Grenze International Journal of Engineering and Technology, Jan Issue

Classification of Covid-19 in Chest X-Ray Image

C. M. Spoorthi [1], H. T. Basavaraju [2], V. N. Manjunath Aradhya [3] and K. Tejaswi [4]
[1, 3,4] Department of Computer Applications, JSS Science and Technology University, Mysuru, India
Email: spoorthichinnu987@gmail.com, aradhya@sjce.ac.in, tejaswi@jssstuniv.in
[2] Department of Computer Science, Yuvaraja's College Mysuru, University of Mysore, India
Email: basavaraju.com@gmail.com

***Abstract*—The entire planet is in danger because of the Covid-19 epidemic. It is now seen by people all around the world as a severe problem in China, as it was previously. It has infected more than 150 million individuals throughout the world and harmed more than 30,60,000 people as of June 1, 2021. Finding answers to questions about covid-19 has turned into a severe burden for the medical field. For the direct diagnosis of COVID-19, chest X-rays are a frequent approach. In this study, the pattern analysis methods are incorporated for the automatic classification of images. LBP, HOG, and Hu Moments are extracted to classify the chest X-Ray images using SVM and Random Forest classifiers. The performance of the proposed model is measured using accuracy. The result shows that method is promising and encouraging.**

***Index Terms*— COVID-19, X-ray, LBP, HOG, Hu Moments, Linear SVM, Random Forest.**

I. INTRODUCTION

The first corona viruses were discovered in the 1960s. Two major outbreaks have been caused by corona viruses [1]. MERS and SARS are two types of respiratory syndromes. It exhibits the same respiratory symptoms as the others, but it can also lead to kidney failure. Viruses are constantly changing, which can result in a new variant. All of a virus's genetic material is stored in something called RN (ribonucleic acid). There are some similarities between RNA and DNA, but they are not the same. When viruses infect your cells, they attach to them, enter them, and make copies of their RNA, which allows them to spread. The RNA is changed if there is a copping mistake. These changes are referred to as mutations by scientists. A new corona virus can infect the upper or lower part of your respiratory tract. The membrane can become inflamed and swollen. In other cases, the infection can spread throughout your alveoli. This might cause your lungs to enlarge, making it difficult for you to breathe [2]. Fever, suffocation, shortness of breath, old age, and hysical pain are some of the symptoms. Headache, Nausea, a loss of taste or smell. Also, this virus can lead to shock, asthma, liver problems, respiratory problems, death, and heart problems.

AI is a quickly expanding discipline of computer science with numerous applications in health care. ML is a subfield of AI that uses neural network algorithms to perform in-depth learning. It can recognize patterns and do difficult computing operations far more quickly and precisely than humans. The separation of images with extremely similar properties is a popular application for ML-based image analysis. The split of the image of interest, the discovery of effective image features integrated into a subdivided location or frequency, and the design of an acceptable mechanical classification system to deliver correct image samples in targeted categories are all part of this strategy. Using ML-based differentiation, we expected that COVID-19 patients' X-Ray images could be reliably distinguished from other kinds of pneumonia. Our findings support the idea that basic ML

Grenze ID: 01.GIJET.8.1.95

classification can be used in conjunction with other tests to assist differentiates COVID-19 patients' chest X-ray pictures. Overall, we believe that our method might be simply adapted to any future outbreak of quick X-Ray image categorization.

II. Literature Survey

Yasar and Ceylan [3] used DT-CWT, LBP, and CNN model for analyzing and classifying the COVID chest X-Ray images into Normal and Abnormal cases. In the experimental stage of the investigation, the participants were trained using the k-fold cross-validation approach. Aradya et al. [4] introduced the one-shot cluster-based method for identifying Covid-19 pneumonia in chest X-ray images with fewer training samples by using Generalized Regression and Probabilistic Neural Networks. Shelke et al. [5] used the conventional VGG16, DenseNet-161, ResNet-18 Deep Learning model for categorizing chest X-rays into four groups i.e, Normal, Pneumonia, tuberculosis (TB), and COVID-19. Furthermore, X-rays demonstrating the COVID-19 can be classed as mild, moderate, or severe depending on their severity. Kaiser et al. [6] created multi-modal data such as health status, close contact with the track, and COVID-19 data, as well as self-monitoring data and data analysis using KNN+K-means, Logistic regression, Bayesian DT. a fuse in notice provided by a fuzzy-neural network approach. Hussaina et al. [7] created 22 layered CNN diagnostic tools using Sigmoid, ReLU, and Leaky ReLU for 2 class classification (COVID and Normal), 3 class classification (non-COVID, pneumonia normal, and COVID), and 4 class classification (Normal, COVID, viral pneumonia, non-COVID and bacterial pneumonia). Ranjan et al. [8] created CNN based model to classify Thoracic Diseases by reducing the X-Ray image resolution using an auto encoder. Rajpurkar et al. [9] used 121 layered CheXNet to detect pneumonia from frontal-view chest X-ray images at a level exceeding practicing radiologists. Basavaraju et al. [10] discussed the elementary features for two dimensional gray scale images to obtain the discriminative features. Siyuan et al. [11] praposed neighboring aware graph neural network (NAGNN) model to detect covid-19 base on chest CT images. Aradya et al.[12] introduced phase by phase learning approach for the classification of covid-19 chest X-Ray image.

III. Proposed Methodology

Corona Virus is a set of linked RNA with a structure that is extremely difficult to decipher. As a result, we proposed a model that divides an X-Ray image into Normal and COVID-19 categories. We have applied some pre-processing to eliminate noise from the X-Ray image, then segment the lungs using LBP, HOG, and Hu Moments to extract the information out of the Chest X-Ray images, and lastly classify them using SVM and Random Forest. Figure 1 shows the methodological architecture in detail.

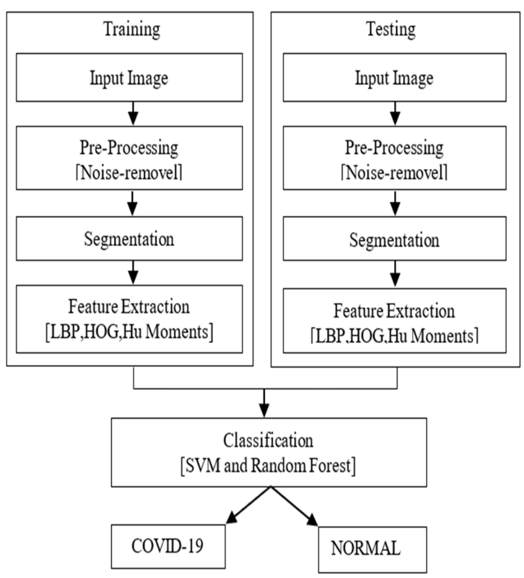


Figure 1. Training and testing architecture of proposed methodology

A. Pre-processing

The term "image pre-processing" refers to manipulations on visuals at the most basic level. It is used to improve image data by removing unwanted distortions and improving some visual features that are important for further processing and analysis [13,14].

- *Gaussian Blur* - The goal of applying filters is to improve or change the qualities of an image, as well as to extract important information from images such as edges, corners, and spaces. The kernel, a tiny table, and each pixel in the image and its neighbours are all defined as filters. Gaussian blur is a technique for reducing image noise and preserving image detail. This will aid in the reduction of background noise in the lungs.

B. Segmentation

The objective of medical image segmentation is the extraction of quantitative information, such as morph-metric data from large amounts of data, textures, and patterns on the organ. The main purpose of which is the removal of the lung field on a chest X-ray.

- *Removing of the spinal cord* - In the spinal cord, Removal, we can remove the spinal cord and reduce the complexity and improve the accuracy.
- *Binarization* - The entire X-ray image is transformed to a binary image using a threshold value depending on light intensity in the range of 127 to 255).
- *Contour mask* - To Contour-mask-image binarization, the next step is to take two of the major contours (or regions), in the lung area. Finally, applying a contour mask on an original x-ray image to obtain a segmented lung region.

C. Feature Extraction

Pattern recognition has benefited significantly from dimensionality reduction approaches. We often achieve a density of input data, reducing storage requirements. In some cases, there is also an improvement in the effects of segregation due to the decrease in noise present in a large proportion of natural images. Feature output is the main component of a computer vision pipeline. In reality, every machine learning model is based on the idea of collecting valuable features that accurately represent the objects in an image. A factor in machine learning is each measurable asset or a physical condition factor. Features are introductions that are feed on machine learning models to exclude predictions or divisions. Some handmade feature sets are:

- *Local Binary Patterns (LBP)*

The local Binary Pattern (LBP) [15] is a simple but powerful texture operator that allows you to define the pixels in an image using a threshold value and the area of each pixel, and then treat the result as a binary number. It was first described in 1994 (LBP), and it has since proven to be a useful texture classification system. With LBP is combined with the histogram, that will send the pictures of the lungs, as in a vector of data. Since LBP is a visual descriptor, which can be seen in equation 1 i.e, *Ge* is the intensity of an image I and *ni (n = 0,..., p-1)* be the intensity of a pixel in the 3 × 3 neighbourhood excluding the center pixel. The LBP feature extraction process is shown in Figure 2.

$$LBP = \sum_{i=0}^{p-1} s(ni - Ge) > i \qquad (1)$$

$$S(x) = \begin{cases} 1 \; if \; x > 0 \\ 0 \; otherwise \end{cases} \qquad (2)$$

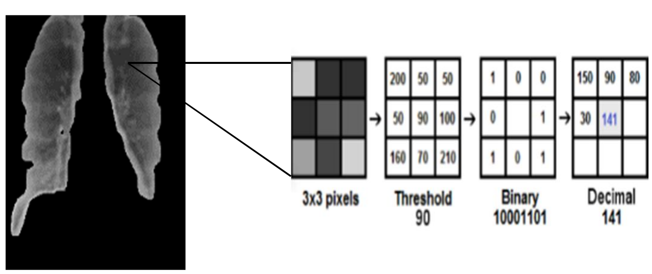


Figure 2. Feature extraction process of Local Binary Pattern

- *Histogram of Oriented Gradients (HOG)*

The Gradient Histogram is a simple function of the extraction procedure, which was developed in the context of the identification of shape in an image. If necessary, with the aims of the image. This leads us to the qualities that resist it, depending on the change of the light, select an image with the two filters, which are sensitive to both horizontal and vertical brightness gradients. They record information about the borders, regions, and textures, divide the image into cells of a certain size, and compute the histogram, over the ratio within the cell, in the normalization of the histogram of each cell by comparing it with a block of neighbouring cells. Further weakened with the lighting effect to the entire image, on the construction of a one-dimensional character vector of the information contained in each cell. To determine the gradient (or change) in the x-direction, we need to subtract the value on the left from the pixel value on the right as shown in equation 3. Similarly, to calculate the gradient in the y-direction, we will subtract the pixel value below from the pixel value above the selected pixel as shown in equation 4. Next, calculate the orientation (or direction) for the same pixel by writing the tan inverse for the angles as shown in equation 5.

$$dx = I\ (x+1, y) - I\ (x, y) \tag{3}$$

$$dy = I\ (x, y+1) - I\ (x, y) \tag{4}$$

$$\theta(x, y) = tan^{-1}\left(\frac{dy}{dx}\right) \tag{5}$$

- *Hu Moments (HuM)*

Hu Moments are usually taken from an image's shadow or outline. By characterizing an item's silhouette or outline, we can extract a shape feature vector (i.e., a list of numbers) to represent its shape. We usually get this after performing some sort of segmentation. [16,17] Hu Moments are a set of seven numbers that are invariant to translation, scaling, rotation, and reflection and are determined using central moments. While hu is in the seventh moment, the indicator for image reflection changes. Where M_1 calculates the area of the region, M_2 calculates the center of the object in the x-direction, M_3 calculates the center of the object in the y-direction, and M_4, M_5, M_6, M_7 calculates spreads of the image in four directions.

$$M_1 = (s_{20} + s_{02}) \tag{6}$$

$$M_2 = (s_{20} - s_{02})^2 + 4s^2_{11} \tag{7}$$

$$M_3 = (s_{30} - 3s_{12})^2 + (3\ s_{21} + s_{03})^2 \tag{8}$$

$$M_4 = (s_{30} + s_{12})^2 + (s_{21} + s_{03})^2 \tag{9}$$

$$M_5 = (s_{30} - 3s_{12})(s_{30} + s_{12})[(s_{30} + s_{12})^2 - 3(s_{21} + s_{03})^2]\ (3s_{21} - s_{03})(s_{21} + s_{03})[3(s_{30} + s_{12})^2 - (s_{21} + s_{03})^2] \tag{10}$$

$$M_6 = (s_{20} - s_{02})\ [\ (s_{30} + s_{12})^2 - (\ s_{21} + s_{03})2] + 4\ s_{11}(s_{30} + s_{12})\ (s_{21} + s_{03}) \tag{11}$$

$$M_7 = (3s_{21} - s_{03})\ (s_{30} + s_{12})\ [(s_{30} + s_{12})^2 - 3(s_{21} + s_{03})^2]\ (3s_{30} - 3s_{12})\ (s_{21} + s_{03})\ [3(s_{30} + s_{12})^2\ (s_{21} + s_{03})^2 \tag{12}$$

D. Classification

Classification is a supervised learning concept in machine learning that divides a set of data into categories. Speech separation, face recognition, text recognition, document separation, and other issues are the most common. It could also be a multi-class or binary split problem. For machine learning classification, there are numerous machine learning algorithms. Random Forest and Linear SVM are two classification algorithms used.

- *Support Vector Machine (SVM)*

SVM may be used to solve a wide range of classification and regression issues. It is still one of the most widely used robust prediction algorithms that may be applied to a variety of classification application cases. To accurately identify 2 or more different classes in a separation problem, SVM finds a well-separated line termed a 'Hyper plane.' The goal is to achieve a complete hyper plane separation by dividing training data using the SVM algorithm. More systematically, the method (SVM) generates a hyperplane with a large amount of space (even if not equally spaced) that aids in isolation, external detection, retrieval, and other tasks. A hyper plane with the greatest distance between the nearest training points provides a fair classification of classes. Because this is a two-class problem (COVID-19 and NORMAL), the linear SVM was taken into account. Because Linear SVM is the best solution for a two-class problem. The SVM is chosen based on where it will classify the X-Ray image based on Hyper plane. It is considered COVID-19 if it is more than the Hyper plane, and NORMAL if it is less than the Hyper plane.

- *Random Forest*

Random Forest is built on the concept of learning together, which is the process of merging multiple divisions to solve a complex problem and improve a model's performance. It is a subset of a database that contains several decision trees in distinct subsets and takes measures to improve the accuracy of the dataset's supposition. The random forest forecasts the outcome by taking predictions from each tree and relying on many predictive votes, rather than relying on a single decision-making tree. The more trees in the forest, the better the accuracy and prevention. Because the random forest combines numerous trees to forecast the dataset's class, some decision trees may correctly anticipate the outcome. The dataset's feature variable contains some actual values, allowing the classifier to predict accurate outcomes rather than a guess. Each tree's predictions must have very low correlations. It comprises two phases: the first is to combine N decision trees to build a random forest, and the second is to generate predictions for each tree created in the first phase.

IV. Experimental Analysis

Images of the covid19-radiography-database were used to obtain chest X-rays. Chest X-rays from Normal (1,341) and COVID-19 (1,200) were collected. The X-ray pictures acquired are divided into two categories: normal and COVID-19. All images are in PNG format and have 1024 x 1024 resolution pixels (552 x 552). The images are greyscale, with pixel intensities ranging from 0 to 255.

Two algorithms for considered for comparative analysis. In the training phase, the accuracy will be calculated with respect to Linear SVM and Random Forest classification algorithms. The highest accuracy obtained from the experimental result is SVM with 96.9% for 80-20 [80 for testing and 20 for training]. We considered 80-20 out of 80-20, 90-10, 50-50, 70,30 because most of the time we go with 80-20. The performance of the proposed model is evaluated using performance measures like accuracy (equation 13). This performance uses certain parameters like True Positive (TP): Covid-19 affected people correctly classified as infected. True Negative (TN): Healthy people correctly classified as healthy. False Positive (FP): Healthy people incorrectly classified as infected and False Negative (FN): Infected people incorrectly classified as healthy.

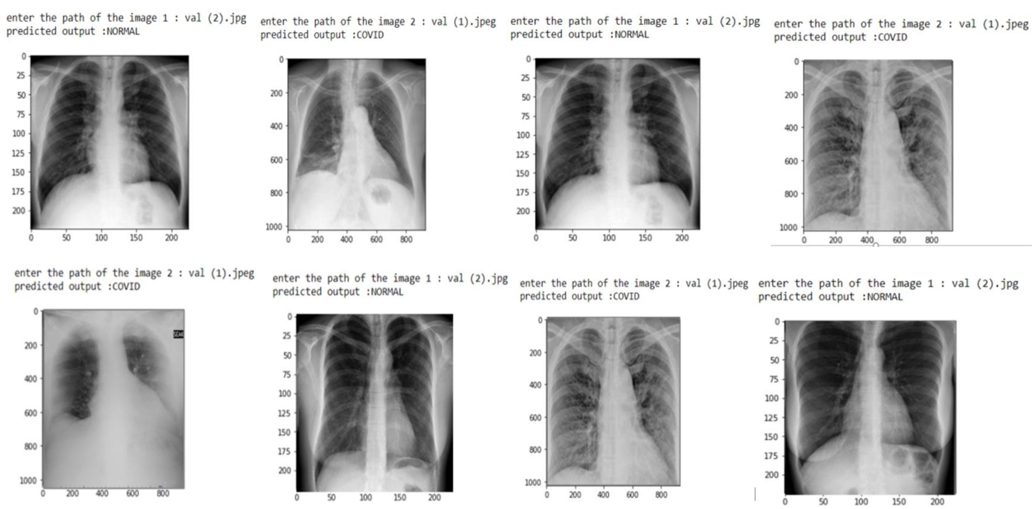
The results are all are shown in Table I.

$$\text{ACCURACY} = \frac{(\text{TP+TN})}{(\text{TP+TN+FP+FN})} \qquad (13)$$

TABLE I. ACCURACY TABLE

Train set-Test set	SVM	Random Forest
80-20	96.94	96.33
70-30	97.14	96.33
50-50	94.45	92.73
90-10	99.18	95.51

A. Illustration of sample results of chest X-Ray images



VI. Conclusion

Image analysis techniques are crucial when it comes to disease treatment and diagnosis. COVID-19 has a terrible influence on our modern life, especially public health systems in the global economy, due to its current prevalence. These techniques, we feel, could be employed as a computer-assisted detection tool. As a result, In this paper, we present a methodology for detecting COVID-19 using chest X-ray images. COVID-19 and Normal are the two groups for which the suggested model may provide an appropriate diagnosis. Based on the use of LBP, HOG, and Hu Moments in extracting features in COVID-19 images, then split the image using Linear SVM and Random Forest. The proposed model has a classification accuracy of 96.9%, which is the best successful accuracy in all of our knowledge in the data sets utilized in the experiment. The preparation of a very big database algorithm for classification algorithms is also described in the paper. Our long-term goal is to enhance lung segmentation accuracy and precision. In addition, new features must be extracted.

References

[1] CoronaVirusHistory:https://www.webmd.com/lung/coronavirus-history#:~:text=Continued-,hCoronavirus%20Evoluton,Seven%20coronaviruses%20can%20infect%20humsns.

[2] Symptoms:https://www.webmd.com/lung/coronavirus

[3] Huseyin Yasar and Murat Ceylan, "A New Deep Learning Pipeline to Detect Covid-19 on Chest X-Ray Images using Local Binary Pattern, Dual-Tree Complex Wavelet Transform and Convolutional Neural Networks", Appl Intell 51, https://doi.org/10.1007/s10489-020-02019-1, pp. 2740–2763,2021.

[4] V. N. M. Aradhya, M. Mahmud, D. S. Guru, B. Agarwal, and M.S. Kaiser, "One-shot cluster-based approach for the detection of COVID–19 from chest X–ray images". Cognitive Computation, pp. 1-9, 2021.

[5] A. Shelke, M. Inamdar, V. Shah, A. Tiwari, A. Hussain, T. Chafekar, and N. Mehendale, " Chest X-ray Classification Using Deep Learning for Automated COVID-19 Screening", SN Computer Science, 2(4). doi:10.1007/s42979-021-00695-5, 2021.

[6] M. Shamim Kaiser, Mufti Mahmud, Manan Binth Taj Noor, Shamim Al Mamun, K. M. Abir Mahmud, Saiful Azad, V. N. Manjunath Aradhya, Punitha Stephan, Thompson Stephan, Ramani Kannan, Mohammed Hanif, Tamanna Sharmeen, Tianhua Chen, and Amir Hussain, "Iworksafe: Towards Healthy Workplaces During Covid-19 With An Intelligent Phealth App For Industrial Settings", (Doi: 10.1109/Access.2021.3050193), pp. 13814-13828, 2021

[7] Emtiaz Hussaina, Mahmudul Hasana, Md Anisur Rahmanb, Ickjai Leec, D. Tasmi Tamanna, A. Mohammad Zavid Parvez, "CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images", vol. 142, pp. 110495, 2021.

[8] Ekagra Ranjan, Soumava Paul, Siddharth Kapoor, Aupendu Kar, Ramanathan Sethuraman, and Debdoot Sheet, "Jointly learning convolutional representations to compress radiological images and classify thoracic diseases in the compressed domain", https://doi.org/10.1145/3293353.3293408, 2018.

[9] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, P. Matthew Lungren, and Y. Andrew Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning", arXiv preprint arXiv:1711.05225, 2017.

[10] H. T Basavaraju, V. N. M. Aradhya and D.S. Guru, "A novel arbitrary-oriented multilingual text detection in images/video In Information and decision sciences", pp. 519-529, Springer, Singapore,2018.

[11] Siyuan Lu, Ziquan Zhu, Juan Manuel Gorriz,Shui-Hua Wang, and Yu-Dong Zhang, "NAGNN: Classification of COVID-19 based on neighboring aware representation from deep graph neural network", https://doi.org/10.1002/int.22686, 2021.

[12] V.N.M. Aradhya,M. Mohmud, M. chowdhury,D. S.Guru,M.S.Kaiser and S.Azed, "Learning through One Shot: A Phase by Phase Approach for COVID-19 Chest X-Ray Classsification", IEEE-EMBS Conference on Biomedical Engineering and Sciences(IECBES), pp. 241-244, dio:10.1109/IECBS48179.2021.9398761, 2021.

[13] M.C. Hanumantharaju, V.N.M. Aradhya, and G. Hemanth Kumar, "Understanding and Analysis of Enhanced COVID-19 Chest X-Ray Image", SpingerBriefs in Applied Science and Technology, Springer, Singapore, dio:10.1007/978-981-15-6572-4_6, 2020.

[14] S.S, Bhat, M.C. Hanumantharaju, and V.N.M. Aradhya," Super Resolution Techniques for COVID-19 Chest X-Ray images", Computer communication,Networking and IOT,Sringer,Singapore,dio:10.1007/978-981-16-0980-0_53,2021.

[15] Tapabrata Chakraborti, Brendan McCane, Steven Mills, and Umapada Pal, "LOOP Descriptor: Local Optimal-Oriented Pattern", pp(99)1-1,2018.

[16] M. Poorani, G. Ravindra, T. Prathiba, "Integrated Feature Extraction for Image Retrieval", IJCSMC, vol. 2, pp. 28-35, 2013.

[17] Ming-Kuei Hu, "Visual pattern recognition by .moment invariants", doi:10.1109/tit.1962.1057692, vol. 8, Issue. 2, pp. 179-187, 1962.