IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DETECTION OF COVID-19 FROM CHEST X-RAY IMAGES USING IMAGE PROCESSING TECHNIQUES

Mrs.P.J.Mercy[1] and J.Micheal Rose[2]
1 Associate Professor, 2 III MCA Student
Department of Computer Applications, Sarah Tucker College, Thirunelveli-7.

ABSTRACT

Novel coronavirus disease (nCOVID-19) is the most challenging problem for the world. The disease is caused by severe acute respiratory syndrome coronavirus-2 (SARS-COV-2), leading to high morbidity and mortality worldwide. The study reveals that infected patients exhibit distinct radiographic visual characteristics along with fever, dry cough, fatigue, dyspnea, etc. Chest X-Ray (CXR) is one of the important, non-invasive clinical adjuncts that play an essential role in the detection of such visual responses associated with SARS-COV-2 infection. However, the limited availability of expert radiologists to interpret the CXR images and subtle appearance of disease radiographic responses remains the biggest bottlenecks in manual diagnosis. In this study, we present an automatic COVID screening (ACoS) system that uses radiomic texture descriptors extracted from CXR images to identify the normal, suspected, and nCOVID-19 infected patients. The proposed system uses two-phase classification approach (normal vs. abnormal and nCOVID-19 vs. pneumonia) using Random Forest classification algorithms. The detection of severe acute respiratory syndrome coronavirus 2 (SARS CoV-2), which is responsible for coronavirus disease 2019 (COVID-19), using chest X-ray images has life-saving importance for both patients and doctors. In addition, in countries that are unable to purchase laboratory kits for testing, this becomes even more vital. In this work, we aimed to present the use of machine learning for the high-accuracy detection of COVID-19 using chest X-ray images.

1. INTRODUCTION

In this Review, we focus on the origin and evolution of SARS-CoV and MERS-CoV. Specifically, we emphasize the ecological distribution, genetic diversity, interspecies transmission and potential for pathogenesis of SARS-related coronaviruses (SARSr-CoVs) and MERS-related corona viruses (MERSr-CoVs) found in bats, as this information can help prepare countermeasures against future spillover and pathogenic infections in humans with novel corona viruses.

A novel coronavirus was identified in patients with SARS. The virus was isolated in cell culture, and a sequence 300 nucleotides in length was obtained by a polymerase-chain-reaction (PCR)-based random-amplification procedure. Genetic characterization indicated that the virus is only distantly related to known coronaviruses (identical in 50 to 60 percent of the nucleotide sequence). On the basis of the obtained sequence, conventional and real-time PCR assays for specific and sensitive detection of the novel virus were established. Virus was detected in a variety of clinical specimens from patients with SARS but not in controls. High concentrations of viral RNA of up to 100 million molecules per milliliter were found in sputum. Viral RNA was also detected at extremely low concentrations in plasma during the acute phase and in feces during the late convalescent phase. Infected patients showed seroconversion on the Vero cells in which the virus was isolated.

SARS

Severe acute respiratory syndrome (SARS) has recently emerged as a new human disease, resulting globally in 435 deaths from 6,234 probable cases (as of 3 May 2003). Here we provide proof from experimental infection of cynomolgus macaques (Macacafascicularis) that the newly discovered SARS-associated coronavirus (SCV) is the aetiological agent of this disease. Our understanding of the aetiology of SARS will expedite the development of diagnostic tests, antiviral therapies and vaccines, and may allow a more concise case definition for this emerging disease.

Automated point-of-care molecular assays have greatly shortened the turnaround time of respiratory virus testing. One of the major bottlenecks now lies at the specimen collection step, especially in a busy clinical setting. Saliva is a convenient specimen type that can be provided easily by adult patients. This study assessed the diagnostic validity, specimen collection time and cost associated with the use of saliva.

In December 2019, a cluster of acute respiratory illness, now known as novel coronavirus–infected pneumonia (NCIP), occurred in Wuhan, Hubei Province, China.1-5. The disease has rapidly spreadfrom Wuhan to other areas. As of January 31, 2020, a total of 9692 NCIP cases in China have been confirmed. Internationally, cases have been reported in 24 countries and 5 continents.6 On January 3, 2020, the 2019 novel coronavirus (2019-nCoV) was identified in samples of bronchoalveolar lavage fluid from a patient in Wuhan and was confirmed as the cause of the NCIP.7 Full-genome sequencing and phylogenic analysis indicated that 2019-nCoV is a distinct clade from the betacoronaviruses associated with human severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS).7 The 2019-nCoV has features typical of the coronavirus family and was classified in the betacoronavirus 2b lineage. The 2019-nCoV has close similarity to bat coronaviruses, and it has been postulated that bats are the primary source. While the origin of the 2019-nCoV is still being investigated, current evidence suggests spread to humans occurred via transmission from wild animals illegally sold in the Huanan Seafood Wholesale Market.8

2. LITERATURE REVIEW

[1]In this Review, we summarize the current knowledge on the origin and evolution of these two pathogenic coronaviruses and discuss their receptor usage; we also highlight the diversity and potential of spillover of bat-borne coronaviruses, as evidenced by the recent spillover of swine acute diarrhoea syndrome coronavirus (SADS-CoV) to pigs.

[2]The severe acute respiratory syndrome (SARS) has recently been identified as a new clinical entity. SARS is thought to be caused by an unknown infectious agent.

[3]Severe acute respiratory syndrome (SARS) has recently emerged as a new human disease, resulting globally in 435 deaths from 6,234 probable cases (as of 3 May 2003). Here we provide proof from experimental infection ocynomolgus macaques (Macaca fascicularis) that the newly discovered SARS-associated coronavirus (SCV) is the aetiological agent of this disease. Our understanding of the aetiology of SARS will expedite the development of diagnostic tests, antiviral therapies and vaccines, and may allow a more concise case definition for this emerging disease.

[4]This study assessed the diagnostic validity, specimen collection time and cost associated with the use of saliva.

[5]This review attempts to give a comprehensive view of the origin of the virus, the mode of its entry and infecting human beings, and further discusses the possibility of new drugs and vaccines against the virus.

[6]In December 2019, novel coronavirus (2019-nCoV)–infected pneumonia (NCIP) occurred in Wuhan, China. The number of cases has increased rapidly but information on the clinical characteristics of affected patients is limited.

[7]Recently, we reported the discovery of three novel coronaviruses, bulbul coronavirus HKU11, thrush coronavirus HKU12, and munia coronavirus HKU13, which were identified as representatives of a novel genus, Deltacoronavirus, in the subfamily Coronavirinae. In this territory-wide molecular epidemiology study involving 3,137 mammals and 3,298 birds, we discovered seven additional novel deltacoronaviruses in pigs and birds, which we named porcine coronavirus HKU15, white-eye coronavirus HKU16, sparrow coronavirus HKU17, magpie robin coronavirus HKU18, night heron coronavirus HKU19, wigeon coronavirus HKU20, and common moorhen coronavirus HKU21. Complete genome sequencing and comparative genome analysis showed that the avian and mammalian deltacoronaviruses have similar genome characteristics and structures. They all have relatively small genomes (25.421 to 26.674 kb), the smallest among all coronaviruses. They all have a single papain-like protease domain in the nsp3 gene; an accessory gene, NS6 open reading frame (ORF), located between the M and N genes; and a variable number of accessory genes (up to four) downstream of the N gene. Moreover, they all have the same putative transcription regulatory sequence of ACACCA.

[8] Here we study a single patient who was a worker at the market and who was admitted to the Central Hospital of Wuhan on 26 December 2019 while experiencing a severe respiratory syndrome that included fever, dizziness and a cough. Metagenomic RNA sequencing4 of a sample of bronchoalveolar lavage fluid from the patient identified a new RNA virus strain from the family Coronaviridae, which is designated here ‘WH-Human 1’ coronavirus (and has also been referred to as ‘2019-nCoV’). Phylogenetic analysis of the complete viral genome (29,903 nucleotides) revealed that the virus was most closely related (89.1% nucleotide similarity) to a group of SARS-like coronaviruses (genus

Betacoronavirus, subgenus Sarbecovirus) that had previously been found in bats in China5. This outbreak highlights the ongoing ability of viral spill-over from animals to cause severe disease in humans.

[9]An epidemic of severe acute respiratory syndrome (SARS) has been associated with an outbreak of atypical pneumonia originating in Guangdong Province, People's Republic of China. We aimed to identify the causative agent in the Guangdong outbreak and describe the emergence and spread of the disease within the province.

[10] Here we report the identification and characterization of a new coronavirus (2019-nCoV), which caused an epidemic of acute respiratory syndrome in humans in Wuhan, China. The epidemic, which started on 12 December 2019, had caused 2,794 laboratory-confirmed infections including 80 deaths by 26 January 2020. Full-length genome sequences were obtained from five patients at an early stage of the outbreak. The sequences are almost identical and share 79.6% sequence identity to SARS-CoV. Furthermore, we show that 2019-nCoV is 96% identical at the whole-genome level to a bat coronavirus. Pairwise protein sequence analysis of seven conserved non-structural proteins domains show that this virus belongs to the species of SARSr-CoV. In addition, 2019-nCoV virus isolated from the bronchoalveolar lavage fluid of a critically ill patient could be neutralized by sera from several patients. Notably, we confirmed that 2019-nCoV uses the same cell entry receptor—angiotensin converting enzyme II (ACE2)—as SARS-CoV.

3. METHODOLOGY

Image Enhancement

The image enhancement generally known as image pre-processing is applied on the images to improve the perception of images data set for automatic pattern detection via image processing techniques. In a two route procedure for taking advantage of both grey distribution of blue channel (dark) and grey scale of the images.

- **Gray Blue**

I= rgb2gray (RGB**)** converts the true color image RGB to the grayscale image I. The rgb2gray function converts RGB images to grayscale by eliminating the hue and saturation information while retaining the luminance.

- **Circular Average Filter**

It takes the average of all elements in a window around each location. Let's say you were looking at an element (pixel) at (100,200) and had a disc of radius 10. So it would take all pixels in a circle from 90 to 110 and 190 to 210, multiply them by the values in the special array, which are 1's in the disc, and 0's in the corners which are not included in the disc, then sum them up, and set the output value at (100, 200) to that sum. This will give the effect of blurring the array(image).

Image Segmentation

Image segmentation is an inseparable part of any image description and recognition technique which can divide image into black and white regions by using pre-defined threshold values [25]. Detection of abnormalities like tumors and tissues classification can be carried out by image segmentation techniques.

- **Hill Climbing Segmentation**

The method is based on a hill-climbing approach and achieves the segmentation by performing two main tasks. First, the hill-climbing algorithm detects local maxima of clusters in the global three-dimensional

Color histogram of an image. Then, the algorithm associates the pixels of an image with the detected local maxima as a result, several visually coherent segments are generated. The segmentation algorithm is simple and fast. Moreover, the whole segmentation process is performed without any hand-tuning of parameters. This method does not assume any a priori knowledge on the number of clusters or the content of an image. KEY WORDS Color-based image segmentation, Hill-Climbing Algorithm.

Feature Extraction

After applying image enhancement and image segmentation, two other techniques namely logical summation and negation are applied where their output. In this step, one may only make the image more clear however, it's not enough since the ROI has not been found automatically. Therefore, it is essential to implement an algorithm to determine the cold nodules (the biggest white regions) which are the regions of interest (ROIs). For this purpose, a simple hill climbing algorithm followed by hundred times of consecutive runs will be performed in order to be sure that the heuristic methodology will finally find the ROI with largest area in terms of number of pixels.

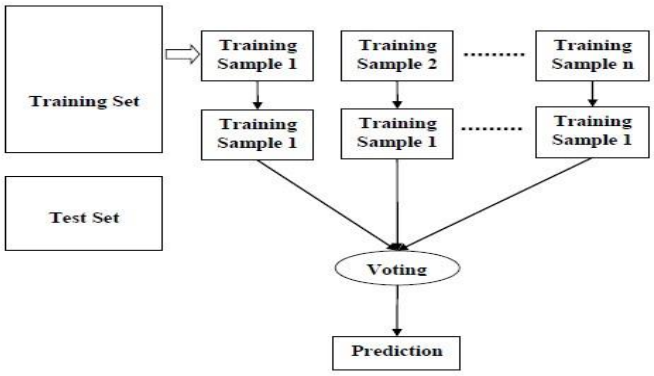
Classification

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

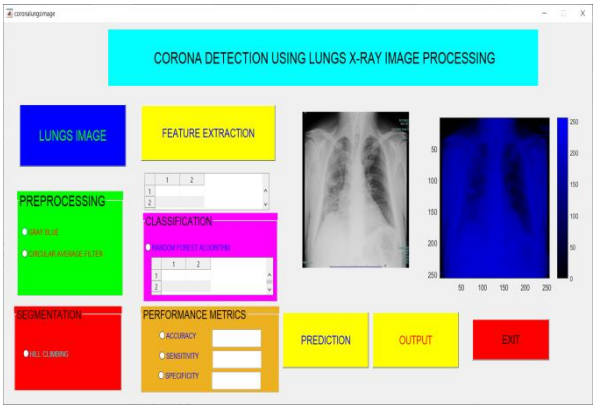
❖ Random Forest Algorithm

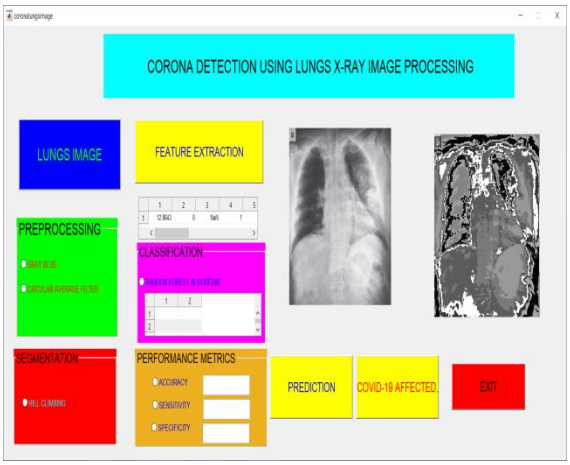
We can understand the working of Random Forest algorithm with the help of following steps,

- 1 − First, start with the selection of random samples from a given dataset.
- 2 − Next, this algorithm will construct a decision tree for every sample. Then it will
- get the prediction result from every decision tree.
- 3 − In this step, voting will be performed for every predicted result.
- 4 − At last, select the most voted prediction result as the final prediction result.



4. RESULTS





5.CONCLUSION

Detection of COVID-19 from chest X-ray images is of vital importance for both doctors and patients to decrease the diagnostic time and reduce financial costs. Artificial intelligence and deep learning are capable of recognizing images for the tasks taught. In this study, several experiments were performed for the high-accuracy detection of COVID-19 in chest X-ray images using ConvNets. Various groups—COVID-19/Normal, COVID-19/Pneumonia, and COVID-19/Pneumonia/Normal—were considered for the classification. Different image dimensions, different network architectures, state-of-the-art pre-trained networks, and machine learning models were implemented and evaluated using images and statistical data. When the number of images in the database and the detection time of COVID-19 (average testing time = 0.03 s/image) are considered using ConvNets, it can be suggested that the considered architectures reduce the computational cost with high performance. The results showed that the convolutional neural network with minimized convolutional and fully connected layers is capable of detecting COVID-19 images within the two-class, COVID-19/Normal and COVID-19/Pneumonia classification respectively.

6. REFERENCES

[1]Cui J, Li F, Shi ZL. Origin and evolution of pathogeni coronaviruses. Nat Rev Microbiol 2019;17(3):181–92.

[2]Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 2003;348(20):1967–76.

[3]Fouchier RA, Kuiken T, Schutten M, van Amerongen G, van Doornum GJ, van den Hoogen BG, et al. Aetiology: Koch's postulates fulfilled for SARS virus. Nature 2003;423(6937):240 .

[4]To KKW, Yip CCY, Lai CYW, Wong CKH, Ho DTY, Pang PKP, et al. Saliva as a diagnostic specimen for testing respiratory virus by a point-of-care molecular assay: a diagnostic validity study. Clin Microbiol Infect 2019;25(3):372–8 .

[5]Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020;395(10223):470–3.

[6]Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020.

[7]Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, et al. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol 2012;86(7):3995–4008.

[8]Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579(7798):265–9.

[9]Zhong NS, Zheng BJ, Li YM, Poon, Xie ZH, Chan KH, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet 2003;362(9393):1353–8.

[10]Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020;579 (7798):270