed a deep learning-based method to identify and localize the normal, bacterial, and viral pneumonia. They used some existing convolutional neural network (CNN) binary classification models which are AlexNet, ResNet18, DenseNet201, and squeezeNet. The proposed work has been trained and tested using a lung X-ray dataset which contains 5247 (1341 normal images, 2561 bacterial pneumonia images, and 1345 viral pneumonia). This work has been done in three trials; the first one, for normal and pneumonia images classification which showed an accuracy test of 98%. While the second trial was for normal, bacterial, and viral

pneumonia images the classification accuracy test was 95%. Lastly, for bacterial, and viral pneumonia images the classification achieved an accuracy test of 93.3%.

The research presented in (Dimpy Varshni et al., 2019, 7) aims to improve the medical adeptness in areas where the availability of radiotherapists is still limited. They used dataset ofChestX-ray14, consisting of 112,120 frontal chest X-rays images, which were retrieved from a publically available platform, Kaggle. The method used in the study was a combination of CNN based feature-extraction (ResNet-50 and DenseNet-169 etc) and supervised classifier algorithms such as Random Forest, Support Vector Machine and Naïve Bayes. The architecture of the proposed model was divided into three different stages - the preprocessing stage, the feature extraction stage, and the classification stage. The optimal reach is the use of the DenseNet-169 followed by SVM classifier for providing better feature representation, it is achieved 0.8002 of AUC.

In (A Sharma et al., 2020, 8) emphasizes upon the extensive use of neural networks in detection of pneumonia in chest X-rays. They used a dataset of X-ray images obtained from Kaggle. To develop the model, the authors used both CNN and a batch normalization technique from Keras, then measured their accuracy using a confusion matrix. Typically, there are three basic building blocks in CNN to preserve the spatial structure: convolution layer, max-pooling layer, and fully connected layer. They had pre-processed the X-ray images and were cropped to the ideal dimensions for computational needs. Then they found the accuracy and value loss using a confusion matrix. And last they did model fitting. The model achieved 0.9806 of accuracy.

The below table illustrates an overview of the presented related work

Author	Sample	Title	Source	Conclusion/ Results
Wu, H. et al	dataset containing 5863 images	Predict pneumonia with chest X-ray images based on convolutional deep neural learning networks.	Journal of Intelligent & Fuzzy Systems.	The average recognition rate of detecting pneumonia is up to 97% by the proposed ACNN-RF.

(T. Rajasenbagam, et al	12,000 dataset images	Detection of pneumonia infection in lungs from chest X-ray images using deep convolutional neural network and content-based image retrieval techniques.	Journal of Ambient Intelligence and Humanized Computing.	classification accuracy was 99.34% in the unseen chest x-ray images.
Khalid El Asnaoui et al.	CT dataset which contain 5856 images	Automated Methods for Detection and Classification Pneumonia based on X-Ray Images Using Deep Learning	University Mohammed V in Rabat.	Inception_ResNet_V2 shows highly satisfactory performance with rate of increase in training and validation accuracy (more than 96%). On the other hand, the results determined that CNN, Xception, VGG16, VGG19, Inception_V3, and DenseNet201 have low performance which was around 84% accuracy.
T. Rahman et al	lung X-ray dataset which contains 5247	Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray.	Applied sciences.	The classification accuracy test of normal and pneumonia images was 98%, while, for normal, bacterial, and viral pneumonia images, it was 95%, and for bacterial, and viral pneumonia, it was found to be 93.3%.
Dimpy Varshni et al	112,120 frontal chest X-rays images	Pneumonia detection using CNN based feature extraction.	IEEE explore.	classifier for providing better feature representation, it

				achieved 0.8002 of AUC.
A Sharma et al	dataset of X-ray images obtained from Kaggle	Detection of Pneumonia using ML & DL in Python.	IOP Conference Series: Materials Science and Engineering.	The model achieved 0.9806 of accuracy.

Table 1 Literature review table

2. Methodology

2.1. Data set

In this work, we were using a chest x-ray pneumonia dataset that was prepared by Kaggle [14]. The dataset comprised of 5866 chest x-ray images, all of them were labeled, and have resolutions varying from 600p to 2056p. Out of 5866 chest x-ray images, 5216 chest x-ray images for train (3875 pneumonia and 1341 normal), 16 chest x-ray images for validation (8 pneumonia and 8 normal), and 624 chest x-ray images for test (390 pneumonia and 234 normal). However, the dataset used in this work includes cases of viral and bacterial infection but does not include any case of viral and bacterial coinfection.

2.2. Pre-processing

Data preprocessing is the transformations applied to your data before feeding it to the model by converting the raw data into a clean data set. In this work, we applied three main functions; augmentation, normalization, and set generation.

- **Augmentation**

We started with data augmentation which is a technique used to increase the size of data used for training a model. We apply augmentation on our data to improve the accuracy. Data augmentation techniques that we used: Rotation, shift, shear, zoom, flip, fill.

- **Normalization**

In this step, we rescaled the pixel values (between 0 and 255) to [0,1] interval; this is called images normalized.

- **Set generation**

Lastly, we generate train_generator, validation_generator, and test_generator by '`flow_from_directory`' function, which reads images directly from directory to resize images to 350 x 350 pixels, make batch size 32, and class mode as binary.

2.3. Model

Keras is a deep learning framework that has two types of models: sequential and functional API, these models written in Python, running on top of the machine learning platform TensorFlow9]. We decided to use Keras framework to perform exploratory data analysis, data pre-processing and then, build a deep learning model and evaluate it [10].

2.3.1. defining the model:

In this project, we were using sequential model with 5 types of layers [11]:

1. **Convolution Layer** '`layers.Conv2D`': A convolution multiplies a matrix of pixels (determine with window size) with a filter matrix and sums up the multiplication values. Then the convolution slides over to the next pixel and repeats the same process until all the image pixels have been covered and the result of this layer will be feature maps.

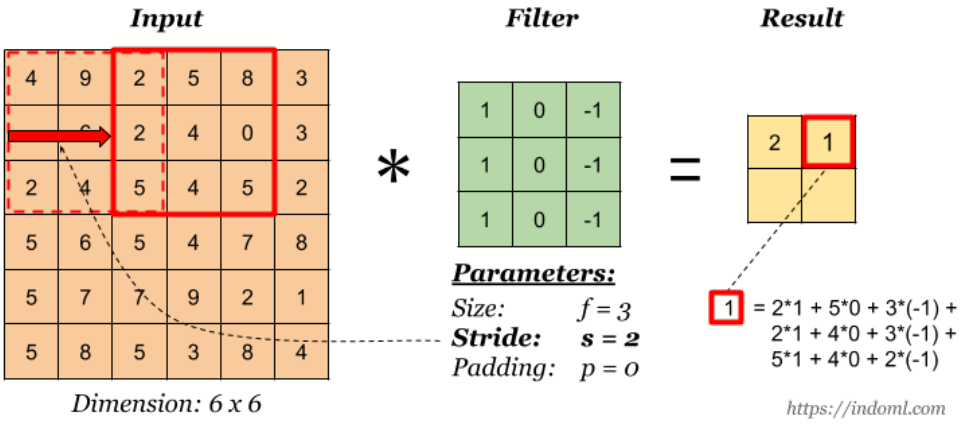

Figure 1 Overview of the convolution layer [15]

2. **Pooling Layer:** which is down-sampling, most often in the form of "max-pooling, `'layers.MaxPooling2D'`, where we select a region from the feature maps (which came from the convolution layer) by determine the window size, and then take the maximum value in that region, and that becomes the new value for the entire region.

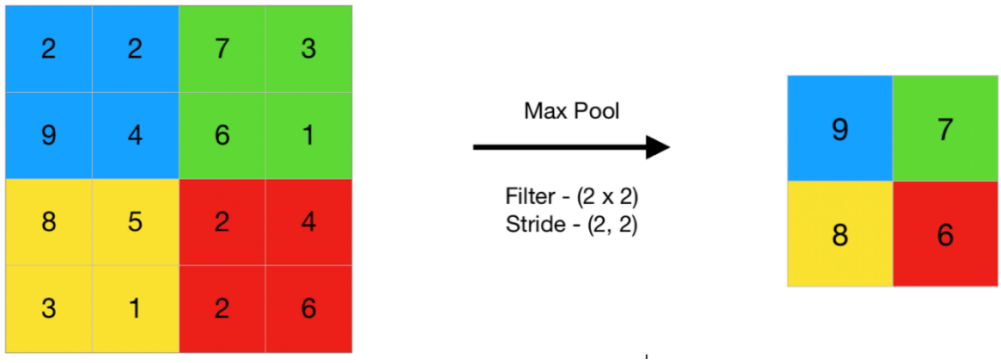

Figure 2 overview of MaxPooling Layer [16]

3. **Flatten Layer** `'layers.Flatten()'`: this layer converts our 3D feature maps(that came from Maxpooling layer) to 1D feature vectors.
4. **Dropout Layer** `'layers.Dropout()'`:prevent neural networks from Overfitting by reducing the capacity or thinning the network during training.
5. **Dense Layer** `'layers.Dense()'`: each neuron in the dense layer receives input from all neurons of its previous layer. Then applying the following operation on the input to provide output [12]:

```
output = activation(dot(input, kernel) + bias)
```

- **input:** represent the input data.
- **kernel**: represent the weight data.
- **dot:** represent numpy dot product of all input and its corresponding weights.
- **bias:** represent a biased value used in machine learning to optimize the model.
- **activation:** represent the activation function.
- **output:** will be passed into the next layer.

On the convolution and denes layers we use parameter called `'activation'` which used to apply activation function. Activation function is useful to introduce nonlinearity into the neural network so that the network can learn the complex relationship between input and output data.

2.3.2. Compiling Model

Before training the model, we should compile the model. To do that, we have to determine these parameters [13]:

1- `'Loss_funcation'`: A lower score indicates that the model is performing better. Different problems require different loss functions to keep track of progress.
2- `'optimizer'`:helps us achieve better results for the loss function.
3- `'metrics'`: used to evaluate the model.

3. Conclusion And Recommendations

Pneumonia is one of the diseases that is increasingly becoming research hotspots in recent years. Detection of pneumonia will save time, reduces effort for doctors, and helps early detection in treatment. This research proposed and developed a Pneumonia detection model using the Deep Convolutional Neural Network and Pneumonia Chest X-ray dataset. In this work, using sequential model form keras framework has been trained and tested in a dataset and reached an accuracy of 91.71% and 85.89% for testing. As a result of the magnificent outcomes accomplished in this work, we recommend that further studies be conducted to improve existing models using Pre-trained Model, since they have proved their effectiveness and efficiency in achieving better and higher accuracy.

Reference

1. A. Imran, "Training a CNN to detect Pneumonia", 2019. Available: https://medium.datadriveninvestor.com/training-a-cnn-to-detect-pneumonia-c42a44101deb. [Accessed 30 June 2021].
2. S. Hwang and H. Kim, "Self-Transfer Learning for Fully Weakly Supervised Object Localization", Lunit Inc., Seoul, Korea, 2016. Available: https://arxiv.org/pdf/1602.01625.pdf. [Accessed 30 June 2021].
3. Wu, H. et al, “Predict pneumonia with chest X-ray images based on convolutional deep neural learning networks", Journal of Intelligent & Fuzzy Systems, 2020. Available: http://sdl.edu.sa.sdl.idm.oclc.org/middleware/Default.aspx?USESDL=true&PublisherID=AllPublishers&BookURL=https://sdl.idm.oclc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bsu&AN=146380088&site=eds-live. [Accessed 16 June 2021].
4. T. Rajasenbagam, S. Jeyanthi and J. Arun Pandian, "Detection of pneumonia infection in lungs from chest X-ray images using deep convolutional neural network and content-based image retrieval techniques", Journal of Ambient Intelligence and Humanized Computing, 2021. Available: https://link.springer.com/content/pdf/10.1007/s12652-021-03075-2.pdf. [Accessed 14 June 2021].
5. K. EL ASNAOUI, Y. CHAWKI and A. IDRI, "Automated Methods for Detection and Classification Pneumonia based on X-Ray Images Using Deep Learning", 2017. Available: https://arxiv.org/ftp/arxiv/papers/2003/2003.14363.pdf. [Accessed 16 June 2021].
6. T. Rahman et al., "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray" journal of Applied sciences, 2020. Available: https://res.mdpi.com/d_attachment/applsci/applsci-10-03233/article_deploy/applsci-10-03233-v2.pdf. [Accessed 16 June 2021].

7. D. Varshni, R. Nijhawan, K. Thakral, A. Mittal and L. Agarwal, " Pneumonia detection using CNN based feature extraction " IEEE explore, 2021. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8869364. [Accessed 16 June 2021].
8. "Detection of Pneumonia using ML & DL in Python", IOP Conf. Series: Materials Science and Engineering, 2021. Available: https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012066/pdf. [Accessed 16 June 2021].
9. "keras-team/keras", GitHub, 2021. [Online]. Available: https://github.com/keras-team/keras. [Accessed: 30- Jun- 2021].
10. J. Brownlee, "How to Use the Keras Functional API for Deep Learning", Machine Learning Mastery, 2021. [Online]. Available: https://machinelearningmastery.com/keras-functional-api-deep-learning/. [Accessed: 30- Jun- 2021].
11. "An intuitive guide to Convolutional Neural Networks", freeCodeCamp.org, 2021. [Online]. Available: https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050/. [Accessed: 30- Jun- 2021].
12. "Keras - Dense Layer - Tutorialspoint", Tutorialspoint.com, 2021. [Online]. Available: https://www.tutorialspoint.com/keras/keras_dense_layer.htm. [Accessed: 30- Jun- 2021].
13. A. Alyousfi, "Deep Learning in Keras - Building a Deep Learning Model", Stack Abuse, 2021. [Online]. Available: https://stackabuse.com/deep-learning-in-keras-building-a-deep-learning-model. [Accessed: 30- Jun- 2021].
14. "Chest X-Ray Images (Pneumonia)", Kaggle.com, 2021. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. [Accessed: 05- Jul- 2021].
15. "57 CNN Basics", Kaggle.com, 2021. [Online]. Available: https://www.kaggle.com/manmohan291/57-cnn-basics. [Accessed: 05- Jul- 2021].

16. "CNN | Introduction to Pooling Layer - GeeksforGeeks", GeeksforGeeks, 2021. [Online]. Available: https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/. [Accessed: 05- Jul- 2021].