Application of Convolutional Neural Networks Method in Predicting Pneumonia Diseases Based on X-Ray Test Results

Abstract. Pneumonia is an infectious disease that attacks the lungs so that the air sacs in the lungs are experiencing inflammation and swelling [1,3]. Pneumonia is usually often known as wet lungs. Pneumonia can attack all walks of life, both adult as well as toddlers. From the phrase Dr. Christina Widaningrum as the head of Sub Directorate (Kasubdit) of Acute Respiratory tract infections (RESPIRATORY) Ministry of health (Kemenkes), Indonesia was ranked 10th in the world in the case of death from pneumonia [2]. To find out the condition of a healthy lung or not from someone patient needs to do an x-ray. However, based on the results of the x-ray lung conditions or exposed to pneumonia by bacteria or viruses can we know based on the explanation from doctors in their field. Therefore, this study is intended to help doctors and patients so that knowing more clearly and more confident again that picture of lung x-ray results included into the category of a healthy lung or exposed to pneumonia by bacteria or viruses. To know it, can be assisted with the use of method or algorithm of Convolutional Neural Networks (CNNs). Why should we use CNNs method? Because CNNs is a deep neural network that's suitable for image data. And the other side, CNNs has a good accuracy value for classification image to predict something. Later, we use the image data from the patient's lung x-ray results so that the CNNs will work by finding an image that corresponds to the category of the lungs of someone who belongs into the lungs are healthy or are exposed to pneumonia because bacteria or viruses. And based on the score showed that the performance of CNNs attained an accuracy rate of 97%.

Keywords: *Pneumonia, Convolutional Neural Networks, X-Ray*

1. Introduction

Disease is an impairment of the normal state of human being that interrupts or modifies its vital functions [5]. Diseases that can affect anyone, both adults, toddlers, and even the elderly. Diseases also have many types, so you need different handlers according to the disease. In almost every country, disease is an important problem, ranging from mild, moderate to severe diseases. One country that has a problem with disease is Indonesia. Indonesian people suffer from this disease. Pneumonia is one of the dangerous diseases that causes the death rate of the Indonesian population to increase.

Pneumonia itself is an infectious disease that attacks the lungs so that the air sacs in the lungs experience inflammation and swelling [1,3]. Pneumonia is commonly known as wet lung. This disease can affect all people. However, the mortality rate due to pneumonia is more dominant in infants and toddlers. This

is in accordance with the expression of the Head of the Sub-Directorate (Head of Sub-Directorate) of the Acute Respiratory Tract Infection (ISPA) of the Ministry of Health, Dr. Christina Widaningrum M.Kes. He said Indonesia was ranked 10th in the world in the case of an attack of a toddler pneumonia killer virus, with 15.5% per 2015 or 554,650 cases of pneumonia. The cause of someone getting pneumonia is through various kinds, some through viruses or bacteria.

Based on Ms. Christina's explanation, this number should not be underestimated. To reduce the death rate from pneumonia, cooperation between the government and doctors is needed. In predicting pneumonia experienced by patients, doctors need to carry out several examinations to diagnose patients. Examination that needs to be done by the patient is a physical examination and other investigations. One of the examinations that can be used as a diagnosis by doctors in predicting patients who have pneumonia or not is by x-ray. Because with the results of the x-ray test, you will see a picture of the shape of the patient's lungs, so that the doctor can provide the right diagnosis to the patient regarding pneumonia. To clarify and help doctors predict the lungs of patients who are healthy or have pneumonia by viruses or bacteria, an accurate method is needed. This method is a CNNs (Convolutional Neural Networks). Why should we use CNNs method? Because CNNs is a deep neural network that's suitable for image data. And the other side, CNNs has a good accuracy value for classification image to predict something. Usually in conducting image classification can be used SVM method, but if we used CNNs method the accuracy value is better than SVM method. And then, this method is suitable to classify images of patients' lungs from x-ray tests, whether the results included in the lungs that are healthy or exposed to viral or bacterial pneumonia appropriately. Therefore, this method can help doctors predict pneumonia suffered by their patients.

2. Research Methodology

2.1. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) is a development of multilayer perceptron (MLP) that is designed to process two-dimensional data in the form of images [8]. CNNs is included in the type of Deep Neural Network because of the high network depth and much applied to image data [7]. Technically, CNNs is an architecture that can be trained and consists of several stages. Input (input) and output (output) of each stage consists of several arrays commonly called feature maps. Each stage consists of three layers, namely the convolution of the main layer and the most important to use. And Then proceed with the pooling layer is a composite layer that is used to extract the average value or maximum value in the pixel section and the layer activation function.

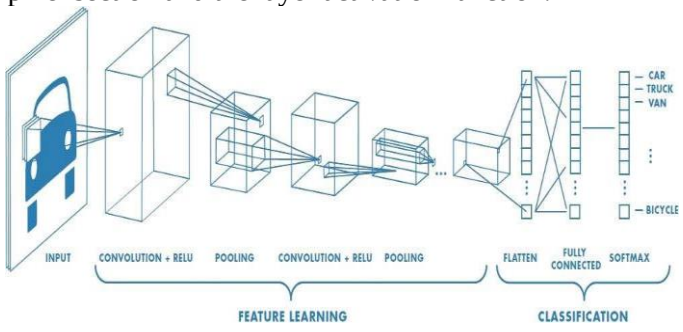

Figure 1. Illustration of CNNs

Different volumes are owned by each input layer that will be entered and will be represented by depth, height and width. The results will be obtained in the form of a number that varies because it's depended by the results of filtering the previous layers and also the number of filters used.

2.1.1 Operation of Convolution

Convolution layer is part of the stage on CNNs architecture. This stage performs a convolution operation at the output of the previous layer. This layer is the most basic main process of the CNN network

architecture. Convolution is a mathematical term where the application of a function to the output of another function is repeated. Convolution operations are operations on two functions of real value arguments. This operation uses the output function as a Feature Map from the input image. The input and output obtained can be seen as two real value arguments. Convolution operations can be written as follows:

$$s(t) = (x * w)(t) \tag{1}$$

where,
$S(t)$ = Single output of future map
x = Input
2 = Weight/Kernel

The Feature Map is the result or single output given from function $s(t)$. The argument consists of two, the first argument (x) is the input and the second argument (w) is the kernel or filter. When seen as input as a two-dimensional image, it can be said t as pixels and can be replaced with i and j. Therefore, operations for convolution to inputs with more than one dimension can write as follows:

$$S(i,j) = (K * I)(i,j) = \sum\sum I(i-m, j-n)K(m,n) \tag{2}$$

Equation above is a basic calculation in a convolution operation where i and j are the pixels of the image. The calculation is commutative and K appears as kernel, I as input and kernel can be reversed relative to input. Or, convolution operations can be seen as multiplication matrices between input images and kernels with the output can be calculated by dot product.

2.1.2 Convolutional Layer

Convolutional Layer consists of neurons arranged in such a way as to form a filter with length and height (pixels) [6]. In accordance with the image channel below, where the first layer in the feature extraction layer is Convolutional Layers with a size of $5 \times 5 \times 3$. The purpose of the $5 \times 5 \times 3$ size is 5 pixels in length, 5 pixels in height and 3 in thickness. These filters will be shifted to all parts of the image. The output of the filter shift is called activation map or known feature map, by performing a "dot" operation between the input and the value of the filter. Consider the following illustration:

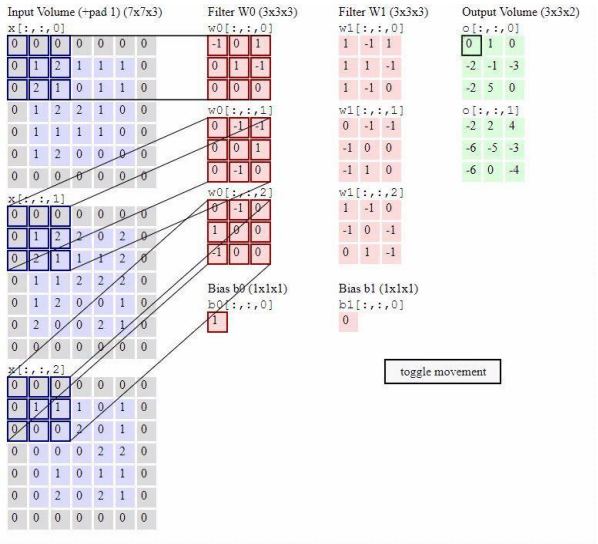

Figure 2. Illustration of Feature Map

The next is stride which is a parameter in determining how many filter shifts in the convolutional layer size. If the step value taken is 1, then the convolutional filter will shift horizontally first as much as 1 pixel, continue vertically as much as 1 pixel. However, the selection of steps or strides has no rules. But when using the smaller stride, more information obtained from an input. However, if the selected stride is smaller, the calculation of “dot product” between the input and the value of the filter is more than

choosing a larger stride. In addition, the small selection of strides does not necessarily get good performance.

Padding or Zero Padding is a parameter that determines the number of pixels (if seen in figure 2. contains a zero number) that will be added to each input side. This is used to manipulate the output dimensions of the convolutional layer (feature map). The purpose of using padding is that the dimensions of the output in the convolutional layer will always be smaller than the input (except, if we use of 1×1 filter with stride 1). The output will be reused to input for the convolutional layer, so that more information will be discarded. With padding, we can adjust the output dimension so that it can remain the same as the input dimension or so that it is not too reduced. So we can use convolutional layers with more successful extracted features. With padding too, the performance of the convolutional filter model will increase and focus on the actual information that is around the zero padding area. In Figure 2., the 5×5 input is the real dimension. If convolution is done with a 3×3 filter and stride 2 is chosen, a 3×3 feature map will be obtained (more information is built). To calculate the dimensions of a feature map, we can use formulas:

$$output = \frac{W-N+2P}{S} + 1 \quad (3)$$

where,
$W =$ long or high of input
$N =$ long or high of filter
$P =$ as a padding
$S =$ as a stride

After convolutional layer, pooling layer is usually done. In principle, the pooling layer consists of a filter with several sizes and stride that moves to follow all areas of the feature map. Usually pooling is often used, namely; max pooling and average pooling. For example, if the max pooling is 2×2 which will be used with stride 2, then for each filter move, the pixel area measuring 2×2 is chosen by its maximum value. If the value chosen is the average value, this is called average pooling. Consider the following illustration:

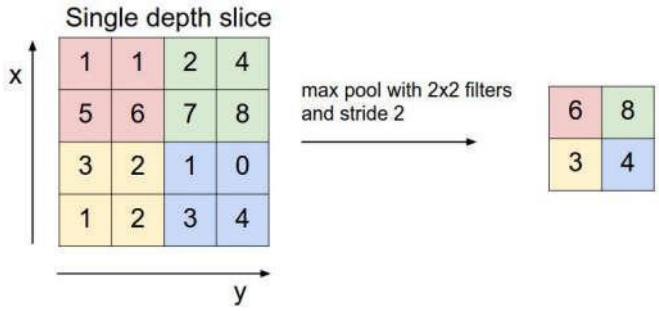

Figure 3. Illustration of Max Pooling

The purpose of using the pooling layer is to reduce the dimensions of the feature map (the other words down sampling), so that it accelerates the computational process because the updated parameters will decrease and overfitting.

2.1.3 Fully-Connected Layer

The feature map is created based on the feature extraction layer, which is still a multidimensional array, so it needs to be flattened or reshaped in the vector feature map which will be used as input from the fully-connected layer [6]. All layers that activate neurons from the previous layer are connected to neurons in the next layer (like a neural network) usually called the fully-connected layer. To be able to connect with all neurons in the layer, each activation of the previous layers needs to be converted first into one-dimensional data.

Fully-Connected layer usually uses a method called Multi-Perceptron Layer. This method has a target in the process of data so that it can carry out classification. The difference between fully-connected

layers and convolution layers is the neurons from the convolution layer are connected to a specific input area, while the fully-connected layer has neuron connections to almost all parts. However, the two layers continue to do “dot product” operations, so the function is not much different.

3. Experimental Results and Discussions

3.1. Dataset

In this dataset, image data were used where the image data was obtained from Guangzhou Women and Children 's Medical Center, Guangzhou. The total images used were 329 images, where 150 samples of lung x-ray test were in the healthy category, 111 samples of lung x-ray test were categorized as bacterial pneumonia and 68 samples of lung x-ray were categorized as disease viral pneumonia. Here are some examples of pictures of the patient's x-ray test:

Figure 4. Normal Lungs

Figure 5. Bacterial Pneumonia

Figure 6. Viral Pneumonia

In each category, 80% of all images are used as training data and 20% as testing data.

3.2 Model

The method used is CNNs in Tensorflow, where Tensorflow is a framework of programs or packages that can help CNNs systems in processing image data. In the chosen CNNs model, 3 convolutional layers are used, namely:

1. First Layer use 32-3 x 3 filters
2. Second Layer use 64-3 x 3 filters
3. Third Layer use 128-3 x 3 filters

Each of these 3 layers will use max-pooling of 2 x 2. After running this CNNs program, predictive accuracy will be seen on the CNNs program.

3.3 Results and Evaluation

With image data that has been obtained as explained in the dataset section, 329 are classified into 3 classes with the output as follows:

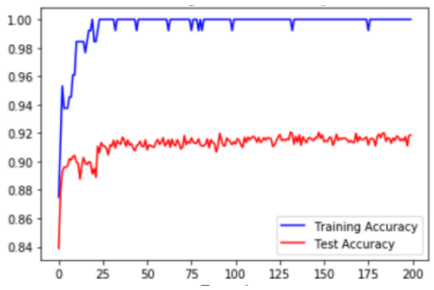

Figure 8. Training and Test Accuracy

Figure 9. Training and Test Loss

From the two plots that know the graph above, we can see in figure 8, the accuracy of the tests almost becomes stagnant after 25-50 epochs and rarely increases in certain epochs. At first, validation accuracy increases linearly with loss, but then does not increase too much.

From loss validation shows that from the two images above there is a sign of overfitting, similar to the accuracy of the test which is linearly decreased but after 25-30 epochs, there is an increase which means that the model tries to memorize the data and succeed.

4. Conclusion

With the Convolutional Neural Networks (CNNs) method we can classify which results of the x-ray test which are the lungs classified as healthy and exposed to pneumonia due to bacteria and viruses. So, this method can help doctors provide the right beliefs and predictions to patients. This is evidenced in the evaluation obtained in Epochs 25 and so on, the accuracy of 0.97 to 1.00 is obtained, meaning that the program from the CNNs method is able to accurately identify the patient's x-ray test images, even though at epochs the accuracy rate is unstable because the program from the method CNNs is trying to memorize the data set image.

5. Acknowledgement

This research is supported financially by University of Indonesia, with a DRPM-UI PITTA 2018 research grant scheme.

REFERENCES

[1] Healthline https://www.healthline.com/health/pneumonia Accessed: 1 June 2019.

[2] Viva https://www.viva.co.id/gaya-hidup/kesehatan-intim/852613-pneumonia-di-indonesia-peringkat-10-dunia Accessed: 1 June 2019.

[3] Alodokter https://www.alodokter.com/pneumonia/penyebab Accessed: 1 June 2019.

[4] Kementrian Kesehatan Republik Indonesia http://www.depkes.go.id/article/view/410/pneumonia-penyebab-kematian-utama-balita.html Accessed: 1 June 2019.

[5] Ecyclopedia Britannica https://www.britannica.com/science/human-disease Accessed: 1 June 2019.

[6] Li H, Lin Z, Shen X, Brandt J and Hua G 2015 A Convolutional Neural Network Cascade for Face Detection *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA, USA) pp. 5325-5334.

[7] Tuti P, Imania A A and Pertiwi B U 2019 Image Classification of Golek Puppet Images using Convolutional Neural Networks Algorithm *International Journal of Advances* **11** 34-45

[8] Muhamad Y, Budhi I and Casi S 2019 Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry's Nail *Journal of Physics: Conference Series* **1201**

[9] Krizhevsky A, Sutskever I and Hinton G. E. 2015 ImageNet Classification with Deep Convolutional Neural Networks *Advances in Neural Information Processing Systems* pp. 1907-1105.

[10] Chellapilla, K, Puri S and Simard P Y 2006 High-Performance Convolutional Neural Networks for Document Processing *Proc. of the 10th Int. Workshop on Frontiers in Handwriting Recognition*.

[11] Kermany, Daniel; Zhang, Khang; Goldbaum, Michael. 2018 Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. Mendeley Data, v2.

[12] Cengil E and Çonar A 2016 Robust Face Detection with CNN *Proceedings of the 4nd ICAT International Conference on Advanced Technology & Sciences* pp. 47-51.

[13] S C Turaga, J F Murray, Viren, F Roth, M Helmstaedter, K Briggman, W Denk and H S Seung 2010 Convolutional networks can learn to generate affinity graphs for image segmentation.*Neural Computation* **22** 511–538.