

Pneumonia Classification in X-ray Images Using Artificial Intelligence Technology

1st Han Trong Thanh
School of Electronics and Telecommunications
Hanoi University of Science and Technology
Hanoi, VietNam
thanh.hantrong@hust.edu.vn

2nd Pham Huong Yen
School of Electronics and Telecommunications
Hanoi University of Science and Technology
Hanoi, VietNam
Yen.PH202228M@sis.hust.edu.vn

3rd Trinh Bich Ngoc
Faculty of Mechanical Engineering
National University of Civil Engineering
Hanoi, VietNam
ngoctb@nuce.edu.vn

Abstract— The article focuses on the research of image classification algorithms, namely the images indicate pathology of pneumonia caused by bacteria and viruses. The proposed method is based on using the VGG16, VGG19, DenseNet169 networks to extract data characteristics and train the model classification. The X-rays are classified including normal people, patients with viral pneumonia, and bacterial pneumonia. The provided source was medical data on chest X-ray images of patients who were manually classified by specialists. However, the accuracy of the classification is highly dependent on the number of images, the resolution of the images, and whether the X-ray image is correctly classified. In this study, the algorithms give relatively positive classification results with an accuracy of approximately 85%.

Keywords— *VGG16, VGG19, DenseNet169, AI (Artificial Intelligence), CNN (Convolution Neural Network).*

I. INTRODUCTION

Pneumonia is a problem of public health globally in recent times. The public health impact of this group of diseases is now more severe than HIV infection, malaria, cancer, and myocardial coloration. Over 30% of antibiotic costs both in the community and in the hospital are used to treat pneumonia. The increased tendency of antibiotic resistance among strains of bacteria that cause community pneumonia poses challenges in the treatment of diseases using antibiotics. In such a context, the identification of the causative agents of pneumonia is of utmost importance [1].

The application of artificial intelligence and image processing technologies to diagnose diseases from medical imaging is a very much-discussed area in which there is a classification of diseases based on X-rays images. From the chest X-ray image, the diagnosis and identification of lung diseases are carried out, then recommend the appropriate treatment. Artificial intelligence technology will help to classify diseases from X-ray images quickly and bring high accuracy in disease diagnosis. Classification of diseases based on X-rays was not too difficult with high precision due to the advent of GPUs (Graphics Processing Units) and image processing based on artificial intelligence (AI).

This research focuses on the employment and development of algorithms using Convolutional Neuron networks (CNN) which is an advanced Deep Learning model to automatically classify normal people and patients with pneumonia caused by agents such as bacteria, viruses. In the

paper, three CNN models used to train the model is VGG16, VGG19, and DenseNet 169. The dataset includes 5864 JPEG images of the lungs in JPEG format hand-classified by specialists into three categories: normal people, X-rays of viral pneumonia, and X-rays of bacterial pneumonia. Comparing all results will evaluate the effectiveness of each model.

The paper is organized as follows. Part II introduces an overview of pneumonia and X-ray images of pneumonia caused by different agents. Part III gives a summary of the algorithms and implementation model. Part IV presents performance results and provides an assessment for each case. Conclusion and development direction in part V.

II. PNEUMONIA AND X-RAY OF PNEUMONIA

A. Pneumonia

Pneumonia is a subclinical disease caused by damage to the lung organs (alveoli, interstitial connective tissue, and terminal bronchioles) such as an inflamed lung, which mainly affects the small air sacs called alveoli. Pneumonia is often caused by infections caused by multiple agents, such as viruses, bacteria, fungi, and other microorganisms such as parasites is less common. Besides, pneumonia can also be caused by toxic chemicals.

Pneumonia affects about 450 million people globally each year and makes about 4 million deaths. In the XIX century, pneumonia was considered by William Osler as "The captain of the men of death", the advent of antibiotic therapy and vaccines in the XX century saved many people. However, in developing countries, pneumonia remains the leading cause of death for the elderly and children.

The detection of pneumonia pathogen to suitable treatments is an extremely important problem in which the X-ray is the most important clinical test in the diagnosis of lung diseases. When reading a chest X-ray, the basal lung lesions will be identified and then the association of that damage to the surrounding organs can be known. Careful analysis of these associations will provide an accurate diagnosis of the disease.

B. Viral Pneumonia

Viral pneumonia can be caused by many types of respiratory viruses but mainly by influenza viruses and respiratory syncytial viruses. The disease usually appears in



Fig. 1. Viral pneumonia wounded a net shape (blurry ray).

the cold season in densely populated communities. Basic lesions are necrosis, bronchial epithelialization, and increased mucus secretion. Mucus and necrotic matter form the mucus nodes in the bronchi. The lung parenchyma is damaged by inflammation spreading from the bronchioles to the surrounding area (interstitial organization, alveoli). After a week, the alveoli are filled with secretions and white blood cells. Perioral bronchial mono-lymphatic infiltrates, in the interstitial organization of the alveolar ductus area and the alveolar wall. Intracellular inclusions can be found in alveoli.

Chest X-ray images of viral pneumonia (in Fig. 1) are common:

- Thick bronchial wall
- Shades around the bronchi (blurred mirror image)
- Blurred rays around the hilum
- Blurred buttons are movable
- In children, images of "shallow air" (hypertonic region), atelectasis (respiratory syncytial virus and other viruses) may be occurred.
- In adults, diffuse bilateral infiltrates (acute respiratory distress syndrome in adults) can be seen in influenza viral pneumonia, Adenovirus.
- Pleural effusion can be seen in pneumonia caused by Adenovirus [2].

C. Bacterial pneumonia- pneumococcal bacteria

Lesions usually occupy an entire lung, while the rest of the lung is completely normal. Pneumococcal pneumonia can also damage multiple foci of bronchitis but is uncommon. Image of alveolar inflammation of fibrin – leukocytes (in Fig. 2, Fig. 3, Fig. 4), progressing through stages:

- Congestive phase (the first day): the most prominent is the congestion of the pulmonary capillaries and the secretion of fibrinolytic secretions containing less white blood cells into the alveolar lumen.
- Stage "red calcification" (days 2 - 3): the cross section of the lung inflammation is dark red, similar to the color of the liver. The alveoli are full of colloidal secretions with a lot of fibrin, erythrocytes, pneumococci, a moderate number of polymorph nuclear leukocytes, and a few monocytes. The pleural lesion is also acutely inflamed with fibrin patches in the visceral leaf.
- Stage "gray calcification" (days 4 - 5): cross-section of the lung is light gray, alveolar lumen contains less red blood cells, but there are many neutrophils. This phase lasts 3 - 4 days, then it moves to the absorption phase. Fibrin secretions are liquefied by enzymes released from granulocytes. This liquid is phagocytosed by alveolar macrophages. Lung organization is completely restored.



Fig. 2. Right middle lobe pneumonia by pneumococcus



Fig. 3. Pneumococcal pneumonia right middle lobe



Fig. 4. CT-Scan image of pneumococcal lobe pneumonia (bronchial image in the bunch of fuzzy)

Chest X-ray: typical and popularity image is a bunch of fuzzy occupy the entire lung lobes, bronchial gas, rarely necrosis or cavernous. The bunch of fuzzy may be unclear in patients who are highly dehydrated or sometimes see lots of congestion, pleural effusion. X-ray symptoms are usually absorbed after 4 weeks [2].

D. Bacterial pneumonia - staph

Pneumonia contains many foci, the center of inflammation is bronchi or bronchioles necrosis, hemorrhage as Fig. 5 and Fig. 6. Inflammatory foci are pus-filled with neutrophils, pus cells, edema, hemorrhage.

When the inflamed foci rupture creates abscesses. The alveolar wall is destroyed in severe cases. Air enters the destroyed alveoli but cannot get out, creating the thin air sacs (pneumatoceles) characteristic of staph pneumonia, common in children.

Chest X-ray: a common picture is pneumonia round, irregular size, on both sides of the lung, and asymmetric.

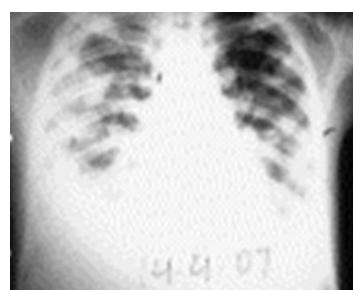


Fig. 5. Staphylococcal pneumonia, multiple infiltrative lesions, abscesses in both lungs, right pleural effusion

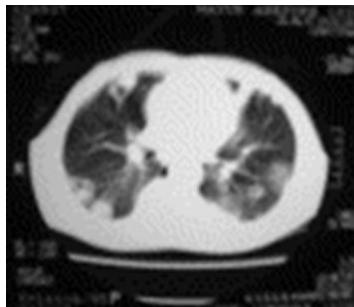


Fig. 6. CT-Scan image shows many abscesses, peripheral infiltrates, close to the visceral pleura

Some abscesses include gas - water. Thin wall pneumatoceles are common in children. Lesions change rapidly [3].

Besides, bacterial pneumonia has many other causes due to streptococci, bacilli ... but less common

III. CNN NETWORKS

Convolutional Neural Networks (CNN) is one of the most popular and most influential deep learning models in the computer vision community. CNN is used in many problems such as image recognition, video analysis, MRI images, or in the field of natural language processing and most of them deal well with these problems [4] [5].

A. CNN and Layers

CNN includes a set of basic classes like the convolution layer, nonlinear layer, pooling layer, fully connected layer, and layers linked together according to the determined order [6] (Fig.8). Firstly, the image goes through the convolution layer and the nonlinear layer, then the computed values go through the pooling layer to reduce the number of operations while preserving the characteristics of the data. Convolution layers, nonlinear layers, and pooling layers can appear one or more times in a CNN network. Finally, the data is passed through fully connected and SoftMax to calculate the probability of object classification.

Convolution Layer: Convolution Layer is the basic and most important component to create a CNN. The Convolution layer is a hidden layer in the neural network. A filter matrix - Convolution Filter will be scanned through input data matrix from left to right, top to bottom; Each value of the input matrix will be multiplied corresponding to this filter matrix and summed up. The result of each of these matrix multiplication is a specific number. This new set of result numbers in turn forms a new matrix, called the feature map. The feature maps will then be lumped together and become the final output after the convolution layer.

With input image matrix of size $[W_1 \times H_1 \times D_1]$, the output size is $[W_2 \times H_2 \times D_2]$ calculated:

- $W_2 = (W_1 - F + 2P) / S + 1$
- $H_2 = (H_1 - F + 2P) / S + 1$
- $D_2 = K$

Where

- F: size of filter or Kernel filter ($N \times N$)
- S: Stride value
- P: Number of zero-padding added to the image border
- K: Number of the image filter

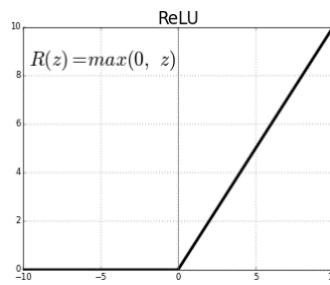


Fig. 7. ReLU Function

ReLU (Rectified Linear Unit) is widely used because of its simple calculation, helping the training of Deep Networks many times faster. The ReLU function has the math formula of $f(x) = \max(0, x)$ in Fig. 7 and is an important computational function because real-world data wants to be learned as non-negative linear values.

Max pooling Layer: Pooling is used to reduce the size of the matrices, thereby reducing computation time. Similar to the Convolution layer, it also includes a filter that slides through each value of the input matrix. However, this filter only selects a value from the values it passes to keep. The most commonly used type is max pooling. At that time, the maximum value will be kept.

An input image matrix of size $[W \times H \times D]$, the output size obtained after pooling layer is calculated in (1) **Error! Reference source not found.** as:

$$\text{output} = \frac{(W - F + 1) * (H - F + 1) * D}{S} \quad (1)$$

Where:

- F size of filter or Kernel matrix
- S is stride value

Full connected Layer: is a layer that aggregates the characteristics of the data that has been trained through the previous layers into vector form, or in other words, it is a layer of found feature layers, converting data from **3D** or **2D** to **1D**, means 1 vector left. There may be 1 or more FC layers before the final FC output, the number of neurons of the FC output depends on the number of outputs wanting to synthesize.

B. VGG16 Network

VGG came out with some improvements, firstly the deeper learning VGG Model, followed by a change in the order of the Convolution layer. VGG uses a sequence of Convolution consecutively Conv-Conv-Conv in the middle and end of the VGG architecture. This will make the calculation longer but the Features will still be retained more than using Max-Pooling after each Convolution. VGG16 is proposed by K. Simonyan and A. Zisserman, in University of Oxford; is a convolutional neural network that is 16 layers deep. The most unique thing about VGG16 is that instead of having a large number of hyper-parameter they focused on having convolution layers of 3×3 filter with a stride 1 and always used the same padding and max-pool layer of 2×2 filter of stride 2. In the end it has 2 full-connected-layers followed by a SoftMax for output [7].

In Fig. 9, the default input image size of VGG-16 is $224 \times 224 \times 3$. After each pooling layer, the size of the feature map is reduced by half. The last feature map before the fully connected layers is 7×7 with 512 channels and it is

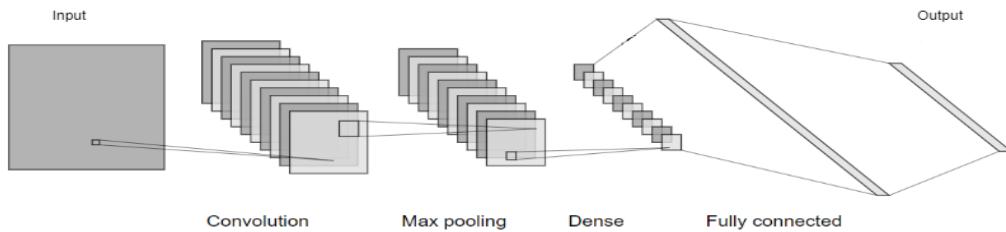


Fig. 8. A basic CNN structure model

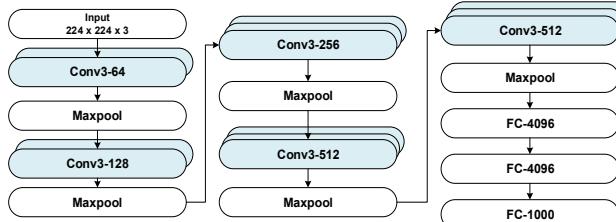


Fig. 9. VGG16 Network model

expanded into a vector with 25,088 ($7 \times 7 \times 512$) channels. Finally, it is synthesized through 2 layers fully connected 4096 nodes, and produces an output of 1000 specific nodes. Because the input size decreases through convolution but the computation depth increases, so the VGG network in general and VGG16 in particular are often used to classify and identify objects.

C. VGG19 Network

Basically, the VGG19 network model has the same structure as VGG16 but has 3 more convolution sets in the last 3 layers (forming 4 conv-stack), so VGG-19 is a convolutional neural network that is 19 layers deep in Fig. 10. As a result, the network has learned rich feature representations for a wide range of images.

The addition of three more conv3x3 sets in the last 3 layers helps the VGG19 network learn more deeply and increase the accuracy when classifying and identifying objects, but the training time will be longer than VGG16.

D. DenseNet Network

DenseNet (Dense Connected Convolutional Network) is one of the newest networks for visual object recognition. DenseNets are built from dense blocks and pooling operations, where each dense block is an iterative concatenation of previous feature maps in Fig. 11. This architecture is similar to Resnet, but with a few differences:

- Parameter efficiency, DenseNets are more efficient in the parameter usage
- Implicit deep supervision, DenseNets perform deep supervision thanks to short paths to all feature maps in the architecture (similar to Deeply Supervised Networks)
- Feature reuse, all layers can easily access their preceding layers making it easy to reuse the information from previously computed feature maps [8].

With traditional CNN, if there is L layer, there will be L connection, while in DenseNet there will be $L(L + 1)/2$ connection [9].

A **Dense Block** is a module used in convolutional neural networks that connects all layers (with matching feature-map sizes) directly with each other. It was originally proposed as

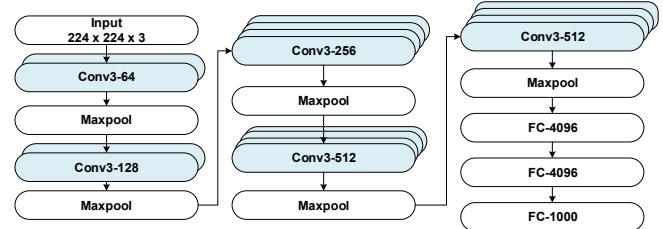


Fig. 10. VGG19 Network model

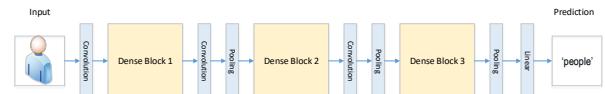


Fig. 11. DenseNet architecture

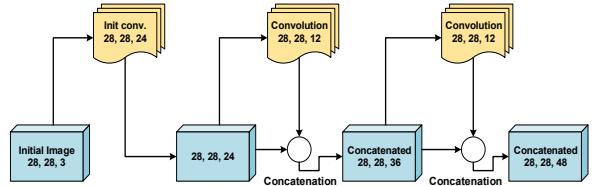


Fig. 12. The conv-layer with concatenated results

part of the DenseNet architecture. To preserve the feed-forward nature, each layer obtains additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers. At each dense block, there will be normalization, nonlinearity, and dropout. To reduce the size and depth of the feature, the transition layer is placed between dense blocks, it consists of conv kernel size = 1, average pooling (2×2) with stride = 2, reducing the output to $[14 \times 14 \times 48]$ while the input image size is $[28 \times 28 \times 3]$ in Fig. 12.

IV. IMPLEMENTATION OF MODELS

With the mentioned CNN networks, they are used to train the model and classify the data into 3 classes as normal people, viral pneumonia patients, or bacterial pneumonia patients.

A. Training model

In the study, the available data of 5864 JPEG images [10] have been divided into 3 folders including: training set 5216 images, validation set 24 images, test set 624 images and in each folder there will be corresponding types (Class_01: corresponding to data of patients who are normal, Class_02: corresponding to data of bacterial pneumonia, Class_03: corresponding to data of viral pneumonia). (Detail in Fig. 13).

Perform model training according to three-CNN networks: DenseNet169, VGG16, VGG19. The CNN network structure and details of the layers of each network

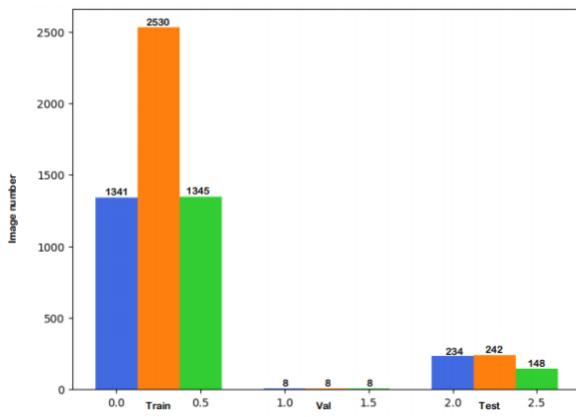


Fig. 13. The number of images in the folder

were written in Chapter III of the study. Here will perform deployment training with $\text{epochs} = 20$ and $\text{batch_size} = 16$.

Training model using DenseNet169 network, the error rate and the accuracy of the model through each epoch are monitored and statistically recorded (in Fig. 14, Fig. 15). Towards the end of the epoch of the model error rate decreases and improved accuracy compared to the first epoch. Training model using VGG16 network with tracking graph of error rate and accuracy rate is shown in Fig. 16, Fig. 17, Fig. 18, Fig. 19, respectively.

During the training and testing model, the Densenet169 network has the fastest training time 38 minutes, but in the last epoch the accuracy is only 75.57%. While the VGG16 network and VGG19 network all achieved the corresponding results of over 80%, though the model training time is longer.

In order to evaluate the accuracy of the classification models, in addition to using accuracy, the $F1 - score$ parameter [11] also is employed. The $F1 - score$ is an overall statically measurement that combines and balances both the precision and recall values, where $Precision$ is the ratio of how much of the predicted is correct and $Recall$ is The ratio of how many of the actual labels were predicted. The basic idea is to compute all precision and recall of all the classes, then $F1_score$ for the entire model is average them to get a single real number measurement.

With the problem of classifying into multi-classes as in this study, $Precision$ and $Recall$ are calculated in the (2) (3) by using the confusion matrix [12]. A confusion matrix is a way of classifying true positives, true negatives, false positives, and false negatives when there are more than 2 classes. This matrix has columns representing actual values and rows representing the predicted values.

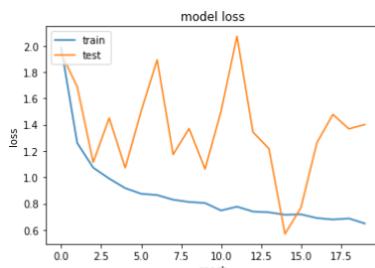


Fig. 14. Error rate per epoch of DenseNet169

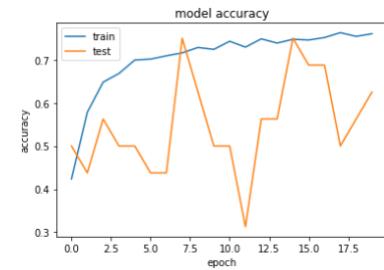


Fig. 15. Accuracy rate per epoch of DenseNet169

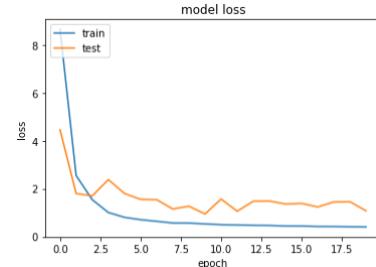


Fig. 16. Error rate per epoch of VGG16

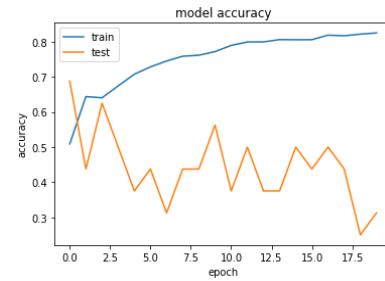


Fig. 17. Accuracy rate per epoch of VGG16

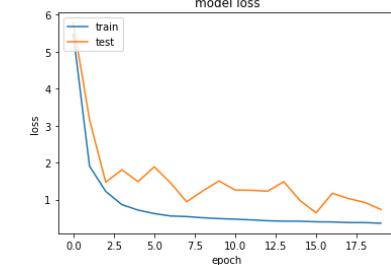


Fig. 18. Error rate per epoch of VGG19

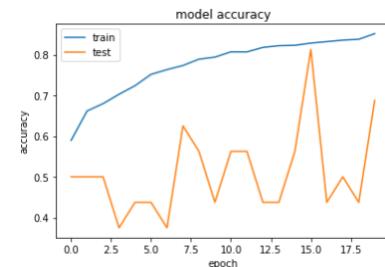


Fig. 19. Accuracy rate per epoch of VGG19

$$Precision = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \cap h(x_i)|}{|h(x_i)|} \quad (2)$$

$$Recall = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \cap h(x_i)|}{|Y_i|} \quad (3)$$

Where n is the number of class; Y_i is the ground truth label assignment of the i^{th} class; x_i is the i^{th} class; $h(x_i)$ is the predicted labels for the i^{th} class.

TABLE 1. RESULTS OF DENSENET169, VGG16 AND VGG19 NETWORKS

Net	Train Accuracy	Train times (min)	Test Precision	Test times (min)	Test Recall	F1-score
Densenet 169	76.57%	38	78.19%	1	78.65%	78.42%
VGG16	82.78%	64	79.88%	1.5	82.17%	81.01%
VGG19	85.25%	71	83.14%	1.5	83.05%	83.09%

Therefore, the **F1 – score** parameter is calculated as:

$$F1_score = \frac{P \times R}{P + R} \quad (4)$$

The classification results of networks with the training, test precision, test recall and **F1 – score** values are given in TABLE 1. In the training, prediction and recall steps, the VGG19 network has the highest results, respectively, 85.25%, 83.14% and 83.05%, the other algorithms achieve lower results in turn: VGG16 and DenseNet169. From these results, the F1-score parameter of the VGG19 achieved the highest result up to 83.09% while the DenseNet169 achieved the lowest result of 78.42%. Based on the execution time of the networks (evaluated above), it can be seen that all three networks have a fairly accurate classification, but VGG19 network is the most optimal and accurate network.

B. Models experiment

To test the model with real data, 10 X-rays of a normal person, 10 X-rays of a patient with viral pneumonia, and 10 X-rays of a patient with bacterial pneumonia were selected. Then, the trained models are used and the weights of the nodes are also saved in **hdf5** file format. Experimental results for each model are shown in TABLE .

By including X-ray images of the patient's lungs in the diagnosis, it is possible to verify the accuracy of the model after training. The results show that the training process is quite good and the disease classification is accurate. Although the medical error is very dangerous, with relatively little medical data and often confidential medical information, the model stops at an acceptable level. From the results achieved at present, the future can aim to build diagnostic models with higher accuracy.

TABLE 2. TEST THE THREE-MODELS

	NOMAL	BACTERIA	VIRUS
Densnet169	10/10	8/10	9/10
VGG16	10/10	9/10	9/10
VGG19	9/10	8/10	9/10

V. CONCLUSION AND DEVELOPMENT

The study focuses on the application of image processing technologies in the classification of normal people and patients with pneumonia due to agents such as

viruses or bacteria. To build a disease classification model, three different CNN networks are used for the training process. The results show that the model works quite effectively and accurately for different diseases. From the results obtained at present, the next development direction of the topic is to complete and develop the model so that it can be applied in practice such as:

- Increase the accuracy and speed up of models
- Develop applications that can be applied to medical examination and treatment in hospitals
- Develop diagnostic models for other diseases, towards the widespread application of AI in the medical field.

REFERENCES

- [1] T. D. T. M. Ha, "Viêm phổi cộng đồng và viêm phổi liên quan đến chăm sóc y tế do các tác nhân vi sinh phát hiện bằng real-time PCR đám", 01/2019.
- [2] "Viem phoi do virus," 10/2015. [Online]. Available: benhvien103.vn. [Accessed 9 2020].
- [3] S. P. a. G. J. Kannoja, "Ensemble of hybrid CNN-ELM model for image classification", in *5th International Conference on Signal Processing and Integrated Networks (SPIN)*, 2018.
- [4] S. e. a. Pang, "A novel fused convolutional neural network for biomedical image classification", in *Medical & biological engineering & computing* 57.1, 2019.
- [5] U. Karn, "An intuitive explanation of convolutional neural networks", 2016. [Online]. [Accessed 10 2020].
- [6] X. e. a. Zhang, "Accelerating very deep convolutional networks for classification and detection", in *IEEE transactions on pattern analysis and machine intelligence* 38.10, 2015.
- [7] "Step by step VGG16 implementation in Keras for beginners," 2019. [Online]. Available: <https://towardsdatascience.com>. [Accessed 10 2020].
- [8] S. e. a. B. Jégou, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation", in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017.
- [9] G. e. a. Huang, "Densely connected convolutional networks", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
- [10] "kaggle," 2017. [Online]. Available: <http://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>. [Accessed 9 2020].
- [11] C. a. E. G. Goutte, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation", in *European conference on information retrieval*, Springer, Berlin, Heidelberg, 2005.
- [12] M.-L. a. Z.-H. Z. Zhang, "A review on multi-label learning algorithms", *IEEE transactions on knowledge and data engineering* 26.8, pp. 1819-1837, 2013.