

# Disease Classification based on Dermoscopic Skin Images Using Convolutional Neural Network in Teledermatology System

I Ketut Eddy Purnama<sup>\*</sup>, Arta Kusuma Hernanda<sup>\*</sup>, Anak Agung Putri Ratna<sup>1</sup>, Ingrid Nurtanio<sup>2</sup>, Afif Nurul Hidayati<sup>3</sup>, Mauridhi Hery Purnomo<sup>\*</sup>, Supeno Mardi Susiki Nugroho<sup>\*</sup>, Reza Fuad Rachmadi<sup>\*</sup>

<sup>\*</sup>Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

<sup>1</sup>Department of Computer Engineering, Universitas Indonesia, Jakarta, Indonesia

<sup>2</sup>Department of Informatics, Universitas Hasanuddin, Makassar, Indonesia

<sup>3</sup>Department of Dermatology and Venereology, Universitas Airlangga, Surabaya, Indonesia

**Abstract** — We have proposed a system of classification and detection of skin diseases that can be applied to Teledermatology. This system will classify skin diseases on dermoscopic images using the Deep Learning algorithm, Convolutional Neural Network (CNN). Dermoscopic image data in this study from MNIST HAM10000 dataset which amounts to 10,015 images and published by International Skin Image Collaboration (ISIC). The dataset is divided into seven class of skin diseases which fall into the category of skin cancer. The image classification process will use two pre-trained CNN models, MobileNet v1 and Inception V3. The model results from the learning process will be applied to a web-classifier. The comparison of predictive accuracy shows that the web-classifier using the CNN Inception V3 model has an accuracy value of 72% while the web-classifier that uses the MobileNet v1 model has an accuracy value of 58%.

**Keywords**—Skin Diseases, Dermoscopic Image, Deep Learning, Convolutional Neural Network.

## I. INTRODUCTION

Skin disease is one of the diseases that often infected the population of Indonesia according to the data of Indonesian Health Profile 2009 which is ranked as the third most disease occurring in the outpatient in hospitals throughout Indonesia. The patients are spread out among the country which consists of thousands of islands. To provide healthcare to the patients including the patient with skin diseases, Primary Healthcare Center are established throughout the region in Indonesia, however this is not balanced with the distribution of medical experts who tend to accumulate in large cities or large provinces. With the advancement of Information technology and communication, health services equalization throughout the country can be done through the so-called telemedicine system, and one of the most applicable telemedicine services is skin disease or teledermatology. This service refers to the utilization of telecommunication technology for medical diagnosis of skin diseases, healthcare, and patients treatment. This service process can be divided into two modes of operation, a real-time mode that allows the patient to directly interact with a medical expert, and a store-and-forward mode that will store patient data before the data is accessed by medical experts at a time. Teledermatology has a great potential to revolutionize medical data delivery to improve skin disease treatment Services (dermatology), provide equitable services to remote areas and allow dermatologist to provide to the patient conducting an advanced examination of the primary dermatological center.

With all the advantages, telemedicine technology is expected to complement conventional diagnostic systems. In conventional dermatological diagnosis systems still require various procedures, such as visual preliminary clinical screening, dermoscopic imaging analysis, biopsies, and histopathological examinations. All benefit of teledermatology system may be added by such a support system for medical expert in making medical diagnosis [2]. This research propose the application of Convolutional Neural Network to make automatic classification of skin lesions appear in dermoscopic images. This automatic classification system can be applied either in store-and-forward mode or in live interactive mode, hence the diagnosis can be made by medical expert confidently and efficiently.

## II. MATERIALS AND METHODS

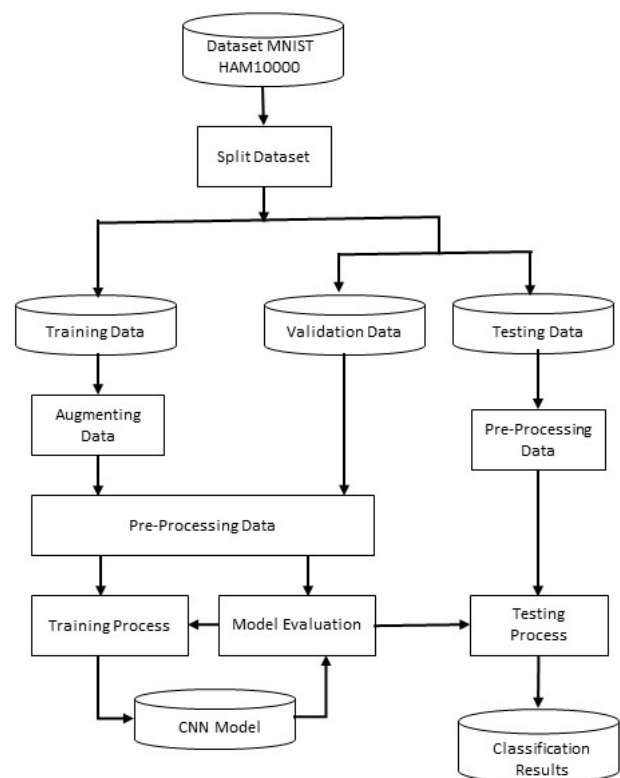


Figure 1. System Design

### A. Research Data

This research uses public dataset MNIST HAM1000 that is available freely on the internet as a large collection of multi-source dermatoscopic images of pigmented lesions. The link of this data is <https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000> [3]. Among several diseases, we only concentrate on seven skin diseases or classes: Melanocytic Nevi (nv), melanoma (mel), Benign keratosis-like lesions (bkl), Basal cell carcinoma (bcc), Actinic keratoses (akiec), Vascular lesions (vasc), dermatofibroma (df).

### B. Methods

The procedures of the classification system using CNN is shown in Figure 1. Research data is divided into training data, and testing data. Part of testing data is taken as data for validation. Then, training data is added with augmented data to increase the number of data and to balance the number of data of all disease classes. Afterwards, the data is pre-processed before entering the training process which include CNN model and model evaluation. In the testing process, the pre-processed data testing are then tested with the results of model evaluation.

### C. Splitting the Data

All dermoscopic image contained in the MNIST HAM1000 dataset has its related classes and features. The classes and features are stored in a formatted file Comma Separated Values(.csv). The MNIST HAM10000 dataset has seven data features with their own names such as lesion\_id, image\_id, dx, dx\_type, age, sex, localization. The lesion\_id feature is an ID that characterizes the patient's identity while image\_id is a numbering ID that characterizes dermoscopic image. The dataset containing duplicate images, in total there are 4,501 duplicated images. We remove them resulted in only 5,514 images remain to be further processed. Then, the processed dataset is then divided into data for the training process, and data for testing and validation. The number of data Training is 9,077 images, the number of data for validation and testing is 938. In detail, Table 1 shows the number of training data and validation of each disease.

Table 1. The data Data Split process

Class Data	Training Data	Validation Data
Melanocytic Nevi (nv)	5,954	751
Melanoma (mel)	1,074	39
Benign Keratosis (bkl)	1,024	75
Basal Cell Carcinoma (bcc)	484	30
Actinic Keratoses (akiec)	301	26
Vascular Lesions (vas)	131	11
Dermatofibroma (df)	109	6
Total	9,077	938

### D. Data Augmentation

Data augmentation is the process to add data so that the number of data to be processed is increasing. From the previous process, it is obtained that the number of dermoscopic image for the training data is 9,077 images and the number of images for validation data and test data are 938 images. In the data augmentation process, new data are only added to the data training. The new data is generated by modifying spatial properties of the images such as horizontal/vertical flips, rotations, variations in the brightness of images, horizontal/vertical shifts and zoom to existing data. Another

aim of data augmentation process is to balance the amount of data from each dataset class. Based on total train data of each class of skin diseases in Table 3, shows the highest number of train data found in the melanocytic class of Nevi (nv) amounting to 5,954 images. The highest amount of data is then used to be the target of adding data to each class that has the amount of data below the target data. The result of adding data per class will be shown in Table 2 .

Table 2. The number of data resulted from data augmentation process

Class Data	Initial	After Data Augmenting
Melanocytic Nevi (nv)	5,954	5,954
Melanoma (mel)	1,074	5,920
Benign Keratosis (bkl)	1,024	5,920
Basal Cell Carcinoma (bcc)	484	5,858
Actinic Keratoses (akiec)	301	5,217
Vascular Lesions (vas)	131	5,290
Dermatofibroma (df)	109	4,410
Total	9,077	38,569

### E. Preprocessing Data

In this stage, pre-processing is done to the training data and validation data. The first step in preprocessing data is to normalize the image size. The original dermoscopic image in the MNIST HAM10000 dataset has a size of 600x450 pixels in the RGB format. The images are then resized to 224x224 pixels to all layers. The RGB values are also normalized into values of -1, 0 and 1.

### F. Training Model

Training model was conducted to retrain pretrained model CNN. Pre-trained models used in this study are MobileNet v1 and Inception V3. Model Training process using Jupyter Notebook environment running on cloud server. In the training Process this model used data train that has passed the process of data augmentation. Augmentation process will increase the amount of data train that originally only 9,077 images to 38,569 images. Data of agumentation results will be used in the training model process while the amount of initial data is still used in train steps determination.

In the process of training model need to determine the size of batch size, train steps and epoch that will be explained as follows:

#### 1. Batch Size

Batch Size determination based on the number of data trains that you want to spread across the neural network in each iteration (train steps). In This training process, the batch size is defined as a number of 10, which means that each step will be spread 10 data train to neural network.

#### 2. Train Steps

Train steps are specified from Equation 1. The number of sample data in that formula is the initial amount of data train before the data augmantation process. From the Equation 1 operation produced the train steps size in the training process model this time is 908.

$$\text{Train Step} = \frac{\text{Sample Data Amount}}{\text{Batch Size}} \quad (1)$$

$$\text{Train Step} = \frac{9.077 \text{ Image}}{10 \text{ Batch Size}} = 908$$

3. Epoch

Epoch is a set of rounds of training steps. Some epoch needed to training a neural network so that errors were found to be close to zero. Epoch on the training model process This is defined a number of 30 times epoch.

G. Pretrained Model

After the size determination of the batch size, train steps and epoch in the training process is to do the process load pretrained model. Pretrained models used this study are MobileNet v1 and Inception V3. Load process pretrained model uses a class function contained in the Keras API. The Class contained in the Keras API will load all layers contained in the architecture of Mobilenet v1 and Inception V3. Here's the code to load all the Mobilenet v1 and Inception V3 layers.

```
MobileNet = keras.applications.mobilenet.MobileNet()
InceptionV3 = keras.applications.inception_v3.InceptionV3()
```

The layer of the MobileNet v1 architecture amounts to 93 layers. Then adjustment layer by removing the last five layers. Next, the addition of two layers are the Dropout layer and Dense layer. The dropout layer is added to prevent overfitting. Dense layer is added to run Fully Connected layer. Dense layer has softmax activation function . The process of removing some layers and adding layers to the model according to the needs of classification is named with Fine tuning. The final model of MobileNet was obtained with a layer of 90 layers. The whole layer is then freeze The weight value except the last 23 layers. Those layers will be the train layer.

Layer on pretrained model Inception V3 numbered 313 layer. Just like the MobileNet model, the model Inception V3 will pass Fine Tuning step by adding two layers of Dropout and two Dense layers. Two layers Dropout added to prevent overfitting. The first Dense layer was added to clarify the first node's output value amounting to 128 using the relay activation function. While on the second Dense layer to clarify the output value of the second node amounting to 7 with the softmax activation function. The layers added on the MobileNet v1 and Inception V3 models.

H. Model Evaluation

After the training model is completed, the next process is to evaluate the training model . This evaluation aims to check the process of training model that has been done. To evaluate the model and validate it used Cross Validation method. This method is a statistical method for evaluating and comparing the learning algorithm. In Cross validation, data train and data validation are crossed in multiple epoch so that each data node has a chance to be validated. In each iteration will use the first fold (K-1) of data to study a single model, and subsequent models learned are required to make predictions about the data. The performance of each learning algorithm on each fold will be known for its accuracy using the confusion matrix. If the model has a low level of accuracy it will be done reconfiguring algorithms used in the training process and then will be retraining. The graph in Figure 2 shows the comparison of the performance of MobileNet v1 and Inception V3 of the training results in the evaluation using Cross Validation.

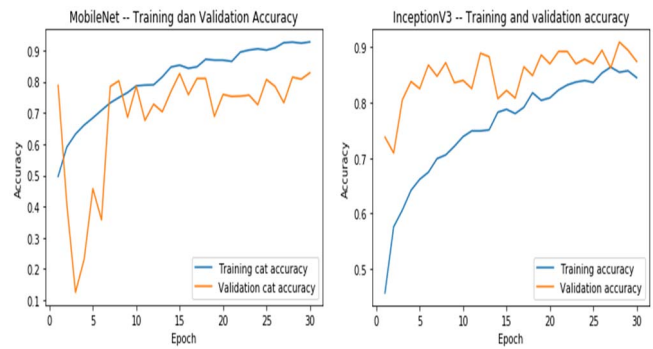


Figure 2. Comparison chart of MobileNet v1 and Inception V3 validation performance

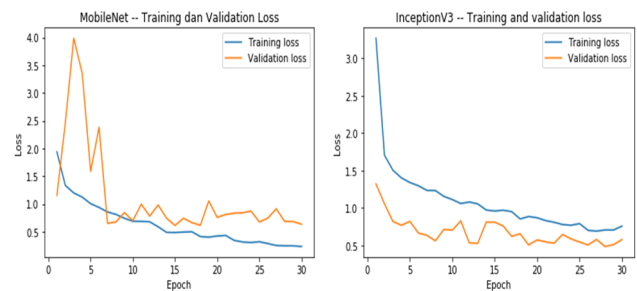


Figure 3. Comparison chart of Value loss during the training process

The chart in Figure 3 shows a comparison of the value of loss that occurred during the training model MobileNet v1 and Inception V3. In Figure 4 indicates that the accuracy value validations on the process training Mobilenet v1 and Inception V3 always increase and in Figure 5 indicate value of loss from both models have decrease during the training so that it canconcluded that MobileNet v1 and Inception V3 doesn't have underfitting and overfitting problem during training process. To clarify the value of cross validation result in each class of disease, then create confusion matrix as in Figure 4 and Figure 5.

I. Implementation

From the design results of the skin disease classification system on dermoscopic image using the Covolutional Neural Network has been designed and undertaken. The system was implemented into a Web-Client-Classifer that runs on teledermatology technology.

**Confusion Matrix MobileNet v1**

akiec	0,46	0,25	0	0,07	0,15	0,07	0
bcc	0,10	0,76	0,03	0,03	0,06	0	0
bkl	0,20	0,08	0,17	0,06	0,28	0,20	0
df	0	0	0	0,50	0,16	0,33	0
mel	0,10	0,07	0	0	0,64	0,17	0
nv	0,007	0,02	0,001	0,05	0,08	0,82	0,001
vasc	0	0	0,09	0	0,09	0,18	0,63
	akiec	bcc	bkl	df	mel	nv	vasc

**Predicted Label**

Figure 4. Confusion matrix training process MobileNet v1

	akiec	bcc	bkl	df	mel	nv	vasc
akiec	0,53	0,34	0,03	0	0	0,07	0
bcc	0,03	0,86	0,03	0	0	0,06	0
bkl	0	0	0,68	0	0,05	0,26	0
df	0	0	0	0,16	0,16	0,66	0
mel	0,02	0,15	0,05	0	0,20	0,56	0
nv	0	0,003	0,01	0	0,001	0,98	0,002
vasc	0	0,09	0,09	0	0	0	0,81

Figure 5. Confusion matrix training process Inception V3

### III. RESULT AND TESTING

The web-classifier in the study is a website used to simulate the implementation of a skin disease classification system on teledermatology technology. There are two types of web-classifier on the design of technology teledermatology, which is "Private Website" which can only be accessed through a network of primary healthcare or clinic that is specifically connected with the server teledermatology and "Public Website" which accessible to anyone who wants to do early detection of the skin diseases they suffer. Figure 6 is the private interface of the website and Figure 7 is a view of the public website.

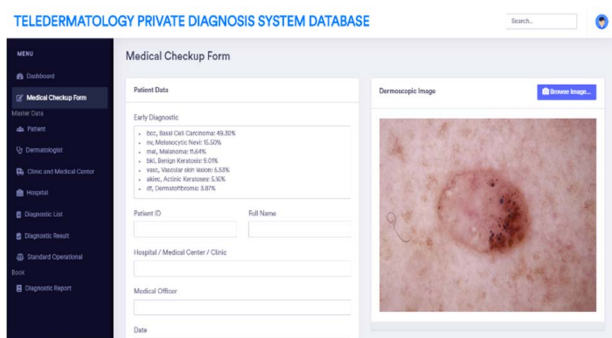


Figure 6. Prediction results are displayed in private website

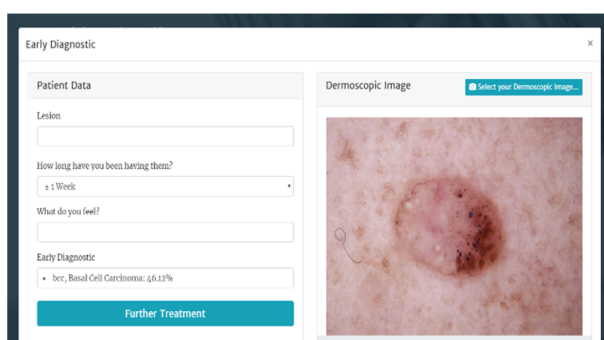


Figure 7. Prediction results are displayed on the public website

In this study also conducted tests on the model of training results that have been implemented in the web-classifier. This test is done by conducting dermoscopic image input by 10 experiments of each class of skin diseases in the web-classifier. Result of web-classifier prediction with MobileNet v1 and web-classifier model with Inception V3 model then compared to its performance. The results is shown in Table 3 and Figure 8.

Table 3. Accuracy Comparison of Inception and MobileNet v1

Disease Class	Inception	MobileNet v1
akiec	90,0	60,0
bcc	80,0	90,0
bkl	80,0	10,0
mel	60,0	40,0
nv	90,0	70,0
df	16,7	70,0
vasc	90,0	66,7

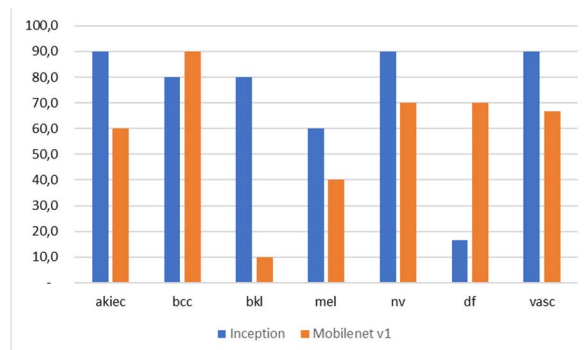


Figure 8. Accuracy Comparison of Inception and MobileNet v1

Performance test of each class of skin diseases on web-classifier mobilenet v1 obtained average value of correct prediction performance only 58% and misclassification value is 42%. While the prediction performance test using Inception V3 was obtained average value of correct prediction performance of 78% and misclassification value only 22%.

### IV. CONCLUSION

We have describe skin disease classification system based on Dermoscopic Skin Images Using Convolutional Neural Network (CNN). We have tested two pre-trained CNN model, Inception and MobileNet v1. Melanocytic Nevi (nv) has the largest dermoscopic image data, which is 5,954 images and becomes the class of skin diseases which has the highest percentage of predictions correctly on the Inception V3 model which is 90%. While, dermatofibroma (df) that has the fewest dermoscopic image data, namely 109 images, becomes the class of skin diseases which has the lowest percentage prediction of 67% in the MobileNet v1 model and 17% in the Inception V3 model. Skin diseases that have a distinctive characteristic of their image data will be easier in predictor for example in vascular skin disease class (vasc). With an original data amount of only 131 images, vascular has an accuracy rate of 70% on the MobileNet v1 model and 90% on the Inception V3 model. Based on the results of testing the application of the CNN model on web classification, the disease classification system can be applied in Teledermatology applications.

### V. ACKNOWLEDGEMENT

We thank to Kemenristekdikti (Kementrian Riset, Teknologi dan Pendidikan Tinggi) for financing the research under the KRUP (Konsorsium Riset Unggulan Perguruan Tinggi) Scheme.

### VI. REFERENCES

- [1] Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.

- [2] Romero, J. Garrido, and M. Garcia-Arpa, "Telemedicine and Teledermatology : Concepts and Applications," *Actas Dermosifiliograficas (English Edition)*, vol. 99, no. 7, pp. 506-522, 2008.
- [3] Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions," *Scientific data*, vol. 5, p. 180161, 2018.
- [4] M. Binder, A. Steiner, M. Schwarz, S. Knollmayer, K. Wolff, and H. Pehamberger, "Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study," *British Journal of Dermatology*, vol. 130, no. 4, pp. 460-465, 1994.
- [5] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical 61 imaging (isbi), hosted by the international skin imaging collaboration (ISIC)," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 168-172, IEEE, 2018.
- [6] Bengio et al., "Learning deep architectures for AI," *Foundations and trends R in Machine Learning*, vol. 2, no. 1, pp. 1127, 2009.
- [7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097
- [9] S. Iovte and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 19, 2015.