OPEN ACCESS
Indo-JC Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 5, Issue. 2, September 2020. pp. 43-54
doi:10.21108/indojc.2020.5.2.454

Pneumonia Classification from X-ray Images using Residual Neural Network

Abdan Hafidh Ahnafi [#1], Anditya Arifianto [#2], Kurniawan Nur Ramadhani [#3]

 # School of Computing, Telkom University
Jl. Telekomunikasi No.1 Terusan Buah Batu Bandung, 40257 Indonesia

[1] abdanhafidh@student.telkomuniversity.ac.id
[2] anditya@telkomuniversity.ac.id
[3] kurniawannr@telkomuniversity.ac.id

Abstract

Pneumonia is a virus, bacterium, and fungi infection disease which causes alveoli swelling and gets worse easily if it is not taken care of immediately. There are symptoms that can be recognized through x-ray images, for example the appearance of white mist in the lungs. A pneumonia classification system has already developed, but it still produced low accuracy. In this research we develop classification system by increasing the depth of CNN architecture using Residual Neural Network to improve accuracy from previous research. The dataset contains 2 classes which are pneumonia and normal, and trained to produce the best learning strategy with various scenarios. The model trained using data train that has been oversampling. The best scenario is achieved by ResNet152 architecture using dropout 0.5. This scenario achieved a result of 0.88 precision, 0.95 recall, 0.92 f1-score, and 0.89 of accuracy. The classification model on this research produces higher accuracy compared to the research of Enes Ayan, et.al. in 2019 which produced 0.87.

Keywords: Residual Neural Network, classification, x-ray, pneumonia.

Abstrak

Pneumonia merupakan penyakit infeksi dari virus, bakteri, dan jamur yang menyerang paru-paru sehingga mengakibatkan kantung udara (alveoli) meradang dan akan menjadi akut dan mematikan jika tidak segera diperiksa. Terdapat gejala yang bisa dikenali melalui citra x-ray, misalnya terdapat kabut putih pada paru-paru Pengembangan sistem klasifikasi pneumonia sudah pernah dilakukan, tetapi masih menghasilkan akurasi yang rendah. Pada penelitian ini dibangun sebuah sistem klasifikasi pneumonia dengan meningkatkan kedalaman pada arsitektur CNN menggunakan Residual Neural Network (ResNet) untuk meningkatkan akurasi pada penelitian sebelumnya. Dataset yang terdiri atas 2 kelas yaitu pneumonia dan normal, dilatih menggunakan ResNet untuk menghasilkan strategi pembelajaran terbaik dengan berbagai skenario. Model dilatih menggunakan data latih yang dilakukan oversampling. Skenario terbaik diperoleh dengan meggunakan arsitektur ResNet152 yang menggunakan dropout 0.5. Hasil klasifikasi terbaik diperoleh dengan presisi sebesar 0.88, recall sebesar 0.95, f1-score sebesar 0.92, dan akurasi sebesar 0.89. Model klasifikasi pada penelitian ini menghasilkan akurasi yang lebih tinggi dibandingkan dengan penelitian Enes Ayan, dkk. pada tahun 2019 yang menghasilkan 0.87.

Kata Kunci: Residual Neural Network, klasifikasi, x-ray , pneumonia.

Received on Jan 2020. Accepted on Aug 2020

I. Introduction

PNEUMONIA is a virus, bacterium, and fungi infection disease which causes alveoli swelling and gets worse easily if it is not taken care of immediately. There have been 922.000 infants deceased caused by Pneumonia in 2015 [1]. The main problem of this infection is its symptoms which are hard to identify. Recently, there is a checkup method using X-ray images which can be diagnosed by the doctors or even the experts. Nevertheless, the result needs to be improved because the misidentification of pneumonia can lead to death [2].

Extreme care when diagnosing Pneumonia is needed since the misdiagnose can increase the probability of death. Pneumonia diagnosed via X-ray images. There are some symptoms that are easy to recognize, for instance, the presence of white cloudy fogs inside the patience's lungs. The unclear lungs is one of the most recognized Pneumonia symptoms[4]. Identifying Pneumonia symptoms is too risky for the ones with no backgrounds of the disease and requires a long time for diagnosis. The classification system is needed to help early diagnosis of pneumonia.

The research for describing Pneumonia symptoms has been done in many classification methods and one of them using Self-Organizing Map (SOM). That method was able to get rid of unnecessary objects on the X-ray images and the result is 0.90 on accuracy [5]. There are also research to classify other lung diseases not just pneumonia, using Convolutional Neural Network, where its accuracy reaches 0.88 [6]. Then in 2019, the CNN method was applied to the pneumonia classification and showed the accuracy number on 0.87 [7]. Current performance can be increased by increasing the depth of CNN architectures such as using ResNet.

This research analyzes the Pneumonia classification using transfer learning method with ResNet architecture which the main purpose is to improve accuracy on previous research with the same dataset.

II. Related Work

Resnet is one of the popular architectures of CNN recently. The ResNet architecture was made by Kaiming He [8] in purpose to introduce the new and deep residual framework as shown in Fig. 1. Residual block in each layer of the architecture of ResNet is the fundamental as its block becomes a shortcut for the identity function explicitly by the training process through one layer or more.

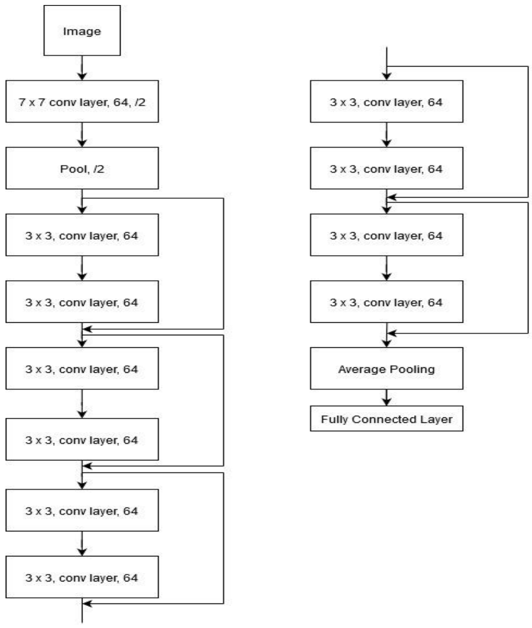

Fig. 1. Residual Neural Network Architecture

The usage of residual block could prevent the vanishing gradient problem. ResNet modules consist of bottleneck layer, made up of a 1 x 1 convolution that reduces number of channels by a factor of 4, followed by a 3 x 3 convolution and finally a 1 x 1 convolution expand the layers as shown in Fig. 2. Bottleneck block can reduce the computational cost of network, since 1 x 1 convolutions are 9 times less expensive than a 3 x 3 convolution, they are used to minimize the number of channels that comes into the 3 x 3 convolution.

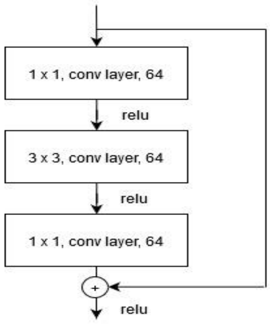

Fig. 2. Resnet Modules

Kaiming, et. al claims that the residual network are easier to optimize than the ordinary network that only stacks the layer . The use of ResNet also easily increased the architecture's accuracy compared to the previous one [8]. ResNet has many types of architectures showing the depth of each layer, which are ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152. In this research, we choose 3 ResNet architecture which is ResNet50, ResNet101, and ResNet152.

III. RESEARCH METHOD

A. Dataset

The data consist of x-ray images taken from infant's lung ranged one to five-year-old at Guangzhou Women and Children's Medical Center in Guangzhou, China as shown in Fig. 3. All the images were captured from some of the patient's routine clinic checks and are available in kaggle.com (https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia). In the dataset, there are 2 classes, pneumonia and normal with the number of 6,504 images on JPEG.

Normal Pneumonia

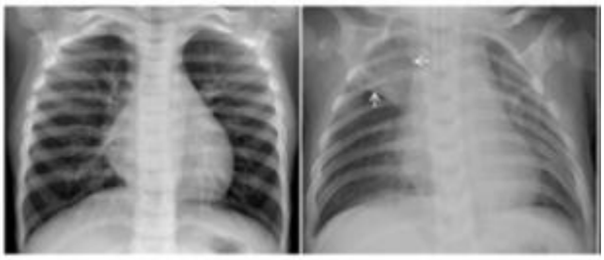

Fig. 3. Pneumonia Dataset

B. *Oversampling*

One of the problems in machine learning methods is the class imbalance. The class imbalance is occurred in the dataset when the minority class is smaller than the majority class [9]. In this research we determined a minority class is the x-ray images which is normal and a majority class is the x-ray images which is infected by Pneumonia. The normal class has fewer numbers than the pneumonia class. If we used imbalanced classes of dataset in the model is produced lower prediction. More information from majority class is dominates, can lead to biased decision in the classification system [10].

In this research, oversampling is used to resolve the problem of class imbalance. Oversampling is performed randomly chosen and duplicated the minority [11]. Oversampling is increasing the number of minority classes by replicating the same information [11]. Therefore, the number of minority and majority classes is equal. The oversampling method is shown in Fig. 4.

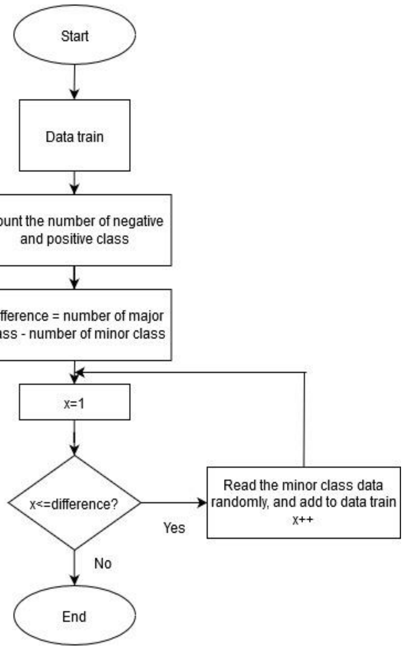

Fig. 4. Random Oversampling

C. *Cross Validation*

To measure the model performance, the cross-validation is used. Cross validation is one of the fundamental methods in machine learning for evaluation in a prediction or learning [12]. In this research we use 3-fold cross-validation. 3-fold cross-validation dividing the images into two sets : training and validation. The process is repeated 3 times with different fold, changing the training and validation subset after the calculation of k. The performance of model is measured by the average accuracy of validation set from each subset. This process takes full advantage of the entire dataset, when the number of data is small [12]. 3-fold cross-validation illustration is shown in Fig. 5.

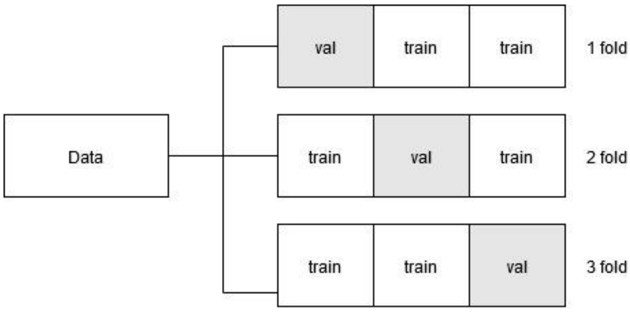

Fig. 5. Three-fold cross-validation illustration

D. *Built System*

The aim of the research is to recognize whether of human lungs is normal or infected with Pneumonia base on x-ray images.

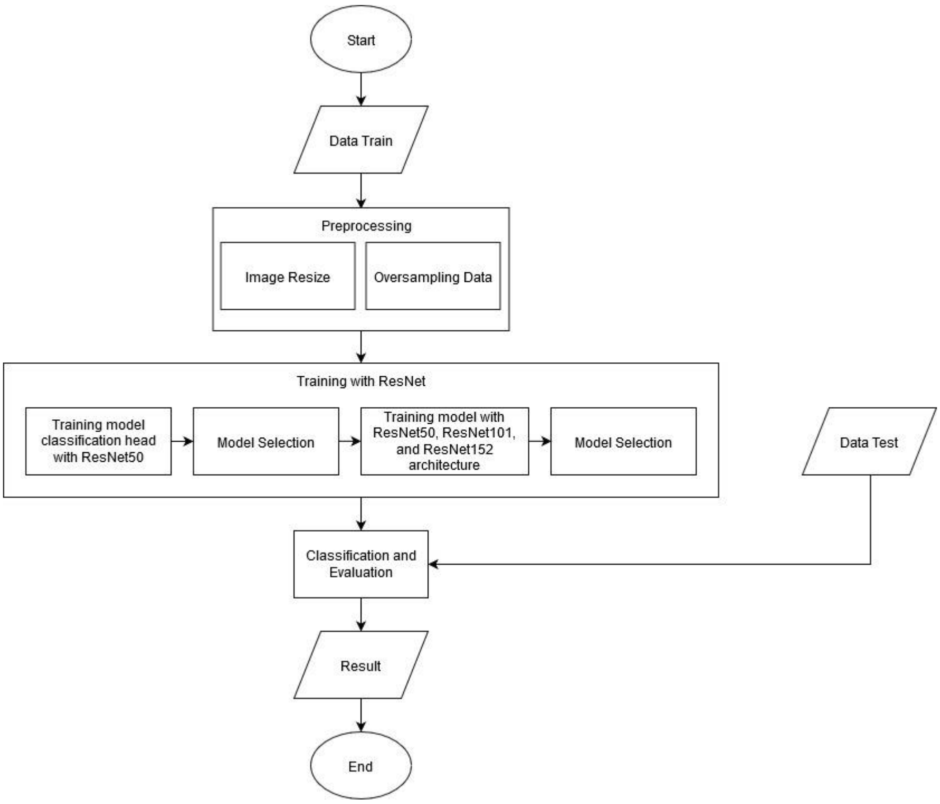

Fig. 6. Flowchart of the system

Fig. 6. shows the training process with ResNet architectures. Before the training process, the image data is resized to 150 x 150 to speed up the process involving the dataset. After the resize is complete, the data train is oversampled to equalize class distribution between normal class and pneumonia class. The number in the data train is 3,875 for pneumonia class and 1,341 for normal class. The number for each class is imbalance, therefore oversampling is used to increase the number of normal class so it has the same number with pneumonia

class. The amount of data train after oversampling is equal which is 3,875 for each class. Next, the image from data train is trained by Resnet50 architecture, and it is evaluated using 3-fold cross-validation to find out how well the model can classify images that has not been encountered during training. With a relatively small amount of data, makes training a model for pneumonia classification is not easy. The transfer learning method can be used to help this problem by using ResNet model that has been trained by ImageNet[13]. The dataset in this research only consist of 2 classes, different from ImageNet dataset that has 1000 classes [14]. To resolve that problem, a new fully connected layer must be created to replace ImageNet classification head. Two main scenarios were designed to find out the best architecture which achieve the best performance. First scenario is to find the best classification head for ResNet50 architecture. The second scenario is to find the best ResNet architecture between ResNet50, ResNet101, and ResNet152. The best architecture will be used to evaluate the data test to find out the classification result.

E. *Evaluation Metrics*

In evaluating the performance of this system, the Confusion Matrix is used as the base to show the difference between the prediction model and the real situation of the data[15].

TABLE I
CONFUSION MATRIX

Predicted	**Actual**	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

In this research we determined a positive class is the x-ray images which is infected by pneumonia. From the table I, TP (True positive) which the number of positive class that were correctly predicted as positive, TN (True Negative) is the number of negative class that were correctly predicted as negative, FP (False Positive) is the number of positive class that were predicted as negative, FN (False Negative) is the number of negative class that were predicted as a positive.

1) *Precision* : Precision is a TP ratio prediction divided by the overall result which was positively predicted by the system. Precision can be calculated by equation (1) [15].

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

2) *Recall* : Recall (or commonly called sensitivity) is TP ratio prediction divided by the overall data which was positive. Recall can be calculated by equation (2) [15].

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

3) *F1-Score* : F1-Score is divided by precision and recall. F1-score can be calculated by equation (3) [15].

$$F1\ Score = 2 * \frac{(Recall * Precision)}{(Recall + Precision)} \quad (3)$$

4) *Accuracy :* Accuracy is percentage of the system that was predict the data correctly. Accuracy can be calculated by equation (4) [15].

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (4)$$

IV. RESULTS AND DISCUSSION

A. First Scenario

The aim of the first scenario is to decide which classification head is the most optimal for the ResNet architecture. We decided to use ResNet50 in this scenario because it has the fewest layer compared to the others. The classification head that we have tried are shown in table II.

TABLE II
SCENARIO FOR CLASSIFICATION HEAD

	Model 1	**Model 2**	**Model 3**	**Model 4**
Layer	Dropout	Dropout	Batch Normalization	Batch Normalization
	Global Average Pooling	Flatten	Global Average Pooling	Flatten
	Dense	Dense	Dense	Dense

The dropout use 0.5 probability and the dense layer all output two classes. All models are trained using 15 epoch and batch size is 64

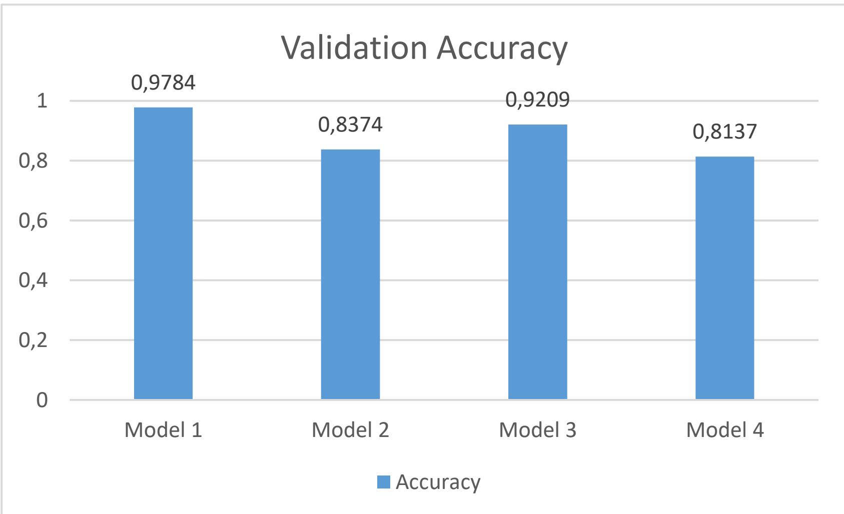

Fig. 7. Observation Result for classification head

Fig. 7. is shown the validation accuracy of training process. Validation accuracy is shown the percentage of the system that was predict the data correctly. From those experiments using the ResNet50 architecture

conclude the validation accuracy as shown in Fig. 7. Among those 4 models, it is shown that the Dropout + Global Average Pooling model is the highest with the validation accuracy of 0.97. It happened because the dropout usage with 0.5 probability can deactivate some neuron while the system runs the training process. With disabling the neurons, the extraction process which has been done by the system becomes fewer and gathering them with Global Average Pooling which prevent the data overfitting by reduce the parameter.

B. Second Scenario

In this scenario the best classification head from the previous scenario is installed into 3 architectures of ResNet to compare which one is the most optimal in classifying Pneumonia. Each architecture is trained with 15 epoch and 64 batch size.

From those 3 architectures, the Fig. 8. shows the comparison of the result.

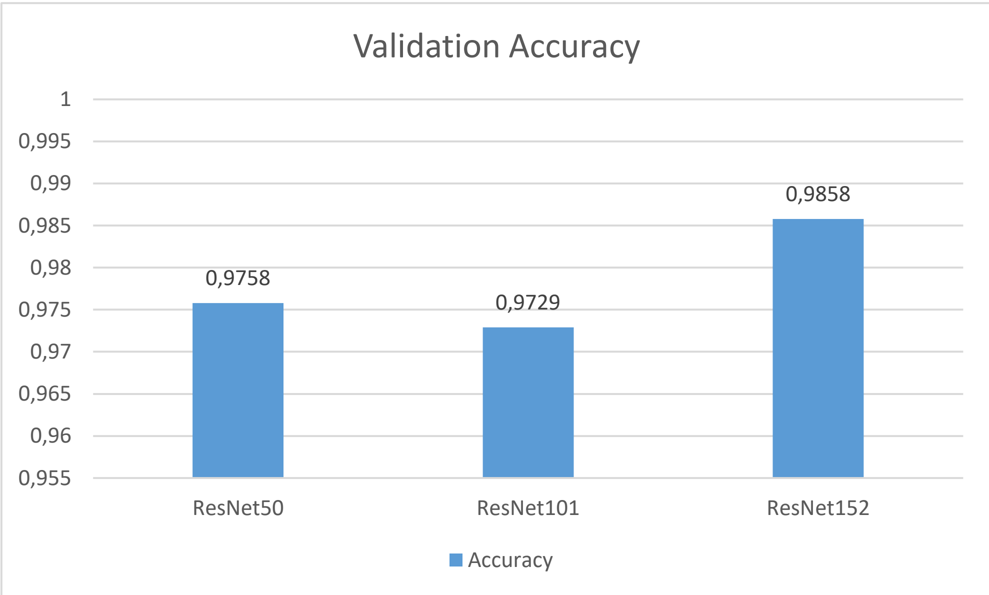

Fig. 8. Observation Result of 3 Resnet Architecture

Fig. 8. Shows the comparison of validation accuracy of three ResNet architecture. ResNet152 architecture has the best validation accuracy result among the others because of the usage of the classification head using dropout + global average pooling, well-combined with the extraction feature from transfer learning. The usage of dropout layer can deactivate some neurons in the training process. It makes the transfer learning weights of ResNet152 architecture that enter the fully-connected layer produce a good distribution of each neuron, so it can produce the best validation accuracy of 0.98.

From the previous result, the best architecture is obtained. This model is used to evaluate the data test. This table III shows the result of confusion matrix.

TABLE III
RESULT OF CONFUSION MATRIX

Predicted	Actual	
	Pneumonia	Normal
Pneumonia	372	51
Normal	18	183

The model was able to classify well. The result of the classification results for pneumonia class obtained from the model are 0.88 precision, 0.95 recall, 0.92 f1-score, and 0.89 of accuracy as shown in table IV.

TABLE IV
RESULT OF ACCURACY

Class	Precision	Recall	F1-score	Accuracy
Normal	0.91	0.78	0.84	0.89
Pneumonia	0.88	0.95	0.92	

The performance of the model with oversampling data is achieved higher accuracy than without oversampling data. The comparison of results between model with oversampling and without oversampling is shown in Fig. 9.

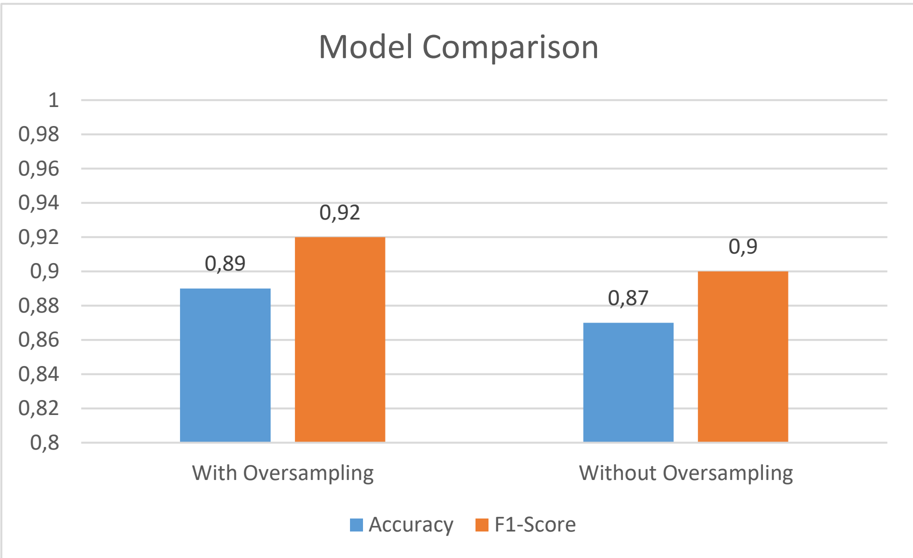

Fig. 9. Result comparison between with oversampling and without oversampling

This research also compares Enes Ayan, et al.[7] which use the VGG16 architecture and showed the accuracy number on 0.87. With increasing the depth of CNN architectures using ResNet152, this research successfully increasing the accuracy to 0.89. Model of this research also produce higher number of true prediction in pneumonia class then previous model. Table V is shown the comparison of this research with the previous research.

TABLE V
Comparison of This Research with Previous Research

Network	Number of layers	Number of true prediction (normal class)	Number of true prediction (pneumonia class)	F1-score	Accuracy
CNN with VGG16 (previously proposed model)	16	201	348	0.90	0.87
CNN with ResNet152 (the model proposed in this study)	152	183	372	0.92	0.89

V. Conclusion

In this study, we built a model for classifying the x-ray images using ResNet with transfer learning. The transfer learning method with ResNet architectures used for training process and for the validation of the model, k-fold cross-validation is used. The oversampling method can increase the number of minority class resolve the problem of class imbalance. Based on the scenario, the use of dropout and global average pooling on the classification head with ResNet152 architecture produces the best classification result with are 0.88 precision, 0.95 recall, 0.92 f1-score, and 0.89 of accuracy. Then it can be concluded that : 1) oversampling method is achieved higher accuracy then without oversampling; 2) The use of dropout with 0.5 probabilities can prevent the overfitting of model. Compare the previous research of Enes Ayan et.al. which achieved a result 0.87 of accuracy, our result increase the accuracy by 0.02.

References

[1] World Health Organization, Pneumonia Factsheets, 2019.
[2] Zar, Heather & Andronikou, Savvas & Nicol, Mark, “Advances in the diagnosis of pneumonia in children”, in *BMJ Clinical Research*, 2017
[3] Muder, Robert & Aghababian, Richard & Loeb, Mark & Solot, Jerald & Higbee, Martin, “Nursing home-acquired pneumonia: An emergency department treatment algorithm’, in *Current Medical Research and Opinion*, 2004
[4] Waterer, Grant & Kessler, Lori & Wunderink, Richard, “Delayed Administration of Antibiotics and Atypical Presentation in Community-Acquired Pneumonia”, in *Chest*, 2006
[5] Al-Hadidi, M. R. A., Dorgham, O., & Razouq, R. S., “Pneumonia Identification Using Self Organizing Map Algorithm”, in *ARPN Journal,* 2016
[6] Abiyev, R. H., & Ma’aitah, M. K. S., “Deep Convolutional Neural Networks for Chest Diseases Detection”, in *Journal of Healthcare Engineering*, 2018
[7] E. Ayan and H. M. Ünver, "Diagnosis of Pneumonia from Chest X-Ray Images Using Deep Learning", in *Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT),* 2019
[8] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian, "Deep Residual Learning for Image Recognition" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015*
[9] Ren, F., Cao, P., Li, W., Zhao, D., and Zaiane, O, “Ensemble based adaptive over-sampling merhod for imbalanced data learning in computer aided detection of Microaneurysm”, in *Computerized Medical Imaging and Graphichs : The Official Journal of the Computerized Medical Imaging Society,* 2016.
[10] Jian, C., Gao, J., and Ao, Y, “A new sampling method for classifying imbalanced data based on SVM ensemble”, in *Neurocomputing,* 2016.
[11] Rajesh. N. and Ravindra, D, “Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier”, in *Biomedical Signal Processing and Control,* 2018.
[12] Rodriguez. J., Aritz. P, and Lozano. J., “Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation”, in *EEE Transactions on Pattern Analysis and Machine Intelligence,* 2010
[13] S. J. Lee, G. Koo, H. Choi and S. W. Kim, "Transfer learning of a deep convolutional neural network for localizing handwritten slab identification numbers," in *Fifteenth IAPR International Conference on Machine Vision Applications (MVA),* 2017

[14] Russakovsky, Olga & Deng, Jia & Su, Hao & Krause, Jonathan & Satheesh, Sanjeev & Ma, Sean & Huang, Zhiheng & Karpathy, Andrej & Khosla, Aditya & Bernstein, Michael & Berg, Alexander & Li, Fei Fei, “ImageNet Large Scale Visual Recognition Challenge”, in *International Journal of Computer Vision*, 2014

[15] Powers, David M W., "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" in *Journal of Machine Learning Technologies*, 2011