Investigation of the performance of Machine Learning Classifiers for Pneumonia Detection in Chest X-ray Images

Conference Paper · July 2020

DOI: 10.1109/EIT48999.2020.9208232

3 authors:

Rabia Al Mamlook
Western Michigan University

Shengfeng Chen
Western Michigan University

Hanin Fawzi Bzizi
Western Michigan University

Investigation of the performance of Machine Learning Classifiers for Pneumonia Detection in Chest X-ray Images

Rabia Emhamed Al Mamlook
Industrial Engineering
Western Michigan University
Kalamazoo, MI, USA
rabiaemhamedm.almamlook@wmich.edu

Shengfeng Chen
Industrial Engineering
Western Michigan University
Kalamazoo, MI, USA
shengfeng.chen@wmich.edu

Hanin Fawzi Bzizi
Biomedical Sciences
Western Michigan University
Kalamazoo, MI, USA
Hanin.f.bzizi@wmich.edu

***Abstract*—Pneumonia is one of the serious diseases that is caused by a bacterial or viral infection of the lungs and has the potential to result in severe consequences within a short period. Therefore, early diagnosis is a key factor in terms of the successful treatment process. Thus, there is a need for an intelligent and automatic system that has the capability of diagnosing chest X-rays and to simplify the pneumonia detection process for experts as well as for novices. This study aims to develop a model that will help with the classification of chest x-ray medical images into normal(healthy) vs abnormal(sick). To achieve this, seven existing state-of-the-art machine learning techniques and well-known convolutional neural network models have been employed to increase efficiency and accuracy. In this study, we propose our deep learning for the classification task, which is trained with modified images, through multiple steps of preprocessing. Experimentally, it demonstrated that the deep learning technique for the classification task performs the best, compared to the other seven machine learning techniques. In this study, we successfully classified chest infection in chest X-ray images using deep leaning based on CNN with an overall accuracy of 98.46%. It achieved a more successful result in detecting pneumonia cases.**

***Index Terms*—Chest X-Ray , Classification Model,Deep Learning, Machine Learning,Pneumonia.**

I. Introduction

Pneumonia is one of the serious diseases which cause most of the deaths in adults globally. According to Health Metrics and Evaluation (IHME), the highest pneumonia mortality rates in 2017 were among people aged 70 and older. More than 1.13 million pneumonia-related deaths are reported every year were in this age group [1]. Pneumonia is a disease of infectious origin that causes inflammation in the air sacs or alveoli of one or both lungs [2]. The air sacs get filled with fluid which can cause difficulty in breathing and an over-generation of mucus and sputum as shown in figure 1. Pneumonia is most commonly caused by viruses or bacteria and, less commonly, other microorganisms [3].

Nowadays, chest X-ray (CXR) imaging is used commonly for health intensive care and analysis of many lung diseases such as pneumonia, cancer because of relatively low prices. Therefore, Chest X-rays are currently the best available tool

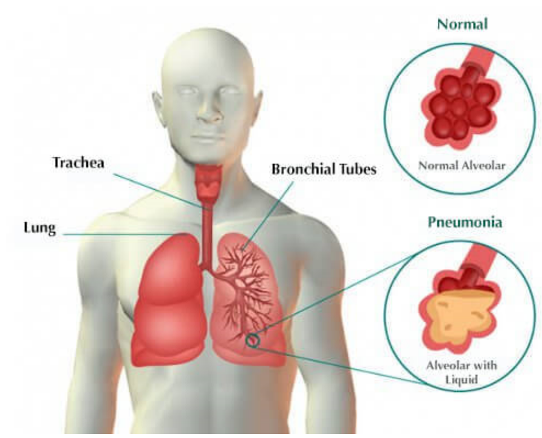

Fig. 1. Pneumonia
[4]

for diagnosing pneumonia [5] which has played a huge role in clinical care [6] and epidemiological research [7]. However, detecting pneumonia in chest X-rays is a challenging task that relies on the availability of expert radiologists. [8]. Accurate image analysis and image interpretation are very crucial for a better diagnosis. Though image interpretation by conventional machine learning algorithms depends mostly on expertly crafted features, computer vision is a very effective machine learning application. However, not all doctors have good diagnosing tools to diagnose patients. As a result, sometimes their diagnosing is not very accurate. It is also much more difficult to judge pneumonia just by looking at the chest X-rays images. Therefore, more specific diagnostic tools must be developed which are cheap and accurate in diagnosing pneumonia.

With the advancements of information technology, machine-learning algorithms have been successfully applied to many healthcare problems and have explained complex relationships, and improved clinical predictions [9]. In recent years, several researchers have proposed different artificial intelligence (AI)-based solutions for different medical problems. Deep learning techniques based on Convolutional neural networks (CNNs) have allowed researchers to obtain successful results in wide medical applications, including classifying skin cancer from skin photographs [10], detecting metastasis on pathological images and disease classification in X-ray images all of which demonstrated expert-level diagnostic accuracy [11]. As a result of this trend, numerous studies have emerged that analyzes the influence of deep learning and convolutional neural networks in the field of health and imaging medical diagnosis. Specifically, due to successes in other fields especially in medical problems, deep learning algorithms such as Convolutional neural networks (CNNs) have recently been applied for image classification tasks. [12]. However, algorithms in those studies have not yet been fully validated in various data sets, limiting the generalization of the results [13].

Classification methods are among the most commonly used techniques in medical imaging [14], where the goal is building classifiers that are capable of predicting whether x-ray images are normal or show the presence of Pneumonia. Especially, this study tries to answer the following research question: Could machine learning algorithms be used to assist in the diagnosis of whether a patient has pneumonia or not by looking at chest X-ray images? This inspires us to use machine learning algorithms to predict pneumonia based on the x-ray images. Moreover, it would provide doctors with immediate information about the patient's condition and risk level to recommend more diagnostic tests without delay. Thus, the investigative and predictive methods, such as machine learning algorithms, are vital to make smart decisions that will help doctors and radiologists to get more information to prevent themselves from misdiagnosing a patient.

The primary objective of this study was to examine and evaluate the performance of machine learning techniques and convolutional neural network (CNN) for the classification of pneumonia, which is based on Chest X-ray Images to achieve high accuracy.Other objective is to provide radiologists and medical experts with a tool which is a lower cost .This tools will Be helping radiologists and medical experts to identify the slit and slow changes among. Moreover, this tool could help them to read their X-ray images and to create a basis for a model to read more complex data like CT images

The remainder of the paper is organized as follows: section 2 discusses the materials and methods,the section 3 explains classification models,section 4 describes model evaluation and validation,section 5 discusses the results and discussion , and section 6 summary and concludes the paper.

II. Materials and Methods

Figure 2 shows the main diagram of our proposed full system. As seen in this figure, the system is composed of several steps including the Data collection and preprocessing description, building the classification models, and extracting the required features. These steps can be divided into four phases: Data Set, data preprocessing, building and validating classification models, and feature extraction. The details of each step are discussed in the following sections.

Fig. 2. Diagram of our proposed full system

A. Data Collection

The chest x-ray image dataset is available at [15]. It was the Chest X-Ray images which include normal chest X-Ray and Pneumonia chest X-Ray. Here, we have taken 5,856 sample images of the dataset to use later on for recognition and classification of pneumonia.1,583 of which are normal X-Ray lung images, and 4,273 of which are X-Ray lung images with Pneumonia from patients. These data set was used for training in deep learning method and testing data using "Pandas" for predicting Pneumonia.The X-ray images were divided into train, test, and validation groups. 80% of data was used for training, 10% used for testing, and 10% used for validations. The data set was randomized in advance to avoid potential bias.We have trained the dataset using Keras on the top of TensorFlow.

B. Data Prepossessing

The most important part of machine learning is data and data must be clean for models to process it. When it comes to image data, there are some preparation methods. Various types of pre-processing tasks such as dimension reduction, image resize, and image cropping is applied. Depending on the original data sets, it showed several problems that need to be addressed before image classification. First, resize the image to the same size as the original image. The image can be resized to get rid of unnecessary information in the background The raw images have different image sizes one from another. the machine learning model requires the data being trained must have the same size. Therefore, an immediate image resizing is essential to prepare for machine learning models as shown in Figure 3.

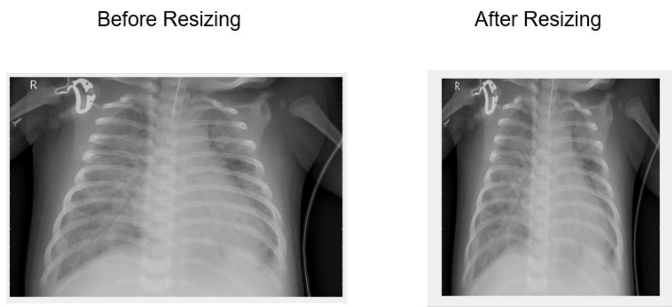

Fig. 3. Image Resizing

Next, observing the X-Ray image samples, most samples were two-dimensional image, while a few samples were three-dimensional image. Dimensional reduction as shown in Figure 4 was needed to convert three-dimensional images into two-dimensional images.

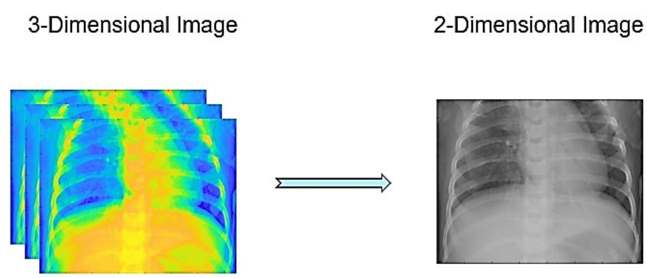

Fig. 4. Image Dimension Reduction

The two preprocessing approaches mentioned above (re-sizing and dimension reduction) were the most important for preprocessing because of the unequal data dimensions. Other pre-processing approaches were also used on individual images samples whose quality was out of control.

III. Classification module

There were various machine learning models utilized for classifying Pneumonia X-Ray images from normal ones. The objective of the module is to provide a binary classification of presence-absence of pneumonia within a chest x-ray and to select the best model for prediction. We were built systems using traditional methods of Machine Learning in which the features form the input were selected and extracted manually and based on the criteria that the engineer considered appropriate.

Figure 5 shows the paradigm shift of the traditional machine learning and deep neural network. The Machine Learning algorithms are applied according to the features selected and extracted manually, while the Deep neural network learns how to extract the most representative or complex features and patterns on its own according to the input. Moving from machine learning to Deep Learning. CNN was considered one of the best models for image recognition and classification in general. Therefore, the following sections were focusing on

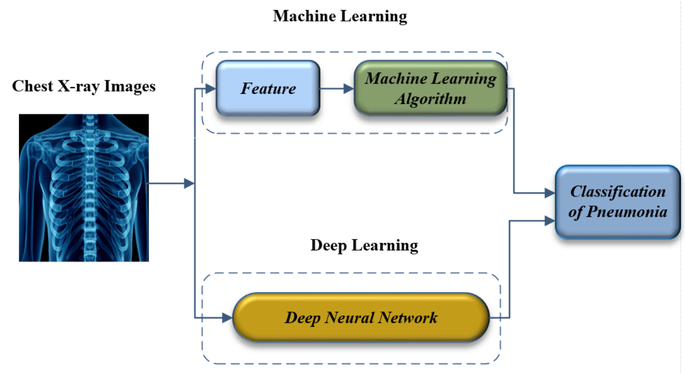

Fig. 5. Machine learning flow versus Deep Learning flow

describing CNN model that was used for this data set The models that were implemented for comparison were Random Forest, Xgboost, K-nearest neighbor, Decision Tree, Gradient Boost, Adaboost, and Convolutional Neural Network (CNN). The brief principle, experiment setup, and experiment results of each machine learning algorithm are described as below:

A. *Decision Tree*

Decision tree algorithms are the most commonly used algorithms in classification [16]. The main goal of the decision tree is to produce a model that calculates the value of a required variable based on numerous input variables [17].

B. *Random Forest*

Random Forest is a model including many decision trees. Each decision tree is built with training data randomly sampled and splitting nodes are selected with subsets of features. Random Forest overcomes the decision tree's drawback that has high variance when fitting the data. Random forest is widely used in classification for its good performance.

C. *K-nearest neighbor(KNN)*

K-nearest neighbor (KNN) technique is a simple algorithm used for both classification and regression problems. It is measured concerning the value of k, which define how many nearest neighbors request to be tested to describe the class of a sample data [18], [19].

D. *AdaBoost*

AdaBoost or Adaptive Boosting has great success in the application. It uses decision trees with a single split. By putting weight on difficult instances, a new decision tree is added with the focus on more difficult patterns.

E. *Gradient Boost*

Similar to AdaBoost which uses a decision tree, Gradient Boost can add trees one at a time while keeping existing trees unchanged. Gradient descent is used for calculating the loss to minimize a set of parameters.

F. XGBboost

XGBoost is a recently developed method that can efficiently implement machine learning algorithms under the Gradient Boosting framework. It provides a parallel tree boosting that can utilize multi-threading in a modern CPU and solve large problems in a fast and accurate way. More information can be found at github.com/dmlc/xgboost.

G. Deep Learning

To compute interesting functions, a non-linearity called an “activation function” is typically inserted between each layer in the neural network. Deep learning steps as following :

Image Convolution

One of the most important machine learning models for image classifications is CNN nowadays because it’s able to extract image features efficiently with the utilization of Image convolution. A good example is shown in figure 6 that an image’s features can be extracted by an image filter with image convolution. Different image filters can extract different image features.

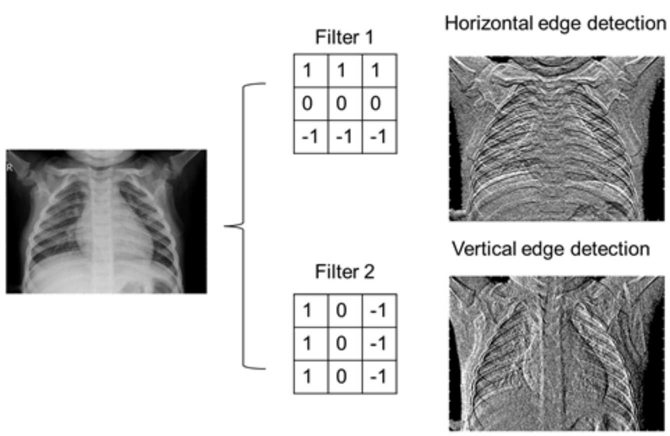

Fig. 6. Image Convolution

For instance, filter 1 in figure 6 extracted horizontal edge from the original image, while filter 2 extracted vertical edge correspondingly. Imagine hundreds or thousands of filters included in CNN can extract the same number of features from a single image. Through a proper training and validation process, CNN is enabled to predict or classify images from rich features being extracted from the sample images.

Image Pooling

Image Pooling is a popular approach in CNN for data dimension reduction. It is responsible for reducing processing time by downsampling the data it receives from the preceding convolutional layers to decrease the feature map’s dimensionality and sharpen the identified features. Max pooling is a typically used algorithm for pooling. Modern images are commonly large which would significantly increase computational complexity. In CNN, a large image can be reduced to a smaller size by image pooling. As Figure 7 suggested, a 4x4 image can be reduced to a 2x2 image by either average pooling or max pooling. This is done by moving a 2x2 pooling window on the image surface and calculating either average value or maximum value within the pooling window for creating the new image surface. This approach not only reduces the image size efficiently but also extract key features (i.e., max value or average value) from the image. It is worth mentioning that max-pooling is more popular and used more often in applications and research.

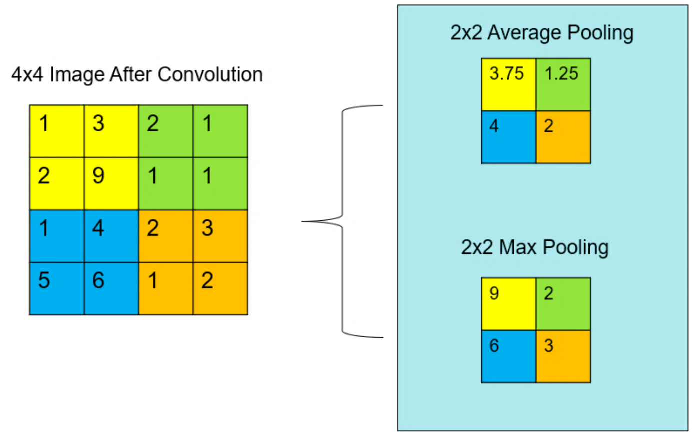

Fig. 7. Image Pooling

Convolutional neural network

CNN is a class of deep neural networks that specializes in analyzing images and thus is widely used in computer vision applications such as image classification and clustering, object detection, and neural style transfer. As Figure 8 displayed, the input images were preprocessed to have a uniform size of 40x40 pixels. The first layer resulted from 32 filters followed by a 2x2 pooling layer (max pooling). The second layer consisted of 64 filters resulting in a 14x14x64 data set, followed by a 2x2 max pooling layer again. The third layer was a fully connected layer in which a three-dimensional data set from the previous layer was vectorized to a single vector. Subsequently, a regular neural network was applied with one extra hidden layer and an output layer for binary classification. The convolutional layer and pooling layer take care of extracting the features maps from the input image. And in the last layers are located the fully-connected layers that carry out the classification function, reaching a prediction from the feature map extracted from the image.

IV. Model evaluation and validation

This section was described as the main metric used to evaluate the performance of both the classification and detection module. Several performance metrics have been used to figure out the performance of the Machine Learning algorithms in this study. As the study deals with classification problems, performance metrics relating to classifications are discussed as follows:

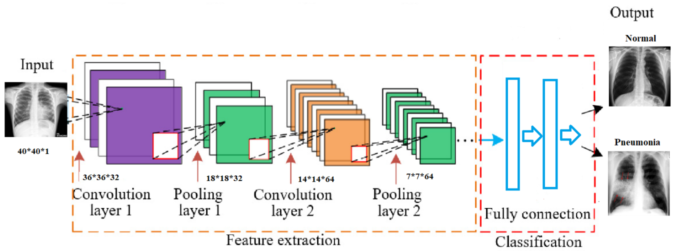

Fig. 8. Convolutional neural network

a. The Confusion Matrix

From the confusion matrix, we computed the following:

- True Positive (TP): It is the correct classification of the positive class, for example, if an image contains pneumonia and the model classifies the presence of pneumonia.
- True Negative (TN): It is the correct classification of the negative class, for example, there is no pneumonia present in the image and the model after classification declares that the image is not pneumonia present.
- False Positive (FP): It is the incorrect prediction of the positives, for example, the image does have pneumonia, but the model classifies that the image does not contain pneumonia in it.
- False Negative (FN): It is the incorrect prediction of the negatives, for example, there is no pneumonia in the image, but the model says an image is pneumonia one.

B. Precision

It checks how precise the model works by checking the correct true positives from the predicted ones.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

C. Recall

It calculates how many actual true positives the model has captured, labeling them as positives.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

D. Accuracy

Accuracy determines that how many true positives TP, True negatives TN, False positive FP and False negatives FN were correctly classified:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

E. F1-Score

F1-Score Precision: It checks how precise the model works by checking the correct true positives from the predicted ones.

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recal} \tag{4}$$

F. Receiver Operating Characteristic Curve (ROC-Curve)

The receiver operating characteristic curve (ROC-curve) represents the performance of the proposed model at all classification thresholds. AUC provides the area under the ROC-curve integrated from (0, 0) to (1, 1). It gives the aggregate measure of all possible classification thresholds. AUC has a range from 0 to 1is a 100% wrong classification. It is attractive for two reasons: first, it is scale-invariant, which means it checks how well the model is predicted rather than checking the absolute values; and, second, it is the classification threshold invariant as it will check the model's performance irrespective of the threshold being chosen.

V. RESULTS AND DISCUSSION

Our proposed model is designed and development to detect and classify pneumonia from chest X-ray images. It contains both image processing and convolutional neural network.We developed a model The algorithm begins by transforming chest X-ray images into sizes smaller than the original. The next step involves the identification and classification of images by the convectional neural network framework, which extracts features from the images and classify them. This work has presented the X-Ray images for Pneumonia detection based on convolutional neural networks and different machine learning. By training a set of strong CNNs on a large scale dataset, we built a model that can accurately predict Pneumonia.During each epoch data is trained over and over again to learn the feature of data. The performance evaluation of the model is estimated by using classification accuracy and cross-validation. We performed 5-fold and 10- fold cross-validation and presented the results in terms of mean and standard deviation(SD). Performances of seven Machine learning classification models are presented in

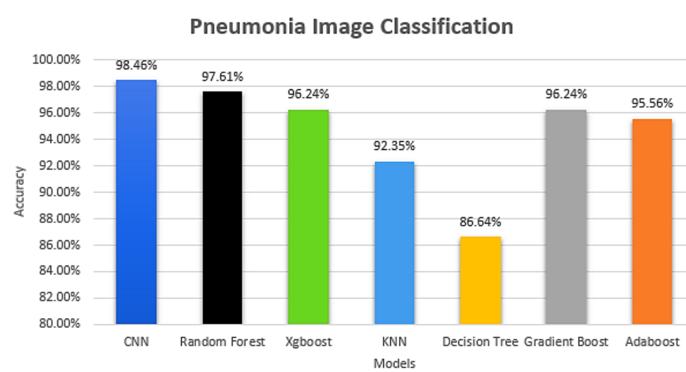

Fig. 9. Accuracy of Pneumonia Image Classification

The results presented in Figure 9, show that accuracy of Pneumonia Image Classification. It can be seen that all machine-learning models, except the decision tree. The highest accuracy was 98.46% to deep learning . On the other hand, the decision tree classifier achieved the lowest performance of 85.64%. An F-test carried out over these results confirms that the performance differences between predictive models are statistically significant. As shown in Figure 10, F-score of Pneumonia Image Classification.The average F- score of the classifier is all-more than 90%, which is all an excellent result. Therefore, in these results of the f-score, the classification accuracy of Deep Learning (DNN) with Feed Forward and Random Forest (RF) is approximately 98.95%, which is large as compared to other classifiers.

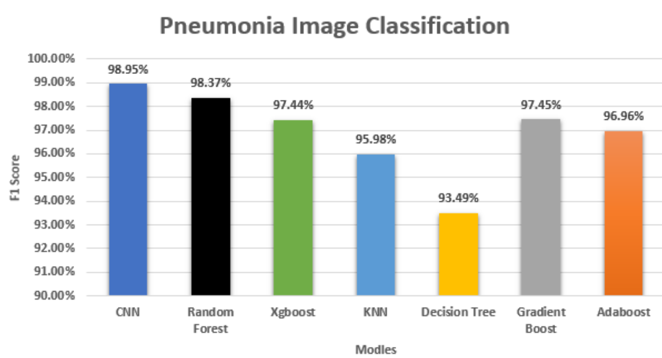

Fig. 10. F1-score of Pneumonia Image Classification

Two-performance metrics are displayed in Figure 9 and Figure 10 respectively. As shown in figures 9 and 10, the performances of the model showed better overall results with the convolutional neural network.Multi-label classification ROC curves is demonstrated for the seven classifiers as shown in Figure 11.We explained that machine-learning models based on the convectional neural network were able to achieve high sensitivity and specificity for classifying pneumonia from chest X-ray images. Generally, we developed the seven models used to predict pneumonia. Our results show that the Deep learning method had the most significant ability to predict pneumonia with an AUC of 98.48% that the overall performance of the model is excellent. In this paper, we have proposed an algorithm that combines the predictions of seven of the most regularly-used self-labeled algorithms, using a voting methodology. Machine learning based on a deep CNN is employed to improve the performance of diagnosing pneumonia in terms of accuracy achieved. Our preliminary numerical experiments present the efficacy of the proposed algorithm and its classification accuracy, therefore illustrating that reliable prediction models could be developed.

VI. Summary and Conclusions

In this paper, we offered a model for detecting and classifying pneumonia from chest X-ray images using machine learning methods based on Convolutional Neural Networks (CNN) .In particular, comparing all seven models based on test accuracy score, F-score and ROC curve, it appears that CNN outperforms all other models by a small margin with test accuracy score 98.46%. Random forest performs surprisingly well with test accuracy score 97.61%. It should note that the CNN model was run with only 100 epochs due to its long-running and waiting time. We believe, with proper parameter tuning and giving a little more time, the CNN model could

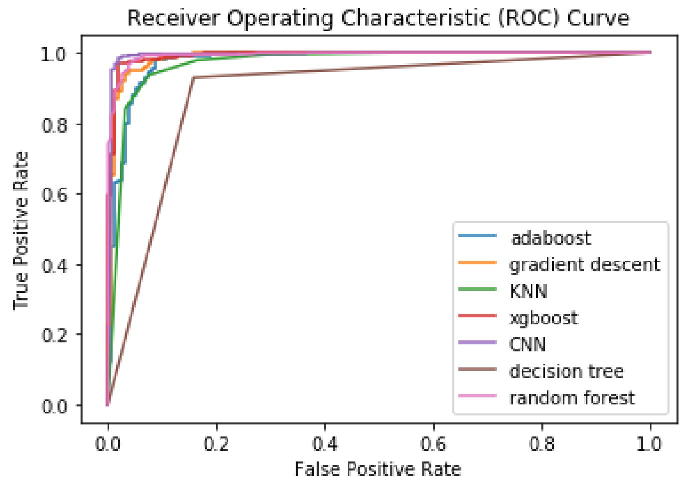

Fig. 11. ROC Curve

achieve an even better result. The result of our model indicates that our model outperforms other methods, which use no extra training data on chest radiography and shows that image training may be sufficient for general medical image recognition tasks. Moreover, the cooperation between a machine learning-based medical system and the detection of Pneumonia will improve the outcomes and bring benefits to clinicians and their patients.This study may use in similar research which will help to develop a diagnostic tool such as COVID-19 diagnostic. future, we will collect X-ray images from hospitals with big data to train and test the system to predict better results. Also, we have plan to work with more complex medical data .

References

[1] M. Roser, H. Ritchie, and B. Dadonaite, "Child & infant mortality," *Our World in Data*, 2013.

[2] C. Turkington and B. Ashby, *The encyclopedia of infectious diseases*. Infobase Publishing, 2007.

[3] R. T. Sousa, O. Marques, F. A. A. Soares, I. I. Sene Jr, L. L. de Oliveira, and E. S. Spoto, "Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs," *Procedia Computer Science*, vol. 18, pp. 2579–2582, 2013.

[4] step to healthy, "Symptoms of pneumonia and remedies," 2018. [Online]. Available: https://steptohealth.com/symptoms-of-pneumonia-complementary-natural-remedies/

[5] W. H. Organization *et al.*, "Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children," World Health Organization, Tech. Rep., 2001.

[6] T. Franquet, "Imaging of pneumonia: trends and algorithms," *European Respiratory Journal*, vol. 18, no. 1, pp. 196–208, 2001.

[7] T. Cherian, E. K. Mulholland, J. B. Carlin, H. Ostensen, R. Amin, M. d. Campo, D. Greenberg, R. Lagos, M. Lucero, S. A. Madhi *et al.*, "Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies," *Bulletin of the World Health Organization*, vol. 83, pp. 353–359, 2005.

[8] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, "Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[9] C. Combi and G. Pozzi, "Clinical information systems and artificial intelligence: Recent research trends," *Yearbook of medical informatics*, vol. 28, no. 01, pp. 083–094, 2019.

[10] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[11] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature medicine*, vol. 24, no. 9, pp. 1342–1350, 2018.

[12] S. Xu, H. Wu, and R. Bie, "Cxnet-m1: Anomaly detection on chest x-rays with image-based deep learning," *IEEE Access*, vol. 7, pp. 4466–4477, 2018.

[13] Z. Li, C. Wang, M. Han, Y. Xue, W. Wei, L.-J. Li, and L. Fei-Fei, "Thoracic disease identification and localization with limited supervision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8290–8299.

[14] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[15] C. X. ray Images dataset, "Chest x-ray images (pneumonia)," 2019. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumoniaIM-0001-0001.jpeg

[16] M. D. Twa, S. Parthasarathy, C. Roberts, A. M. Mahmoud, T. W. Raasch, and M. A. Bullimore, "Automated decision tree classification of corneal shape," *Optometry and vision science: official publication of the American Academy of Optometry*, vol. 82, no. 12, p. 1038, 2005.

[17] G. Kesavaraj and S. Sukumaran, "A study on classification techniques in data mining," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*. IEEE, 2013, pp. 1–7.

[18] N. Settouti, M. E. A. Bechar, and M. A. Chikh, "Statistical comparisons of the top 10 algorithms in data mining for classification task," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 1, pp. 46–51, 2016.

[19] S. Chen, R. Almamlook, and Y. Gu, "Offline handwritten digits recognition using machine learning."