2020 IEEE Bombay Section Signature Conference (IBSSC) | 978-1-7281-8993-2/20/$31.00 ©2020 IEEE | DOI: 10.1109/IBSSC51096.2020.9332158

Covid19 Identification from Chest X-Ray Images using Local Binary Patterns with assorted Machine Learning Classifiers

Sudeep D. Thepade, Ketan Jadhav

Pimpri Chinchwad College of Engineering, Savitribai Phule Pune University, Pune, India

sudeepthepade@gmail.com, jadhavketan14@gmail.com

Abstract — **The novel corona virus caused by the SARS-CoV2 virus originated in Wuhan, China and spread globally. The massive outbreak of the virus resulted in millions of people being infected. Early detection of the virus is crucial in the complete recovery of the patient but can be fatal if detected in the later stages. The symptoms of the virus being similar to flu make it difficult to detect. This paper attempts an automated system for identification of the Covid19 virus infected images of chest X-Ray. The proposed method uses a dataset which has human chest X-Rays of non infected people as well as patients suffering from pneumonia and Covid19 virus infection. Local binary patterns with variations in its input parameters are used for feature extraction. The resulting feature sets are classified using several machine learning algorithms and ensembles of these individual models. Results of experimentation are obtained across 10 fold cross validation testing. Evaluation metrics accuracy, positive predictive value (PPV), sensitivity and f-measure are used to compare performance. Observations of the results show that the ensemble of RTree-RForest-KNN gives the best classification performance while ensemble models perform better than most individual classifiers. Comparing the input parameters of the LBP, the best performance is given by parameters R=6 (P=48) and R=7 (P=56) gives the best performance for the average of metrics for 10 fold cross validation in the proposed Covid19 identification method from chest X-Ray images.**

Keywords — **Covid19, Local Binary Pattern, Machine Learning, Histogram**

I. Introduction

Covid19 also known as the novel corona virus is caused in humans by the SARS-CoV2 virus (Severe Acute Respiratory Syndrome Corona virus). This virus is predominantly found in animals, especially bats but due to its zoonotic nature, it can affect humans as well. Zoonotic nature of a virus means that this virus can affect other animals which in turn can affect humans. The Covid19 virus first emerged in the Wuhan region, China in December 2019 and has rapidly spread globally, causing deaths all over the world. By May 2020, over 5 million people had been infected by the virus which has caused more than three hundred thousand casualties globally.

The symptoms for this disease include fever, fatigue, difficulty in breathing, sore throat, fever, cough and headache. The massive outbreak of the virus has caused it to be declared as a pandemic by the WHO. Studies show that early detection of the disease can lead to the complete recovery of the patient but it becomes difficult to treat in the later stages of detection. The symptoms of the virus are similar to that of flu and hence medical experts are heavily reliant on other methods of detection like medical images. Computer aided diagnosis has played a huge role in the diagnosis of the disease. X-Ray images of the chest are being used for the identification of the virus infection by healthcare professionals globally.

The paper is structured in the following way; the Literature Survey conducted for the existing chest X-ray analysis methods is presented in Section 2. Section 3 discusses the proposed Covid 19 infection identification method. Experimentation environment used for the proposed method are given in Section 4. The observed results obtained from experimentation are elaborated in Section 5. Section 6 compiles the observations as the conclusion for this paper.

II. Literature Survey

Covid19 virus has rapidly spread over the globe and infected millions of people. Studies can conclude that this virus uses respiratory droplets and close contact [1] as modes of transmission in humans. As of June 1, 2020, more than 5.5 million people had been infected with the virus resulting in the deaths of more than three hundred thousand people globally [2]. Inadequate resources and lack of expertise has made it difficult to diagnose the virus in humans. X-Ray images of the human chest are used to detect the virus but the images look very similar to those with pneumonia which makes it even more difficult to correctly identify the virus.

Machine learning has been widely used in the past to classify images in the medical field. Various feature extraction methods and classifiers have been attempted recently to classify chest X-Ray images. Local Binary Patterns (LBP) [3] are used in the past for the feature extraction process to classify X-Ray images. Several variations of LBP like CS-LBP [4] are also attempted. The advancements in deep learning have made a major contribution in image processing and classification. Neural networks have been observed to obtain a higher accuracy for classification. But these systems require large datasets and availability of data is a major issue. Also, neural networks take a long time to train which decrease the feasibility of their use. Pneumonia detection using chest X-Ray images has been attempted using several techniques of deep learning in the recent past [5]. Convolutional Neural Networks (CNN) are used for detecting abnormalities in chest images [6]. Variations of CNNs are also been used to classify Covid19 virus recently [7]. Several machine learning algorithms have been used in the past for classification of medical images. K-Nearest Neighbors (KNN) and Support Vector machines (SVM) are widely used to classify image.

III. Proposed Covid19 Identification technique using Machine Learning Classifier with Histogram of Luminance Chroma features

The proposed Covid 19 infection identification method uses the human chest X-Ray images of patients. There are two main phases in the identification; feature extraction and classification. Local Binary Patterns (LBP) are used for extracting feature sets from the input chest X-Ray images. Later, the feature sets are classified using supervised learning algorithms. Various machine learning models and ensembles of these models are trained on the feature sets for classification purposes. The following figure demonstrates the process used in the proposed method.

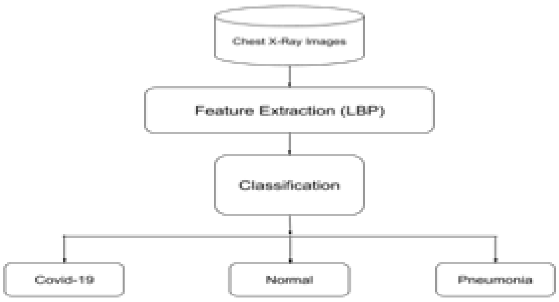

Figure 1: Flowchart representing covid-19 identification using the proposed method.

A. *Feature Extraction from chest X-ray image using Local Binary Patterns..*

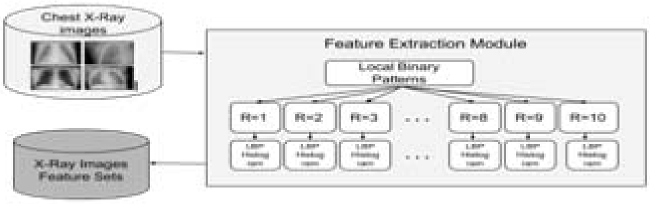

Figure 2: Feature extraction from X-Ray images after applying local binary patterns.

Local Binary Patterns (LBP) is a robust method for feature extraction that focuses on the texture features of an image. It operates on the thresholding of a pixel with each of its neighboring pixels. LBP is widely used for feature extraction in classification problems. Each pixel value of an image is compared with all its neighboring pixel values and a binary string is generated using the comparison results. The proposed method constructs an histogram using all the binary strings obtained after comparison for each pixel. Feature sets are generated using this LBP Histogram.

The LBP algorithm uses a simplistic approach to generate the binary string for each pixel. A simple comparison of pixel value generates 1 if it is greater than or equal to and 0 if it is less than the neighbouring pixel value. While reconstructing the image, this binary string is converted into decimal which becomes the updated value of this pixel. The LBP algorithm used takes two parameters as input along with the image; radius and number of neighbours.

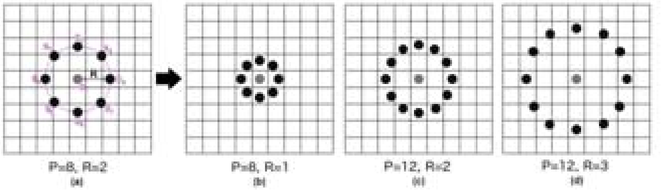

Figure 3: Representation of LBP input parameters; radius and number of points.

Figure 3 demonstrates the radius (R) and number of neighbours (P) parameters of the LBP descriptor. Radius determines how far the neighbouring pixels are from the central pixel used for calculations. The number of points determines how many points are to be considered for comparison. The value of P is the length of the binary string generated. This p-bit binary string is computed for each pixel in the image and a histogram is plotted using these strings in the proposed method. Various values of the input parameters R and P are experimented with in this method.

B. *X-ray Image Classification using MLP for Covid 19 Identification:*

The feature sets extracted using variations of local binary patterns are further used for classification. Several supervised learning algorithms are trained on the feature sets obtained. The proposed method uses the various machine learning algorithms for classification alias Naive Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Trees and Random Forests.

Along with these individual learning methods, ensemble learning is also used to generate and train ensemble models from these individual models. It is usually more reliable to use more than one model in designing a classifier. Ensemble learning is a method in which more than one model works on a training set simultaneously. These models are compared with each other and the best performing one is selected. There are several methods used to select the best performing model. The proposed method uses the majority voting method. The ensembles experimented with in the proposed method are ‘Random Forests-Random Trees-SVM’, ‘Random Forests-Random Trees-KNN’ and ‘Random Forests-SVM-KNN’.

IV. Experimentation Environment

The dataset used for evaluation contained X-Ray images highlighting the lungs and chest region of patients. The dataset contained 68 X-Ray images of coronavirus infected patients and 79 X-Ray images of normal lungs and 158 X-Ray images of pneumonia patients in jpeg format. Feature extraction was performed in python and the Waikato Environment for Knowledge Analysis (WEKA) tool was used for classification of the extracted features.

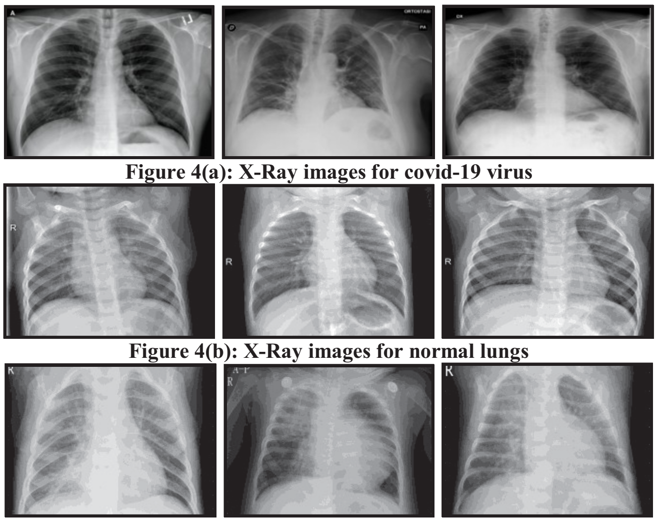

Figure 4(a): X-Ray images for covid-19 virus

Figure 4(b): X-Ray images for normal lungs

Figure 4(c): X-Ray images for pneumonia

Figure 4: Sample X-ray images from Covid19 Image Data Collection[8][9].

Various supervised learning algorithms were used for classification and accuracy, positive predictive value (PPV), sensitivity and f-measure of classification were used as performance metrics to evaluate the methods. Based on the confusion matrix obtained for each classifier, accuracy, positive predictive value (PPV), sensitivity, f-measure and Matthews Correlation Coefficient (MCC) can be calculated using the equations (1), (2), (3), (4) and (5). Accuracy, PPV, sensitivity, f-measure and MCC for Covid19 class as well as for weighted average of all classes is considered in the proposed method.

$$Accuracy = TP + TN / (TP + FP + TN + FN) \tag{1}$$

$$Positive predictive value (PPV) = TP / (TP + FP) \tag{2}$$

$$Sensitivity = TP / (TP + FN) \tag{3}$$

$$F - Measure = (2 * PPV * sensitivity) / (PPV + sensitivity) \tag{4}$$

$$MatthewsCo \quad rrelationC \quad oefficient \; (MCC) = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

where TP, TN, FP, FN denote the count of true positive, true negative, false positive and false negative respectively. The PPV, sensitivity, f-measure and MCC are considered for the Covid19 class and also for the weighted average of all classes.

V. Results and Discussion

The dataset of chest X-Ray images is passed to the feature extraction module where Local Binary Patterns (LBP) with variations in its input parameters (Radius and Number of points) to obtain feature sets for classification. The LBP Histogram obtained is considered for classification. The table demonstrates the relation between the input parameters of the LBP descriptor for discussion of results.

TABLE 6
REPRESENTATION OF INPUT PARAMETERS RADIUS(R) AND NUMBER OF POINTS(P) USED IN THE PROPOSED METHOD

R	1	2	3	4	5	6	7	8	9	10
P	8	16	24	32	40	48	56	64	72	80

The feature sets generated in the first phase are then used for classification in the next phase. Various machine learning algorithms are experimented with for classification of the feature vectors. Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbors (KNN), Random Tree, Random Forests as well as three ensembles of the algorithms are used for classification and the results are used for discussion.10 fold cross validation testing is used for experimentation.

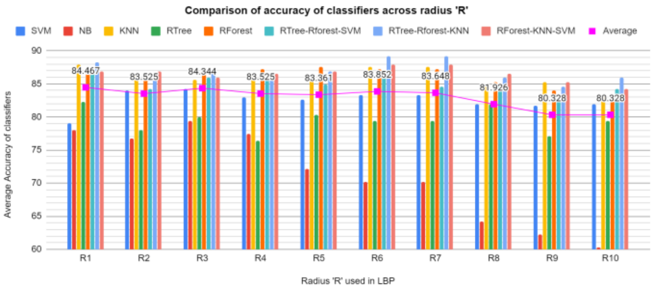

Figure 5: Comparison of average accuracy of classifiers across radius ‘R’ in LBP

Figure 5 demonstrates the comparison of the average accuracy of various classifiers applied over different feature sets of the LBP input parameter ‘R’. It can be observed from the chart that the feature sets extracted with R=1 (P=8) have a better average performance compared to others. An average accuracy of 84.467 can be observed across classifiers for feature sets with R=1. Comparing all classifiers individually, the highest classification accuracy is achieved by the ensemble RTree-RForest-KNN for R=6 and R=7 which is 89.180. Comparison of classifiers show that the ensemble methods perform better than individual classifiers.

PPV is the probability of the Covid19 classified instances being actually labelled Covid19. Figure 6 represents the comparison of PPV for Covid19 class across radius ‘R’ for LBP. It can be observed that SVM performs the best among other classifiers for average PPV. SVM has an average PPV of 0.936 across the radius parameter. Considering individual values of ‘R’, SVM has the best individual PPV for R=1 and R=2

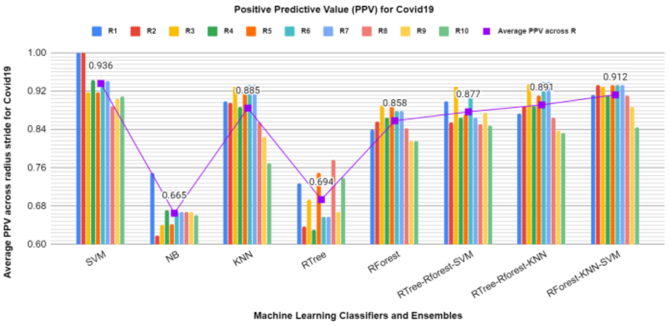

Figure 6: Comparison of average Positive Predictive Value for Covid19 class for machine learning classifiers

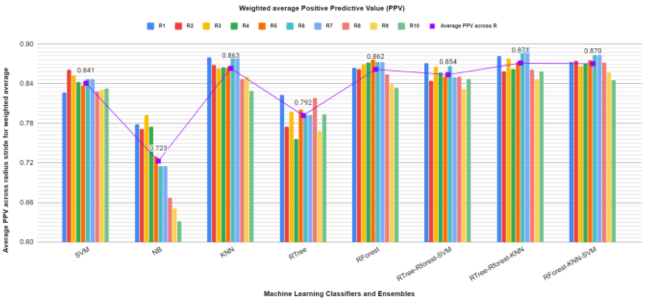

Figure 7: Comparison of weighted average Positive Predictive Value for machine learning classifiers

Figure 7 represents the comparison of weighted average of positive predictive value for all classes across radius parameters. It is seen that the ensemble Random Tree-Random Forest - KNN gives the top average performance of 0.871 among all classifiers. Comparing the individual radius values, the ensemble Random Tree- Random Forest - KNN has the highest PPV of 0.896 for R= 6 and R= 7.

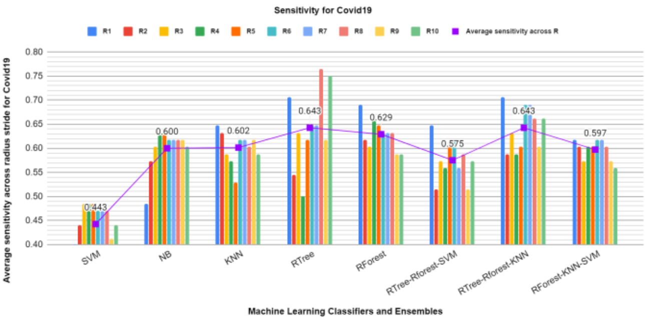

Figure 8:Comparison of average Sensitivity for Covid19 class for machine learning classifiers

Figure 8 represents the comparison of average sensitivity for various classifiers across the radius parameter of LBP for the Covid19 class. It can be seen that Random Tree and the ensemble of Random Tree - Random Forest - KNN has the highest average sensitivity of 0.643 among classifiers. Considering individual radius values, Random Tree for R=8 has the highest sensitivity of 0.765.

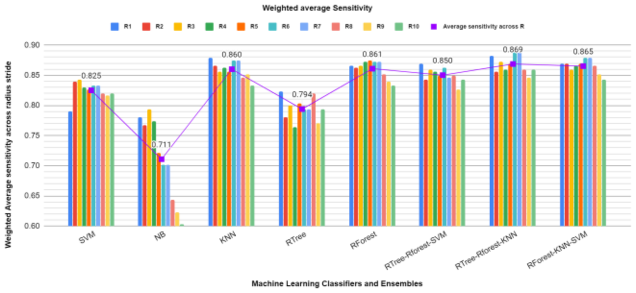

Figure 9: Comparison of weighted average Sensitivity for machine learning classifiers

Comparison of average sensitivity for weighted average of all classes across radius ‘R’ for various classifiers is demonstrated by figure 9. It can be observed from the figure that the ensemble Random Tree - Random Forest - KNN has the highest average sensitivity of 0.869. Considering individual radius values, the ensemble Random Tree - Random Forest - KNN has the highest sensitivity of 0.892 for R=6 and R=7.

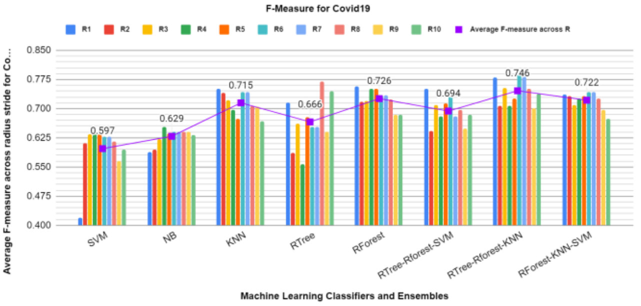

Figure 10: Comparison of average F-measure for Covid19 class for machine learning classifiers

Figure 10 demonstrates the comparison of the average F-measure for Covid19 class across radius ‘R’. It can be seen from the figure that the ensemble Random Tree - Random Forest - KNN performs the best with an average F-Measure of 0.746 among other classifiers. Comparing individual radius values, the ensemble Random Tree - Random Forest - KNN for R=6 and R=7 have the highest F-Measure of 0.797.

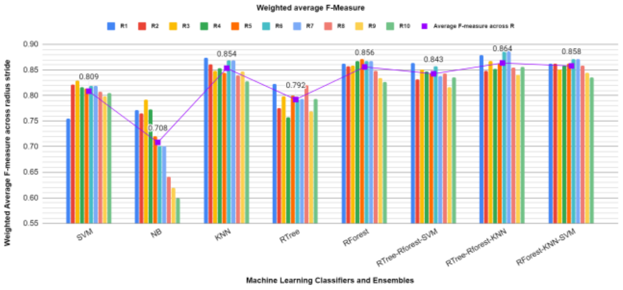

Figure 11: Comparison of weighted average F-Measure for machine learning classifiers

Comparison of F-Measure for the weighted average of all classes is demonstrated by figure 11. It can be observed that for an average across radius 'R', the ensemble Random Tree - Random Forest - KNN outperforms other classifiers with an average F-Measure of 0.864. The ensemble performs the best even after comparing individual radius values for R=6 and R=7 with a F-Measure of 0.888.

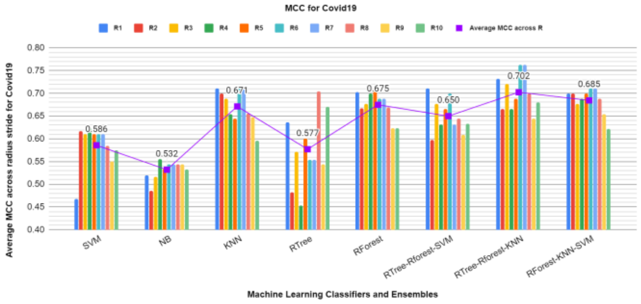

Figure 12: Comparison of average MCC for Covid19 class for machine learning classifiers

Comparison of average Matthews Correlation Coefficient (MCC) for the Covid19 class is represented by figure 12. It can be seen that the ensemble Random Tree - Random Forest - KNN has a superior average MCC of 0.702 compared to other classifiers. Also comparing the individual radius values, the ensemble Random Tree - Random Forest - KNN has the highest MCC of 0.763 for R=6 and R=7.

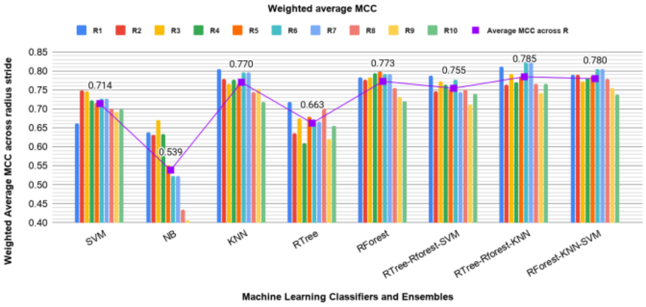

Figure 13: Comparison of weighted average MCC for machine learning classifiers

Figure 13 represents the comparison of the Matthews Correlation Coefficient (MCC) for the weighted average of all classifiers across radius 'R'. For individual radius values, the ensemble Random Tree - Random Forest - KNN has the highest MCC of 0.825 for R=6 and R=7. Even after considering the average MCC across 'R', the ensemble performs the best with an MCC of 0.785.

Figure 14 represents the comparison of the average taken for performance metrics PPV, Sensitivity, F-Measure and MCC for the Covid19 class. It is evident from the figure that the ensemble Random Tree - Random Forest - KNN performs the best among classifiers with the average of 0.746. Also, ensemble learning is seen to perform better than individual machine learning algorithms.

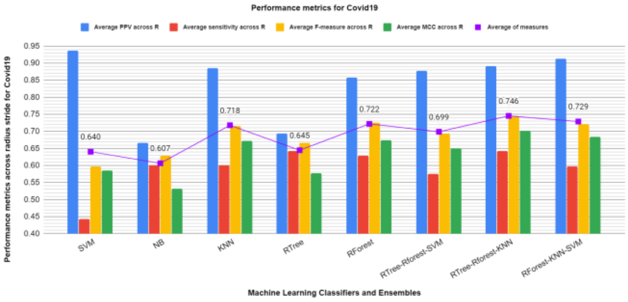

Figure 14: Comparison of average of performance metrics of machine learning classifiers for Covid19 class

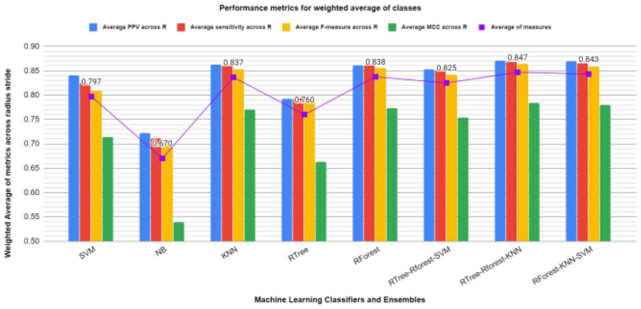

Figure 15: Comparison of average of performance metrics of machine learning classifiers for weighted average of all classes**.**

The comparison of performance metrics for the weighted average of all classes is represented by figure 15. It can be seen that the ensemble Random Tree - Random Forest - KNN has the highest average of 0.847 among classifiers. Furthermore, ensembles perform better than individual machine learning algorithms.

Overall, from all the performance metrics used in the proposed, the ensemble methods perform considerably better than individual machine learning classifiers for the Covid19 class and also the weighted average of all classes.

VI. Conclusions

The SARS-CoV2 virus or corona virus has affected millions worldwide and caused a lot of deaths. Medical imaging and computer aided diagnosis has helped researchers and doctors in the diagnosis and identification of a lot of diseases. Introduction of Machine Learning in classification has automated the identification process. This paper focuses on use of chest X-Ray images to identify the Covid19 virus infection in humans. The images were first passed to a feature extraction module where with the help of local binary patterns (LBP) feature sets were extracted. LBP histograms were generated using variations of the input parameters (radius and number of points) of the LBP descriptor and used as feature sets. These feature sets were used for classification where various machine learning algorithms as well as ensembles of the algorithms were trained on the input data across different testing methods. The results of experimentation can conclude that the ensemble of RTree-RForest-KNN performs the best on the data and the ensemble methods perform better than most individual classifiers. Comparing the LBP input parameters, the parameter R=6 (P=48) and R=7 (P=56) gives the best performance for the average of metrics across the 10 fold cross validation.

References

[1] Covid19 Disease Information https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendation (last referred on 10 may 2020)

[2] Covid19 Disease Spread and Diagnosis Difficulties https://covid19.who.int/ (last referred on 10 may 2020)

[3] S. Kim, J. Lee, B. Ko and J. Nam, "X-ray image classification using Random Forests with Local Binary Patterns," 2010 International Conference on Machine Learning and Cybernetics, Qingdao, 2010, pp. 3190-3194, doi: 10.1109/ICMLC.2010.5580711

[4] Ko BC, Kim SH, Nam JY. X-ray image classification using random forests with local wavelet-based CS-local binary patterns. J Digit Imaging. 2011;24(6):1141-1151. doi:10.1007/s10278-011-9380-3

[5] Rajpurkar, P., et al., CheXNet: Radiologist-Level Pneumonia Detection on Chest XRays with Deep Learning. arXiv preprint arXiv:1711.05225, 2017.

[6] T. I. Mohammad, A. A. Md, T. M. Ahmed, and A. Khalid, “Abnormality detection and localization in chest x-rays using deep convolutional neural networks,” 2017, http://arxiv.org/abs/1705.09850.

[7] Singh D, Kumar V, Vaishali, Kaur M., “Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks”, [published online ahead of print, 2020 Apr 27]. Eur J Clin Microbiol Infect Dis. 2020;1-11. doi:10.1007/s10096-020-03901-z

[8] Cohen J.P., Morrison P., Dao L. COVID-19 Image Data Collection. arXiv 2020, arXiv:2003.11597.

[9] Cohen, J.P.; Morrison, P.; Dao, L. COVID-19 Image Data Collection. Available online: https://github.com/ieee8023/covid-chestxray-dataset (last referred on 10 May 2020).