IFERP
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 9, Issue 11, November 2022

A Filter Approach for COVID 19 Prediction using X-Ray Images

[1] Yasovar Tammareddy, [2] Revanth Reddy B, [3] Chetan Reddy M
[1][2][3] Birla Institute of Technology and Science Pilani (Hyderabad), India
Email: [1] tyasovar@gmail.com

Abstract*— With the onset of the pandemic in 2020, X-Rays have found their ways in being a means to detect what is known as the Covid-19 virus. This means being able to predict the presence of the Sars-CoV-2 virus based on a Single X-Ray, and not resorting to multiple tests etc. Being able to accurately report the presence of the SarS virus based on an X-Ray would be a breakthrough and would be cost efficient as the steps to find the presence of the virus are now simplified to just a single step. To automate the process of detecting the presence of a virus, we use Deep Learning. As a part of the project, instead of resorting to traditional CNN modelling techniques where we apply CNN directly on the images, here, we first use a Filter approach on the image followed by training using CNN models to see how the model performs.

Index Terms: - Convolutional Neural Networks, X-Rays, Filter, Covid-19, Image Classification

I. INTRODUCTION

The virus Sars-CoV-2 also famously known as the Covid-19 virus, which The World Health Organization declared as the disease-causing virus that began the pandemic since March 2020. The pandemic has affected millions of people around the world, taking lives and livelihoods away from everyone globally. Covid-19 seriously affects the lungs such as the other disease known as pneumonia.

X-ray is an electromagnetic wave of high energy density which can go through human body parts to create images of internal body parts. Chest X-rays aid in the detection of pulmonary disorders by giving an image of the thoracic cavity, which includes the chest and spine bones, as well as soft organs, blood vessels, and lung’s airways which provides sufficient information to predict whether the person is suffering from a pulmonary disease or not.

In this project a model based on CNN is used for automatic diagnosis of a pulmonary disease. The filter approach used primarily is used in generalizing convolutional neural networks from low dimensional regular grids such as images. The CNN models are finally evaluated with many various performance indicators. Accuracy, precision, ROC AUC, sensitivity, specificity, and F1 score are among the measures used.

II. LITERATURE REVIEW

Existing literature sheds light on the ongoing research and results of various models trying to solve the problem of Covid Detection. Most of them prove that the usage of CNN has in-fact created a reputable solution to the detection of the virus. Many papers have used several models and refined the process of using Deep Learning for detection.

Rubina Sarki et al. used traditional CNN methods and obtained an accuracy of 93.75%in detecting COVID-19 virus. However, this is only a tertiary classification and is meant to be used as an additional step in confirming the presence of SarS Virus and hence is not really robust.

Aijaz Ahmed Rashi et al. used six convolutional layers in their CNN model and also applied data augmentation to increase their model performance. Key findings showed that their model had the ability of invariance, but due to a lack of dataset, accurate results in image classification were not obtained, and more work was needed.

Areej A. Wahab used the famous CheXnet Algorithm, developed by Stanford professors to detect pneumonia, and extended it to detect the presence of SarS-virus and reported an accuracy of 89.7%.

They made use of the loss function to predict errors in their model and fit the CNN model accordingly.

Azher Uttal et al. Used a fully custom CNN model, which included pre-processing as well as post processing the X-ray images before classifying them as Covid or normal. They implemented classification and feature selection techniques for the purpose of data classification. Vgg16, MobilenetV2 performed well on the dataset used by the authors. The accuracy of the customized CNN model turned out to be around 97%.

III. DATASET AND PREPROCESSING

The dataset used consists of 7200 images which consists of 3600 Covid-19 and 3600 Normal chest X-ray images and then the model is tested using test images. The data has been taken from COVID 19 RADIOGRAPHY DATABASE where several thousands of chest X-ray images of Covid 19 patients are available.

We first resize the image to 50x50 images from which we remove a value of 127.5 so as to have max and min values of -127.5 and 127.5 respectively where -127.5 meaning no light and 127.5 meaning maximum light.

IFERP
connecting engineers... developing research
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 9, Issue 11, November 2022

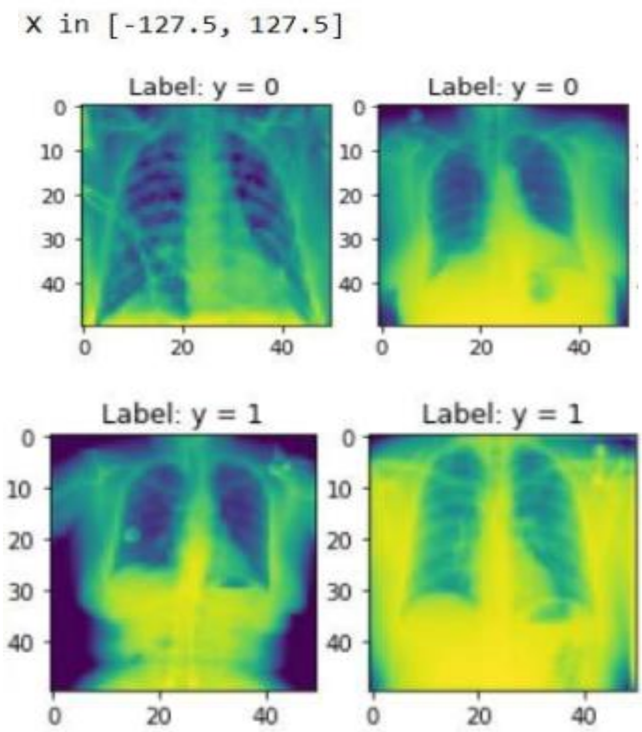
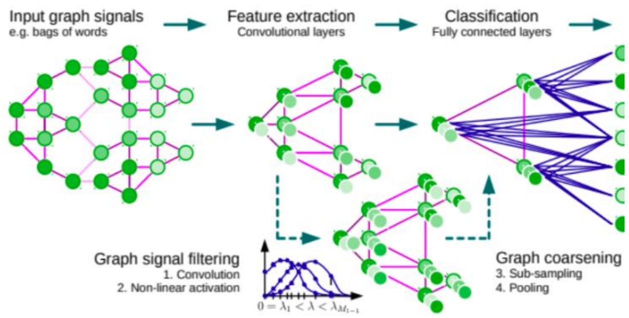


Figure 1: Chest X-Rays

IV. FILTER APPROACH

The filter approach is a subset of the feature selection technique in machine learning. It is a criteria-based model that uses statistical methods to evaluate the relevance of our model. This is done by keeping only those predictors of the model that pass some criterion.

In this project we are trying to use recursive formulation for fast filtering. The Lanczos algorithm generates the orthonormal basis of the Krylov subspace and appears appealing due to the coefficient independence. The pooling operation or coarsening needs utilizable neighborhoods on the graph, where similar vertices are clustered together. For pooling to be fast and efficient we first create a balanced binary tree and rearrange the vertices. This arrangement makes the process very efficient and helps in utilization of parallel GPU architecture for accessing memory as its local instead of fetching matched nodes.



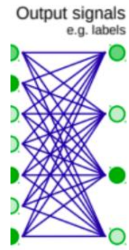


Figure 2: Use Filter Approach on the CNN

V. MODEL

We have taken a filter approach to our project. The filter taken is lanczos filter. The filter technique is mostly utilized in generalizing CNN from low-dimensional regular grids/ matrices like pictures to high-dimensional matrices like graphs.

A proper ML project requires us to give good inputs. This need becomes even more apparent when the number of features we are considering is extremely large. We need not use every single feature that is possible available to us to create an algorithm. We can help reduce noise and improve our algorithm by selecting the important features and feeding only these to the computer. This too can be done by a filter approach.

After the pre-processing, we trained our model with the help of 5 variations/ models of CNN. CNNs are a kind of DL System that uses a grid type of structure to handle data. CNNs are DL algorithms that analyze data that has some kind of spatial link. CNNs are comparable to any other neural networks, however, but they have an additional degree of complexity because of their convolutional layers.

The following models of CNN have been selected after considering their widespread use as well as appropriateness of the data:

i) AlexNet

AlexNet is a CNN which consists of eight layers - 5 of these are convolutional are 3 are fully connected. When this network was published, it was considered as one of the largest convolutional neural networks.

ii) LeNet

LeNet is a pioneering CNN, and it has a simple architecture of 2 convolutional layers and 3 fully connected layers. Generally, this is used to process low resolution images as higher clarity images need bigger and larger layers than the ones available here.

iii) VGG

VGG has 16 layers in total. 13 of which are convolutional and 3 are fully-connected layers, using the ReLU back from AlexNet.

iv) SqueezeNet

SqueezeNet was released in 2016 and was developed by a

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 9, Issue 11, November 2022

team of researchers at DeepScale, Stanford, and UCB. This CNN is 18 layers deep.

v) ResNet50

ResNet addressed the problem of accuracy getting saturated by just increasing the number of layers by skip connection (shortcut connections) while building deeper models. This has 26M parameters

VI. RESULTS

Model	Accuracy	F1-Score	AUC
AlexNet	85.37%	0.8399	0.9525
LeNet	93.12%	0.9327	0.9831
VGG	85.32%	0.8532	0.94989
SqueezeNet	88.27%	0.8885	0.9606
ResNet	86.16%	0.8517	0.9538

VII. CONCLUSION

As we can see from our results above, LeNet provides us with the highest f1 score and accuracy. It has achieved an accuracy of 93.12% despite being the simplest network used by us. As such this could point to the fact that this project does not need a bigger and deeper neural network model to improve its performance. Instead, more effort could be made in the pre-processing part to improve our model.

With this idea, we can work forward and keep pushing our model's efficiency higher. Additional data could also help us with the project to improve program performance and account for the noise in our inputs.

REFERENCES

[1] Aijaz Ahmad Reshi, Furqan Rustam, Arif Mehmood, Abdulaziz Alhossan, Ziyad Alrabiah, Ajaz Ahmad, Hessa Alsuwailem, Gyu Sang Choi, and Átila Madureira Bueno. 2021. An Efficient CNN Model for COVID-19 Disease Detection Based on X-Ray Image Classification. Complex. 2021 (2021). https://doi.org/10.1155/2021/6621607
[2] Susnjara, A., Perraudin, N., Kressner, D., & Vandergheynst, (2015). Accelerated filtering on graphs using Lanczos method.
[3] X-ray(Radiography)-Chest,2022, https://www.radiologyinfo.org/en/info/chestrad?PdfExport=1
[4] Kayalibay, B., Jensen, G., & Smagt, (2017). CNN-based Segmentation of Medical Imaging Data.
[5] Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering.
[6] Mikhail Belkin, & Partha Niyogi (2008). Towards a theoretical foundation for Laplacian-based manifold methods. Journal of Computer and System Sciences, 74(8), 1289-1308.
[7] Henaff, M., Bruna, J., & LeCun, Y. (2015). Deep Convolutional Networks on Graph-Structured Data.