RESEARCH ARTICLE OPEN ACCESS

COVID-19 Detection from Chest X-Ray Images Using CNN

Dr.P.Bhaskar Naidu [1].P.Mary Harika [2].J.Pavani [3], B.Divyadhatri [4], J.Chandana [5]

[1] Professor, Department of Computer Science and Engineering

[2], [3], [4], [5] B.Tech., Scholars, Department of Computer Science and Engineering

QIS College of Engineering and Technology, Ongole

ABSTRACT

COVID-19 global pandemic affects health care and lifestyle worldwide, and its early detection is critical to control cases' spreading and mortality. The actual leader diagnosis test is the Reverse transcription Polymerase chain reaction (RT-PCR), result times and cost of these tests are high, so other fast and accessible diagnostic tools are needed. Inspired by recent research that correlates the presence of COVID-19 to findings in Chest X-ray images, this papers' approach uses existing deep learning models (VGG19 and U-Net) to process these images and classify them as positive or negative for COVID-19. The proposed system involves a preprocessing stage with lung segmentation, removing the surroundings which does not offer relevant information for the task and may produce biased results; after this initial stage comes the classification model trained under the transfer learning scheme; and finally, results analysis and interpretation via heat maps visualization. The best models achieved a detection accuracy of COVID-19 around 97%.

I. INTRODUCTION

Coronavirus illness is a disease that comes from Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS). A novel coronavirus, COVID-19, is the infection caused by SARS-cov-2. In December 2019, the first COVID-19 cases were reported in Wuhan city, Hubei province, China. World Health Organization (WHO) declared COVID-19 a pandemic on March 11 2021, up to July 13 of 2021 there are 188,404,506 reported cases around the world, which have caused 4,059,220 deaths.

These diseases cause respiratory problems that can be treated without specialized medicine or equipment. Still, underlying medical issues such as diabetes, cancer, cardiovascular and respiratory illnesses can make this sickness worse. Reverse transcription Polymerase chain reaction (RT-PCR), gene sequencing for respiratory or blood samples are now the main methods for COVID-19 detection. Other studies show that COVID-19 has similar pathologies presented in pneumonic illness, leaving chest pathologies visible in medical images. Research shows RT-PCR correlation with Chest CT , while others study its correlation with X-ray chest images. Typical opacities or attenuation are the most common finding in these images, with ground-glass opacity in around 57% of cases. Even though expert radiologists can identify the visual patterns found in these images, considering monetary resources at low-level medical institutions and the ongoing increase of cases, this diagnostic process is quite impractical. Recent research in Artificial Intelligence (AI), especially in Deep Learning approaches, shows how these techniques applied to medical images performed well.

There are only a few large open access datasets of COVID-19 X-ray images; most of the published studies use as a foundation the COVID-19 Image Data Collection, which was constructed with images from COVID-19 reports or articles, in collaboration with a radiologist to confirm pathologies in the pictures taken. Past approaches use different strategies to deal with small datasets such as transfer learning, data augmentation or combining different datasets, finding good results in papers as using a VGG16 with 86% accuracy; with a Dark Covid Net presents 87% accuracy classifying three classes in which is included Covid; This paper presents a new approach using existing Deep Learning models. It focuses on enhancing the preprocessing stage to obtain accurate and reliable results classifying COVID-19 from Chest X-ray images.

The preprocessing step involves a network to filter the images based on the projection it is (lateral or frontal), some common operations such as normalization, standardization, and resizing to reduce data variability, which may hurt the performance of the classification models, and a segmentation model (U-Net) to extract the lung region which contains the relevant information, and discard the information of the surroundings that can produce misleading results (de Informática, 2020). Following the preprocessing stage comes the classification model (VGG16-19), using the transfer learning scheme that takes advantage of pre-trained weights from a much bigger dataset, such as imagenet, and helps the training process of the network in performance and time to convergence. It is worth noting that the dataset used for this research is at least ten times bigger than the ones used in previous works. Finally, the visualization of heatmaps for different images provides helpful information about the regions of the images that contribute to the prediction of the network, which in ideal conditions should focus on the appearance of the lungs, backing the importance of lung segmentation in the preprocessing stage. After this section, the paper follows the next order: first, the Methodology applied for these approaches, followed by the experiments and results obtained, a discussion of the products, and lastly the conclusions.

Our methodology consists of three main experiments to evaluate the performance of the models and assess the influence of the different stages of the process. Each experiment follows the workflow shown in . The difference between experiments is the dataset used. In all instances, the same images for COVID-19 positive cases were used. Meanwhile, three different datasets for negative cases were used. In that order, Experiment 1 and 2 consists of evaluating positive vs. Negative cases datasets, and Experiment 3 involves Pre-COVID era images (images from 2015-2017).

II. RELATED WORK

The images from the COVID-19 datasets have a label corresponding to the image projection: frontal (posteroanterior and anteroposterior) and lateral. Upon manual inspection, several mismatched labels were found, affecting model performance, given the difference between the information available from the two views and that not every patient had both views available. In order to automate the process of filtering the images according to the projection, a classification model was trained on a subset of BIMCV-Padchest dataset (Bustos et al., 2020), with 2481 frontal images and 815 lateral images. This model allowed us to filter the COVID-19 datasets efficiently and keep the frontal projection images that offer more information than lateral images.

Finally, to train COVID-19 classification models, the positive dataset (BIMCV-COVID19+), once separated, has 12,802 frontal images. In Experiment 1, images from BIMCV-COVID-dataset were used as negative cases, with 4610 frontal images. BIMCV-COVID — was not organized; also, some of the patients from this dataset were confirmed as COVID-19 positive in a posterior evaluation. Therefore, the models trained on this data could have a biased or unfavorable performance based on dataset size and false positives identified by radiologists. Experiment 2 used a curated version of BIMCV-COVID — for negative patients to avoid this bias, by eliminating patients' images that correlate with the positive dataset, a total of 1370 images were excluded. Finally, Experiment 3 used a Pre-COVID dataset of images collected from European patients between 2015 and 2017. There are 5469 images; this dataset was obtained from BIMCV, but it has not been published yet.

Lung segmentation

Three datasets were used to train the U-Net models for these segmentations: Montgomery dataset (Jaeger et al., 2020) with 138 images, JSTR (Shiraishi et al., 2020) with 240, and NIH (Tang et al., 2020) with 100. Despite the apparent small amount of data, the quantity and variability of the images was enough to achieve a useful segmentation model.

Image separation

For the classification task, data were divided into a train (60%), validation (20%), and test (20%) partitions, following the clinical information to avoid

images from the same subject in two different partitions, which could generate bias and overfitting in the models. Accordingly, the data distribution was as follows:

- For the classification model to filter images based on the projection, the data was composed of frontal images, 1,150, 723, and 608 for train, test, and validation partitions. In contrast, in the same partitions, the separation of lateral images was 375, 236, and 204 images.
- For the COVID-19 classification model, the positive cases dataset has 6475 images for train, 3454 for test, and 2873 for the validation set. Meanwhile, for the negative cases datasets, the BIMCV-COVID dataset is divided into 2342, 1228, and 1040 images for train, test, and validation. After the BIMCV-COVID-dataset was curated, there were 1645 images, 895, and 700 for the train, test, and validation sets. Finally, the Pre-COVID era dataset was divided into 2803 images, 1401, and 1265 for the train, test, and validation sets.
- For the COVID-19 comparison with previous works, the COVID cases dataset has 286 images for train, 96 for the test, and 96 for the validation set. Meanwhile, for the negative cases datasets, Normal images are divided into 809 for training, 270 for test and validation sets. For Pneumonia, there are 2329 images for the train, 777 for the other two groups each.
- The image quantity was considerably less for the segmentation task, so creating a test dataset was avoided, leaving the distribution of 80% (382 images) for the train set and 20% (96 images) for validation data.

III. EXISTING SYSTEM

COVID-19 is an infectious disease caused by the emergence of the new coronavirus in Wuhan, China, in December 2019. Four to five days after a person contracts the virus, symptoms typically appear. However, in some cases, the onset of symptoms can take up to two weeks. Some individuals never even exhibit any symptoms. The most common symptoms of COVID-19 are fever, cough, shortness of breath, fatigue, shaking chills, muscle pains, headaches, sore throats, runny or stuffy noses, and issues with taste or smell. If a patient has some of the symptoms presented, they are asked to test immediately. RT-PCR is also called a molecular test.

Drawbacks

- Rt PCR is Time taking process
- Rapid antigen is error prone

IV. Proposed System

The proposed algorithm was trained before and after classification while compared to traditional Convolutional Neural Network (CNN). After, the process of pre-processing and feature extraction, the CNN strategy was adopted as an identification approach to categorize the information depending on Chest X-ray recognition. These examples were then classified using the CNN classification technique. The testing was conducted on the COVID-19 X-ray dataset, and the cross-validation approach was used to determine the model's validity. The result indicated that a CNN system classification has attained an accuracy of 98.062 %.

V. CONCLUSIONS

This approach shows how existing models can be helpful for multiple tasks, especially if it is considered that the changed U-Net models do not have better performance. Also is shown how image noise can generate bias in the models. Most metrics show the images without segmentation as better for classifying COVID disease. Further analysis shows that even if metrics are better, these models are based on visible pathologies across lungs as clear evidence of COVID, so real accurate models must center on lungs parts for classifying. In this case, segmentation

is needed for reliable results by reducing this bias. Transfer learning was vital for the results presented.

REFERENCES

1. Rothan, H.A., Byrareddy, S.N.: The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J. Autoimmun. 109, 102433 (2020)
2. Xu, X., et al.: Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2. Eur. J. Nucl. Med. Mol. Imaging 47(5), 1275–1280 (2020).
3. Razai, M.S.: Coronavirus disease 2019 (covid-19): a guide for UK GPs. Bmj 368:m800 (2020)
4. Umakanthan, S., et al.: Origin, transmission, diagnosis and management of coronavirus disease 2019 (COVID-19). Postgrad. Med. J. 96(1142), 753–758 (2020)
5. Pan, F., et al.: Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. Radiology 295, 715–721 (2020)
6. Jacobi, A., et al.: Cardiothoracic imaging portable chest X-ray in coronavirus disease-19 (COVID-19): a pictorial review. Clinical Imaging 64, 35–42 (2020)
7. Jahmunah, V., et al.: Future IoT tools for COVID-19 contact tracing and prediction: a review of the state-of-the-science. Int. J. Imaging Syst. Technol. 31(2), 455–471 (2021)
8. Moitra, D., Mandal, R.K.: Classification of non-small cell lung cancer using one-dimensional convolutional neural network. Expert Syst. Appl. 159, 113564 (2020)
9. Maghdid, H.S., et al.: (2021) Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. In Multimodal Image Exploitation Learning 11734, 117340E (2021). (International Society for Optics and Photonics)
10. Yildirim, M., Cinar, A.C.: A deep learning based hybrid approach for COVID-19 disease detections. Traitement du Signal 37(3), 461–468 (2020)
11. El Asnaoui, K., Chawki, Y.: Using X-ray images and deep learning for automated detection of coronavirus disease. J BiomolStructDyn 39, 1–12 (2020)
12. Baltruschat, I.M., et al.: Comparison of deep learning approaches for multi-label chest X-ray classification. Sci. Rep. 9(1), 1–10 (2019)