JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AN INTELLIGENT SYSTEM FOR THE PREDICTION OF COVID -19

Dr Julie M David [1]MES College Marampally Kerala,India
Neenu[2] ,Riya Azeez [3] [1]MES College Marampally Kerala,India

Abstract - The Novel Corona Virus 2019 is a respiratory problem that similar to pneumonia. It 's one of the most infectious diseases in the 21st century. This paper proposes the methodology is based on a hybrid model which combines the convolutional neural networks techniques and random forest model with extrapolations based on recurrent neural networks. This combination provides understandable results in relation of a coefficient that varies with the limitation measures, which may be further refined by expert rules that capture the expected changes in such measures. Random forest will help to reduces overfitting and helps to improve the accuracy. We propose an improved hybrid classification approach for COVID-19 images by combining the strengths of CNN to extract features to select the most relevant features. The suggested method was tested on two publicly available COVID-19 X-ray datasets, and it achieved great performance while reducing computing complexity.

Keywords:Covid-19,Convolutional network, Random forest.

1. Introduction

Coronavirus disease (COVID-19) is a contagious disease. It is caused by the SARS-CoV-2 virus. In December 2019, the virus was first recognized in Wuhan, China [1]. Over 198 million people have been affected with COVID-19 to globally, with more than 4.2 million losses of life. The affected peoples and death cases is quietly increasing on a daily basis which is aggressive on our society and health field [2]. Since it is a highly contagious disease, To manage the expansion of the disease, the most desirable preventive method is to identify patients. Therefore, the early diagnosis is crucial for correct treatment in order to possibly reduce the stress in the health field. Time consuming clinical testing is one among the crucial element behind the quick expansion of the corona virus [4].

The random forest is a good classification algorithm entailing of many decision trees. When creating each individual tree, it employs bagging and feature randomization in an attempt to generate an uncorrelated forest of trees whose committee prediction is more accurate than any one tree. Random forest is great with high dimensional data as we were working with subsets of data. To easily work with hundreds of features it is faster to train than decision trees as we are working only in a subset of features in our model.

COVID-19 diagnosed utilizing always one among the three tests in most cases. Assessment based on Reverse Transcription Polymerase Chain Reaction test, Chest X-Ray and Computed Tomography images [8].

Among the three ways of diagnosis, more robust is CXR because it is cheaper, quicker and widespread. Therefore, the aim of this project is to evolve an application for classifying COVID-19 affected lungs and healthy lungs using only CXR images. In machine learning, classification refers to a predictive modeling problem where a class label is predicted for a given example of input

This study will classify the dataset into two. Healthy Lungs and Covid-19. The system proposed involves pre-processing, feature extraction, and classification. For feature extraction, Convolution Neural Network is used. Three distinct Machine Learning techniques, such as CNN, Decision Tree, and Random Forest, are utilized to efficiently categorize COVID-19 and normal chest X-ray pictures during the classification step. Here I also compare classification using these three algorithms with and without feature extraction and the algorithm which provides the best results are used in the application.

2.Literature Review

Machine (SVM), and Convolutional Neural Network (CNN) approaches are occupied in T Siswantining et al suggested system[2].After that an ensemble learning mechanism called stacking is applied. The Stacking model, with a 95 percent accuracy, provides the greatest results.

Shui-Hua Wang et al proposed an artificial intelligence-based tool called FGCNet model to diagnose chest CT pictures. Cnn is used to study individual image-level presentations. The cnn employs rank-based average pooling and multiple-way data augmention. Relationaware representations were learned from graph convolutional networks (GCN). To fuse individual image level features and relation-aware features from both GCN and CNN,

Deep feature fusion(DFF) was designed, which improves performance.

VaralakshmiPerumal et al. [3] presented a transfer acquiring method. Texture feature extraction is performed utilizing Haralick features to identify COVID-19 by statistical analysis. The proposed transfer learning model has a 93 percent accuracy rate.

El-Sayed M. El-kenawy et al suggested different feature selection and classification algorithms for COVID-19. AlexNet, a Convolutional Neural Network (CNN), is primarily used to acquire characters. The Guided Whale Optimization task is adopted to pick features(Guided WOA).To make predictions, a suggested voting classifier, guided WOA based on Particle Swam Optimization is used. The system has a 95% accuracy rate. Şaban ÖZTÜRK et al proposed a technique based on optimization of convolutional neural network (CNN) architecture to detect COVID-19 from CT images. COVID-19 images were trained using ResNet-50 and VGG-16 architectures. After that, the features from Feature fusion are used to integrate the last layer of these two designs. For COVID detection, these novel image attributes are categorized. For the classification process, a multi-layer perceptron (MLP) structure optimized by the whale optimization algorithm is used and 88.06% accuracy is achieved.

Rohith N Reddy et al. [1] proposed a method to detect the COVID-19 using chest X-Ray pictures. A CAD system is created using a Support Vector Machine (SVM) Classifier for classifying the common and COVID-19 cases. The present system is being trained and, trial on a small image subset of COVID-19 cases. An accuracy of 57.1% in classifying the COVID-19 images is obtained.

Muhammad Imad et al proposed a method to efficiently classify the COVID-19 infected patients based on chest X-ray radiography using Machine Learning techniques. To begin, the image is pre-processed to boost the contrast . The Histogram of Oriented Gradients (HOG) is used to extract the discriminant characteristics. The method of classification employs Machine Learning methods like Support Vector Machine (SVM), K-Nearest Neighbors, Random Forest, Naïve Bayes algorithm, and Decision Tree. SVM has the precision, with a 96 percent accuracy rate.

To expand a quick identification method for COVID-19, Jose David Bermudez Castro et al. explained a method that uses deep learning (DL) approaches utilizing X-ray pictures.

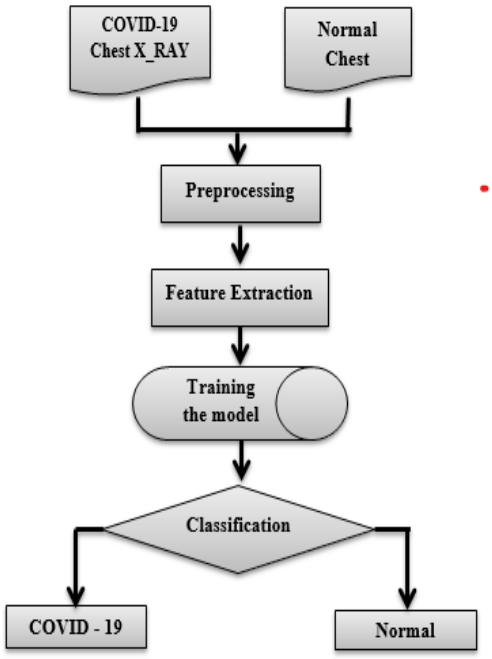
In the newly proposed technique, two deep learning technologies are used. X-Ray and non-X-Ray pictures formulated on Mobile-Net architecture. The second one is for identifying chest X-Ray images with characteristics of COVID-19 depending on the Dense Net architecture. The system achieved 99.3% accuracy.

MdMamunurRahamana et al. proposed an automated CAD technique designed for classifying COVID-19 samples from healthy and pneumonia cases with CXR pictures. To identify the most suitable one, 15 different pre-formed CNN models were examined. 89.3% accuracy is obtained.

2. Proposed Methodology

pre-processing, feature extraction and classification are all coming us part of our suggested hybrid system. Convolution Neural Network is being used for feature extraction. A hybrid intelligent system incorporates multiple intelligent technologies. In this classification step, three distinct Machine Learning algorithms, such as CNN, Decision Tree and Random Forest, are utilized to efficiently categorize COVID-19 from normal chest X-ray pictures. For preprocessing, the withdrawal of important features is attained by a convolutional neural network . A correlation of classifiers like CNN , random forest and decision tree is also done. Feature extraction with and without different metrics such as accuracy, precision, recall, specificity and F1 score are utilized to examine the outputs. The examination of the results reveals that CNN has the most categorization accuracy

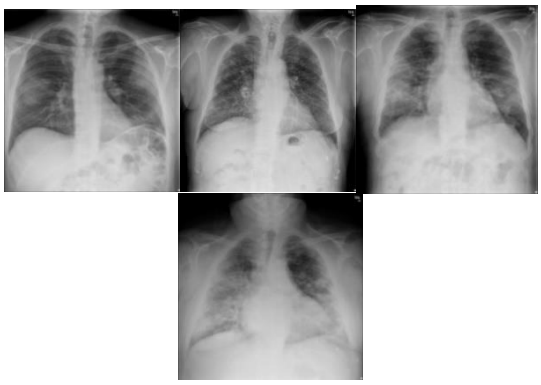
2.1. Proposed Framework



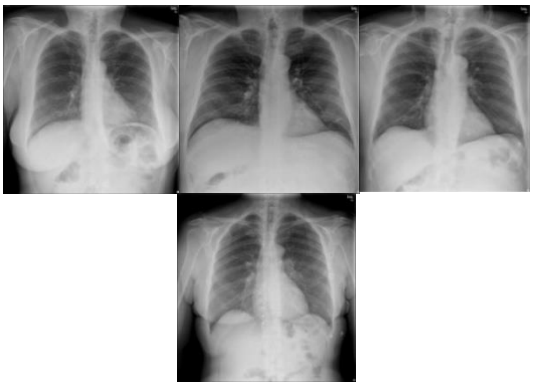
2.2. Dataset

In this proposed system we are using a balanced and homogeneous database. It contains 852 PA (Posterior-Anterior) CXR images. In this, 426 are Covid-19 positive and 426 are Covid-19 negative. From these 396 positives and 396 negative CXR images are taken as the train dataset. The remaining 30 positive and 30 negative CXR images are taken as the test dataset. The figures for training and testing datasets are properly labeled with negative or positive.

Some original samples from the dataset



X-RAY image of COVID-19 affected victim



X-RAY image of normal person

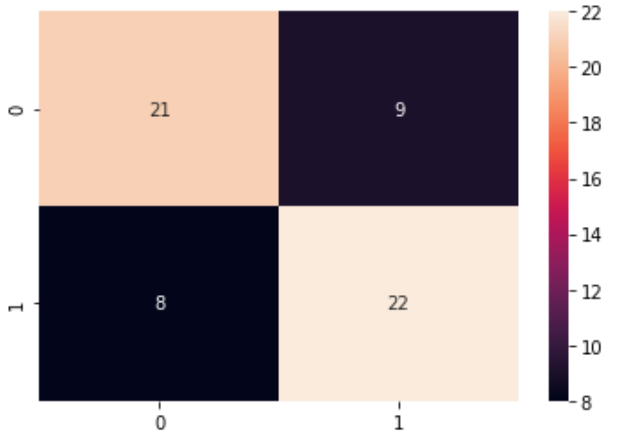
2.3. Preprocessing

The preprocessing stage prepares the data to the model, making picture analysis and processing easier. Images are preprocessed with the specific technique in a context that input images to the model will be easy to process and we can extract the best features from it. The preprocessing consists of normalize the images ,image cropping and converting all image size into standard ratio. The covid CXR pictures in the dataset are in JPEG format. The image size is ranging from 508 × 500 to 4248 × 3480 pixels. All the pictures are in gray scale. Keep the goal size to 128 by 128 pixels for the experimental setup. The "preprocess input" function in keras is used to alter the input image to meet the model's requirement. Keras Preprocessing is the data preprocessing and data augmentation module of the Keras deep learning library. It provides utilities for working with image data, text data, and sequence data.

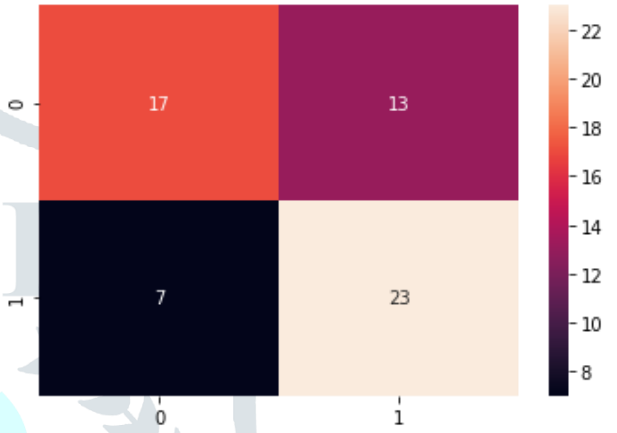
2.4. Feature Extraction

Feature Extraction is the method of extracting new characteristics from the existing ones to decrease the features of a dataset. It is a dimensionality reduction procedure. These extracted features are simple to analyze, But they can accurately and uniquely characteristic the actual data set.

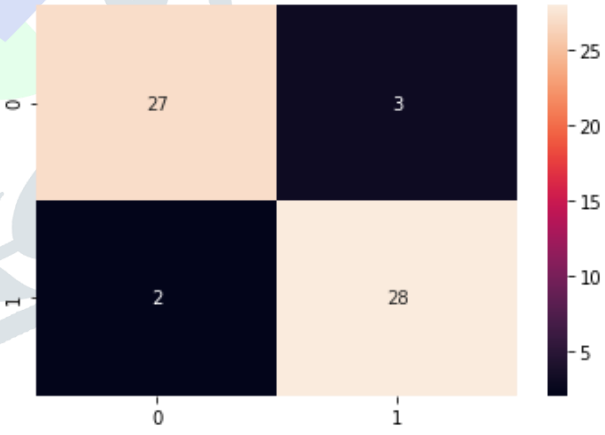
Machines perceive visuals as a matrix of numbers. This matrix's size is determined by the amount of pixels in the supplied image. Pixel values represent the pixel's brightness or intensity. An image's number of features will be the same as its quantity of pixels. Therefore, the count of pixels must be lowered to minimize the features. Here feature extraction is achieved by Convolution Neural Network. The concept used here is Transfer learning. Transfer learning is a machine learning method where an existing model is reused for a second task. Feature selection is the primary method to increase the performance neural network because of the unnecessary attributes and the high amount of original data sets..



Random forest(with feature extraction)



Random forest
(without feature extraction)



CNN with feature extraction

2.5. Classification

Classification is a data mining task that allocate items in a collection to target categories or classes. The aim of classification is to accurately predict the target class for each cases in the data. A classification is an ordered set of related categories used to group data according to its similarities. It includes codes and descriptions, and permits survey feedbacks to develop into meaningful categories in to produce useful data.

Following the process of extraction of feature, classification is done using several machine learning

methods such as Convolution Neural Network (CNN), Decision Tree, and Random Forest.

2.5.1. Convolution Neural Network

A deep learning artificial neural network is a convolutional neural network or CNN . using this method images can be recognized and processed. It is a technique for picture classification and image recognition in neural networks that is specifically intended to process pixel data. CNN uses numerous layers of arrays to process the data. CNN is utilized in image recognition and facial recognition applications. The three layers of Convolutional Neural Networks are Convolutional, Pooling, and Fully Connected. Convolution layer is the initial layer . It extracts characteristics of the input image. The convolutional layer preserves the link connecting pixels by identifying visual attributes with a small square of input data. It's a mathematical procedure with two inputs: image matrix and kernel/any filter. when the image is too huge , the second layer, the pooling layer minimizes the count of parameter. Pooling is the process of downscaling an image obtained from previous layers. The pooling layer reduces the density of a picture by shrinking it. The fully connected layer, the third, is a layer that combines and compresses the input from other layers in to a vector. It will divide the output into as many classes as we want in the network.

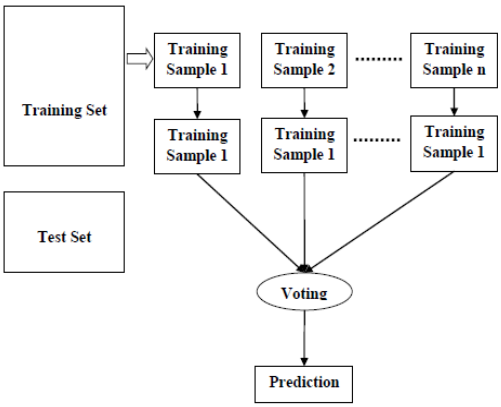
Transfer learning is a machine learning method in which an existing model which has been trained on one job is utilized on a second task that is linked to it. By keeping the weights in more than one layers fixed, fine-tuning them, or completely altering the weights when training the new model, the previous model can be reused. A neural network model 's training time is cut in half with transfer learning. It also cuts on error. [6,9] Several CNN models exist. VGG16 stands for Visual Geometry Group and is being used as the procedure for the CNN method for classification. VGG as the convolutional network is utilized for face recognition and picture categorization . The RGB image of 224x244 is being used as the VGG's input. The average RGB value for all photos while training set is determined, and the image is then fed in to the VGG convolution network.

The pre-trained VGG-16 model is being used in the experimental design procedure. The input layer input image is (128,128,3). There are 13 convolution layers and 5 pooling layers of the model. In the generated model, the activation function is sigmoid. The optimizer ADAM is used. The loss function utilized by the optimizer is binary cross entropy. The epoch is set to 30 after the data has been organized in the experimental process. In the back-propagation procedure, we find the loss function value together.

2.5.2 .Random Forest

To classify and predict data, a supervised learning method called random forest is used. The Random forest method constructs decision trees for each data sample, extracts prediction from each, and votes on the best answer. It is a whole method that is higher level to a single decision tree since it averages the result to reduce overfitting. The procedure entails choosing random specimens from a dataset, building a decision tree for each sample, getting the prediction output from the decision tree, voting and selecting the highest opted prediction output. Random forests works by training and resulting the class, that is the technique of classes product by individual trees while generating numerous decision trees. It outperforms single decision trees and is far more efficient than typical machine learning algorithms, particularly when dealing with enormous datasets. It has the capacity to handle thousands of explanatory variables. It can describe complex interactions among variables and is convenient for displaying nonlinear effects of variables. Outliers are hardly a problem for random forest.

The diagram below will demonstrate how its works.



2.6. Performance Evaluation

In machine learning, the confusion matrix is most typically used to compute the precision of a classification model. In a classification problem, the count of accurate and incorrect results are tallied and the result is correlated to the reference data. Accuracy, Precision, Recall, Specificity, and F1-score are few of the usual matrices. To solve the confusion matrix, four statistical indices have been used that is a true positive (TP), true negative (TN), false positive (FP), false negative (FN) were calculated, which is given in equation (2) to (5).True Positive (TP) refers to predicted yes that they have infected by COVID-19, True Negative(TN) refer to predicted no, that they don't have infected by COVID-19, False Positive(FP) refer to predicted yes, but they don't have infected by COVID-19, False Negative (FN) refer to predicted yes, but they have actually infected by COVID-19.

- Accuracy =(TN+TP)/TN+TP+FN+FP (2)
- Precision = (TP)/(TP+FP) (3)
- Recall = (TP)/(TP+FN) (4)
- F1Score=2∗(Precision∗Recall)/(Precision+Recall) (5)

Accuracy indicated that how frequent the classification is correct, while precision indicates how often it will correct classify during prediction. However, recall indicates that from the all positive class how much predicted correctly while specificity evaluates that the negatives that are identified correctly. The F1-score indicates the mean of precision and recall

	Accuracy	Precision	Recall	F1- Score
CNN	0.92	0.93	0.90	0.91
Random Forest	0.72	0.73	0.71	0.72
Decision Tree	0.70	0.65	0.83	0.73
Random Forest (without feature extraction)	0.67	0.70	0.70	0.69
Decision Tree (without feature extraction)	0.66	0.67	0.63	0.65

The total conduct of the process is represented by the confusion matrix. The confusion matrix for each algoritham can be found below

Following the feature extraction process, classification is performed using separate machine learning methods like Convolution Neural Network (CNN), Decision Tree, and Random Forest.

2.7. Result analysis

A chest X-ray dataset was used to forecast coronavirus (COVID-19) affected victims and normal victims in this study. Several machine learning algorithms is been used. The dataset has been trained and tested on chest X-Ray images depending on machine learning algorithms were utilized. Machine learning methods like Convolution Neural Network (CNN), Decision Tree, and Random Forest [3] are used to instruct and fix the dataset on chest X-ray pictures. Before and after feature extraction, these algorithms are implemented to the dataset. Table 1 displays the output of the various machine algorithms. The outputs implies that when classification is done after feature extraction, better results are produced. The outputs are extremely poor without feature extraction. The CNN model delivered the highest accuracy of 92 percent. Random forest and decision tree produce similar results. Table 1 shows that, when correlated to other machine learning models, the CNN received better prediction results using several metric measures like accuracy, precision, recall and f1 score. Table 1. The following are four performance results acquired using several machine learning algorithms.

Table 1 . Performance results are obtained from several Machine Learning algorithm.

	Accuracy	Precision	Recall	F1- Score
CNN	0.92	0.93	0.90	0.91
Random Forest	0.72	0.73	0.71	0.72
Decision Tree	0.70	0.65	0.83	0.73
Random Forest (without feature extraction)	0.67	0.70	0.70	0.69
Decision Tree (without feature extraction)	0.66	0.67	0.63	0.65

The total conduct of the technique is considered by the confusion matrix. The following is the confusion matrix for each algorithm.

3. Conclusion & future scope

The COVID-19 coronavirus infection is exceedingly contagious, putting the lives of thousands of people in danger. Early diagnosis is critical in the cure of COVID-19, that should be quicker and less expensive. In this analysis, we utilized the advantages of utilizing chest x-ray figures, multiple machine learning techniques were utilized to automatically identify between the COVID-19 affected victims and usual chest X-ray pictures. This research focuses on x-ray pictures, which are easier to obtain, more widely available around the globe, and less expensive. We discovered that CNN is the best option for classification and feature extraction when using CNN and Random. Accuracy, precision, recall and F1 score are used to evaluate each model. F1 score. The output of the various methodologies demonstrate that CNN outperformed the other machine learning techniques in aspects of accuracy.

This technique gives as a better option to the problem of how beneficial chest X-Rays are for diagnosing COVID – 19. It will assist doctors in making high- quality decision in clinical field. If a posterior-anterior chest X-Ray is taken, this model may efficiently distinguish COVID-19 positive lungs from normal lungs. This can be really useful in diagnosing COVID-19 with technology.

4. References

[1] Rohith N Reddy. “COVID-19 Detection using SVM Classifier”

[2] T Siswantining , R Parlindungan. “Covid-19 classification using X-Ray imaging with ensemble learning”

[3] VaralakshmiPerumal ,Vasumathi Narayanan ,Sakthi Jaya SundarRajasekar. “Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features”

[4] Tao Ai, Zhenlu yang, Hongyan hou, Chenao Zhan, Chong Chen, Wenzhi LV, Qian Tao, Ziyong Sun “Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases”

[5] Yoo S.H., Geng H., Chiu T.L., Yu S.K., Cho D.C., H eo J., Choi M.S., Choi I.H., CungVan C., Nhung N.V ., Min B.J., Lee H. “Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging”.

[6] Sethy P.K., Behera S.K., Ratha P.K., Biswas P. “Detection of coronavirus disease (COVID-19) based on deep features and support vector machine”

[7] Minaee S., Kafieh R., Sonka M., Yazdani S., Jamalip our Soufi G. “Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning”.

[8] Khan A.I., Shah J.L., Bhat M.M. “CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images”

[9] Jain G., Mittal D., Thakur D., Mittal M.K. “A deep learning approach to detect Covid-19 coronavirus with X-Ray images”.

wikipedia
https://www.mygreatlearning.com/blog/feature-extraction-in-image-processing/
https://www.javatpoint.com/convolutional-neural-network-in-tensorflow
https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2