Pneumonia Detection

by Lakshmi Soundarya CHAKKA

Submission date: 25-Jun-2023 06:05PM (UTC+0100)

Submission ID: 209436334

File name: 150621_Lakshmi_Soundarya_CHAKKA_Pneumonia_Detection_2140659_1003614139.docx (149.29K)

Word count: 3278

Character count: 18113

FINAL YEAR DISSERTATION

Dissertation Title

Detection of Pneumonia from Chest X-Ray Images using Convolutional Neural Network and Transfer Learning

In Partial Fulfillment
of the Requirements for the Degree of
Master's in Computer Science

Student Name	:	Chakka Lakshmi Soundarya
Student ID	:	2301563/1
Supervisor	:	Elias Eze

Abstract

The diagnosis of the disease need lot of analysis and the doctor has to check all the insights of the x-ray image. The x-ray image will be taken by the scanner with ultra sound. The generated x-ray images are different that are compared with the normal images.Inorder to obtain the insights in a more accurate in less time lot of research is going on medical records. The deep learning methods are easily used so that the patterns from the x-rays will be easily scanned by the neural network algorithm then only the doctor will survive the patient from the disease. We proposed classification between the normal cases and the lung opacity cases. The CNN architecture with different layers will have the total parameters will be used it as the trainable parameters and in our architecture there is no non trainable parameters. The hyper parameter tuning will increase the results to more than 95%. To classify the diseases we have taken the machine learning library keras. The base model will be developed using the Inception v3 and the custom deep convolution neural networks. The Adam Optimizer is to reduce the connections between the layers to improve the accuracy in order to produce the best results. The categorical cross entropy is the loss function is used to test the trained model with the different set of images whether it is classify the disease correctly or not. So our model provides the accuracy to classify the x-ray images as normal or the lung opacity disease. The automatic classification of the medical x-rays will be done effectively then there is a chance of the low death rates because the patients will get the treatment in time.

Contents

1 Introduction

Background

It is difficult for the un-trained persons to analyze the x-ray images. As the population is getting increasing day by day the patients are also increasing then there is a need of the skilled professionals to do the diagnosis. The radiologists will provide the analysis report to the doctor and the doctor will take the report and perform analysis to start the treatment. Sometimes the confusion may be happen. To avoid such we are introducing the deep learning on the medical imaging x-rays.

Problem statement

The lung functionality is inhale and exhale of breathing can be taken. The total lung capacity is 6000ml of oxygen for humans. The swelling or inflammation of the bronchial tubes and irritation of the bronchial tubes causes diseases. The lung diseases are categorized into emphysema, cystic fibrosis/ bronchiectasis, asthma, lung cancer pneumonia and plural effusion etc. Among all the diseases the 27% of the effected people (young or old) with the disease of pneumonia. The pneumonia is nothing but accumulation of water or fluid in the lungs and it is caused by the bacteria named streptococcus pneumonia. The diagnosis of the disease can be done through the chest X-ray and blood tests like CBC (Complete Blood Count).The physiological parameters like pulse Oxymeter readings are also used to diagnose the disease. The pneumonia Chest X-ray gives the information of white spots in the lungs which infiltrates to identify the lung infection. The infected area depicts like the white spots in the black background of the chest X-ray.

Research Question and Objectives

This section introduces the motivation/research questions, and objective of the project. For example:

Research question: Can a model detect whether the person is having the pneumonia or not. The modification of the neural network in terms of freezing the hidden layers gives the accurate prediction?

Objectives:

1. Detected Pneumonia from Chest X-Ray images using Custom Deep Convolutional Neural Network and by tuning the pre-trained model with InceptionV3 with 5856 images of X-ray (1.15GB).

2. For tuning we are removing the output layers and store the first few layers to fine-tuned model for two new label classes (Pneumonia and Normal).

3. The Custom Deep Convolutional Neural Network attained with the testing accuracy 90%.

2 Literature Review

[1] T. Rahman, M. E. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, et al., "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection using Chest X-ray," Applied Sciences, vol. 10, p. 3233, 2020.

In this work author proposed that Normal vs. pneumonia, bacterial vs. viral pneumonia, and normal, bacterial, and viral pneumonia are the three categorization methods. Normal and pneumonia image classification accuracy, bacterial and viral pneumonia image classification accuracy and normal, bacterial, and viral pneumonia image classification accuracy the rates of pneumonia were 98%, 95%, and 93.3%, respectively. This is the highest accuracy in any scheme compared to the stated accuracies in the literature. As a result, the planned study may aid in the speedier diagnosis of pneumonia by radiologists and the screening of pneumonia patients at airports.

[2] T. Ishiguro, Y. Kobayashi, R. Uozumi, N. Takata, Y. Takaku, N. Kagiyama, et al., "Viral pneumonia requiring differentiation from acute and progressive diffuse interstitial lung diseases," Internal Medicine, vol. 58, pp. 3509-3519, 2019.

There are several papers in which the authors experimented with altering the parameters of deep layered CNN for pneumonia identification. Pneumonia can cause alveolar or interstitial diffuse pacifications on a lung radiograph. Patients having alveolar infiltration on chest radiographs, in particular those with lobar infiltrates exhibit bacterial infection evidence in the lab. Interstitial infiltrations on radiographs, once again, may be associated with viral pneumonia.

[3] M. Aydogdu, E. Ozyilmaz, H. Aksoy, G. Gursel, and N. Ekim, "Mortality prediction in community-acquired pneumonia requiring mechanical ventilation; values of pneumonia and intensive care unit severity scores," Tuberk Toraks, vol. 58, pp. 25-34, 2010.

In this paper author reported as Pneumonia is an infection of the lungs caused by bacteria or viruses. Fortunately, antibiotics and antiviral medicines can effectively cure this bacterial or viral infectious illness. Nonetheless, earlier detection of viral or bacterial pneumonia and subsequent application of appropriate treatment medication can considerably help to prevent the patient's health from deteriorating and finally leading to death.

[4] M. I. Neuman, E. Y. Lee, S. Bixby, S. Diperna, J. Hellinger, R. Markowitz, et al., "Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children," Journal of hospital medicine, vol. 7, pp. 294-298, 2012.

According to the author work X-ray pictures of pneumonia are blurry and frequently misclassified as other illnesses or benign abnormalities. Furthermore, the pictures of bacterial or viral pneumonia are occasionally misclassified by specialists, which results in patients receiving incorrect treatment and thereby worsening their illness [4-6]. There are significant subjective variations in radiologists' judgments in diagnosing pneumonia. There is also a scarcity of trained radiologists in low-income countries (LRC), particularly in rural regions. As a result, there is an urgent need for computer-aided diagnosis (CAD) systems that may assist radiologists in recognising various kinds of pneumonia from chest X-ray images soon after collection.

[5] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," Radiology, vol. 284, pp. 574-582, 2017.

Sundaram et al. obtained an Area under the Curve (AUC) of 0.94-0.95 by combining AlexNet and GoogLeNet with data augmentation. Using a modified two-network ensemble, the same authors achieved AUC of 0.99. Architecture In the above-mentioned literatures, the maximum accuracy recorded in categorising normal vs pneumonia patients and bacterial vs viral pneumonia utilising X-ray pictures using deep learning algorithms was 96.84% and 93.6%, respectively. As a result, there is a lot of room for improvement, either by using different deep learning algorithms, modifying the existing outperforming algorithms, or combining several outperforming algorithms as an ensemble model to produce a better classification accuracy, especially when it comes to classifying viral and bacterial pneumonia.

[6] Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning (2017). arXiv preprint https://arxiv.org/abs/1711.05225 .

One of the most amazing recent advancements is the automated identification of pneumonia in chest X-ray pictures. Many studies have been conducted in which researchers have attempted to
Deep CNN models are used to diagnose pneumonia. CheXNet is a 121-layer CNN model built by Rajpurkar et al. CheXNet was trained using 100,000 chest X-ray pictures from 14 different illnesses. The suggested model was evaluated using 420 chest X-ray pictures and the findings were compared to professional radiologists. As a consequence, the deep learning-based CNN model outperformed the average radiology performance in identifying pneumonia.

[7] Stephen, O.; Sain, M.; Maduh, U.J.; Jeong, D.-U.: An efcient deep learning approach to pneumonia classifcation in healthcare. J. Healthc Eng. 2019.

A CNN model was presented by Stephen et al. They trained, in contrast to other systems that rely primarily on transfer learning or conventional handmade procedures. The CNN model was created from scratch to extract properties from a particular chest X-ray picture, and it was used to assess whether a person was sick with pneumonia or not.

[8] Toğaçar, M.; Ergen, B.; Cömert, Z.; Özyurt, F.: A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. IRBM. 41, 212–222 (2020).

For the feature extraction phase of the pneumonia classification challenge, Togacar et al. employed three well-known CNN models. They trained each model individually using the same data and acquired 1000 features from each CNN's final fully connected layers. 1000 features were produced using the minimal redundancy maximum relevance (mRMR) feature selection approach and used as input to machine learning classifcation algorithms for the pneumonia classification challenge.

[9] Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damaševičius, R.; de Albuquerque, V.H.C.: A novel transfer learning based approach for pneumonia detection in chest X-ray images. Appl Sci. 10, 559 (2020).

Chouhan et al. utilised ImageNet to train four pre-trained CNN models for the identification of pneumonia. In this work also presented a two-stage approach for distinguishing between bacterial and viral pneumonia. The suggested technique comprises of identifying lung regions with a fully convolutional network (FCN) and classifying pneumonia categories with a deep convolutional neural network (DCNN).

[10] Perez L.; Wang, J.: The efectiveness of data augmentation in image classifcation using deep learning (2017). arXiv preprint https://arxiv.org/abs/1712.04621).

This work contains several computer vision challenges, insufficient data is frequently obtained. Obtaining and labelling data for medical concerns, in particular, is a time-consuming and difficult task. It is a time-consuming job. Fortunately, many ways have been developed to address this issue. One of these strategies, data augmentation, enhances the model's generalisation capacity, minimises overfitting, and boosts model accuracy.

3 Methodology

This section describes the proposed system's architecture, data collection, preprocessing, feature extraction and selection, model development, performance evaluation metrics, and implementation details.

- The data is collected from the kaggle (Chest X-Ray Images (Pneumonia) | Kaggle)
    - Dataset Name : Chest X-Ray Images (Pneumonia)
    - Number of Class : 2 (Normal X-Rays and Pneumonia X-Rays)
    - Number/Size of Images : 5856 (1.15 Gigabyte (GB))
        - Training : 5216 (1.07 Gigabyte (GB))
        - Validation : 320 (42.8 Megabyte (MB))
        - Testing : 320 (35.4 Megabyte (MB))
- Explain the machine learning algorithms used for model development.
    - Model Parameters
        - Machine Learning Library: Keras
        - Base Model : InceptionV3 && Custom Deep Convolutional Neural Network
        - Optimizers : Adam
        - Loss Function : Categorical_Crossentropy
    - For Custom Deep Convolutional Neural Network :
        - Training Parameters
            - Batch Size : 64
            - Number of Epochs : 30

- Detail the evaluation metrics and methods

Data collection and preprocessing

The data collection is used to download from the kaggle and the data dependencies are the token API that is needed to download the data. Once the data is downloaded to check the classes are balanced if no the synthetic data generation is needed with increasing the major classes to minor classes like up sampling and the reducing the minor classes to major classes down sampling. But our dataset is balanced hence no bias.

Data Loading:

- directory — The first parameter used is the path of the train, test & validation folder that we defined earlier.
- target_size — input image size, the target size earlier as 500 x 500.
- color_mode —If the image is either black and white or grayscale set to “grayscale” or if the image has three colour channels set to “rgb.” We’re going to work with the grayscale, because it’s the X-Ray images.
- batch_size — Number of images to be generated by batch from the generator. We defined the batch size as 16 earlier. We choose 16 because the size of the images is too large to handle the RAM.
- class_mode — It Sets to “binary” if there are only multiple classes to predict the outcome, not set to “categorical,” developing an Auto encoder, the input and the output are supposed to looks like the same image, set to “input” in this case. In this we’ve mentioned as 2 classes as per the dataset to predict.

Data Augmentation:

- rescale —a pixel with a value between 0 and 255. 0 in black, 255 in white.
- shear_range — The shape of the image is the transformation of the shear.
- zoom_range — The image is enlarged by a zoom of less than 1.0. The image is more than 1.0 zoomed out of the picture.
- horizontal_flip — images flipped horizontally
- vertical_flip — images flipped vertically at random

- roataion_range — image is rotated its degree in the range 0 to 180.
- width_shift_range — Shifts the image horizontally.
- height_shift_range — Shifts the image vertically.
- brightness_range — brightness of 0.0 no brightness, and 1.0 is maximum brightness
- fill_mode — gaps the lost value of the image to the closest value

ML/AI model development

The Domain that the technical concepts that are needed in developing the model is Computer Vision, Machine Learning. It is used to identify the features using the sub-domain called the Deep Learning or the Image Recognition. In order to obtain the features and assigned some of the weights to the features. The architecture that are needed to classify the images are belongs to normal or the pneumonia. The techniques that are used is inception v3 and the Application is the Image Recognition, Image Classification, Medical Imaging

Evaluation of the proposed system

The model performance will be evaluated by the metrics that leads to get the accuracy along the loss during the training of the model. During training the set of images will be divided into test and train dataset that are found at kaggle. The model prediction metrics are precision and the recall. The precision is used to identify the quality of the positive prediction made by the model and the recall is used to identify the relevant occurrences made by the model.F1 score is an evaluating metric that measures the model accuracy by taking the mean of precision and recall. Support is number of samples that are related to each class which predicts correctly. If the chest x-ray image contains the data points that are nearer to the class that is predicting the same class is true.

4 Expected Outcomes

1. Summary

   Loss : 0.41

   Accuracy: 89.53%

2. Derived Report

   Precision : 88.37%

   Recall : 95.48%

   F1-Score : 91.79%

5 Required Resources

Languages : Python

Dataset : Kaggle | Pneumonia

Tools/IDE : Anaconda/Google Colab

Libraries : Keras, TensorFlow, Inception, ImageNet

6 Prerequisite knowledge/skills required

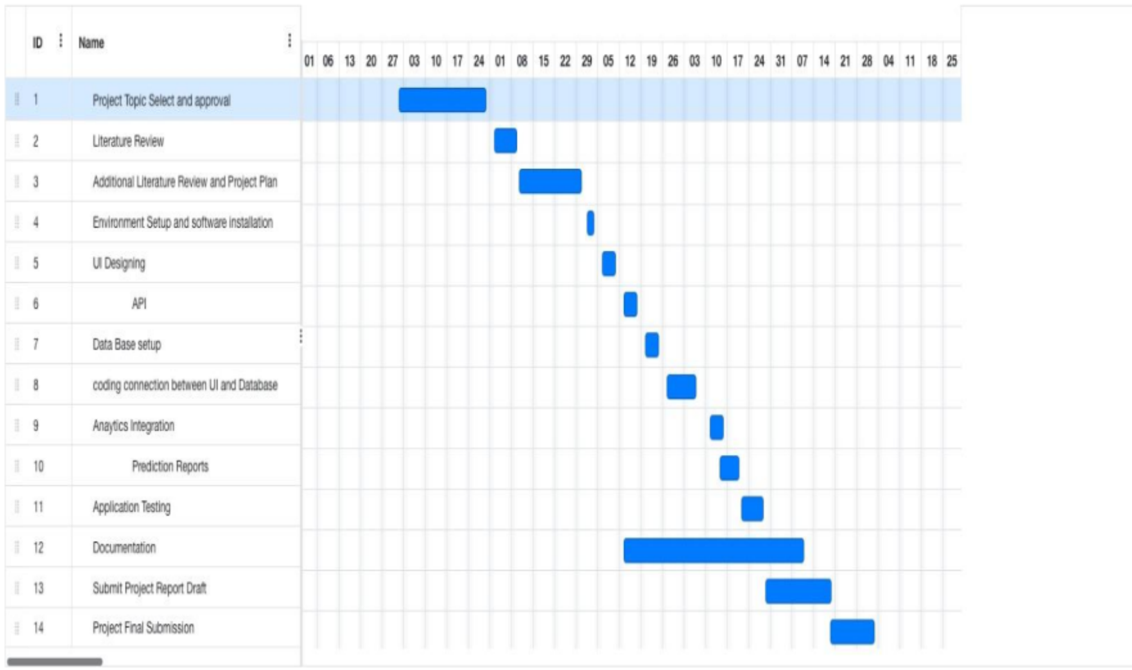
Knowledge of the research topic: The medical imaging is one of the trendy so that every skill person knows about how to learn the patterns or information from the images. This information is very useful in taking the decisions whether the particular person belongs to a particular class. The knowledge that are needed to read the images and convert the image into the binary so that the neural network layers will be used to identify the patterns. The deep learning algorithms are properly known to the developers so that the choosing of the particular algorithm to solve the problem that belong classification or the prediction.

Data analysis skills: The Analytical skills that are needed is how to identify the patterns that are inside the chest x-ray images. The patterns will be correctly classified or not.

Programming with Python: Implementing the neural network algorithm using keras framework leads to create the pipeline of Max pooling, conv2D layers. Freeze the hidden layers

7 Progress Analysis and Gantt Chart

This section contains the progress analysis and Gantt chart.


ID
Name
01 06 13 20 27 03 10 17 24 01 08 15 22 29 05 12 19 26 03 10 17 24 31 07 14 21 28 04 11 18 25
1 Project Topic Select and approval
2 Literature Review
3 Additional Literature Review and Project Plan
4 Environment Setup and software installation
5 UI Designing
6 API
7 Data Base setup
8 coding connection between UI and Database
9 Anaytics Integration
10 Prediction Reports
11 Application Testing
12 Documentation
13 Submit Project Report Draft
14 Project Final Submission

8 Conclusion

The deep learning is used to solve the real time problems in the areas like Banking, Finance and the Healthcare domains. The data may be hybrid in terms of images or textual. Our project purely focus in the image pattern recognition and classify the classifier based on the multi class prediction problem. To identify the patterns or features the each and every pixel of the image is converted into weight to the neural network so that the prediction is possible. The generated model is heavily depend on the transfer learning. The callbacks and early stopping of the training of the model does not get any new generated features.

References

[1] T. Rahman, M. E. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, et al., "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection using Chest X-ray," Applied Sciences, vol. 10, p. 3233, 2020.

[2] T. Ishiguro, Y. Kobayashi, R. Uozumi, N. Takata, Y. Takaku, N. Kagiyama, et al., "Viral pneumonia requiring differentiation from acute and progressive diffuse interstitial lung diseases," Internal Medicine, vol. 58, pp. 3509-3519, 2019.

[3] M. Aydogdu, E. Ozyilmaz, H. Aksoy, G. Gursel, and N. Ekim, "Mortality prediction in community-acquired pneumonia requiring mechanical ventilation; values of pneumonia and intensive care unit severity scores," Tuberk Toraks, vol. 58, pp. 25-34, 2010.

[4] M. I. Neuman, E. Y. Lee, S. Bixby, S. Diperna, J. Hellinger, R. Markowitz, et al., "Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children," Journal of hospital medicine, vol. 7, pp. 294-298, 2012.

[5] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," Radiology, vol. 284, pp. 574-582, 2017.

[6] Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning (2017). arXiv preprint https://arxiv.org/abs/1711.05225

[7] Stephen, O.; Sain, M.; Maduh, U.J.; Jeong, D.-U.: An efcient deep learning approach to pneumonia classifcation in healthcare. J. Healthc Eng. 2019.

[8] Toğaçar, M.; Ergen, B.; Cömert, Z.; Özyurt, F.: A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. IRBM. 41, 212–222 (2020).
[9] Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damaševičius, R.; de Albuquerque, V.H.C.: A novel transfer learning based approach for pneumonia detection in chest X-ray images. Appl Sci. 10, 559 (2020).
[10] Perez L.; Wang, J.: The efectiveness of data augmentation in image classifcation using deep learning (2017). arXiv preprint https://arxiv.org/abs/1712.04621).

Appendix A

Abbreviations

- CNN-Convolution Neural Network
- FPR-False Positive Rate
- TPR-True Positive Rate
- FNR-False Negative Rate
- FPR-False Positive Rate
- ML -Machine Learning
- CXR- Chest X-Ray

Pneumonia Detection

ORIGINALITY REPORT

50%	44%	33%	40%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	web.archive.org Internet Source	11%
2	link.springer.com Internet Source	10%
3	towardsdatascience.com Internet Source	7%
4	Submitted to University of East London Student Paper	5%
5	github.com Internet Source	4%
6	mdpi-res.com Internet Source	2%
7	Ayushi Pandey, Tushar Mehrotra, Ravin Kumar, Sandeep Saxena. "An Analysis of Pneumonia Prediction Approach Using Deep Learning", 2023 International Conference on Disruptive Technologies (ICDT), 2023 Publication	2%

8	Santiago Gómez, David Romo-Bucheli, Fabio Martínez. "A digital cardiac disease biomarker from a generative progressive cardiac cine-MRI representation", Biomedical Engineering Letters, 2021 Publication	1 %
9	www.philstat.org.ph Internet Source	1 %
10	arxiv.org Internet Source	1 %
11	Submitted to Universiti Malaysia Sabah Student Paper	1 %
12	Submitted to National Institute Of Technology, Tiruchirappalli Student Paper	1 %
13	Tawsifur Rahman, Muhammad E. H. Chowdhury, Amith Khandakar, Khandaker R. Islam et al. "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray", Applied Sciences, 2020 Publication	1 %
14	res.mdpi.com Internet Source	<1 %
15	www.researchgate.net Internet Source	<1 %

16	www.kit.nl Internet Source	<1%
17	www.mdpi.com Internet Source	<1%
18	"Computer Vision and Image Processing", Springer Science and Business Media LLC, 2023 Publication	<1%
19	pdfs.semanticscholar.org Internet Source	<1%
20	Olalekan J. Awujoola, Francisca N. Ogwueleka, Philip O. Odion, Abidemi E. Awujoola, Olayinka R. Adelegan. "Genomic data science systems of Prediction and prevention of pneumonia from chest X-ray images using a two-channel dual-stream convolutional neural network", Elsevier BV, 2023 Publication	<1%
21	hdl.handle.net Internet Source	<1%
22	inass.org Internet Source	<1%
23	mafiadoc.com Internet Source	<1%
24	www.hoffmanandcasi.com Internet Source	<1%

25	"Smart Computing and Informatics", Springer Science and Business Media LLC, 2018 Publication	<1 %
26	"Proceedings of International Conference on Artificial Intelligence and Applications", Springer Science and Business Media LLC, 2021 Publication	<1 %
27	Aras M. Ismael, Abdulkadir Şengür. "Deep learning approaches for COVID-19 detection based on chest X-ray images", Expert Systems with Applications, 2021 Publication	<1 %
28	Enes Ayan, Bergen Karabulut, Halil Murat Ünver. "Diagnosis of Pediatric Pneumonia with Ensemble of Deep Convolutional Neural Networks in Chest X-Ray Images", Arabian Journal for Science and Engineering, 2021 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		