Journal of Harbin Engineering University Vol 45 No. 4
ISSN: 1006-7043 April 2024

Prediction of Covid and Pneunomia Using Ensemble Based Machine Learning Approach

T.Sundaravadivel,
Assistant Professor/Programmer, Department of Information Technology, FEAT, Annamalai University, Annamalai Nagar, Tamilnadu

Abstract
Since the release of the novel Covid-19, numerous research projects have been launched worldwide in an effort to accurately forecast it. Since many Covid-19 patients passed away from severe chest congestion, the previous lung illness pneumonia is intimately linked to the latter (pneumonic condition). Even for medical professionals, it might be difficult to distinguish between pneumonia and Covid-19 lung illnesses. The most accurate technique for predicting lung disease is chest X-ray imaging. In this proposed work, the dataset are collected from kaggle. The primary goal is to combine several classifiers to get superior performance over each classifier working alone. In this work local binary pattern are used as the feature extractor and then combines two distinct machine learning models—random forest and KNN classifiers—before conducting training and testing to provide the desired outcomes using stacking which gives the highest accuracy of 82%.
Keywords - KNN (KNN – K Nearest Neighbour), Random Forest, LBP (Local Binary Pattern),

Introduction
All respiratory and lung diseases and disorders that impact human respiration [1]. All breathing-related structures and organs, such as the nasal cavity, throat, larynx, trachea, bronchi, and bronchioles, as well as lung tissue and the respiratory muscles in the chest, can be impacted by respiratory disorders.
The emergence of COVID-19, which originated in China in November 2019 and eventually spread throughout the world, has posed a serious threat to human existence. According to reports, the global infection count now stands at over 63.2 million, with an estimated 1.47 million fatalities. The World Health Organization (WHO) consistently gives countries the knowledge they need to defend against COVID-19[2].
If left untreated, pneumonia is an infection that inflames the lethal. The most common way to identify pneumonia, and it needs to be carefully examined by a specialist. Coughing and dyspnea can result from inflammation of the lungs caused by pneumonia, particularly the alveoli, which swell with fluid or pus [5]. A bacterial or viral lung infection can result in pneumonia, a potentially fatal illness. An early diagnosis of pneumonia is crucial since it can be fatal if treatment is delayed. [6].


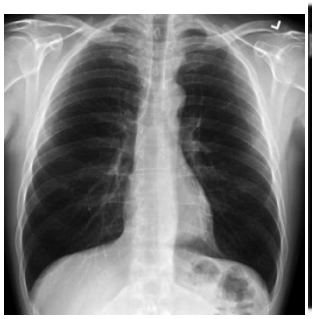


Fig 1. COVID 19 Fig 2. Normal Image Fig 3. Pneunomia Image

Fever, coughing, and dyspnea are among the mild to moderate symptoms seen by Covid-19 infected individuals. However, serious pneumonic infections in their lungs also claimed the lives of other people [3]. Due to a significant decline in oxygen levels, the majority of Covid-19 fatalities had significant chest congestion (pneumonia), which led to a devastating heart attack [4].

Nevertheless, pneumonia is an additional form of pulmonary disease that results in inflammation within the human lung's tiny air sacs.

Literature Survey

To assess [7]InceptionResnet_v2, Resnet50, and MobileNet_v2 are optimized for the classification of pneumonia based on the performance of individual and group learning models. A dataset including 6087 chest x-ray pictures was gathered. Ultimately, a single model was identified wherein InceptionResNet_V2 yielded 93.52% of the F1 score. Furthermore, an ensemble consisting of ResNet50, MobileNet_V2, and InceptionResNet_V2 produced good results.

In this work, a dependable and comprehensible deep learning model capable of very accurate detection of such anomalies is developed [8]. This study suggests using chest X-ray pictures to automatically diagnose pneumonia in children using a neural network-based diagnostic tool. A deep convolutional neural network model with transfer learning approach was suggested for this procedure. Pre-processing is done using techniques including filtering, gamma correction, equalization, and compression before sending the image to the model. The accuracy, recall, precision, and ROC scores of the suggested model are compared to those of ResNet, ImageNet, Xception, and Inception. A conventional X-ray dataset that was gathered from the Women's and Children's Medical Center was used for the experiment. Our suggested algorithm outperforms other traditional algorithms, as demonstrated by the experimental findings.

FEATURE EXTRACTION

Local Binary Pattern

The LBP approach involves these steps.

Intially the image has to be converted into grayscale. It works with pixels in a fixed 3 × 3 neighbourhood. The central pixel is taken into consideration, and its limit is set against its encompassing eight pixels.

On the off chance that the center pixel's escalated is more prominent than or break even with to that of its neighbor, we keep the esteem at 1 or set it to 1. TheLBPvalueofthecenterpixeliscalculated clockwise or counterclockwise [9].Binary tests are performed on 8 neighbors in a 3 x 3 neighborhood. Binary test results are stored in an 8-bit array and converted to decimal values. If there is no transition from LBP to 0-1 or 1-0, it is considered. LBP is defined as no more than 0-1 or 1-0 transitions. For each pixel in the input image, the thresholding, storing of the values point by point in the LBP sequence routine are repeated. Then compute the output LBP sequence and histogram. There are 256 different patterns in a 3 x 3 area, so the minimum is 0 and the maximum is 255.

As a result, we generate a histogram of 256 bins of LBP.

Finally, the output array contains the calculated values.

$$LBP(P,R)=\sum_{p=0}^{p-1} f\left(g_p - g_c\right)2^p$$

P is the number of neighboring pixels specified at radius R. Here gp and gc represent the pixel power of the current pixel and neighboring pixels.

Random Forest Algorithm

Decision trees acting individually and as an ensemble form a random forest [10].The class thatreceives the highest votes determines the model'spredictions in a random forest where individual trees share class predictions.Integratingalargenumber of uncorrelated trees toworktogether to provide aprediction is the main principle of random forests. To prevent the behavior of onetreefrombecomingtoolinkedto the behaviorof other trees in the model, RandomForest uses her two tactics [11].

KNN – K Nearest Neighbour

KNN is a popular machine learning technique that mostly deals with problems related to regression and classification. KNN uses the similarity metric of the distance function to classify datasets. The fields of pattern recognition, intrusion detection, image processing, and data mining can all benefit from the application of this non-parametric supervised learning technique. The data in KNN is classified based on the majority of votes cast for the nearest neighbor [12]. The number of neighbors has the most influence on the KNN classification process. Furthermore, because the

method uses all of the training data during testing, it does not require specific data points to generate the training model. Using the KNN technique, the k nearest data points are selected, and the points are then categorized based on which of the k neighbors won the majority of the votes.

STACKING

Clustering methods integrate various classification models using meta tags. Every model communicates its predictions to the top layer, which bases its judgments on the forecasts of the lower layers. There are several layers above [13]. The smallest layer model is assigned the input features of the initial data set. The output of the lower layer is used by the upper layer model to make predictions. During the collection process, the raw data is put into various models. After calculating each model and its weights, we use meta-classification to estimate each model and its inputs and outputs. The best model is selected and the remaining models are discarded.

PERFORMANCE MEASURES

To asses the performance of Ensemble classification techniques of Random Forest and KNN using stacking technique a set of metrics namely Accuracy, Precision, Recall and F-Score are used in this work.

EXPERIMENTAL RESULTS

DATASETS

A total 251 x-ray images were collected from kaggle datasets in which 80% of images were used for training and 20% were used testing. 111 images of COVID, 70 images of normal and 70 images of pneumonia

.

Performance of Stacking

Classification Report for stacking technique

Inputs	**Precision (in %)**	**Recall (in %)**	**F- Score (in %)**	**Accuracy (in %)**
Normal	55	54	44	82
COVID 19	74	71	72	
Pneunomia	89	95	90	

Fig 1: Classification of Stacking Technique

Conclusion

In this work the performance of COVID 19 is analyzed using the ensemble classification of two machine learning models namely KNN and Random Forest techniques which used using stacking gives the highest accuracy of 82%.

References

1. Mitton, B., Rule, R., & Said, M, "Laboratory evaluation of the BioFire Film Array Pneumonia plus panel compared to conventional methods for the identification of bacteria in lower respiratory tract specimens: a prospective cross-sectional study from South Africa. Diagnostic Microbiology and Infectious Disease", Vol 99(2), 2021.
2. Goyal, S., & Singh, R. (2023). Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques. Journal of Ambient Intelligence and Humanized Computing, 14(4), 3239-3259.
3. Padma, T., & Kumari, C. U. (2020, September). Deep learning based chest x-ray image as a diagnostic tool for covid-19. In 2020 international conference on smart electronics and communication (ICOSEC) (pp. 589-592). IEEE.
4. Chen, P., Nirula, A., Heller, B., Gottlieb, R. L., Boscia, J., Morris, J., ... &Skovronsky, D. M. (2021). SARS-CoV-2 neutralizing antibody LY-CoV555 in outpatients with Covid-19. New England Journal of Medicine, 384(3), 229-237.

5. Harsh Sharma, Jai Sethia Jain, Priti Bansal, Sumit Gupta, “Feature Extractrion and Classification of Chest X-ray images using CNN to detect Pneumonia”, 10th International Conference on Cloud Computing, Data Science and Engineering, IEEE, 2020.
6. Tawsifur Rahman , Muhammad E. H. Chowdhury, Amith Khandakar , Khandaker R. Islam , Khandaker F. Islam , Zaid B. Mahbub , Muhammad A. Kadir and Saad Kashem, “Transfer Learning with Deep Conventional Neural Network (CNN) for Pneumonia detection using Chest X-ray ”, Applied Sciences, Vol 10, pp 1-17, 2020.
7. El Asnaoui, K. (2021). Design ensemble deep learning model for pneumonia disease classification. International Journal of Multimedia Information Retrieval, 10(1), 55-68.
8. Chhikara, P., Singh, P., Gupta, P., & Bhatia, T. (2020). Deep convolutional neural network with transfer learning for detecting pneumonia on chest X-rays. In Advances in Bioinformatics, Multimedia, and Electronics Circuits and Signals: Proceedings of GUCON 2019 (pp. 155-168). Springer Singapore.
9. Poola, R. G., Lahari, P. L., &Yellampalli, S. S. (2023, June). Optimizing Pneumonia Diagnosis through Local Binary Pattern and 2D-Wavelet Transform Based Feature Extraction and Classification. In International Conference on Advanced Communication and Intelligent Systems (pp. 127-139). Cham: Springer Nature Switzerland.
10. Basavaraju, N. S., &Ganesarathinam, S. (2021). Early Detection of Diabetic Retinopathy Using K-means Clustering Algorithm and Ensemble Classification Approach. International Journal of Intelligent Engineering & Systems, 14(6).
11. Alshayeji, M. H., & Sindhu, S. C. (2023). Two-stage framework for diabetic retinopathy diagnosis and disease stage screening with ensemble learning. Expert Systems with Applications, 225, 120206.
12. Piri, S., Delen, D., Liu, T., &Zolbanin, H. M. (2017). A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble. Decision Support Systems, 101, 12-27.
13. Indirani, G., et al. "Prediction of Early Heart Attack for Post-COVID-19 Patients Using IoT Sensors and Machine Learning." Clinical Practice and Post-Infection Care for COVID-19 Patients. IGI Global, 2024. 190-206.