

Impact Of Machine Learning Models In Pneumonia Diagnosis With Features Extracted From Chest X-Rays Using VGG16

G. Jignesh Chowdary^a, G. Suganya^{b*}, M. Premalatha^c, S. Ganapathy^d

^aSchool of Computer Science and Engineering, Vellore Institute of Technology Chennai, Chennai, Tamil Nadu, India.

^{b*}School of Computer Science and Engineering, Vellore Institute of Technology Chennai, Chennai, Tamilnadu, India. E-mail: suganya.g@vit.ac.in

^cSchool of Computer Science and Engineering, Vellore Institute of Technology Chennai, Chennai, Tamilnadu, India.

^dSchool of Computer Science and Engineering, Vellore Institute of Technology Chennai, Chennai, Tamilnadu, India.

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: Pneumonia is a viral, bacterial, or fungal infection that leads to the accumulation of pus or fluids in the alveoli of lungs causing breathlessness, lung abscess, or even death at later stages. Pneumonia is affecting a huge population across the globe. A quite large number of child deaths due to pneumonia are recorded which is significantly greater than death due to AIDS, malaria, and measles. Pneumonia diagnosis is considered one of the high priority research areas in Biomedicine. In this paper, a detailed comparative study was performed using various machine learning algorithms namely Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM). These models are trained with features extracted by a pre-trained deep convolutional neural network (DCNN), VGG16 for the diagnosis of pneumonia from chest x-rays. The combination of VGG16 along with Machine learning models witnessed a considerable improvement in accuracy with reduction in time consumed for training against the usage of DCNN models for prediction. The results of various machine learning models are fine-tuned by modifying the hyper parameters. By comparison, SVM with RBF kernel is identified to perform better than other classifiers.

Keywords: Pneumonia Diagnosis, VGG16, Feature Extraction, Support Vector Machine, Random Forest, Logistic Regression.

1. Introduction

Pneumonia is a lung infection caused due to **the accumulation of fluids or pus in the alveoli (air sacs) of the lungs**. Alveoli are the structures in the lungs where the exchange of carbon dioxide and oxygen takes place during the process of exhalation and inhalation. Accumulation of pus or fluids in alveoli causes difficulty in breathing, chest pain, headaches, fatigue, vomiting, fever, and loss of appetite. Pneumonia is caused due to microorganisms like viruses, bacteria, or fungi. Bacterial and viral pneumonia is contagious, as they can spread from person to person. There are more than 250000 people hospitalized in the United States (US) due to pneumonia out of which 50,000 thousand of them die each year [1]. The child mortality rate of pneumonia is greater than the sum of malaria, AIDS and measles [Rudan et-al 2008] [Adegbola RA]. According to a study, it was estimated that around 17 lakhs of children will die by 2030 due to pneumonia [4] and also stated that around 4 million people could have saved if a proper diagnosis was present. Thus an accurate and précised diagnosis is more important to reduce the death toll due to pneumonia. Pneumonia can be diagnosed with chest x-rays, blood tests, bronchoscopy, pulse oximetry, CT scan, and lung ultrasound [5].

Chest x-rays are the most widely used imaging techniques than CT scans for the diagnosis of pneumonia [6 2001]. **Chest x-rays are preferred more than chest CT scans because x-ray imaging takes less time than CT imaging and high-resolution CT scans may not be available in all regions across the globe.** On the contrary, x-ray imaging is the most widely used technique that is playing an important role in epidemiological studies and clinical care [Cherian et-al 2005, [8] 2001]. There are several places across the globe where there is a scarce of experienced medical professionals who can diagnose these diseases [Naicker S et-al 2009]. With the advancements in computerized methods like deep learning (DL) and machine learning (ML), there were robust improvements in the automatic diagnosis of diseases. With the use of these methods diagnosis of diseases **can be made available to a large population** with reduced cost. Deep Convolutional neural networks (DCNN) and machine learning algorithms are now being used for the diagnosis of many diseases like arrhythmia [Alarsan et-al 2019, Hannun et-al 2019, Acharya et-al 2017], breast cancer [Celik et-al 2020, Cruz-Roa et-al 2014, Dhahri et-al 2019], brain diseases [Talo et-al 2019], diabetic retinopathy [G. Varun et-al 2017], and many more. Qin C

et-al [Qin et-al 2018] performed a survey on using ML and DL techniques to diagnose diseases from chest x-rays.

In this paper, a study was performed on using machine learning algorithms namely random forest (RF), logistic regression (LR) and support vector machine (SVM) trained with features extracted by a DCNN named VGG16 to diagnose pneumonia from chest x-ray images. The chest x-ray dataset used in this study consists of 4273 x-ray images diagnosed with pneumonia and 1583 x-ray images of healthy people. The study was performed by implementing these classifiers by varying specific parameters like kernel function (for SVM), penalty (for LR), and the number of trees for (RF). In this study, SVM with RBF kernel reported the highest performance than other model configurations.

2. Related Works

With the increasing need for accurate and précised diagnosis in medicine, machine learning and deep learning methods are being used. Several methodologies were proposed by researchers across the globe to cater to this need. Like, Shubhangi Khobragade et-al [S. khabragade et-al 2016] proposed a feed-forward neural network for the diagnosis of tuberculosis (TB), pneumonia, and lung cancer. Another neural network-based model was proposed by Ronald Barrientos et-al [R. Barrientos et-al 2016] to diagnose pneumonia from lung ultrasound images. Stephen et-al [Stephen et-al 2019] proposed a six-layered DCNN for the diagnosis of pneumonia. In this work, the first 6 layers of the DCNN are the convolutional layers activated with ReLU activation function and the last two layers are the fully connected layers activated with ReLU and softmax function. This DCNN reported an accuracy of 94%. Oliveira et-al [Oliveira et-al 2008] developed a model called Pneumo-CAD for the diagnosis of pneumonia in children. In this model, wavelet transform coefficients were used for feature extraction and K-nearest neighbor with K=15 and distance-dependent weighting algorithms are used for classification. These two algorithms reported an AUC of 0.97 and 0.94 respectively. Abhishek Sharma et-al [A. Sharma et-al 2017] used the Otsu thresholding method to segment the lung regions that are infected by pneumonia and also computed the area of the lung that is affected by pneumonia.

A novel feature-based classification approach was proposed by N. Deepa et-al [Deepa NV et-al 2018] for the diagnosis of pneumonia from x-ray images. In this approach, 14 texture-based features are extracted from the images, and based on these features a decision rule is proposed to classify normal and pneumonia x-ray images. Haralick parameters, GLCM features, and Congruency parameters are used by some researchers to diagnose pneumonia [H. Ebrahimian et-al 2014]. Keegan Kosasih et-al [K. Kosasih 2015] proposed a novel approach to diagnose pneumonia using the cough sounds of patients. For this work, 815 cough sounds from 91 patients were used. In this work, wavelets and other mathematical analysis methods like non-Gaussianity index and Mel Cepstral coefficients are used for feature extraction. Logistic regression algorithm is used for the diagnosis of pneumonia and this algorithm reported a specificity of 63% and sensitivity of 94%. Chhikara P et-al [Chhikara P et-al 2020] proposed a DCNN for the diagnosis of pneumonia. This DCNN was developed using the transfer learning of InceptionV3 and this model reported an accuracy of 90.1% and precision of 90.7%. Gaobo Liang and Lixin Zheng [Liang G et-al 2020] proposed a transfer learning model using residual networks for the diagnosis of pediatric pneumonia. Their model reported an f1-score of 92.7%. Abdullah et-al [A.A, Abdullah 2011] used cellular neural networks to detect pneumonia symptoms in an x-ray image. For this work, Candy software is used as a neural network simulator and this model reported good performance in diagnosing pneumonia and segmenting the regions of lung infected with pneumonia. Jiansheng et al [L. Jiansheng et-al 2009] developed a DL model for the diagnosis of pneumonia. This DL model is built using fuzzy neural network based Kohonen network and their model reported an accuracy of 86%. Huang et al [J.S. Huang 2014] developed a predictive model using a support vector machine (SVM) with RBF kernel to predict the readmission rate of cured pneumonia patients after discharge from the hospital. This model reported accuracy of 83.85% and 82.24% and also regarded as an important tool to spot pneumonia patients at high risk.

With this motivation, a comprehensive study of using ML algorithms namely Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM) with different parametric values trained with features extracted by VGG16 DCNN is presented in this paper.

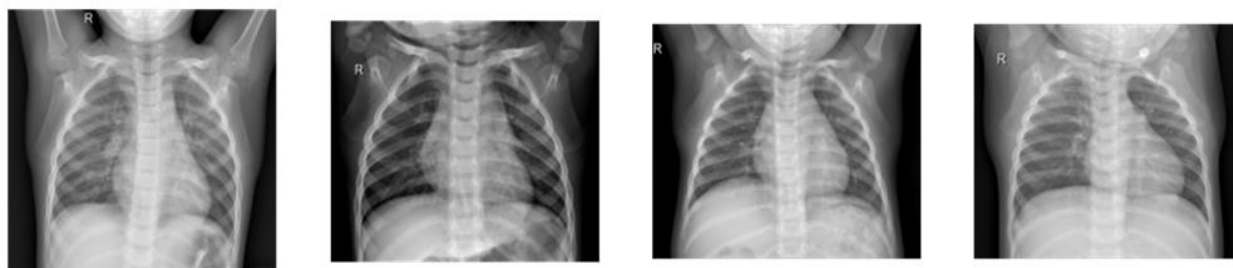
3. Dataset Description and Image Preprocessing

The chest x-ray images used for this study are extracted from the ChestXRy2017 dataset developed by Daniel Kermany et-al [Kermany et-al 2018]. This dataset consists of 5863 images classified into two classes namely pneumonia and normal. There are 4273 chest x-ray images diagnosed with pneumonia and 1583 chest x-rays of healthy persons. Few images from the dataset are shown in figure 1. The chest x-ray images are pre-

processed by converting to PNG (portable network graph) format, grey-scale representation, and resizing to 224 x 224 pixels. After preprocessing the images, the dataset is split in the ratio 90:10 for training and testing the classifiers. Few images from the dataset are shown in Figure 1.



(a) Pneumonia chest x-ray images



(b) Normal chest x-ray images

Figure 1. Sample Images from the Dataset

4. Methods

In this paper, the performance of machine learning classifiers trained with features extracted by VGG16 DCNN diagnose of pneumonia from chest x-ray images is studied. This section is divided into two subsections where subsection 4.1 describes the feature extraction methodology and subsection 4.2 describes the classification algorithms used for the study.

4.1. Feature Extraction

In this study, VGG16 DCNN **with weights trained on the ImageNet database** is used for extracting features from the images. ImageNet database consists of more than 14 million images categorized into 1000 classes. Since pre-trained models like VGG16, InceptionV3, etc have already learned to extract features from the images and also to distinguish images of different classes, these models have shown magnificent performance when applied on datasets of similar domains. VGG16 DCNN has shown excellent performance than other pre-trained models in medical imaging [Yadav et-al 2019], so it is used for feature extraction in this study. VGGNet or VGG16 DCNN contains 16 convolutional layers with receptive fields of size 3 x 3, 5 Max-Pooling layers of pool size 2 x 2 for spatial pooling, and 3 fully connected layers with the last layer activated with Soft-max function and it has 144 million parameters. In this DCNN the hidden layers are activated with Rectified nonlinearity (ReLU) activation function and dropout regularization is used in the fully connected layers. The structure of VGG16 architecture is shown in Figure 2. For this study, in order to use VGG16 as a feature extractor the fully connected layers of the DCNN are removed (highlighted with red color in figure 3). The features extracted by the VGG16 in the form of feature vectors are used for training ML classifiers namely Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM). The structure of the proposed study is shown in Figure 3.

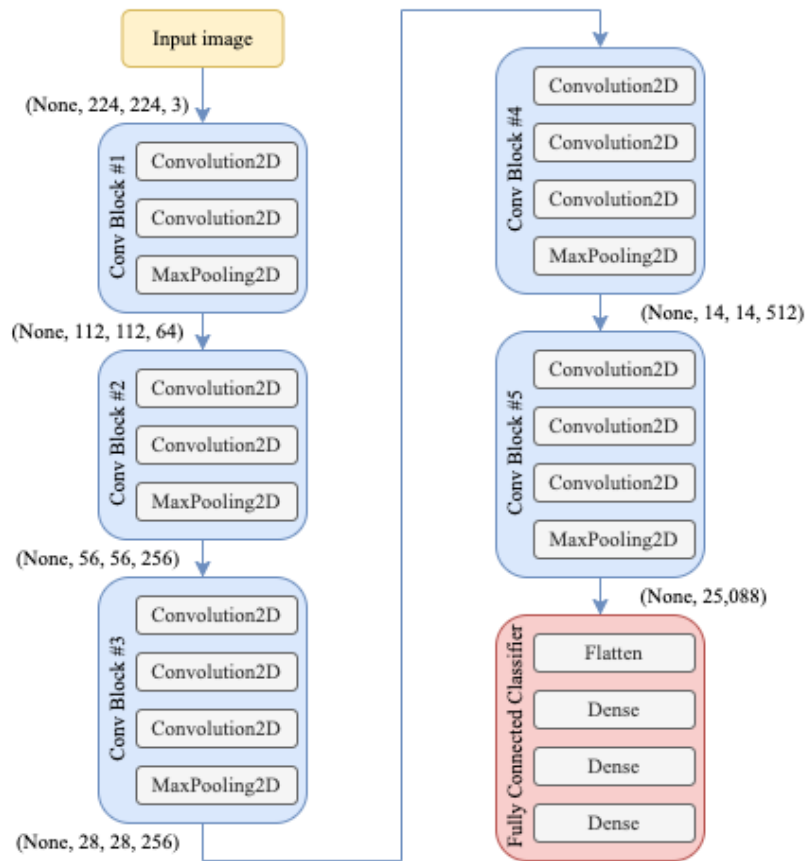


Figure 2. Architecture of VGG16 DCNN

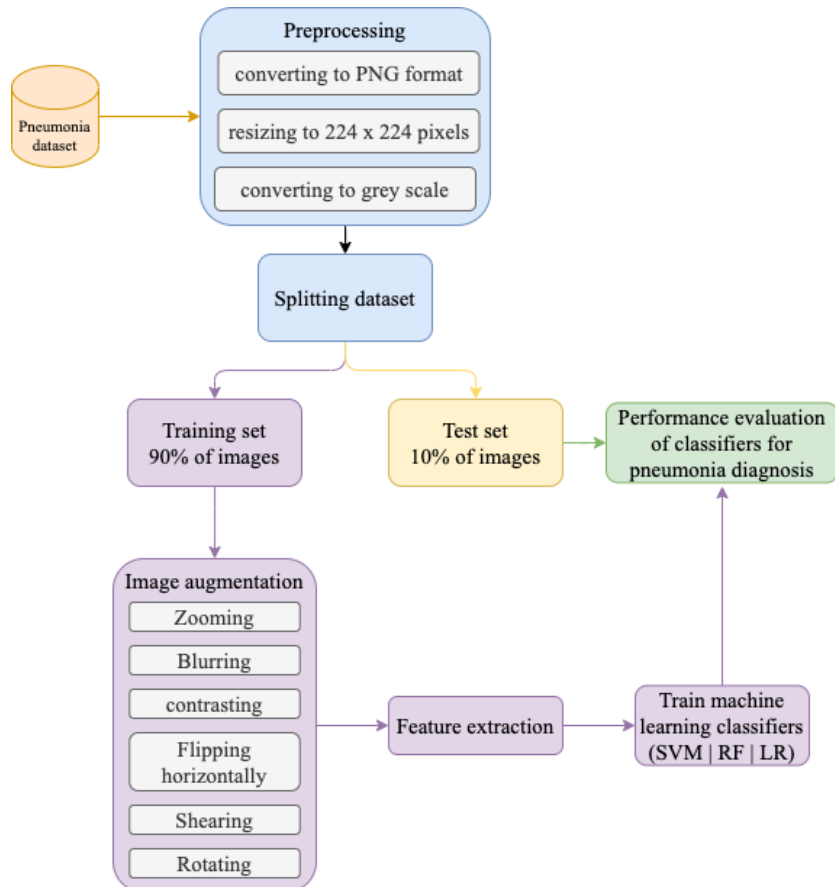


Figure 3. Proposed Study

4.2. Classification Algorithms

4.2.1. Support Vector Machine

Support Vector Machine (SVM) is a machine learning algorithm developed by Cortes and Vapnik [CortesC, Vapnik V 1995] used to solve regression and classification problems. In this study, SVM is used for binary classification. To solve a classification problem, SVM plots the data items in the dataset in an N-dimensional space where N is equal to the number of attributes or features in the data and then finds an optimal hyper-plane(line) to separate these two classes of data. SVM works well for linearly separable data, but the image data used in this study are non-linearly separable. So SVM's have to be tuned using kernel functions.

Kernels are nothing but mathematical functions that transform the non-separable training data into separable data of different classes. The kernel functions used in this study are 'linear', 'RBF', 'polynomial', and 'sigmoid'. The equations of these functions are given below.

Linear kernel function: $K(x, y) = x^T y + c$

Gaussian kernel Radial Basis Function (RBF) kernel function: $K(x, y) = \exp\left(\frac{-\|x-y\|^2}{2\sigma^2}\right)$

Polynomial kernel function: $K(x, y) = \tanh(\gamma \cdot x^T y + r)^d$

Sigmoid kernel function: $K(x, y) = \tanh(\gamma \cdot x^T y + r)$

In the above equations, K is the kernel function, $x = (x_1, \dots, x_n)$ is the features, $y = (y_1, \dots, y_n)$ is the labels of the training data, σ is an adjustable parameter greater than zero, γ is the slope, d is the degree of the polynomial and c is a constant.

4.2.2. Random Forest

Random Forest (RF) or Random decision forests is an ensemble learning algorithm developed by Tin Kam Ho [Ho, T.K 1995] used for regression and classification tasks. The training of RF is done using a technique known as bootstrap aggregation.

RF algorithm works by constructing many decision trees during the training time and this algorithm outputs the value that is mean of values predicted by the individual trees in the forest for regression task and for classification this algorithm outputs the class that is the mode of classes predicted by the individual trees. For this study, the performance RF constructed with 50, 100, 200, and 400 trees trained with features extracted by VGG16 DCNN is studied.

4.2.3. Logistic Regression

Logistic Regression (LR) is a statistical model developed by Tolles J and Meurer WJ [Tolles J, Meurer WJ 2016] used for regression and classification. This algorithm works based on the logistic function. Logistic function or sigmoid function is an S-shaped curve that takes real-valued input and maps it between 0 and 1, but never 0 or 1. The equation of the sigmoid function is shown in equation 2.

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

In the above equation, k represents steepness, L represents the curve's maximum value, x is a real number and x_0 is the x value of the sigmoid midpoint. In this study, LR with 'L1' and 'L2' penalty is studied. The only difference between LR with 'L1 regularization' and 'L2 regularization' is that L2 regularization adds squared magnitude of coefficient $\lambda \sum_{j=1}^p |\beta_j^2|$ as penalty term to the loss function and L1 regularization adds an absolute value of the magnitude of coefficient $\lambda \sum_{j=1}^p |\beta_j|$ as penalty term to the loss function.

5. Experiments and Results

In this study, the performance of DCNN transfer learning for the diagnosis of pneumonia from x-ray images is evaluated and analyzed. The VGG16 DCNN is used for extracting features from the chest x-ray images. To

Impact Of Machine Learning Models In Pneumonia Diagnosis With Features Extracted From Chest X-Rays Using VGG16

use VGG16 DCNN as a feature extractor the fully connected layers of the VGG16 are removed. For classification, ML algorithms namely Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM) is employed. The reason behind choosing these algorithms for classification is that they have shown promising results in medical diagnosis. For an extensive study, specific parameters of the classifiers are varied within reasonable limits. For example, for SVM the experiments were conducted by comparing ‘RBF’, ‘linear’, ‘Polynomial’ and ‘sigmoid’ kernels, for experiments related to RF the number of trees was varied from 50 to 200, and LR is compared based on ‘L1’ and ‘L2’ penalizations. Except for the parameters stated above the other parameters of the classifiers are fixed at their default values. The base model configurations that were evaluated in this study for the diagnosis of pneumonia are shown below (the parameters that were changed for the study are enclosed in parenthesis).

TL+RF = Random forest classifier trained with features extracted by VGG16 DCNN for pneumonia diagnosis (number of trees: 50, 100, 200, 400).

TL+SVM = Support Vector Machine trained with features extracted by VGG16 DCNN for pneumonia diagnosis (kernels=’polynomial’, ‘RBF’, ‘linear’ and ‘sigmoid’).

TL+LR = Logistic Regression trained with features extracted by VGG16 DCNN for pneumonia diagnosis (penalty=’L1’, ‘L2’).

These models are evaluated based on accuracy, error rate, sensitivity, F1-Measure, precision, and specificity. The performance of these classifiers is shown in Table 1 and are pictorially represented in Figure 4.

Table 1. Performance of Tuned Machine Learning Classifiers Trained with Features Extracted by VGG16 DCNN for the Diagnosis of Pneumonia

Final Classifier	Models	Accuracy	Precision	Specificity	Sensitivity	F1-Measure	Error rate
Logistic Regression (LR)	TL+LR(L1)	96.36	93.52	97.58	93.08	93.3	3.64
	TL+LR(L2)	96.53	93.2	97.48	93.94	93.57	3.47
	TL+SVM (RBF)	96.61	93.2	97.48	94.24	93.72	3.39
	TL+SVM (Poly)	96.61	93.04	97.42	94.39	93.71	3.39
Support Vector Machine (SVM)	TL+SVM (Sigmoid)	95.24	89.88	96.27	92.37	91.11	4.76
	TL+SVM (Linear)	96.48	92.89	97.36	94.08	93.48	3.52
	TL+RF (100)	94.77	89.57	96.14	91.01	90.28	5.23
Random Forest (RF)	TL+RF (200)	95.2	89.25	96.06	92.77	90.98	4.8
	TL+RF (400)	95.07	89.25	96.05	92.32	90.76	4.93

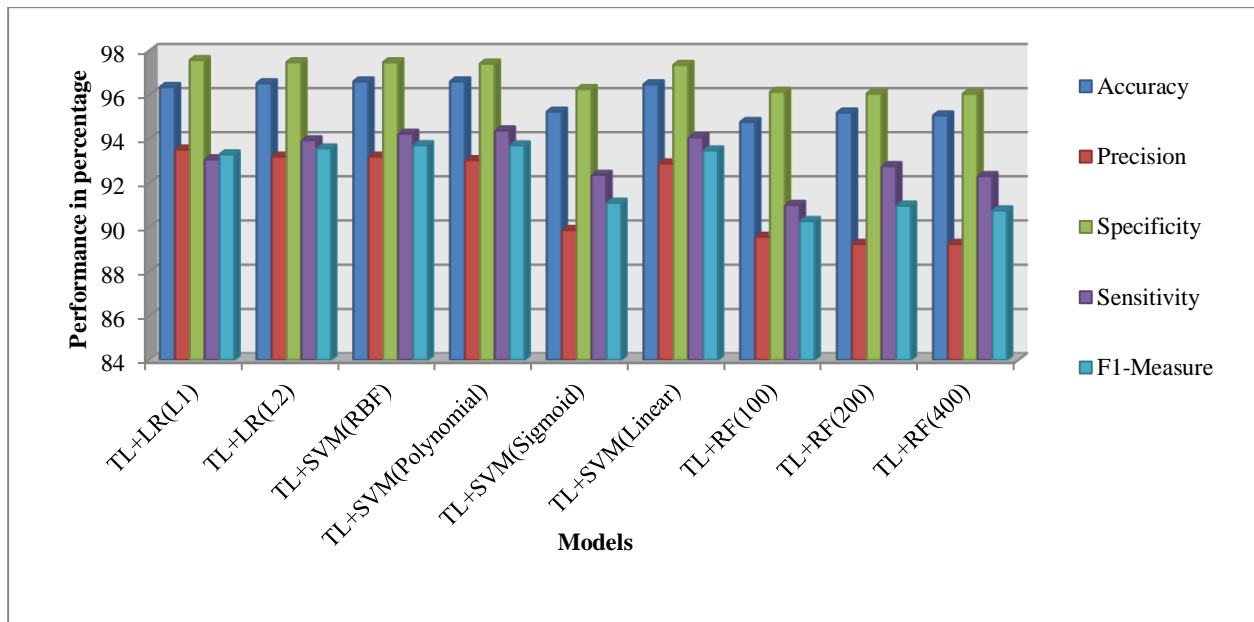


Figure 4(a). Accuracy, Precision, Specificity, Sensitivity and F1-Measure of Classifiers

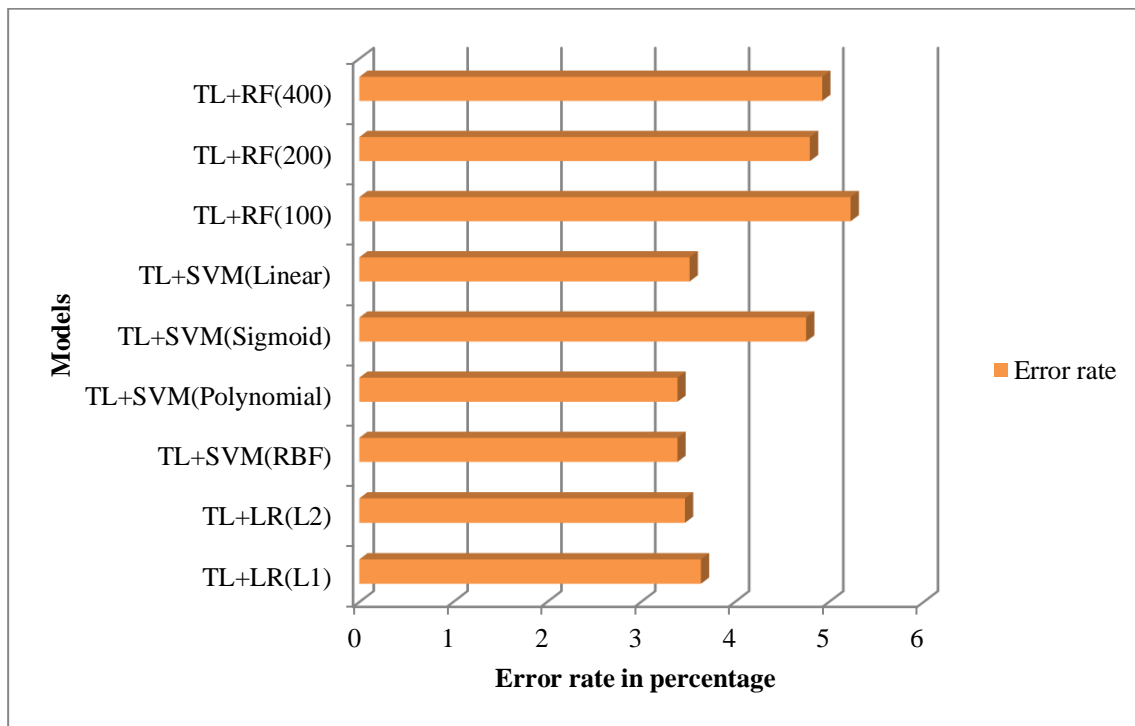


Figure 4(b). Error Rate of Classifiers

The three best performance classifiers that are selected for further investigation are TL+LR(L2), TL+SVM(RBF), and TL+RF (200). Among these three classifiers, Support vector machine (SVM) with RGB kernel trained with features extracted by VGG16 DCNN reported higher performance then followed by Logistic regression with the L1 penalty and Random Forest with 200 trees. This is further confirmed by Area-under-ROC-curve (AUC) values, ROC curve, Intersection over the Union (IoU), and Kappa score. The confusion matrices of the best three classifiers are shown in Figure 5. The AUC, IoU, and Kappa scores of the selected configuration of classifiers are shown in Table 2 and the ROC curves are shown in Figure 6. The comparison of these models based on the positive predicted value (PPV), negative predicted value (NPV), and balanced accuracy is shown in Table 3. Thus the AUC, IoU, Kappa score, accuracy, sensitivity, F1-Measure, and the error rate of TL+SVM(RBF) seems satisfactory for this study.

Table 2. AUC, IoU and Kappa Score of the Best Selected Classifiers

Final configuration of classifier	AUC	IoU	Kappa score
TL+LR(L2)	0.8792	0.9120	3.47
TL+SVM (RBF)	0.8819	0.9141	3.39
TL+RF (200)	0.8345	0.8771	4.8

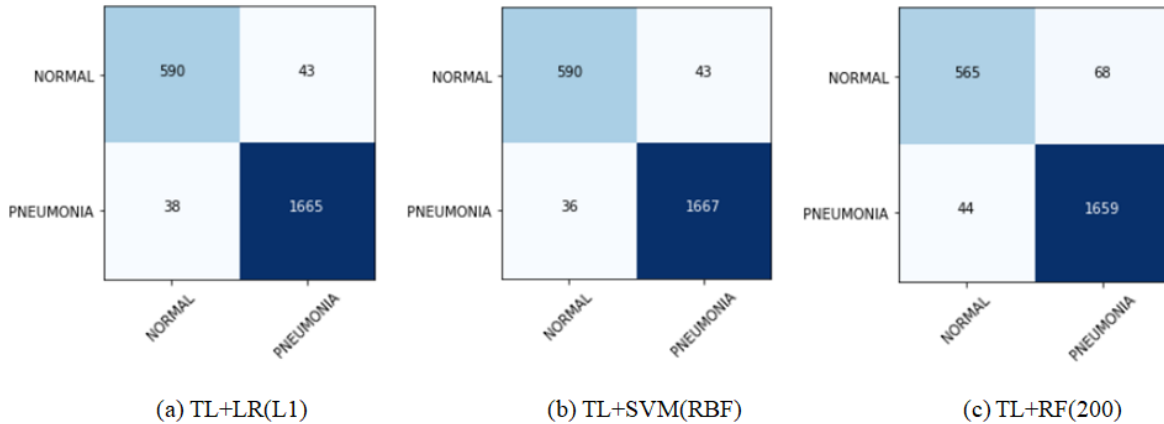


Figure 5. Confusion Matrix of Selected Classifiers

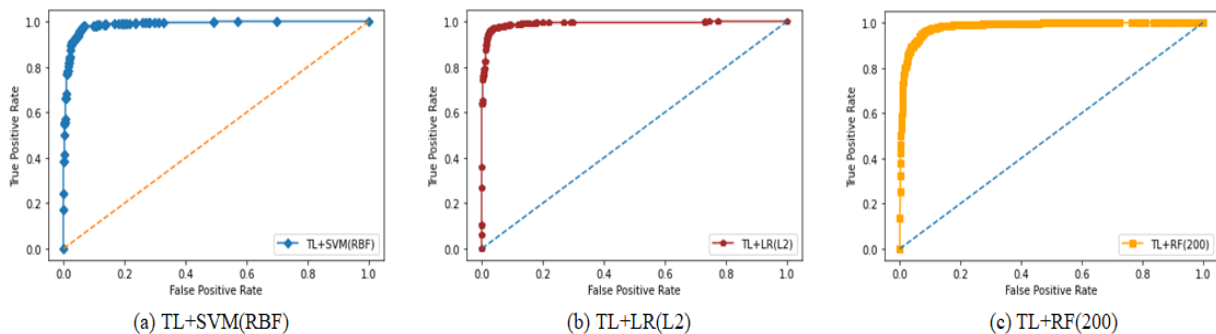


Figure 6. ROC Curves of the Best Selected Classifiers

Table 3. PPV, NPV and Balanced Accuracy of best Configuration of Classifiers

Models	Positive predicted value	Negative predicted value	Balanced accuracy
TL+LR(L1)	0.9320	0.9776	0.9571
TL+SVM(L2)	0.9320	0.9788	0.9586
TL+RF (200)	0.8925	0.9741	0.9441

6. Conclusion

Pneumonia is a pulmonary infection affecting the alveoli of a single or both the lungs. Pneumonia causes severe illnesses like shortness of breath, bacteremia (bacteria in the bloodstream), pleural effusion (fluid accumulation around the lungs), lung abscess, and even deaths. So an accurate and précised diagnosis of pneumonia is very important in today’s world. Considering this as the highest priority, a comparative study was performed using Logistic Regression, Support Vector Machine, and Random Forest trained with features extracted by VGG16 DCNN for the diagnosis of pneumonia. For an extensive study, the performance of the machine learning algorithms are compared by varying specific parameters within reasonable limits like for SVM kernel functions (‘linear’, ‘polynomial’, ‘sigmoid’ and ‘RBF’) are used, for LR penalty term (‘L1’ and ‘L2’) is used and for RF the number of trees (50, 100, 200 and 400) are used. From the study, it was found that SVM with RBF kernel trained with features extracted by VGG16 DCNN reported the highest performance than other classifiers.

References

1. CDC, 2017. <https://www.cdc.gov/dotw/pneumonia/index.html>.
2. Rudan, I., Boschi-Pinto, C., Biloglav, Z., Mulholland, K., & Campbell, H. (2008). Epidemiology and etiology of childhood pneumonia. *Bulletin of the world health organization*, 86, 408-416B.
3. Adegbola, R.A. (2012). Childhood pneumonia as a global health priority and the strategic interest of the Bill & Melinda Gates Foundation. *Clinical infectious diseases*, 54(suppl_2), S89-S92.
4. Online, FE. 17 Lakh Indian Children to Die Due to Pneumonia by 2030; Here's How It CanBe Averted. *The Financial Express*, The Financial Express. www.financialexpress.com/economy/17-lakh-indian-children-to-die-due-to-pneumonia-by-2030-heres-how-it-can-be-averted/1380390/.
5. "Pneumonia." *Mayo Clinic*, Mayo Foundation for Medical Education and Research. www.mayoclinic.org/diseases-conditions/pneumonia/diagnosis-treatment/drc-20354210.
6. World Health Organization. (2001). *Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children* (No. WHO/V&B/01.35). World Health Organization.
7. Cherian, T., Mulholland, E.K., Carlin, J.B., Ostensen, H., Amin, R., Campo, M.D., & Steinhoff, M.C. (2005). Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. *Bulletin of the World Health Organization*, 83, 353-359.
8. Franquet, T. (2001). Imaging of pneumonia: trends and algorithms. *European Respiratory Journal*, 18(1), 196-208.
9. Naicker, S., Plange-Rhule, J., Tutt, R.C., & Eastwood, J.B. (2009). Shortage of healthcare workers in developing countries--Africa. *Ethnicity & disease*, 19(1), 60-64.
10. Alarsan, F.I., & Younes, M. (2019). Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. *Journal of Big Data*, 6(1), 1-15.
11. Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., & Ng, A.Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1), 65-69.
12. Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A., & San Tan, R. (2017). A deep convolutional neural network model to classify heartbeats. *Computers in biology and medicine*, 89, 389-396.
13. Celik, Y., Talo, M., Yildirim, O., Karabatak, M., & Acharya, U.R. (2020). Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. *Pattern Recognition Letters*.
14. Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., & Madabhushi, A. (2014). Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. *In Medical Imaging 2014: Digital Pathology*. International Society for Optics and Photonics, 9041, 904103.
15. Dhahri, H., Al Maghayreh, E., Mahmood, A., Elkilani, W., & Faisal Nagi, M. (2019). Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms. *Journal of Healthcare Engineering*, 2019.
16. Talo, M., Yildirim, O., Baloglu, U.B., Aydin, G., & Acharya, U.R. (2019). Convolutional neural networks for multi-class brain disease detection using MRI images. *Computerized Medical Imaging and Graphics*, 78, 101673.
17. Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., & Webster, D.R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22), 2402-2410.
18. Qin, C., Yao, D., Shi, Y., & Song, Z. (2018). Computer-aided detection in chest radiography based on artificial intelligence: a survey. *Biomedical engineering online*, 17(1), 113.
19. Khobragade, S., Tiwari, A., Patil, C.Y., & Narke, V. (2016). Automatic detection of major lung diseases using Chest Radiographs and classification by feed-forward artificial neural network. *In IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, 1-5.
20. Barrientos, R., Roman-Gonzalez, A., Barrientos, F., Solis, L., Correa, M., Pajuelo, M., & Zimic, M. (2016). Automatic detection of pneumonia analyzing ultrasound digital images. *In IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI)*, 1-4.
21. Stephen, O., Sain, M., Maduh, U.J., & Jeong, D.U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. *Journal of healthcare engineering*, 2019.
22. Oliveira, L.L.G., e Silva, S.A., Ribeiro, L.H.V., De Oliveira, R.M., Coelho, C.J., & Andrade, A.L.S. (2008). Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. *International journal of medical informatics*, 77(8), 555-564.
23. Sharma, A., Raju, D., & Ranjan, S. (2017). Detection of pneumonia clouds in chest X-ray using image processing approach. *In Nirma University International Conference on Engineering (NUiCONE)*, 1-4.

24. Deepa, N.V., Krishna, N., & Kumar, G.H. (2017). Feature Extraction and Classification of X-Ray Lung Images Using Haralick Texture Features. *In International Conference on Next Generation Computing Technologies*, 899-907.
25. Ebrahimian, H., Rijal, O.M., Noor, N.M., Yunus, A., & Mahyuddin, A.A. (2014). Phase congruency parameter estimation and discrimination ability in detecting lung disease chest radiograph. *In IEEE conference on biomedical engineering and sciences (IECBES)*, 729-734.
26. Kosasih, K., Abeyratne, U.R., Swarnkar, V., & Triasih, R. (2014). Wavelet augmented cough analysis for rapid childhood pneumonia diagnosis. *IEEE Transactions on Biomedical Engineering*, 62(4), 1185-1194.
27. 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*, 155-168.
28. Liang, G., & Zheng, L. (2020). A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Computer methods and programs in biomedicine*, 187, 104964.
29. Abdullah, A.A., Posdzi, N.M., & Nishio, Y. (2011). Preliminary study of pneumonia symptoms detection method using cellular neural network. *In International Conference on Electrical, Control and Computer Engineering 2011 (InECCE)*, 497-500.
30. Jiansheng, L., Jinliang, H., & Jianjing, S. (2009). Research on diagnostic models of pneumonia syndromes in TCM based on fuzzy-neural net. *In Chinese Control and Decision Conference*, 933-937.
31. Huang, J.S., Chen, Y.F., & Hsu, J.C. (2014). Design of a clinical decision support model for predicting pneumonia readmission. *In International Symposium on Computer, Consumer and Control*, 1179-1182.
32. Kermany, D., Zhang, K., & Goldbaum, M. (2018). Labeled optical coherence tomography (OCT) and Chest X-Ray images for classification. *Mendeley data*, 2.
33. Yadav, S.S., & Jadhav, S.M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, 6(1), 113.
34. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
35. Ho, T.K. (1995). Random decision forests. *In Proceedings of 3rd international conference on document analysis and recognition*, 1, 278-282.
36. Tolles, J., & Meurer, W.J. (2016). Logistic regression: relating patient characteristics to outcomes. *Jama*, 316(5), 533-534.