Research Article

Diagnosis of COVID-19 using Optimized PCA based Local Binary Pattern Features

Malathy Jawahar[1], Anbarasi LJ[2], Prassanna J[2], Jackson CJ[2], Manikandan R[3], Alzubi JA[4], Dhanya D[5]

[1]Leather Process Technology Division, CSIR-Central Leather Research Institute, Adyar, Chennai 600020, India; [2]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600 127, India; [3]School of Computing, SASTRA Deemed University, Thanjavur, Indi; [4]Faculty of Engineering, Al-Balqa Applied University, Jordan; [5]Department of Computer Science, Mar Ephraem College of Engineering and Technology, Marthandam, India.

ABSTRACT

Introduction: COVID-19 is a pandemic disease affecting the global mankind since December 2019. Diagnosing COVID-19 using lung X-ray image is a great challenge faced by the entire world. Early detection helps the doctors to suggest suitable treatment and also helps speedy recovery of the patients. Advancements in the field of computer vision aid medical practitioners to predict and diagnosis disease accurately.

Objective: This study aims to analyze the chest X-ray for determining the presence of COVID-19 using machine learning algorithm.

Methods: Researchers propose various techniques using machine learning algorithms and deep learning approaches to detect COVID-19. However, obtaining an accurate solution using these AI techniques is the main challenge still remains open to researchers.

Results: This paper proposes a Local Binary Pattern technique to extract discriminant features for distinguishing COVID-19 disease using the X-ray images. The extracted features are given as input to various classifiers namely Random Forest (RF), Linear Discriminant Analysis (LDA), k-Nearest Neighbour (kNN), Classification and Regression Trees (CART), Support Vector Machine (SVM), Linear Regression (LR), and Multi-layer perceptron neural network (MLP). The proposed model has achieved an accuracy of 77.7% from Local Binary Pattern (LBP) features coupled with Random Forest classifier.

Conclusion: The proposed algorithm analyzed COVID X-ray images to classify the data in to COVID-19 or not. The features are extracted and are classified using machine learning algorithms. The model achieved high accuracy for linear binary pattern with random forest classifier.

Key Words: COVID-19, X-ray images (Lungs), Computer Vision, Machine Learning, Local Binary Pattern, Random Forest

INTRODUCTION

The infectious disease caused by coronavirus, subsequently named as COVID-19, discovered in the year 2019, started its outbreak in china but has now spread globally. The protein spikes on the surface of the coronavirus give the appearance of a crown and hence got the Latin name "corona" meaning "crown". The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus is the main cause of this newly discovered disease and are genetically similar to SARS Coronavirus (SARS-CoV) which was first identified in the year 2003. Coronaviruses like SARS-CoV-2, SARS-CoV, Middle East respiratory syndrome Coronavirus (MERS-CoV), were initially circulated among a range of animals and transmitted to humans due to an increase in contact between animals and humans or mutations in the virus. The disease can spread rapidly among humans through an infected person's respiratory droplets expelled during coughing, talking, or sneezing when close in contact with other persons. The severity of illness in an infected person varies from mild and moderate symptoms include fever, cough, sore throat, and tiredness to critical cases such as pneumonia, shortness of breath, organ failure, and death.[1] The complications caused due to this infectious disease are high for people with underlying medical problems such as chronic respiratory disease, diabetes, high blood pressure, heart diseases, or cancer, and older ones.

Corresponding Author:
Manikandan R, School of Computing, SASTRA Deemed University, Thanjavur, India.
Email: srmanimt75@gmail.com

The diagnosis of Coronavirus disease is currently done using a test called RT-PCR (Reverse Transcriptase Polymerase Chain Reaction) which help in detecting the virus's genetic material. There are also some antibody or serology tests available for checking the antibodies against the virus. Diagnostic tests such as RT-PCR are time-consuming and sometimes outputs false-negative results. Medicines against the virus and vaccines to prevent this disease are currently under development. Early diagnosis of this deadly infectious disease using a fast and accurate approach is crucial for controlling the current pandemic situation. Recent studies have reported that COVID-19 can be diagnosed using a chest computed tomography (CT) scan with the help of radiologists and can use as an initial screening technique to identify the infected patients.

Machine Learning algorithms and Deep Neural networks have achieved wide popularity in medical image processing over the past few years. This paper's main focus is to implement an automated and less time-consuming computer vision method to diagnose COVID-19 infected patients. The aim is to build a machine learning model capable of classifying infected patients' chest CT scan images from non-infected ones. The proposed model uses Local Binary Pattern (LBP) to extract features from the processed CT images and classify them accurately. The extracted LBP features are reduced using Principal Component Analysis (PCA). The reduced features are classified using various classifiers namely Random Forest (RF), Linear Discriminant Analysis (LDA), k-Nearest Neighbour (kNN), Classification and Regression Trees (CART), Support Vector Machine (SVM), Linear Regression (LR), and Multi-layer perceptron neural network(MLP).[2-4] The performance metrics such as confusion matrix, sensitivity, and specificity are evaluated to determine the proposed system's classification accuracy.

Tuncer et al.[5] presented an intelligent computer vision approach to automatically diagnose coronavirus using images of Lung X-rays of infected patients using machine learning classification algorithms such as linear discriminant (LD), support vector machine (SVM), subspace discriminant (SD), and k nearest neighbor (kNN), Decision tree (DT). Residual Exemplar Local Binary Pattern (ResExLBP) and iterative ReliefF (IRF) method were used for feature generation and selection. Ardakani et al.[1] suggested an efficient Artificial Intelligence technique for distinguishing 108 COVID-19 infected patients from pneumonia infected patients by taking 1020 CT slices. The most popular ten Convolutional deep neural networks were used for classification. Among them, ResNet-101 and Xception were achieved the best performance. Togacar et al.[6] proposed a deep learning model with a dataset that contains X-ray images of Coronavirus and pneumonia infected patients and normal non infected ones. The dataset was preprocessed using the Fuzzy color technique and the combining and classification of efficient features were done with the help of Support Vector Machines (SVM).

The proposed model is then trained using deep learning models such as MobileNetV2 and SqueezeNet. A fast and alternative screening approach using an efficient Convolutional deep neural network 'nCOVnet' was modelled by[7] for identifying the corona infected patients using a dataset containing X-ray images of infected ones.

The proposed model used a Convolutional neural network for feature extraction and classification of X-ray images of the chest. Tanvir et al.[8] designed an automated COVID-19 and other Pneumonia related disease identification method using images of infected patient's X-rays of the chest. A Convolutional based deep neural network architecture named "CovXNet "was used to extract the important features from images of X-rays and classifying them accordingly. Many CovXNet architecture forms were used for training different datasets and obtained a very satisfactory accuracy rate for all the classifications. CoroNet, a Convolutional deep neural network model uses images of X-rays of chest to diagnose coronavirus disease.[9] The popular Xception architecture is used as a base. The ImageNet dataset is used for pre-training the model and then trained with two publicly available datasets containing images of X-rays of corona infected patients' pneumonia infected ones. A three-fold approach to diagnose corona disease is from the images of X-rays of chest.[10] The initial model identifies the X-rays of patients infected from the normal ones. The second model classifies the images of X-rays of corona infected from Pneumonia. The third model provides a visualization of coronavirus infected images. A deep transfer-based learning technique was suggested for classifying the corona infected patients with Computed Tomography images.[11] Most important features from the CT images are extracted with the help of ResNet-50 architecture. Automated diagnosis of COVID-19 infected patients using a deep transfer learning-based approach is using chest X-rays and the proposed model is a modified Inception model.[4,12] The model performance is evaluated with various performance metrics and compared with various existing competitive models like VGGNet, ResNet, Alexnet, Googlenet, and Inceptionnet. Sousa et al[13] implemented a machine learning model to diagnose pneumonia in infants with the help of radiographic images. Machine learning algorithms such as K-Nearest Neighbour (KNN), Support Vector Machines (SVM), Naive Bayes, are used for image classification. The SVM classifier gave the best results the others. Gozes et al[14] developed Artificial Intelligence-based 2D and 3D deep learning models to automatically detect and track coronavirus disease from thoracic CT scan images.

PROPOSED WORK

Local Binary Pattern

Local Binary Pattern is one of the most popular and effectively used visual descriptors in the field of computer vision.

They can automatically detect and classify the patterns and textures from images and are widely used for various real-time applications such as remote sensing, texture classification, facial recognition, facial expression recognition, etc. Due to its discriminatory power and simplicity in computation. The proposed work uses the LBP texture descriptor for obtaining feature vectors by taking each target window of the image and is processed for extracting an LBP code. The initial step is to convert the original image into a greyscale image. The target window is divided into various cells and the center pixel value of each target window is compared with its neighbor's pixel value for computing a threshold. The value '0' is assigned to each neighbor's whose value is less than or equal to the center pixel and assigns "1" to those who is having a value greater than the center pixel. This results in a binary number, and the decimal value corresponding to this binary number is calculated for the center pixel and is then stored in an output LBP two-dimensional array. The same process is repeated for all pixels. The histogram is then computed based on all the pixel values stored in the LBP array, and finally, the normalized histograms are combined and summed to get the feature vectors from the input image. The LBP histogram can be calculated as:

$$His_i = \sum_{n,m} I\{G_i(n,m) = i\}, i = 0,1\ldots.n-1$$

where 'n' represents the LBP produced labels number. The value 'A' is true for I (A) = 0 and the value of 'A' is false for I (A) = 1.

The original implementation of LBP is restricted to extract features only from small structures and is limited to a fixed scale of 3x3 matrixes. In 2002, Ojala et al[15] proposed an approach that is an extension to the original LBP implementation and is capable of handling neighbourhood of variable sizes. This approach is considered a circular neighborhood instead of a square neighborhood for each pixel with varying boundaries consisting of data points 'P' and radius 'R' of the circle. The neighbourhood is represented using the notation (P, R). This work focuses on implementing a greyscale and rotation invariant-based LBP to identify the "uniform" patterns so that the length of the feature vector can be reduced. The LBP pattern obtained by thresholding containing a maximum of two 0-1 or 1-0 transitions is called "uniform" and those with more than two 0-1 or 1-0 transitions are non-uniform patterns. The histogram contains individual bins for each uniform pattern and all the non-uniform patterns are labelled to a single bin for computation.

Principle Component Analysis

Principle Component Analysis is considered as the most popular and commonly used unsupervised and dimensionality reduction algorithms. The main function PCA is to identify the correlation among variables. The algorithm helps to decrease the dimensionality of a larger dataset and project it onto a dimensionally smaller subspace without information loss. This approach makes the analysis and visualization of data easy and helps the machine learning algorithms process faster with a smaller dataset. PCA's wide applications include visualization, noise filtering, gene data analysis, feature extraction, etc.

The PCA algorithm's initial step is to standardize the data for transforming all the variables to the same scale. This helps in reducing the immense differences in the range of variables and overcomes biased results. Standardization can be done by taking the mean and standard deviation. The next step is to compute the covariance matrix to learn the correlation between the input dataset variables. The data's principal components can be determined from the eigenvectors and eigen[16-19] values computed using covariance matrix or single value decomposition. The eigenvalues obtained from the covariance matrix are sorted in descending order and choose the eigenvectors corresponding to the largest eigenvalues.[20-23] A projection matrix or feature vector is constructed from the selected eigenvectors, and the original dataset is transformed into a smaller dimensional feature space represented by principal components with the help of feature vector.[24]

Simulation Results

The overall flow structure of the proposed COVID-19 classification process is given in the Figure 1.
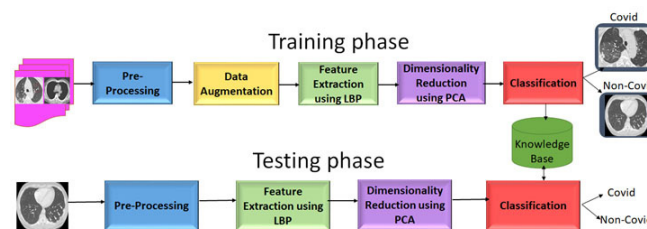

Figure 1: Proposed COVID-19 Classification System using X-ray images.

The initial step is to convert CT images of X-rays into greyscale. Processed images are shown in the Figure 2. The local binary pattern features are generated from the images. The features are reduced using the principal component analysis. The reduced features are classified using Classification and Regression Trees (CART), Linear Discriminant Analysis (LDA), Random Forest (RF), K- Nearest Neighbourhood (KNN), Support Vector Machine (SVM), Linear Regression (LR), and Multi-Layer Perceptron Neural based Network (MLP) for classifying given input images of X-ray into corona virus-infected and non-infected classes. Kaggle included CT images of Lung X-rays are taken, where there are 349 CT images belonging to 216 patients. Nearly 40,000 LBP features were generated from 746 images. An optimal number of components selected using in the PCA is 10. 596 CT images were

taken for training the model and 150 images were used to test the model. The accuracy achieved using LR, LDA is 66%, KNN is 75%, CART is 74%, RF is 77%, SVM 73% and MLP is 74%. Random Forest outperforms the classification accuracy. Figure 3 shows the Precision and Recall of Random Forest, figure 4 shows the confusion matrix and Figure 5 shows the LBP – Machine algorithm Comparison.
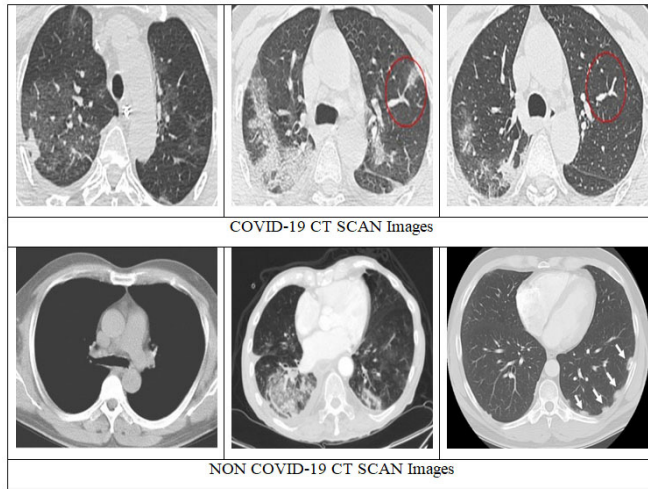

Figure 2: COVID-19& NONCOVID-19 Images.

Table 1: Classification accuracy of different classifiers.

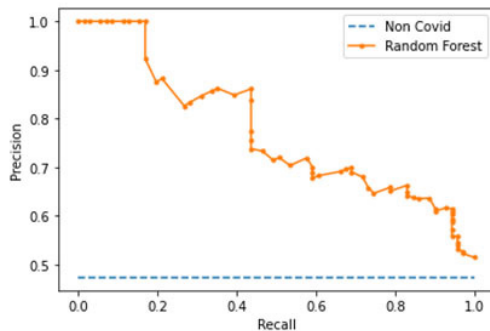| LR	LDA	KNN	CART	RF	SVM
0.663495 (0.051115)	0.662180 (0.055513)	0.752108 (0.034819)	0.741369 (0.065962)	0.773658 (0.063622)	0.738613 (0.047420)

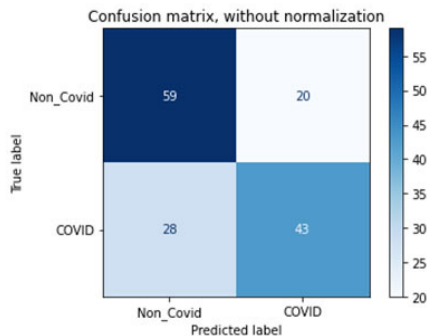**Figure 3:** Precision and Recall of Random Forest.
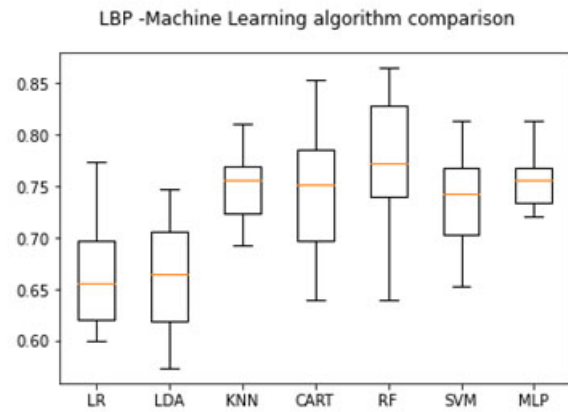

Figure 4: Confusion Matrix.

Figure 5: LBP – Machine algorithm Comparison.

CONCLUSION

A significant problem in the current scenario is classifying the COVID-19 disease. Identification of salient features that distinguish between the images of COVID and NON-COVID is an important task. A CT image-based machine learning approach with LBP and PCA is proposed to differentiate COVID-19 infection from normal chest lung x-ray image. In this study 40,000 local binary features were extracted from the COVID-19 and Non-COVID CT scan images. The optimal features are identified using the principal component analysis. The optimal features are trained with different classifiers namely LR, LDA, KNN, CART, RF, SVM and MLP. Random Forest classifier achieved an accuracy of 77% for the reduced features.

ACKNOWLEDGEMENT

REFERENCES

1. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Comp Bio Med 2020;121:103795.

2. Sharon JJ, Anbarasi LJ. Diagnosis of DCM and HCM Heart Diseases Using Neural Network Function. Int J Appl Engi Res 2018;13(10):8664-8668.

3. Markopoulos PP, Kundu S, Chamadia S, Pados DA. Efficient L1-norm principal-component analysis via bit flipping. IEEE Transact Sign Proces 2017;65(16):4252-4264.

4. Garcia P, Saxena S, Gupta A, "Frontal Sinus Variability as a Tool in Forensic Identification- A Pilot Study Using Radiographic Images and Software Analysis", Int J Cur Res Rev. Vol 11 Issue 08, April, 2019, 08-12.

5. Tuncer T, Dogan S, Ozyurt F. An automated Residual Exemplar Local Binary Pattern and iterative ReliefF based COVID-19 detection method using chest X-ray image. Chemometr Intel Lab Syst 2020;203:104054.

6. Toğaçar M, Ergen B, Cömert Z. COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. Comp Bio Med 2020;121:103805.

7. Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Singh V. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. Chaos Solitons Fractals. 2020;138:109944.

8. Mahmud T, Rahman MA, Fattah SA. CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. Computers in biology and medicine. 2020 Jul 1;122:103869.

9. Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Computer Methods and Programs in Biomedicine. 2020 Nov 1;196:105581.

10. Brunese L, Mercaldo F, Reginelli A, Santone A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. Computer Methods and Programs in Biomedicine. 2020 Nov 1;196:105608..

11. Pathak Y, Shukla PK, Tiwari A, Stalin S, Singh S. Deep transfer learning-based classification model for COVID-19 disease. IRBM. 2020 May 20.

12. Chouhan V, Singh SK, Khamparia A, Gupta D, Tiwari P, Moreira C, Damaševičius R, De Albuquerque VH. A novel transfer learning based approach for pneumonia detection in chest X-ray images. Applied Sciences. 2020 Jan;10(2):559.

13. Sousa RT, Marques O, Soares FA, Sene Jr II, de Oliveira LL, Spoto ES. Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs. Procedia Computer Science. 2013 Jan 1;18:2579-82.

14. Gozes O, Frid-Adar M, Greenspan H, Browning PD, Zhang H, Ji W, Bernheim A, Siegel E. Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. arXiv preprint arXiv:2003.05037. 2020 Mar 10.

15. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence. 2002 Aug 7;24(7):971-87.

16. Barnett TP, Preisendorfer R. Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. Monthly Weather Review. 1987 Sep 1;115(9):1825-50.

17. Hsu D, Kakade SM, Zhang T. A spectral algorithm for learning hidden Markov models. Journal of Computer and System Sciences. 2012 Sep 1;78(5):1460-80.

18. Anbarasi LJ, Mala A. EPR hidden medical image secret sharing using DNA cryptography. International Journal of Engineering and Technology. 2014;6:1346-56.

19. Chachlakis DG, Prater-Bennette A, Markopoulos PP. L1-norm tucker tensor decomposition. IEEE Access. 2019 Nov 22;7:178454-65.

20. Markopoulos PP, Karystinos GN, Pados DA. Optimal algorithms for $ L_{1} $-subspace signal processing. IEEE Transactions on Signal Processing. 2014 Jul 10;62(19):5046-58.

21. Ke Q, Kanade T. Robust l/sub 1/norm factorization in the presence of outliers and missing data by alternative convex programming. In2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 2005 Jun 20 (Vol. 1, pp. 739-746). IEEE.

22. Prasanna J, Venkataraman N. Adaptive regressive holt–winters workload prediction and firefly optimized lottery scheduling for load balancing in cloud. Wireless Networks. 2019 Jul 15:1-9.

23. Prasanna J, Venkataraman N. Threshold based multi-objective memetic optimized round Robin scheduling for resource-efficient load balancing in cloud. Mobile Networks and Applications. 2019 Aug;24(4):1214-25.

24. Farzaneh MJK, Afzalipour R, Vardan M, Shandiz MS, Zarei M. Study of the reasons of the radiographic images repetition in Sistan and baluchestan's treatment centers. Int J Curr Res Rev 2012;04(08):185-188.