Melanoma Cancer Classification Using ResNet with Data Augmentation

1st Arief Budhiman
School of Computing
Telkom University
Bandung, Indonesia
ariefbudhiman1@gmail.com

2nd Suyanto Suyanto
School of Computing
Telkom University
Bandung, Indonesia
suyanto@telkomuniversity.ac.id

3rd Anditya Arifianto
School of Computing
Telkom University
Bandung, Indonesia
anditya@telkomuniversity.ac.id

***Abstract*—Melanoma skin cancer is cancer that difficult to detect. In this study, have been done melanoma cancer classification using Convolutional Neural Network (CNN). CNN is a class of Deep Neural Network (Deep Learning) and commonly used to analyzing images data. A lot of data used on CNN can greatly affect accuracy. In this study, the objective is to get best ResNet model for classifying melanoma cancer and normal skin images. The dataset that used is ISIC 2018. ResNet is used because the model winning the ILSVRC competition at 2015. ResNet architecture model that used are ResNet 50, 40, 25, 10 and 7 models. The architecture trained using data train that has been augmented and undersampling. The validation result on each model calculated using F1 Score. After validation and F1 Score result from the model obtained, the result compared each other to select the best model. The best architecture is ResNet 50 without augmentation that gives a validation accuracy of 0.83 and f1 score of 0.46.**

Keywords—data augmentation, dropout, fully-connected layer, ResNet.

I. INTRODUCTION

At present, CNN has achieved very good performance in the field of computer vision, such as object detection, image recognition, classification, etc. [1]. Convolutional Neural Network (CNN) is a branch of the Deep Neural Network (Deep Learning). CNN has the ability to learn the characteristics of the input provided using layers and training. It take this name from mathematical linear operation between matrixes called convolution. CNN have multiple layers; including convolutional layer, non-linearity layer, pooling layer and fully-connected layer [2]. The traits that can be learned by CNN are not necessarily recognizable by humans, so to prove the traits that are recognized by CNN are quite difficult.

Melanoma cancer is cancer that is difficult to detect in an ordinary way. Besides being a person with melanoma cancer does not feel pain, the form of melanoma cancer is also similar to ordinary moles. In the case of melanoma cancer, the damage to DNA is caused by overexposure to ultraviolet rays (UV), and the affected cells are the melanocytes that produce melanin (pigmentation of the skin). Usually, the first tumor that develops is found on the skin. If melanoma is not detected, it grows and spreads along the first layer of skin before penetrating into the deeper layers and finally, comes into contact with the lymph vessels and the blood [3]. Because of the difficulty of detecting melanoma cancer CNN is used to classify melanoma skin cancer.

Research on the classification and detection of melanoma cancer by various methods has been carried out. In 2012 a melanoma cancer detection study was conducted using SVM and obtained an accuracy of 0.77 [4]. In 2013, a similar study was conducted using the SVM, KNN and Neural Network methods with the best accuracy of 0.8-0.9 [3] . In 2017, a study was conducted regarding the classification of melanoma cancer using KNN, Decision Tree and SVM [5] with an accuracy of 0.78. In the same year, there was a study using ResNet-88 with a dataset from ISIC 2017 with an accuracy of 0.823 [6]. In addition, in 2017 there were also studies using the Bayesian Classifier method [7] with an accuracy of 0.91 and there were also other studies using the SVM method [8] with an accuracy of 0.93 using a dataset from Pedro Hispano Hospital. In addition, in 2016 there was a paper entitled "Deep Residual Learning for Image Recognition" [9] using the ResNet architecture. The paper was a winner at the 2015 ILSVRC (Imagenet competition).

In this study, the objective is to get best ResNet model for classifying melanoma cancer and normal skin images. The dataset that used is ISIC 2018. ResNet is used because the model winning the ILSVRC competition at 2015.

A. Deep Residual Network

Deep Residual Network (ResNet) is an Artificial Neural Network that is created with the aim of overcoming the problem of lower accuracy when creating a plain ANN with a deeper layer than a shallower ANN [10]. In other words, the purpose of the Deep Residual Network is to make ANN with deeper layers with high accuracy [9]. The concept of the Deep Residual Network is to make ANN that can update the weight to a shallower layer (reduce degradation gradient). The concept is implemented using a "shortcut connection". The concept of a residual network is shown in Fig. 1.

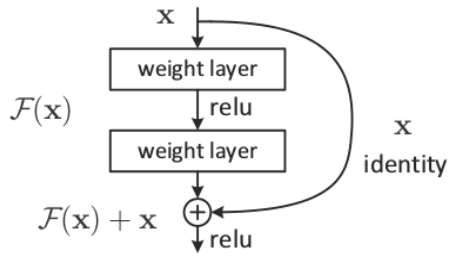

Fig. 1. Residual learning: a building block [5]

B. Fully Connected Network

Transfer learning is commonly used to carry out similar tasks [11]. This technique is used to shorten the model training time. The model only needs to be trained in the fully connected layer.

978-1-7281-4520-4/19/$31.00 ©2019 IEEE

C. Data Augmentation

Data augmentation is one of the regularization methods. The purpose of regularization is to avoid overfitting the created model. Regularization will improve the performance of the model on data that has never been seen. Data augmentation is a data manipulation technique without losing the core or essence of the data. Data augmentation is done to get more data than the original data (synthetic data).

In [12], augmentation was carried out by rotating 18° for 20 times on images with the tags "melanoma" and "seborrheic keratosis", and rotating 45 ° 8 times for data with the tag "Nevus".

D. Dropout

Like a data augmentation, dropout is also a regularization method. At each iteration, the dropout shuts off a number of random neurons at the specified layer and does not use the neurons in forward propagation and back-propagation [13].

E. Confusion Matrix

Confusion matrix is a matrix that describes the performance of an algorithm [14]. Confusion matrix contains information about the actual class and predictions of the model [15].

F. Melanoma skin cancer

Melanoma skin cancer is the type of skin cancer that is the most dangerous. Melanoma occurs in the melanosit cell, which produces the melanin (pigment that gives skin color). The exact cause of melanoma is still uncertain, but exposure to ultraviolet (UV) radiation can increase the risk of melanoma. The risk of melanoma is higher in women aged less than 40.

Knowing the signs of melanoma cancer will increase the chances of curing melanoma cancer. Since the percentage of success of cancer treatment is higher when it is detected earlier.

Clinical diagnosis of earlier melanoma cancer is illustrated by Fig. 2., with five steps [16]:

1) *Asymetry*: melanoma has an unstructure form and cannot be equally divided into two areas;
2) *Border*: melanoma has uneven and rough edges, in contrast to normal moles;
3) *Color*: melanoma is usually a mixture of two or three colors.
4) *Diameter*: melanomas usually have a diameter of more than 6 mm, and are different from ordinary moles.
5) *Enlargement*/evolution: moles that change shape and size after a while will usually become a melanoma.

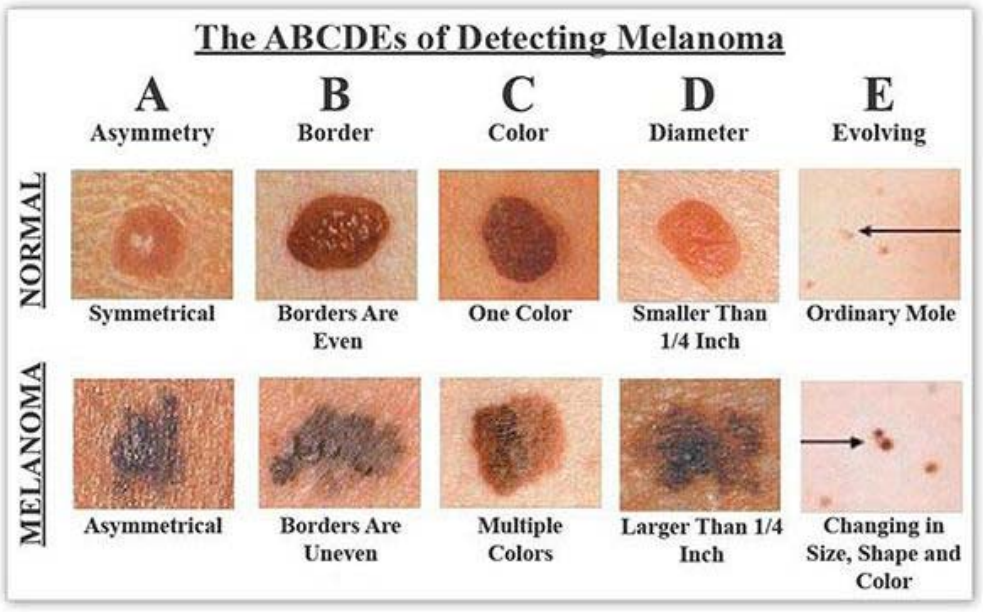

Fig. 2. ABCDE method to detect a melanoma

II. Developed Model

A ResNet model is used to create a system to classify dermoscopy images. The system built is ResNet with depths of 50, 40, 25, 10 and 7. We cut ResNet 50 into depth 40, 25, 10 and 7 for this study.

A. Dataset

The dataset is obtained from the ISIC 2018 page. The type of dataset used is only malignant and benign image data, so to get the required data, ISIC Archive Downloader is used. The dataset obtained from ISIC 2018, where the total number of data is 21,659 with benign data of 19,373 and malignant data of 2,286.

The dataset was then resized with a size of 224x224 with the aim to speed up the process involving the dataset beside at 2015, ResNet using the same size when winning ILSVRC. After getting a dataset that has been resized, the data is then cleaned by removing data that is not good to use. Then the dataset is divided into training data, validation data and test data with 80:10:10 division.

The training data is copied for augmentation. The malignant image data is augmented 2 times, with random rotations of -90 ° to 90 °.

B. Experiment

Experiments are conducted with ResNet Architecture with layer depth of 50, 40, 25, 10, and 7 with the aim to get the best layer depth with the aim of classification of melanoma skin cancer. We use ResNet 50 transfer learning and cut it into 40, 25, 10 and 7 layers. Then, the best model is selected and tested at the fully connected layer with the aim to get the best number of neurons to classify melanoma cancer. The number of neurons observed was 2048, 1024, 512, 256, 128, and 64.

There are 4 schemes used in this study: undersampling without dropout, undersampling with dropout, oversampling without dropout, oversampling with dropout. Dropout is used in fully-connected layer that drop the neuron score randomly to prevent updating weight. The Dropout that used is 0.5, that means 50% of the neuron that pass through dropout will not be updated. Undersampling is done by deleting datasets for superior data, so as to get balanced data. Whereas in oversampling, the data is augmented on the inferior data 2 times with the aim to make the data almost balanced. For each schemes, we use 10 epoch to train the models.

III. RESULTS AND DISCUSSION

Table I shows the validation accuracy of the ResNet model that was created.

TABLE I. RESNET VALIDATION ACCURACY

No	Architecture	Validation Accuracy
1	ResNet 50	0.785
2	ResNet 40	0.835
3	ResNet 25	0.807
4	ResNet 10	0.822
5	ResNet 7	0.824

Validation accuracy shows that much data is predicted correctly by the model compared to the total data. The validation accuracy obtained in the graph above is the validation accuracy of the data that has been cleaned, without oversampling and undersampling. So to get the best model, it is more suitable to use F1 score. The f1 score for each model is shown in Table II.

TABLE II. RESNET F1 SCORE

No	Architecture	F1 Score
1	ResNet 50	0.477
2	ResNet 40	0.394
3	ResNet 25	0.402
4	ResNet 10	0.074
5	ResNet 7	0.094

F1 score can be interpreted as the average weight of precision and recall, where F1 score has the best value of 1 and the worst value of 0. For the case of unbalance data, F1 score becomes the main factor for choosing the best model.

Thus, in this study, the model used is ResNet with a layer depth of 50. Next will be displayed regarding the selection of many neurons in the fully connected layer in Table III.

TABLE III. F1 SCORE NEURON

No	Depth	F1 Score
1	2048	0.454
2	1024	0.420
3	512	0.477
4	256	0.327
5	128	0.456
6	64	0.410

In the graph above, it appears that ResNet 50 with 512 neurons in the fully connected layer produces the highest F1 score. The number of neurons in each task given to the architecture will get different results. So to get the best value, an analysis must be carried out on the fully connected used. In this study, the neurons used do not need to be too high, because the purpose of this study is to carry out binary classification.

After getting the best model, the model will be trained using a dataset that has been done undersampling and oversampling, as illustrated by Table IV.

TABLE IV. F1 SCORE UNDERSAMPLING AND OVERSAMPLING

Method	Undersampling	Oversampling
Without dropout	0.419	0.471
Dropout	0.421	0.328

After testing using oversampling and undersampling data, it is concluded that the model produces the highest f1 score when using oversampling data without dropout. Furthermore, testing results om the test data is displayed on Table V. In the test data, obtained an accuracy of 0.83, f1 score of malignant data of 0.43, and f1 score of benign data of 0.87. The confusion matrix of the model is shown in table VI.

TABLE V. BEST ACCURACY AND F1 SCORE FOR DATA TESTING

F1 Score	Accuracy
0.434	0.829

TABLE VI. CONFUSION MATRIX FOR BEST MODEL

755	54
115	65

It appears that the best model does not produce a high F1 score. This is caused by a dataset that is not good. The f1 score from the data without cleaning the data and oversampling is 0, illustrated using confusion matrix in Table VII and Table VIII. When compared with the results of the confusion matrix of data without data augmentation, a fairly high increase occurs.

TABLE VII. CONFUSION MATRIX RESNET50 WITHOUT DATA AUGMENTATION AND WITHOUT CLEANING DATA

3099	0
365	0

TABLE VIII. CONFUSION MATRIX RESNET 50 WITHOUT DATA AUGMENTATION

754	54
140	40

IV. CONCLUSION

This research is looking for the best ResNet model with a scheme without data augmentation, with data augmentation, and with data augmentation and dropout. The best models obtained in this study using 10 epoch training is the ResNet 50, with data augmentation but without dropout, that gives accuracy of 0.83, f1 score for malignant class 0.43 and f1 score for benign class of 0.87. In the future, the model can be improved using a high accuracy segmentation or some optimizers.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[2] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6.

[3] H. R. Mhaske and D. A. Phalke, "Melanoma skin cancer detection and classification based on supervised and unsupervised learning," in *2013 International conference on Circuits, Controls and Communications (CCUBE)*, 2013, pp. 1–5.

[4] C. Doukas, P. Stagkopoulos, C. T. Kiranoudis, and I. Maglogiannis, "Automated skin lesion assessment using mobile technologies and cloud platforms," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, no. c, pp. 2444–2447, 2012.

[5] N. C. Lynn and Z. M. Kyu, "Segmentation and Classification of Skin Cancer Melanoma from Skin Lesion Images," in *2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, vol. 2017–Decem, pp. 117–122, 2017.

[6] Y. Li and L. Shen, "Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network," *Sensors*, vol. 18, no. 2, p. 556, Feb. 2018.

[7] R. S. Soumya, S. Neethu, T. S. Niju, A. Renjini, and R. P. Aneesh, "Advanced earlier melanoma detection algorithm using colour correlogram," *2016 Int. Conf. Commun. Syst. Networks, ComNet 2016*, no. July, pp. 190–194, 2017.

[8] S. Joseph and J. R. Panicker, "Skin lesion analysis system for melanoma detection with an effective hair segmentation method," *Proc. - 2016 Int. Conf. Inf. Sci. ICIS 2016*, pp. 91–96, 2017.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, vol. 4, pp. 770–778.

[10] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," no. November, 2015.

[11] M. Sabatelli, M. Kestemont, W. Daelemans, and P. Geurts, "Deep transfer learning for art classification problems," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11130 LNCS, no. September, pp. 631–646, 2019.

[12] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding Data Augmentation for Classification: When to Warp?," in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–6, 2016.

[13] S. Park and N. K. B, "Computer Vision – ACCV 2016," vol. 10111, pp. 189–204, 2017.

[14] A. K. Santra and C. J. Christy, "Genetic Algorithm and Confusion Matrix for Document Clustering," *IJCSI Int. J. Comput. Sci. Issues,* vol. 9, no. 1, pp. 322–328, 2012.

[15] O. Abuzaghleh, B. D. Barkana, and M. Faezipour, "Automated skin lesion analysis based on color and shape geometry feature set for melanoma early detection and prevention," *2014 IEEE Long Isl. Syst. Appl. Technol. Conf. LISAT 2014*, 2014.

[16] H. Tsao *et al.*, "Early detection of melanoma: Reviewing the ABCDEs American Academy of Dermatology Ad Hoc Task Force for the ABCDEs of Melanoma," *J. Am. Acad. Dermatol.*, vol. 72, no. 4, pp. 717–723, 2015.