Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13110
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Hair Disease Classification Using Convolutional Neural Network (CNN) Algorithm with VGG-16 Architecture

Ichwanul Muslim Karo Karo[1)*], Dedy Kiswanto[2)], Suvriadi Panggabean[3)], Adidtya Perdana[4)]
[1,2,3,4)]Universitas Negeri Medan
[1)]ichwanul@unimed.ac.id, [2)] dedykiswanto@unimed.ac.id, [3)] suvriadi@unimed.ac.id, [4)] adidtya@unimed.ac.id

Abstract: Hair diseases are common and can be caused by a variety of factors, including genetics, stress, nutritional deficiencies, as well as exposure to sunlight and air pollution. Accurate diagnosis of hair diseases is important for proper treatment, but can be challenging due to overlapping symptoms. The development of the healthcare world has widely utilized machine learning and deep learning approaches to assist in the healthcare field. This research aims to develop hair disease classification using Convolutional neural network (CNN). The CNN-based approach is expected to help health professionals diagnose hair diseases accurately and provide targeted treatment. This research involves an experimental design with three main stages: identifying the research problem, conducting a literature review, and collecting data. The research uses a dataset of hair disease images obtained from Kaggle, which are annotated and organized based on different hair disease types. After the image data is collected, the image dataset will go through the image preprocessing stage. Experiments were conducted using hair disease image data with 15 epochs on a CNN Deep Learning model with VGG-16 architecture, and resulted in an accuracy of 94.5% and a loss rate of 18.47%, with a testing epoch time of 9 hours 48 minutes. The results of this study show that CNN with VGG-16 architecture can successfully classify 10 types of hair diseases.

Keywords: CNN; Hair Diseases; VGG-16; Accuracy; Loss Rate

INTRODUCTION

Hair is a part of the human body that has an important role in determining the beauty and appearance of a person. Head hair has a shape similar to strands and grows on the human scalp with a lot of keratins, and can emerge from the epidermis layer (Karuna & Petrus, 2023). However, just like any other part of the body, hair is also susceptible to various kinds of diseases and damage. Diseases and damage to the hair can be caused by many factors, such as improper care, sun exposure, air pollution, genetics, stress, and lack of balanced nutritional intake. Some examples of hair diseases include alopecia, telogen effluvium, and scalp psoriasis.

Treatment of hair diseases requires an accurate and precise diagnosis in order to provide the right treatment. However, the diagnosis of hair diseases can be a challenge because often the symptoms of hair diseases are similar to each other. Therefore, a hair disease classification system is needed that can help doctors and health professionals diagnose hair diseases with high accuracy.

Machine learning is a technology or algorithm that has been widely implemented in various fields to help solve various problems in various fields. In its application, machine learning algorithms are trained to recognize patterns, classify data and make predictions based on the data entered. Deep learning is a method of applying machine learning that uses artificial reasoning networks to mimic the way the human brain works (Bertoni et al., 2022). Using algorithms as its 'neurons', deep learning is able to determine and analyze the characteristics of a data set. Furthermore, deep learning models can recognize complex patterns in images, text, sound, and other data to generate accurate insights and predictions(Chang et al., 2020). Thus, it is also possible to identify hair diseases using the image of the disease.

LITERATURE REVIEW

Recentl, deep learning technology, especially Convolutional neural network (CNN) has been widely implemented and brought revolutionary success to help the healthcare world in detecting various deadly diseases.. Research by (Aytaç Korkmaz & Binol, 2018) developed early detection of stomach cancer with the help of stomach cancer images obtained from the Department of Pathology, Faculty of Medicine, Fiat University. The features

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13110
e-ISSN : 2541-2019
p-ISSN : 2541-044X

used are Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG). Artificial neural network (ANN) and Random Forest (RF) classifiers were used to classify stomach cancer images with these new lower feature sizes. New medical systems have been developed to measure the effect of these dimensions by obtaining features in different dimensions with dimensionality reduction methods. When all the developed methods are compared, it has been found that the best accuracy results are obtained with the LBP and ANN method.

CNN algorithm has also been widely used to identify skin cancer. Based on the database on Mendeley repository, there are at least 132 articles that discuss the application of the CNN algorithm to identify skin cancer in 2023. One of them is research by (Nandhini et al., 2023). Their research combines CNN algorithm with VGG-16 Architecture. In the experiment's training set, more than 1697 photos were used, resulting in early detection with 96% accuracy. 510 images were used to train the random forest model. A random forest algorithm is also used for testing, and the most accurate method is chosen based on accuracy.

Other deadly diseases that have been identified by CNN are and brain tumors. A study by (Seetha & Raja, 2018) proposes automatic brain tumor detection using Convolutional Neural Networks (CNN) algorithm. Deeper architectural design is done by using small kernels. The neuron weights are made small. The experimental results show that the accuracy rate of CNN reaches 97.5% with low complexity and compared with all existing methods.

Research by (Aslan et al., 2022) investigated thirteen different deep CNN experiments and evaluations, each using 80-20% of lung X-ray images for training and testing. Analysis using INCA feature selection in the VGG-16 network yielded the best prediction values. The average values of accuracy, sensitivity, F-score, precision, MCC, dice, Jaccard, and specificity were 99.14%, 97.98%, 99.58%, 98.80%, 97.81%, 98.83%, 97.68%, and 99.56%, respectively. This proposed study demonstrates the usefulness of the deep CNN model in classifying COVID-19 in X-ray images.

Based on the exposure of several literature studies above, it is possible to identify hair diseases with the CNN algorithm with VGG-16 architecture. This approach is expected to help doctors in making an accurate diagnosis and providing more targeted treatment. In addition, this method can also make it easier for people to detect hair diseases and take the necessary precautions.

METHOD

This section presents the research flowchart and the explanation. The research flowchart adopts previous research (Karo Karo et al., 2023) and shown in Fig. 1. There four fundamental process and a supporting process on this study. Fundamental process is collecting dataset, image preprocessing, deploy model and evaluation process. In addition, supporting process is splitting dataset into data training and data testing.

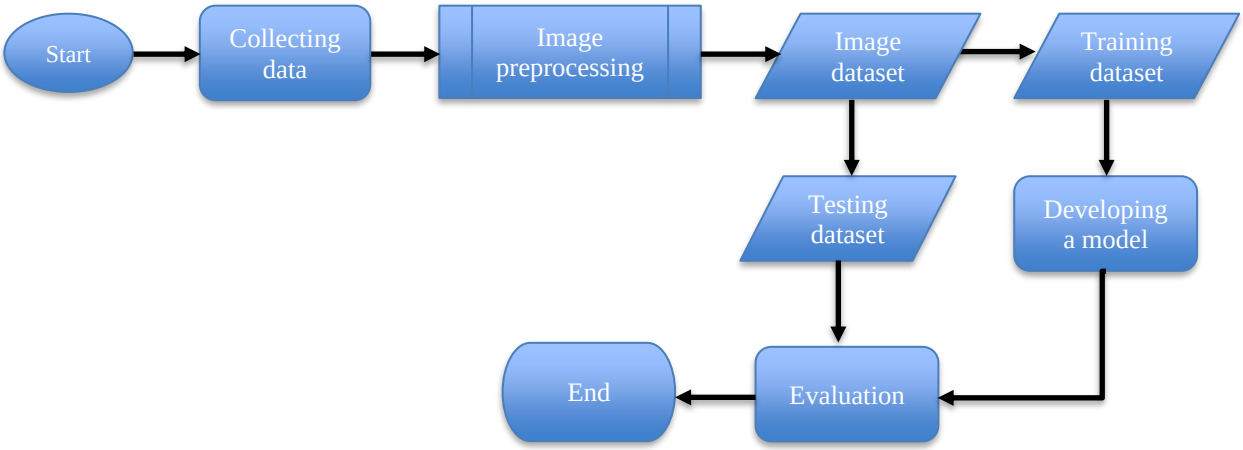


Fig. 1 Research flowchart

Collecting Data

This study extracts images of Hair Diseases from Kaggle. There are 10 types of hair diseases that are successfully obtained in the dataset (Table 1), these are alopecia areata, contact dermatitis, folliculitis, head lice, lichen planus, male pattern baldness, psoriasis, Seborrheic Dermatitis, Telogen Effluvium, dan Tinea Capitis. Examples of hair diseases dataset can be seen in Table 1. Each disease type has 400 images as dataset, so the total image dataset is 4000.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13110
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1 Example of dataset

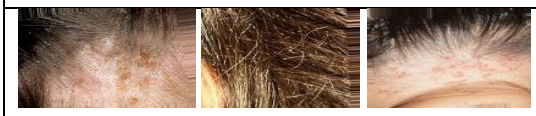
Image	Type	Description of diseases
	Alopecia Areata	It is known as hair loss, a condition in which the immune system attacks the body's own healthy cells. The disease attacks the hair follicles (Żeberkiewicz et al., 2021).
	Contact Dermatitis	Skin inflammation characterized by a reddish itchy skin rash, which arises due to irritation after direct contact with certain substances, or due to an allergic reaction to certain substances (Johansen et al., 2022) .
	Folliculitis	a condition in which hair follicles become inflamed (Chang et al., 2020)
	Head Lice	Head lice are tiny insects that feed on blood from the human scalp. The insects usually spread through direct transfer from the hair of one person to the hair of another (Darmadi et al., 2018)
	Lichen Planus	inflammation of the skin, nails, or mucous membranes due to abnormalities in the immune system (Lewis, 2022).
	Male Pattern Baldness	Male scalp hair loss, also known as androgenetic alopecia, occurs due to changes in hormone levels as a person ages. Genetic factors also affect this condition (Hillmer et al., 2008)
	Psoriasis	inflammation of the skin that causes scaly, thickened, flaky skin, and sometimes itchy skin (Li et al., 2021).
	Seborrheic Dermatitis	It is a common skin disease that occurs on the scalp. This condition causes redness, itching, scaling and dandruff (Moreno-Vázquez et al., 2021).
	Telogen Effluvium	a condition where hair falls out excessively and suddenly. This hair loss is generally not permanent. Telogen effluvium can result from hormonal changes, stress, or the use of certain medications (Rebora, 2019).
	Tinea Capitis	a superficial fungal infection that affects hair follicles and hair shafts, including head hair, eyebrows, and eyelashes, caused by dermatophytes of the genus Trichophyton and Microsporum (Heath & Usatine, 2022).

Image Preprocessing

In this research, image preprocessing goes through two stages. The first stage resizes the hair disease image from its previous size. The entire image is uniformed to 224 x 224 pixels. The second stage is the coloring stage. Coloring process on this study follows the procedure of previous research (Karo Karo et al., 2023). Coloring process is useful for changing the color of the image so that it clarifies the object of hair disease that you want to process.

Model CNN

Convolutional Neural Network (CNN) is a class of deep neural networks, most commonly applied to analyze visual imagery. CNN is a convolution operation consisting of several stages or processing layers inspired by

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13110
e-ISSN : 2541-2019
p-ISSN : 2541-044X

biological neural systems (Bertoni et al., 2022). Convolutional Layer is the core of the Deep Learning Convolutional Neural Network model, convolution kernels are used to extract features from the input image. This process converts the next layer of pixels into the receptive field of the connected area of any convolution kernel on the input image. The Pooling Layer performs down-sampling to improve image feature extraction by minimizing the dimension of the feature map and the number of parameters in the network as well as speeding up computation to control overfitting. This study used CNN algorithm that has been designed on previous study (Karo Karo et al., 2023).

VGG-16 Architecture

VGG16 refers to the VGG model, also called VGGNet. It is a convolution neural network (CNN) model supporting 16 layers. VGG-16, as its name suggests, is a 16-layer deep neural network. VGG-16 is thus a relatively extensive network with a total of 138 million parameters—it's huge even by today's standards. However, the simplicity of the VGG-Net16 architecture is its main attraction. Fig. 2 illustrates the VGG-architecture used on this study. The architecture adopted from previous research (Bezdan & Bačanin Džakula, 2019).

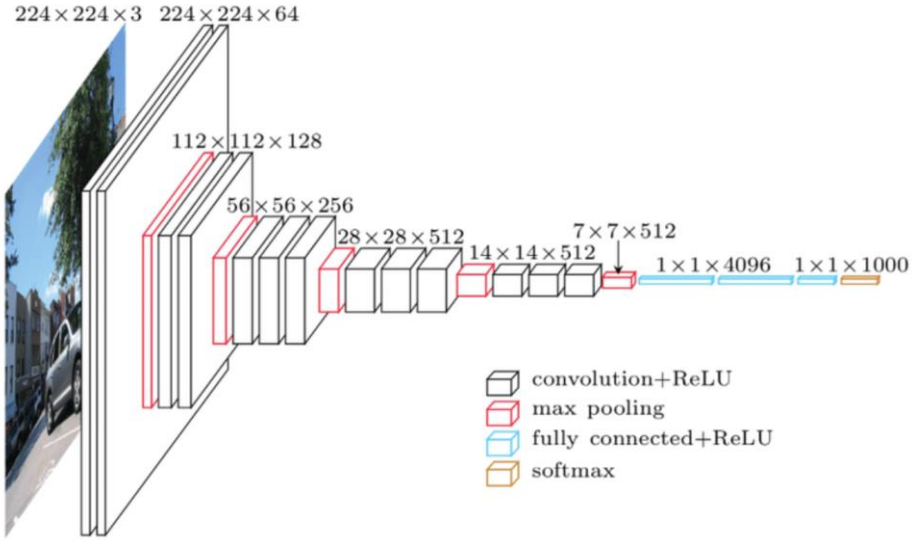


Fig. 2 VGG-16 Architecture (Bezdan & Bačanin Džakula, 2019)

Softmax activation is another description of Logistic Regression that can be utilized for clustering more than two classes [13]. Softmax is useful for converting the output of the last layer into its basic probability distribution. Softmax could be calculated by equation (1).

$$f_i(\vec{x}) = \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_i}} \quad (1)$$

RESULT

There are twelve parameter setups on CNN algorithm. This study sets the initial parameter setup. Table 2 shows the parameter setup on CNN algorithm in this study. By these parameters, the CNN algorithm trains the dataset to provide a model. This study just uses Rel U activation function.

Table 2
Initial Parameter of CNN Algorithm

Layer (type)	Output Shape	Parameters
conv2d (Conv2D)	(None, 150, 150, 32)	2432
max_pooling2d (Maxpooling2D)	(None, 75, 75, 32)	0
conv2d_1 (COnv2D)	(None, 75, 75, 64)	18496
max_pooling2d_1 (Maxpooling2D)	(None, 37, 37, 64)	0
conv2d_2 (COnv2D)	(None, 37, 37, 96)	55392
max_pooling2d_2 (Maxpooling2D)	(None, 18, 18, 96)	0
conv2d_3 (COnv2D)	(None, 18, 18, 96)	83040
max_pooling2d_3 (Maxpooling2D)	(None, 9, 9, 96)	0

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13110
e-ISSN : 2541-2019
p-ISSN : 2541-044X

flatten (Flatten)	(None, 7776)	0
dense (Dense)	(None, 512)	3981824
activation (Activation)	(None, 512)	0
Dense_1 (Dense)	(None, 5)	2565

Result of Image Preprocessing

All hair disease image data from research shoots undergo data preprocessing such as data resizing and data coloring. The results of image preprocessing can be seen in Fig. 3.

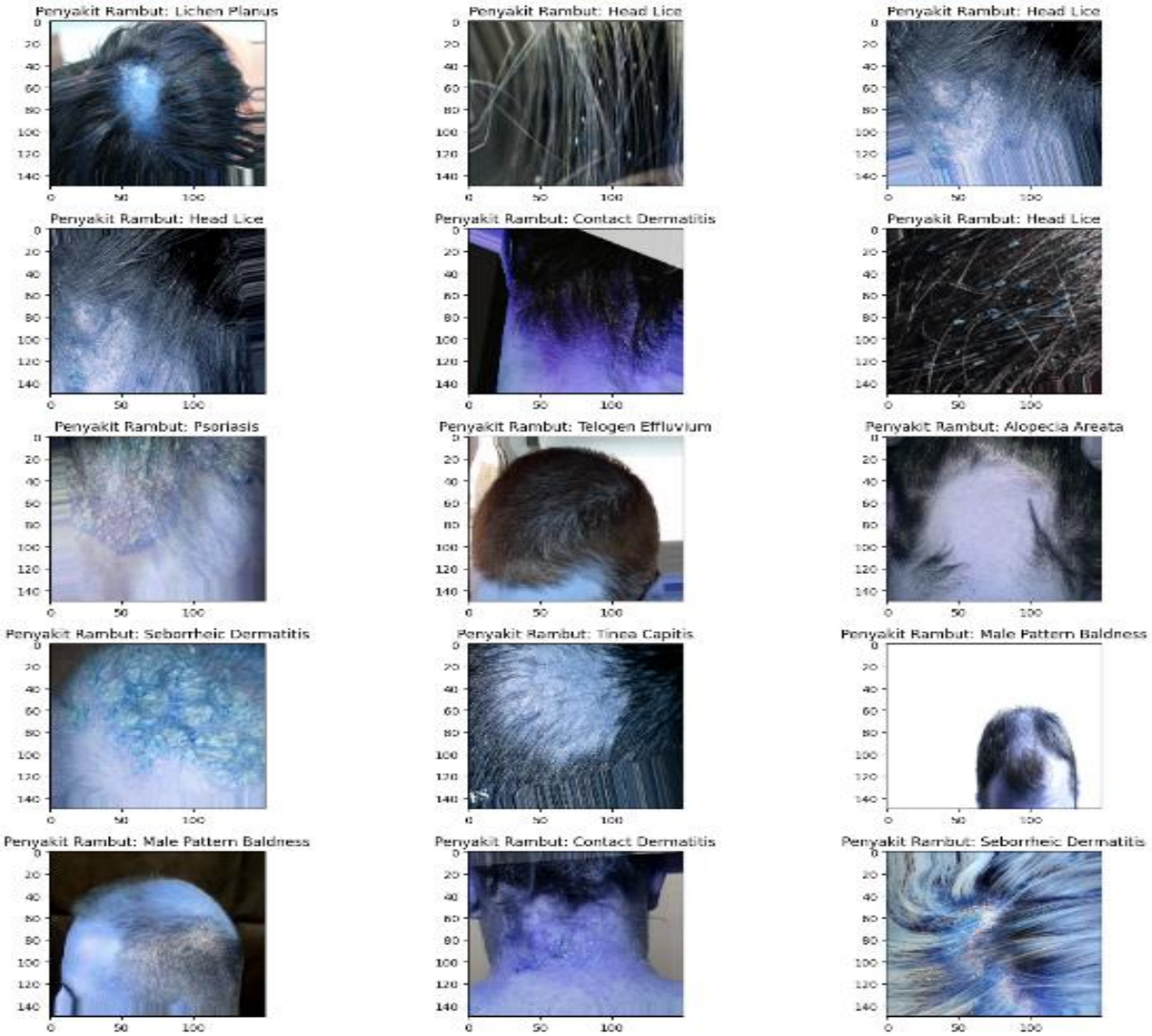


Fig. 3 Image Preprocessing Result

Result of Model CNN with VGG-16

It can be seen from the fig. 4, the prediction results reached 94.5% accuracy, with a misclass of 0.55%. From the picture above it can also be seen that for Alopecia disease does not experience misclass or correctly predicted data. For Contact Dermatitis disease experiencing misclass and predicted Head Lice disease as much as 2 and 38 other data correctly predicted. Folliculitis disease does not experience misclass or correctly predicted data. Head Lice disease does not experience misclass or correctly predicted data. Lichen Planus disease has a misclass and predicted Contact Dermatitis disease as much as 1, predicted Seborrheic Dermatitis disease as much as 1, predicted Tinea Capitis disease as much as 4 and 34 other data are correctly predicted. Male Pattern Baldness disease has a misclass and predicted Telogen Effluvium disease as much as 1, and 39 other data are correctly predicted. Psoriasis disease has a misclass and predicted Male Pattern Baldness disease as much as 1, and 39 other data are correctly predicted. Seborrheic Dermatitis disease has a misclass and predicted Alopecia disease as much as 1, predicted

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13110
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Folliculitis disease as much as 1, predicted Head Lice disease as much as 2, predicted Lichen Planus disease as much as 1, predicted Psoriasis disease as much as 1 and 34 other data correctly predicted. Telogen Effluvium disease has a misclass and predicted Contact Dermatitis disease as much as 1, predicted Tinea Capitis disease as much as 1 and 38 other data are correctly predicted. Tinea Capitis disease has a misclass and predicted Contact Dermatitis disease as much as 2, predicted Psoriasis disease as much as 1, predicted Seborrheic Dermatitis disease as much as 1 and 36 other data correctly predicted

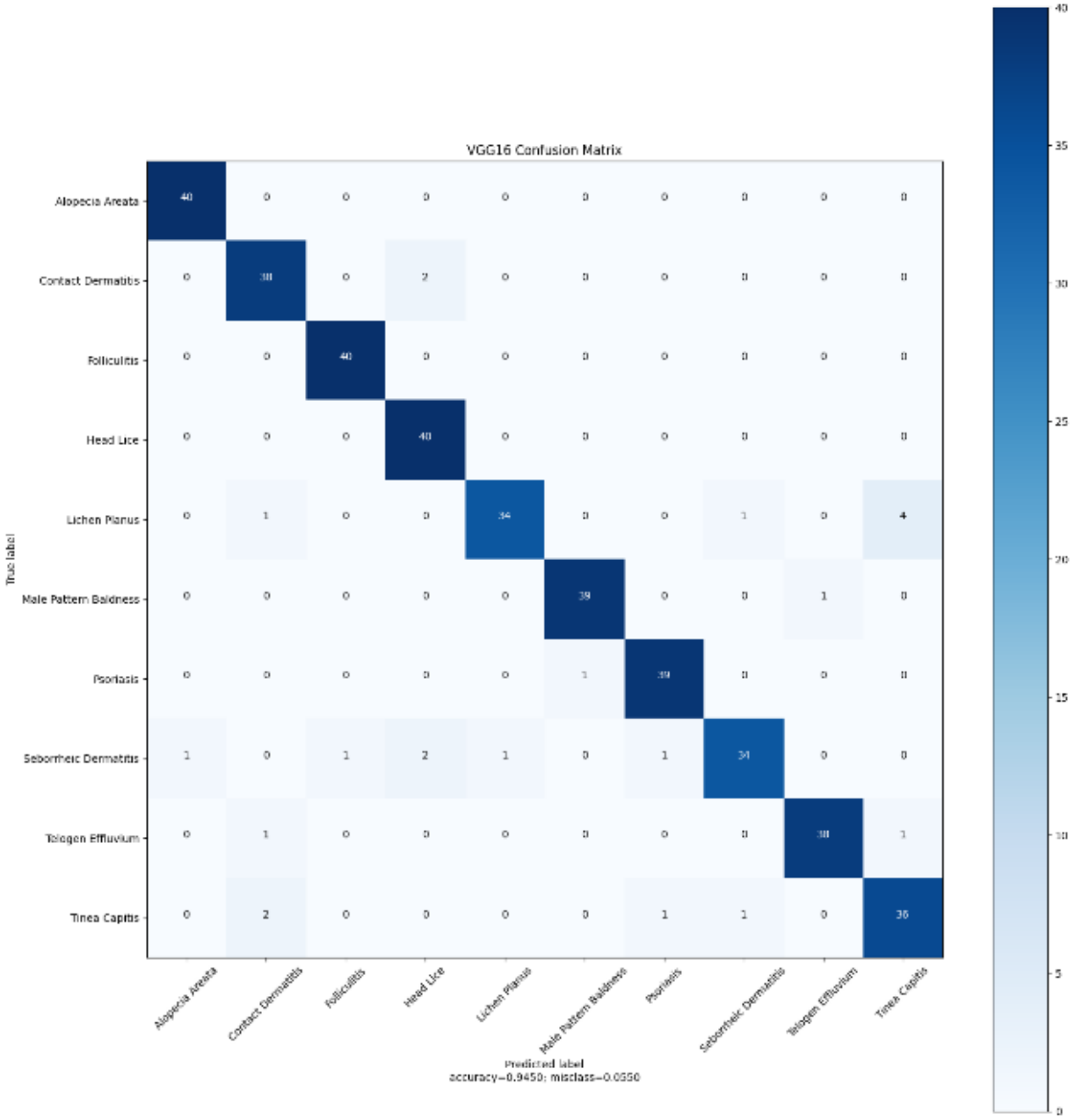


Fig. 4 Performance of CNN Model with Adam Activation Function

.DISCUSSIONS

This study also analyzes each experiment from the epoch aspect. The results can be seen in Fig. 7. Experiments conducted using hair disease image data with an epoch of 15 on the Deep Learning CNN model with VGG-16 architecture and the resulting accuracy value is 94.5% and a loss rate of 18.47% with an epoch testing time of 9 hours 48 minutes.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13110
e-ISSN : 2541-2019
p-ISSN : 2541-044X

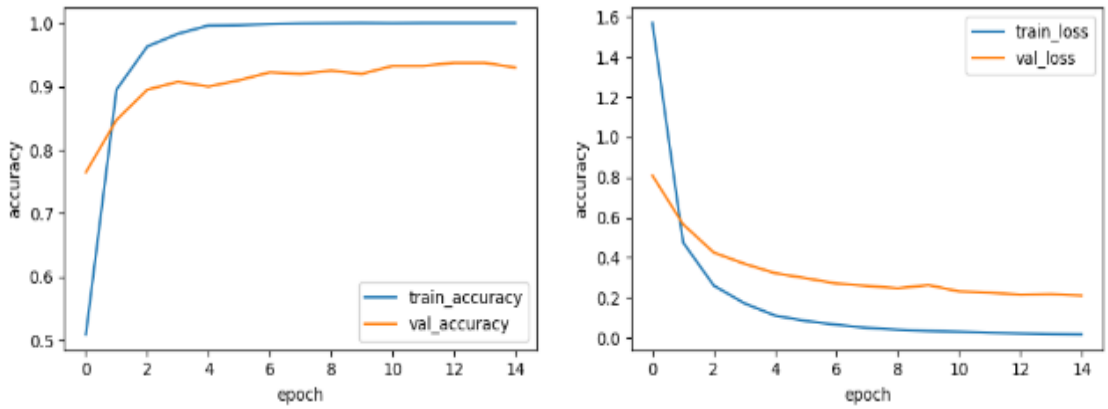


Fig. 5 Accuracy and Loss of Model

CONCLUSION

Classification of Hair Disease Types applies CNN algorithm with VGG-16 architecture. This study uses 10 types of hair diseases, namely Alopecia Areata, Contact Dermatitis, Folliculitis, Head Lice, Lichen Planus, Male Pattern Baldness, Psoriasis, Seborrheic Dermatitis, Telogen Effluvium and Tinea Capitis. The hair disease data is obtained from Kaggle.com as much as 4000 data, which is divided into 3, training data, testing data and validation data. From the application of the CNN algorithm with VGG-16 architecture, this study obtained an accuracy of 94.5%, a misclass of 0.55% with an epoch testing time of 9 hours 48 minutes.

REFERENCES

Aslan, N., Ozmen Koca, G., Kobat, M. A., & Dogan, S. (2022). Multi-classification deep CNN model for diagnosing COVID-19 using iterative neighborhood component analysis and iterative ReliefF feature selection techniques with X-ray images. *Chemometrics and Intelligent Laboratory Systems*, *224*. https://doi.org/10.1016/j.chemolab.2022.104539

Aytaç Korkmaz, S., & Binol, H. (2018). Classification of molecular structure images by using ANN, RF, LBP, HOG, and size reduction methods for early stomach cancer detection. *Journal of Molecular Structure*, *1156*. https://doi.org/10.1016/j.molstruc.2017.11.093

Bertoni, F., Citti, G., & Sarti, A. (2022). LGN-CNN: A biologically inspired CNN architecture. *Neural Networks*, *145*. https://doi.org/10.1016/j.neunet.2021.09.024

Bezdan, T., & Bačanin Džakula, N. (2019). *Convolutional Neural Network Layers and Architectures*. https://doi.org/10.15308/sinteza-2019-445-451

Chang, W. J., Chen, L. B., Chen, M. C., Chiu, Y. C., & Lin, J. Y. (2020). ScalpEye: A Deep Learning-Based Scalp Hair Inspection and Diagnosis System for Scalp Health. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3010847

Darmadi, D., Pradhasumitra, D., & Setiawan, S. E. (2018). EFEKTIFITAS EKSTRAK KULIT DUKU (Lansium domesticum corr) TERHADAP MORTALITAS PEDIKULUS HUMANUS CAPITIS SEBAGAI PENYEBAB PEDIKULOSIS PADA ANAK. *JOPS (Journal Of Pharmacy and Science)*, *1*(2). https://doi.org/10.36341/jops.v1i2.487

Heath, C. R., & Usatine, R. P. (2022). Tinea capitis. *The Journal of Family Practice*, *71*(8). https://doi.org/10.12788/jfp.0493

Hillmer, A. M., Brockschmidt, F. F., Hanneken, S., Eigelshoven, S., Steffens, M., Flaquer, A., Herms, S., Becker, T., Kortüm, A. K., Nyholt, D. R., Zhao, Z. Z., Montgomery, G. W., Martin, N. G., Mühleisen, T. W., Alblas, M. A., Moebus, S., Jöckel, K. H., Bröcker-Preuss, M., Erbel, R., … Nöthen, M. M. (2008). Susceptibility variants for male-pattern baldness on chromosome 20p11. *Nature Genetics*, *40*(11). https://doi.org/10.1038/ng.228

Johansen, J. D., Bonefeld, C. M., Schwensen, J. F. B., Thyssen, J. P., & Uter, W. (2022). Novel insights into contact dermatitis. *Journal of Allergy and Clinical Immunology*, *149*(4). https://doi.org/10.1016/j.jaci.2022.02.002

Karo Karo, I. M., Karo Karo, J. A., Ginting, M., Yunianto, Y., Hariyanto, H., Nelza, N., & Maulidna, M. (2023). Comparison of Activation Functions on Convolutional Neural Networks (CNN) to Identify Mung Bean Quality. *Sinkron*, *8*(4), 2757–2764. https://doi.org/10.33395/sinkron.v8i4.13107

Karuna, E., & Petrus, J. (2023). Penentuan Tingkat Kerontokan Rambut Kepala Pria dengan Metode Fuzzy Inference System Sugeno. *Jurnal Algoritme*, *3*(2). https://doi.org/10.35957/algoritme.v3i2.4204

Lewis, F. M. (2022). Lichen Planus. In *Ridley's The Vulva*. https://doi.org/10.1002/9781119755173.ch24

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Li, X. Q., Chen, Y., Dai, G. C., Zhou, B. B., Yan, X. N., & Tan, R. X. (2021). Abietic acid ameliorates psoriasis-like inflammation and modulates gut microbiota in mice. *Journal of Ethnopharmacology*, *272*. https://doi.org/10.1016/j.jep.2021.113934

Moreno-Vázquez, K., Calderón, L., & Bonifaz, A. (2021). Seborrheic dermatitis. An update. In *Dermatologia Revista Mexicana* (Vol. 65).

Nandhini, V., Sam Christopher, S., Shivashnee, B., & Kumar, C. R. (2023). Early Detection of Melanoma Using Convolutional Neural Network and Random Forest Algorithm. *2023 9th International Conference on Advanced Computing and Communication Systems, ICACCS 2023*. https://doi.org/10.1109/ICACCS57279.2023.10112814

Rebora, A. (2019). Telogen effluvium: A comprehensive review. In *Clinical, Cosmetic and Investigational Dermatology* (Vol. 12). https://doi.org/10.2147/CCID.S200471

Seetha, J., & Raja, S. S. (2018). Brain tumor classification using Convolutional Neural Networks. *Biomedical and Pharmacology Journal*, *11*(3). https://doi.org/10.13005/bpj/1511

Żeberkiewicz, M., Rudnicka, L., & Malejczyk, J. (2021). Immunology of alopecia areata. In *Central European Journal of Immunology* (Vol. 45, Issue 3). https://doi.org/10.5114/CEJI.2020.101264

*name of corresponding author