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. 55-66
doi:10.21108/indojc.2020.5.2.455

Classifying Skin Cancer in Digital Images Using Convolutional Neural Network with Augmentation

Zeyhan Aliyah #1, Anditya Arifianto #2, Febryanti Sthevanie#3

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

1 zeyhana@student.telkomuniversity.ac.id
2 anditya@telkomuniversity.ac.id
3 sthevanie@telkomuniversity.ac.id

Abstract
Skin cancer is a hazardous disease that can induces death if it is not taken care of immediately. The disease is hard to identified since the symptoms have similarities with other disease. An automatically classification system of skin cancer has been developed, but it still produced low accuracy. We use Convolutional Neural Network to enhance the accuracy of the classification. There are 2 main scenarios conducted in this research using HAM10000 dataset which has 7 classes. We compared ResNet and VGGNet architectures and obtained ResNet50 with augmentation as the best model with the accuracy of 99% and 99% macro avg.

Keywords: Convolutional neural network, classification, digital image, skin cancer

Abstrak
Kanker kulit merupakan penyakit berbahaya yang dapat menyebabkan kematian jika terlambat ditangani. Penyakit ini sulit diidentifikasi karena gejala tersebut memiliki kemiripan satu sama lain. Sistem klasifikasi kanker kulit secara otomatis telah dikembangkan, tetapi masih menghasilkan akurasi yang rendah. Kami menggunakan *Convolutional Neural Network* untuk meningkatkan akurasi klasifikasi. Terdapat 2 skenario utama yang diterapkan pada makalah ini dengan menggunakan dataset dari HAM10000 yang berisi 7 kelas. Kami membandingkan arsitektur ResNet dan VGGNet dan memperoleh ResNet50 dengan augmentasi sebagai model terbaik dengan 99% akurasi dan 99% rata-rata makro.

Kata Kunci: Convolutional neural network, klasifikasi, citra digital, kanker kulit.

I. Introduction

SKIN disorder is commonly faced in society and becoming one of the most dangerous disease [1]. The disease is triggered by the damaged DNA cells which mostly caused by UV lights contamination continually. The unhealed wound also causes the skin cells to grow rapidly which causes skin cancer and deadly disease [2][3]. Based on the data, in 2012 there has been 7.230 infants deceased caused by skin cancer in US[4].

Actually, skin cancer can be detected earlier but the Indonesian people is too reluctant to consult with the experts, which makes the condition even worse. The way doctors recognize skin cancer is by perfoming a biopsy which is removing tissues to discover the presence of the disease. However this process takes a quite a long

Received on Jan 2020. Accepted on Aug 2020

time because skin diseases have similarities with each other. This process also has a risk of infection and bleeding which makes the patient are afraid to do a biopsy[5]. Therefore, an simple way to early identify skin cancer is needed to be developed.

An automatic skin cancer classification system was created to help the normal way of diagnosing the disease. Yuexiang Li et. al. use Convolutional Neural Network for classifying melanoma skin cancer with only 2 classes and achive 79.3% accuracy [6]. Another research by Aryan Mobiny et. al. used Bayesian Neural Network method to classify 7 classes of skin cancer which are Melanoma, Melanocytic Nevi, Basal Cell Carcinoma, Actinic Keratoses and Intraepithelial Carcinoma, Benign Keratosis, Dermatofibroma, and Vascular. The study has reached the accuracy of 83.59%. This classification result could be improved by implementing CNN. Thus, we perposed an implementation of skin cancer classification system using CNN and improvement the accuracy from previous research with the same dataset [7].

II. Related Work

A. VGGNET

VGGNET is a CNN architecture that was developed and trained by the Visual Geometry Group, Oxford. It has achieved a remarkable performance on the ImageNet dataset. VGGNET is quite accessible to implement repetitive structures in coding a modern deep learning framework. The blocks that have built on VGGNET very much alike to the classic convolutional networks with the layer sequences: (i) convolutional layer (with padding to maintain resolution), (ii) nonlinear such as ReLU. One VGG block consists of a convolutional layer sequence, followed by a max-pooling layer for down sampling. The layer size which is commonly used in the VGGNET architecture is 3x3 for the convolution and 2 x 2 for max pooling by cutting down half of the resolution afterward on each block. VGGNET also has many types based on the layer is going to be used. One of the most popular types is VGG-16 and VGG-19 producing a fairly good classification accuracy [8].

Fig. 1. VGGNET Architecture

B. *ResNet*

ResNet (Residual Neural Network) is an architecture of CNN which has developed and trained by Kaiming He, with excellent accuracy result by training deep networks. The difference in this architecture is the use of residual blocks. Residual block made by the appearance of each several layers that are stacked directly as a mapping and producing a "shortcut connection". The shortcut connections will pass through one or more layers as shown in Fig. 3 The use of residual blocks can overcome the degradation problem where training errors boost as the depth increased. ResNet also has many types based on layers that are commonly used, such as ResNet34, ResNet50, ResNet101, and ResNet 152 [9]. In this research, the type of ResNet that is used is ResNet50.

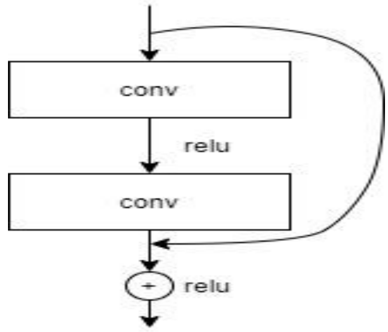


Fig. 2. Residual Block

III. RESEARCH METHOD

A. Dataset

The data use is HAM10000 (Human Against Machine with 10.015 images) taken from kaggle.com. The data use are the same as previous research[7]. The dataset consist of Dermatoscopy images with different populations and modality as shown in Fig. 1. There are 7 classes in this dataset which are: actinic keratosis and intraepithelial carcinoma/Bowen disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), nevi melanocytic (nv), and lesi vascular (vasc).

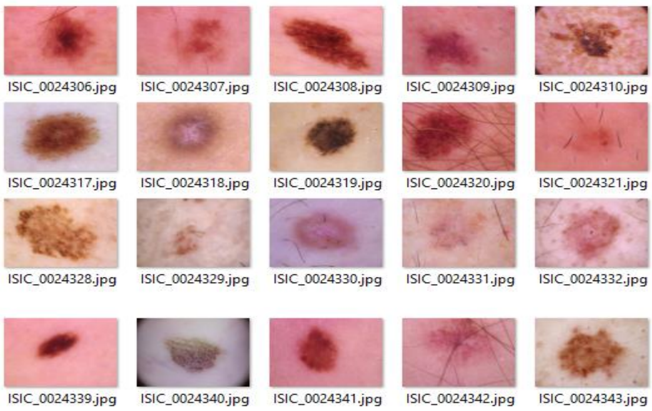


Fig. 3. HAM10000 dataset of skin cancer

Before the training started, the data set is resized from the initial size of 600x450 to 100x75, which aims to speed up during computational process. The dataset is divided into 70% data train and 30% data test. For the data train is taken 10% for validation data is used to determine the required model can classify images that has not been seen before in training. Train data is 6309, test data is 3005, and validation data is 701.

B. Data Augmentation

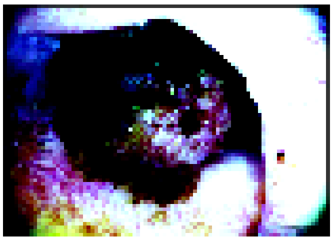
The augmentation is a manipulation technique of an image without lessening any data information [10]. Augmentation can improve the accuracy of CNN model since the model will get additional data which is going to properly generalize. The Table I shows data augmentation parameter of this research.

TABLE I
Table of Data Augmentation

Type	Factor
Feature wise center	True
Feature wise std normalization	True
Rotation range	20
Zoom range	0.1
Width shift range	0.2
Height shift range	0.2
Horizontal flip	True
Vertical flip	True

1) Feature Standardization : Feature Standardization process is standardize pixel value for all datasets. This process uses feature wise center and feature std normalization performance. Feature wise center performed the feature standardization by setting the input of mean to 0 over the dataset. On the other hand, feature std normalization performed the feature standardization by divide the images inputs from the std of the datasets. The transformation can be seen that the image color becomes different, as shown in fig. 4a and fig. 4b.

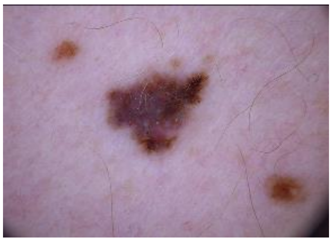
(a) Original Image (without feature standardization)



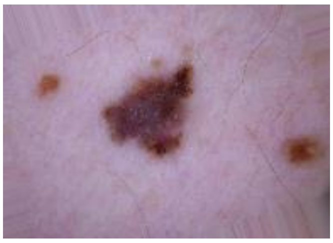
(b) Image with feature standardization

Fig. 4. Visualization of Feature Standardization Process

2) Rotation Range : The rotation range process is performed randomly rotation image from the dataset during training according to the degree input. In this process, 20 degrees is applied for rotation range and the difference is shown as the fig. 5a and fig. 5b.



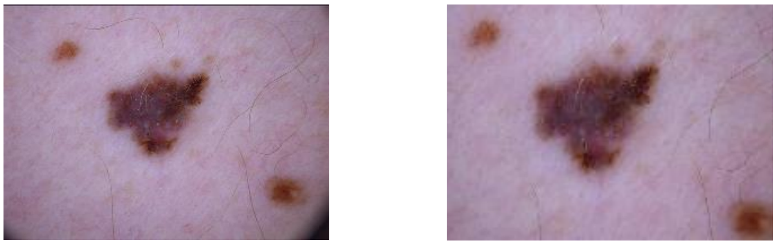
(a) Original Image (without rotation range)



(b) Image with rotation range

Fig. 5. Visualization of Rotation Range Process

3) Zoom Range : Zoom range process is performed randomly zoom image from the dataset during training according to the parameter input. In this process, 0.1 zoom is applied for the range parameter and the difference is shown as the fig. 6a and fig 6b.



(a) Original Image (without zoom range) (b) Image with zoom range

Fig. 6. Visualization of Zoom Range Process

4) Random Shifts: The random shifts process is performed randomly shift image from the during training according to input parameters. This process uses a width shift range and a height shift range as the random shifts. The images slide 20% randomly in vertical and horizontal which applied 0,2 parameter. The difference is shown as the fig. 7a, fig. 7b, and fig. 7c.

(a) Original Image (without random shift) (b) Image with width shift range (c) Image with height shift range

Fig. 7. Visualization of Random Shifts Process

5) Random Flips: Random flips process is performed randomly flip image from the dataset during training according to the input parameters. This process performed a horizontal flip and vertical flip. The difference is shown as the fig. 8a, fig. 8b, and fig. 8c.

(a) Original Image (without random flips) (b) Image with horizontal flip (c) Image with vertical flip

Fig. 8. Visualization of Random Flips Process

C. *CNN Method Training*

The system in this model is a system that can classify the skin from digital images. The system flowchart can be seen in Fig. 9.

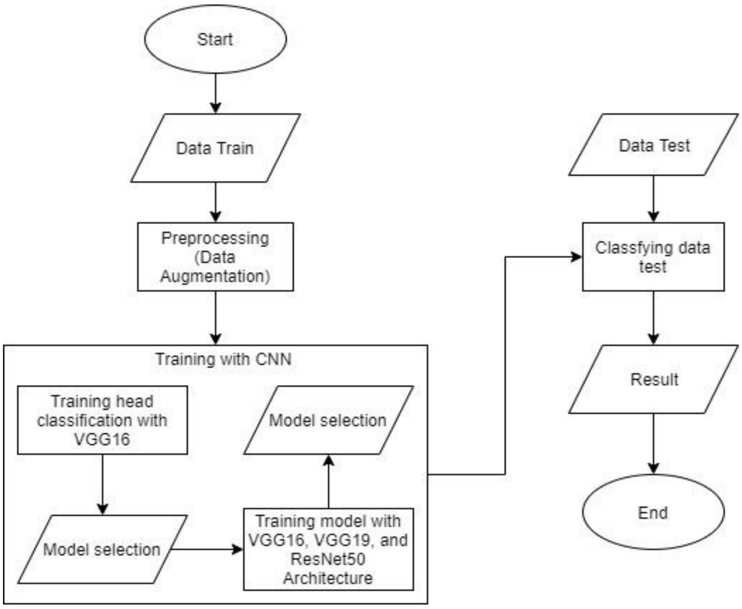


Fig. 9. Flowchart of classification system

Fig. 9 shows the flowchart of this training process. Before the training started, the dataset is divided into 3 parts, which are training data, testing data, and validation data. Next, those data will be processed using augmentation method to add more images to be trained. This training uses transfer learning concept which takes the extraction feature from the trained CNN architecture and replace the head classification. We apply 2 scenarios to decide which head classification architecture has the best result. The best head classification is then installed into different CNN architecture and to find the best training outcome. In the training process, the model is trained using the training data and also would be evaluated by the validation data.

D. *Model Evaluation*

The accuracy is obtained by calculating the amount of the predicted data with the total of the data usage as shown in equation (1)[11].

$$Accuracy = \frac{Number\ of\ Correct\ Prediction}{Number\ of\ Test\ Data} \quad (1)$$

IV. RESULTS AND DISCUSSION

A. *Observation Scenarios for the Best Classification Head*

This scenario aims to discover which head classification is the best based on the accuracy of the training produced. The classification head is installed into VGG16 body. We use VGG16 because this architecture has been used commonly and has fewer layers. The VGG16 body model is already available on ImageNet [12]. The classification head observed in this study consist of several layers such as Dropout, Flatten, Batch Normalization, and Global Average Pooling 2D as shown in Table II.

TABLE II
MODEL OF CLASSIFICATION HEAD

Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Dropout	Batch Normalization	Batch Normalization	Dropout	Flatten	Global Average Pooling
Flatten	Flatten	Global Average Pooling	Global Average Pooling	Dense	Dense
Dense	Dense	Dense	Dense	Dense	Global Average Pooling
Dropout	Batch Normalization	Batch Normalization	Dropout	-	Dense
Dense	Dense	Dense	Dense	-	-

The dropout used in this observation were all 0.3 and 0.15 probability, and the dense layer all output 7 classes. All of these architectures are trained in 30 epoch batch sizes of 128. The result from the observation is shown in Fig. 10.

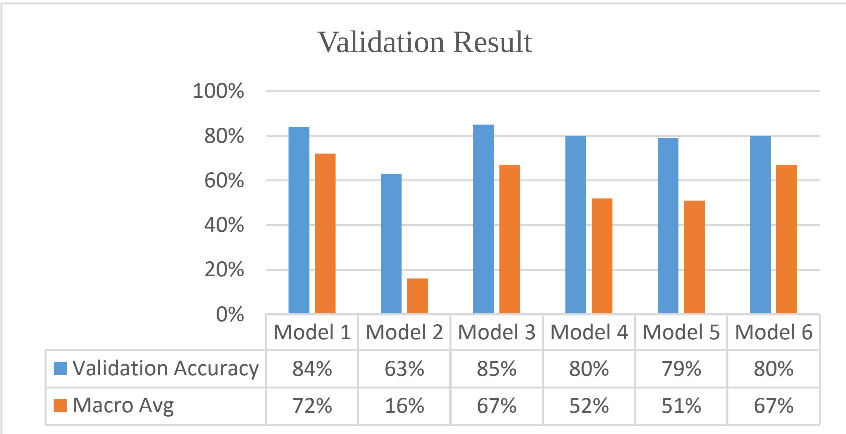


Fig. 10. Result of observation scenario for the best classification head

From the comparison on the Fig. 10, shows that model 3 which contains the Normalized Batch + Global Average Pooling layer produces the validation accuracy of 85% and 67% macro avarage from the total of validation data with 701 images. Macro avarage is the calculation of evaluation reports on the confusion matrix of each label. It shows that the combination of batch normalization layer with global average pooling produce a higher accuracy compared to the dropout layer with global average pooling combination. It happens because the probability dropout 0.3 causes many neurons on the classification head deactivates. Therefore, the training process into the activate function will change. It appears because every learning process has many neurons, which are deactivated, to create the distribution changes of the transfer learning quality. On the other hand, using batch normalization can create higher accuracy since batch normalization is able to boost the training performance. The way to do the boosting is adapting each neuron before put them into function activation and gathering them with global average pooling layer, which diminished the data overfit by reducing the total parameter of the model.

B. Observation Scenario for the Best Architecture

In this scenario, the result of the best classification head is taken from the previous scenario and applied to 3 architectures of CNN (VGG16, VGG19, and ResNet50) and to be retrained. The result is to compared to conclude which architecture producting the best validation accuracy. All those architectures are trained in 30 epoch, 128 batch size and used early stopping to stop the training when it overfit. The result of this scenario is shown by Fig. 11.

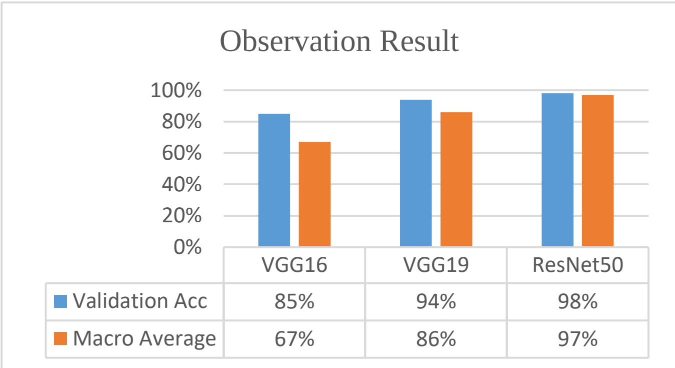


Fig. 11. Result of observation scenario for the best architecture

Based on the Fig. 11, the VGG16 architecture has its validation accuracy of 85%, whereas VGG19 achieved 94% validation accuracy. ResNet50 has the best result with 98% of the validation accuracy. It happened because ResNet architecture utilized the residual block which does not exist in VGG. The residual block on the ResNet reduce the problem of the gradient disappearance on training. Although ResNet has more layer than VGG, The residual block anticipates training errors caused by the depth of layer. The use of the batch normalization layer and the global average pooling help boosting the training process by the ResNet50. It produces a positive impact that the training accuracy can increase significantly. Therefore, this model is used to evaluate the testing data. The confusion matrix result can be in the Table III.

TABLE III
CONFUSION MATRIX RESULT

akiec	**86**	0	0	0	0	0	0
bcc	0	**157**	0	0	0	1	0
bkl	0	0	**324**	0	5	3	0
df	0	0	0	**31**	0	0	0
nv	1	0	1	0	**2000**	1	0
mel	0	0	5	0	13	**341**	0
vasc	0	0	0	0	1	0	**37**
	akiec	**bcc**	**Bkl**	**df**	**nv**	**mel**	**vasc**

Based on the confusion matrix, the largest number of error predictions is the melonama class which predicted to be the nv class. It happened since the mel and the nv class have almost the same shape and color, red and brown. The model is good enough to classify the data with the results are 99% accuracy and 99% macro average. The detail of the classification report is shown in Table IV.

TABLE IV
DETAIL OF CLASSIFICATION REPORT

	Precision	**Recall**	**F1-Score**
Akiec	0.99	1.00	0.99
Bcc	0.98	0.99	0.99
Bkl	1.00	0.98	0.99
Df	1.00	1.00	1.00
Nv	0.99	1.00	0.99
Mel	0.99	0.96	0.97
Vasc	1.00	0.97	0.99
Accuracy			**0.99**
Macro Avg	**0.99**	**0.99**	**0.99**

With the augmentation, the accuracy of the testing process is higher than without augmentation. The comparison of testing accuracy is shown in fig. 12 with its augmentation accuracy of 99%

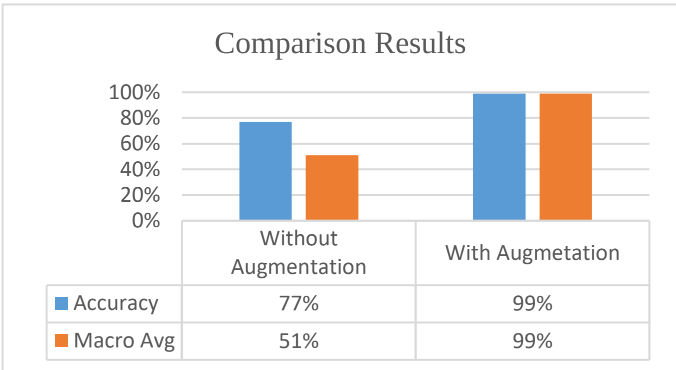


Fig. 12. Comparison Result Between with Augmentation and Without Augmentation

The efficiency of the model with augmentation is well-performed to predict more correct number of the data. It happened because the usage of augmentation created the additional data for the model which made the model generalize properly. Therefore, the accuracy is increased. The comparison is shown in Table V.

TABLE V
COMPARISON OF TRUE PREDICT NUMBER BETWEEN WITH AUGMENTATION AND WITHOUT AUGMENTATION

No.	Class	Number of correct predictions	
		Without Augmentation	With Augmentation
1.	Akiec	29	86
2.	Bcc	84	157
3.	Bkl	158	324
4.	Df	6	31
5.	Nv	1856	2000
6.	Mel	157	341
7.	Vasc	25	37

This research is compared with previous studies that used Bayesian neural networks with the accuracy of 83.59%[7]. We use ResNet50 architecture successfully increasing accuracy into 99%. The detail of comparison this research with previous research shown in Table VI.

TABLE VI
DETAIL OF COMPARISON THIS RESEARCH WITH PREVIOUS RESEARCH

Network	Accuracy	Percentage of True Predicted Label						
		Mel	Nv	Bcc	Akiec	Bkl	Df	Vasc
Bayesian Neural Network	83.59%	0.56	0.93	0.66	0.38	0.79	0.39	0.83
Convolutional Neural Network	99%	0.99	0.99	0.98	0.99	1.00	1.00	1.00

V. CONCLUSION

This research conducted classification on skin cancer using CNN with augmentation. Data augmentation can increase the accuracy of classification model because the use of augmentation created the additional data for the model which made the model generalize properly. Based on the scenario that has been done, the classification head that contains the batch normalization + global average pooling 2D is able to produce the best training accuracy. The model is applied into the ResNet50 architecture and produces the highest validation results of 98%. The architecture of ResNet50 is able to outperform VGG architecture, which has an accuracy of 85% for VGG16 and 94% for VGG19. The model is good enough to classify the data with the results are 99% accuracy and 99% macro average. The use of batch normalization layer can create higher accuracy and can prevent the data overfitting. Therefore, it can be concluded that ResNet architecture produce better accuracy than VGG because the residual block anticipates training errors caused by the depth of layer. This research successfully increasing accuracy by 15,41% from previous research.

REFERENCES

[1] Djuanda Adhi, Hamzah Mochtar, Aisah Siti, “Morfologi Dan Cara Membuat Diagnosis; Rata IGA. Tumor Kulit”, in Buku Ilmu Penyakit Kulit dan Kelamin. Edisi ke-IV.Jakarta, 2005, Badan Penerbit Fakultas Kedokteran Universitas Indonesia.

[2] Buljan Marija, Bulana Vedrana, and Sandra Stanic. “Variation in Clinical Presentation of Basal Cell Carcinoma”, in University Department of Dermatology and Venereology Zagreb Croatia, 2008.

[3] W. Howard, A. Martin, R. Steven, "Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the US Population, 2012", in Journal of the American Medical Association, 2015.

[4] Skin Cancer Foundation, Skin Cancer Facts and Statistics, 2019.

[5] E. Nasr-Esfahani et al., "Melanoma detection by analysis of clinical images using convolutional neural network," in 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016.

[6] Li Y, Shen L., “Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network Sensors”, in Abbreviated Name of Conf., City of Conf., Abbrev. State, 2018.
[7] Mobiny A, Singh A, Van Nguyen H., “Risk-Aware Machine Learning Classifier for Skin Lesion Diagnosis” in J Clin Med., 2019.
[8] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge” in IJCV, 2015.
[9] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian. "Deep Residual Learning for Image Recognition", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[10] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in International Interdisciplinary PhD Workshop (IIPhDW), 2018.
[11] Wen Zu, Nancy Zheng, Ning Wang, "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations", in NESUG proceedings: health care and life sciences, 2010.
[12] 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.