Classification of Herbal Plants Based on Leaf Images using Convolutional Neural Network

Kana Saputra S[1], Debi Yandra Niska[2], Insan Taufik[3], Mhd Hidayat[4], Dinda Farahdilla Dharma[5]

{kanasaputras@unimed.ac.id[1], debiyandraniska@unimed.ac.id[2], insantaufik@unimed.ac.id[3], mhdhidayat@mhs.unimed.ac.id[4], dindafarahdilla@mhs.unimed.ac.id[5]}

Study Program of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, North Sumatera, Indonesia[1,2,3,4,5]

Abstract. Indonesia is an agricultural country that is famous for its wealth of spices and herbal plants. Herbal plants themselves have thousands of species. There are 40,000 species of herbal plants that have been known in the world, and around 30,000 species to be in Indonesia. Herbal plants are a source of new active compounds that have pharmacological and therapeutic effects, both when used directly and through various extraction processes. Herbal plants can be distinguished from the shape of the leaves because each type of plant has different leaf features. Laboratory-based testing also requires skills in sample processing and data interpretation, in addition to time-consuming procedures. Therefore, a simple and reliable herbal plant recognition technique is needed to quickly identify herbs, especially for those who are unable to use expensive analytical instrumentation. This study aims to identify types of herbal plants based on leaf images quickly and accurately using the Convolutional Neural Network method which is part of Deep Learning. This study uses several architectural models of Convolutional Neural Network to classify types of herbal plants. The best accuracy value with the VGG16 architecture is 90% with 93% precision, 90% recall, and 90% F-measure. The VGG16 architecture used epoch = 20, batch_size = 32, and validation_split = 0.2. The result show that CNN Algorithm with the VGG16 architecture is able to classify types of herbal plants with good accuracy.

Keywords: Convolutional Neural Network, Herbal plants, Leaf Images.

1 Introduction

Indonesia is an agricultural country that is famous for its wealth of spices and herbal plants. Herbal plants themselves have thousands of species. There are 40,000 species of herbal plants that have been known in the world, and around 30,000 species to be in Indonesia [1]. Herbal plants have been proven to be able to cure various diseases since old time in various countries. [2]–[4] almost all over the country, herbal plants have been used extensively throughout history for the treatment of various diseases with minimal or no side effects. In addition,

ICIESC 2022, October 11, Medan, Indonesia

DOI 10.4108/eai.11-10-2022.2325271

herbal plants are the materials that are easily available, effective, and affordable without causing a very bad impact on human health.

Although, nowdays most of the modern medicine and pharmaceuticals have replaced traditional medicine as the main treatment for human diseases, herbal medicine is still widely practiced throughout the world. Because when humans consume it, the compounds of herbal plants are well placed to interact with human protein targets, or to alter the growth of commensal organisms, pathogens or parasites that live in the human body, which in turn affects human health and well-being disease [5]. In addition, humans today tend to switch to herbal plants because herbal plants have relatively few side effects compared to modern chemical medicines [6].

Especially in Indonesia, many processed herbal plants that have been developed by ancestors for treatment are called “Jamu”. However, herbs or herbal preparations are not easily created to cure various diseases. Herbal plants are a source of new active compounds that have pharmacological and therapeutic effects, even when used directly and through various extraction processes [7]. Selection of herbal plants is not just done, but through various processes and classification of the right types of plants, one of which is by classifying the leaves on herbal plants.

Herbal plants can be distinguished from leaf shape because each type of plant has different leaf features [8]. Besides that, leaves are easier to obtain because they do not depend on the season and position on the plant part. The position of the leaves on the part of the plant does not require special tools when sampling data in the field and does not disturb and damage the plants. Plant systematics can be classified and recognized based on their reproductive system (flowers) and leaf morphology [9].

Laboratory-based testing also requires skills in sample processing and data interpretation, other than time-consuming procedures [10]. Therefore, a simple and reliable herbal plant recognition technique is needed to quickly identify herbal plants, especially for those who are unable to use expensive analytical instrumentation [9]. The use of Deep Learning in classifying types of herbal plants which was developed to assist humans in recognizing and identifying unknown types of herbal plants more quickly.

Currently, the best image classification results are obtained using a Deep Learning algorithm based on Convolutional Neural Networks (CNN), which has thousands or even millions of adjustable parameters [11]. Deep Learning is a subfield of Artificial Intelligence (AI) that imitates the work of the human brain in processing data and generating patterns for use in decision making [12]. Deep Learning is also powerful for feature extraction because it is superior in providing deeper image information [13]. Based on this explanation, the researcher will apply the Convolutional Neural Network (CNN) method which is part of Deep Learning to classify types of herbal plants quickly and accurately.

2 Research Method

2.1 Data Collection

The data is image of the leaves of herbal plants which are still believed by the Indonesian people to have usefull for various types of diseases or can increase immunity. The types of herbal plants used were 105 leaf images with different sizes. Leaf image data was obtained directly from the location of herbal plant cultivation in Medan Marelan, North Sumatera, Indonesia.

2.2 Data Preprocessing

Data preprocessing is carried out before the classification process. The preprocessing is cropping. Cropping is taking an important part of the leaf image [14].

(a) Before Cropping

(b) After Cropping

Fig. 1. Cropping Process of Leaf Image.

2.3 Data Split

Before the classification process is carried out, first the data must be divided into two part for training data, data validation, and data testing. Training data is used to form the model, validation data is used to evaluate the performance of the model, while testing data is used to test the classification accuracy of the model that has been formed. The percentage of data is 80% for training data and 20% for data testing [15]. The training data and validation data will be used for the modeling stage using the CNN model. Testing data will be used for the stage of testing the formed model.

2.4 Implementation of Convolutional Neural Network

Deep Learning is part of Artificial Intelligence based on Artificial Neural Network (ANN). Deep Learning can be implemented on a computer to classify data retrieval in the form of images, sound, text, or video. The model is formed from the results of testing and training using data sets that have labels and large number. The data is converted into pixel values in the image to be used as an internal representation or feature vector and then classification is obtained and used for detection to classify patterns in the input process [16].

CNN consists of input layer, convolutional, pooling, fully connected layers, and output layer components which are arranged in a stack as shown in Figure 2 [17]. The convolutional layer is the core of the CNN method. This layer is used to extract information from the input used filters which are automatically taught to detect certain features in an image. Layer pooling aims to reduce the number of parameters and computations that will be calculated by doing downscaling. This reduces network complexity and the possibility of overfitting. The fully connected layer is a layer where each neuron is connected to each other with an activation function in the next layer [18].

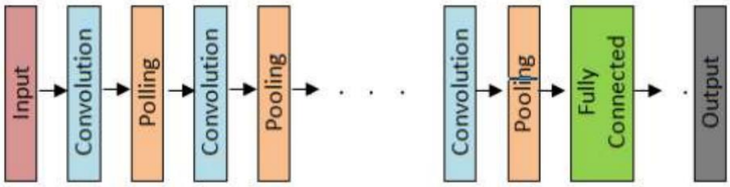


Fig. 2. Components of Convolutional Neural Network.

Modeling for the classification of herbal plant species was carried out using the CNN. The output of the training process is a model that can identify types of herbal plants. The CNN architecture used is LeNet-5, AlexNet [19], VGG16 [20], VGG19, Inception-V3, and ResNet50 [21] which will be trained separately. The activation function used in the output layer uses the softmax function. The cost function used is categorical cross entropy [22].

2.5 Evaluation

The model generated by CNN was evaluated using the calculation of Accuracy, Recall, Precision, and F-Measure. The calculation is based on True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) [23]. The definition and calculation formula for Accuracy, Recall, Precision, F-Measure can be seen in Table 1 [24].

Table 1. Definition & Formula of Accuracy, Recall, Precision, F-Measure.

Size	Definition	Formula
Accuracy (A)	describes how accurate the model is in classifying correctly	A = (TP+TN) / (Total no of samples)
Precision (P)	the number of positive categorize samples was classified correctly divided by the total samples classified as positive samples	P = TP / (TP+ FP)
Recall (R)	the number of samples classified as positive divided by the total samples in the testing set in the positive category	R = TP / (TP+FN)
F-Measure	harmonic mean of precision and recall	F = 2*(P*R) / (P+R)

3 Result and Analysis

The types of herbal plants used are Binahong, Cincau Rambat, Jambu, Keji Beling, Lada Hitam, Sirih, and Som Jawa. An example of herbal plant leaf image data can be seen in Table 2.

Table 2. Image Data of Herbal Plant Leaves.

No	Example	Name	Number
1		Binahong	15
2		Cincau Rambat	15
3		Jambu	15
4		Keji Beling	15
5		Lada Hitam	15
6		Sirih	15
7		Som Jawa	15

This study uses a scenario of 80%:20% data sharing, where 80% of the data is used as train data and 20% is used as test data. The distribution of data for each type of plant can be seen in Table 3.

Table 3. Data Split.

Type of plant	Train Data	Test Data
Binahong	12	3
Cincau Rambat	12	3
Jambu	12	3
Keji Beling	12	3
Lada Hitam	12	3
Sirih	12	3
Som Jawa	12	3
Total	84	21

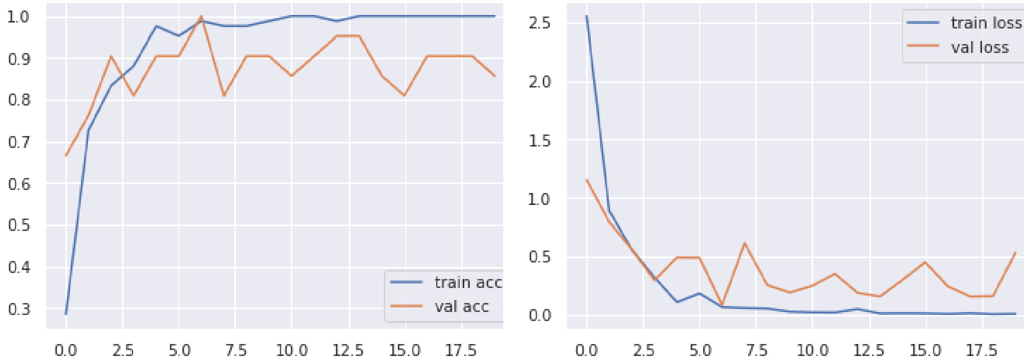
From the experiments and training that have been carried out on several architectures, the values of accuracy, precision, recall, and F-measure can be compared as shown in Table 4.

Table 4. Comparison of CNN Model Test Results.

Architecture	Input Shape	Precision	Recall	F-Measure	Accuration
AlexNet	227 x 227 x 3	0.61	0.71	0.59	0.67
LeNet5	32 x 32 x 1	0.62	0.74	0.66	0.71
VGG16	224 x 224 x 3	0.93	0.90	0.90	0.90
VGG19	224 x 224 x 3	0.91	0.86	0.86	0.86
Inception-V3	229 x 229 x 3	0.89	0.81	0.80	0.81
ResNet50	224 x 224 x 3	0.76	0.79	0.74	0.76

Based on these results, it can be seen that the AlexNet architecture has the lowest accuracy value compared to other architectures. The test results show that the best accuracy value with the VGG16 architecture is 90% with 93% precision, 90% recall, and 90% F-measure. The VGG16 architecture used epoch = 20, batch_size = 32, and validation_split = 0,2.

The graph of accuracy and loss obtained from the VGG16 architecture in the learning process can be seen in Figure 3.



(a) Graph of Learning Process Accuracy (b) Graph of Learning Process Loss

Fig. 3. Graph of Learning Process Accuracy and Loss for VGG16 Architecture

Figure 3 showed that there is a correlation or relationship between the accuracy value and the loss value in the data train with the number of epochs. The correlation that occurs in the accuracy value shows a positive correlation or has a unidirectional relationship. The larger the epoch used, the higher the accuracy value on the data train. Inversely proportional to the accuracy value, the correlation between the number of epochs and the loss value is a negative correlation. The larger the epoch used, the lower the loss value generated in the training data.

The results using the VGG16 architecture with an accuracy of 90% there are still classification errors. Classification errors occurred in Binahong as much as 1 and Lada Hitam as much as 1 as shown in Figure 4.

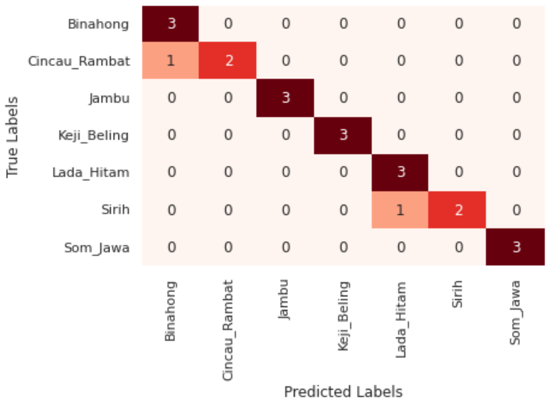


Fig. 4. Confusion Matrix for VGG16 Architecture Test Results

4 Conclusion

This study uses several architectural models of Convolutional Neural Network (CNN) to classify types of herbal plants. The CNN architecture used is LeNet-5, AlexNet, VGG16, VGG19, Inception-V3, and ResNet50. The AlexNet architecture has the lowest accuracy value compared to other architectures. The best accuracy value with the VGG16 architecture is 90% with 93% precision, 90% recall, and 90% F-measure. The VGG16 architecture used epoch = 20, batch_size = 32, and validation_split = 0.2. Classification errors occurred in Binahong as much as 1 and Black Pepper as much as 1. Therefore, the CNN Algorithm with the VGG16 architecture is able to classify types of herbal plants with good accuracy.

Acknowledgments

The authors would like to give deep gratitude to Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Negeri Medan for funding this research through the DIPA 2022.

References

[1] A. H. Krisnawan, F. F. Alkindi, D. Muttaqin, and E. S. Wahyudi, “Pemberdayaan Masyarakat dalam Pemanfaatan Tanaman Herbal Indonesia sebagai Minuman Fungsional Peningkat Imunitas Tubuh,” *CARADDE J. Pengabdi. Kpd. Masy.*, vol. 4, no. 1, pp. 163–172, 2021.

[2] K. Dhama *et al.*, “Medicinal and Therapeutic Potential of Herbs and Plant Metabolites / Extracts Countering Viral Pathogens – Current Knowledge and Future Prospects,” *Curr. Drug Metab.*, vol. 19, pp. 236–263, 2018.

[3] U. Ghani *et al.*, “Advancement and Future Directions towards Herbal Treatment for Various Diseases,” *Saudi J. Med. Pharm. Sci.*, vol. 05, no. 11, pp. 931–941, 2019.

[4] I. Süntar, “Importance of Ethnopharmacological Studies in Drug Discovery: Role of Medicinal Plants,” *Phytochem. Rev.*, vol. 19, no. 5, pp. 1199–1209, 2020.

[5] F. S. Li and J. K. Weng, “Demystifying traditional herbal medicine with modern approaches,” *Nat. Plants*, vol. 3, no. July, pp. 1–7, 2017.

[6] S. Sarno, “Pemanfaatan Tanaman Obat (Biofarmaka) Sebagai Produk Unggulan Masyarakat Desa Depok Banjarnegara,” *Abdimas Unwahas*, vol. 4, no. 2, pp. 73–78, 2019.

[7] D. Subositi and S. Wahyono, “Study of the Genus Curcuma in Indonesia used as Traditional Herbal Medicines,” *Biodiversitas*, vol. 20, no. 5, pp. 1356–1361, 2019.

[8] E. Budianita, T. Ulfadhyani, and F. Yanto, “Implementasi Algoritma Canny Dan Backpropagation Untuk Mengklasifikasi Jenis Tanaman Mangga,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, vol. 11, no. November, pp. 13–21, 2019.

[9] M. A. F. Azlah, L. S. Chua, F. R. Rahmad, F. I. Abdullah, and S. R. W. Alwi, “Review on Techniques for Plant Leaf Classification and Recognition,” *Computers*, vol. 8, no. 4, 2019.

[10] V. Pooja, R. Das, and V. Kanchana, “Identification of Plant Leaf Diseases using Image Processing Techniques,” *Proc. - 2017 IEEE Technol. Innov. ICT Agric. Rural Dev. TIAR 2017*, vol. 2018-Janua, no. Tiar, pp. 130–133, 2018.

[11] D. Argüeso *et al.*, “Few-Shot Learning Approach for Plant Disease Classification using Images Taken in the Field,” *Comput. Electron. Agric.*, vol. 175, no. January, 2020.

[12] S. S. A.-N. Belal A.M. Ashqar, Bassem S. Abu-Nasser, “Plant Seedlings Classification Using

Deep Learning," *Int. J. Acad. Inf. Syst. Res.*, vol. 46, no. 3, pp. 745–749, 2019.
[13] J. W. Tan, S. W. Chang, S. Abdul-Kareem, H. J. Yap, and K. T. Yong, "Deep Learning for Plant Species Classification Using Leaf Vein Morphometric," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 17, no. 1, pp. 82–90, 2020.
[14] M. Farid Naufal, "Perbandingan, Analisis Svm, Algoritma Untuk, dan CNN," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, pp. 311–318, 2021.
[15] S. Kana Saputra, I. Taufik, M. Hidayat, and D. F. Dharma, "Pneumonia identification based on lung texture analysis using modified k-nearest neighbour," *J. Phys. Conf. Ser.*, vol. 2193, no. 1, 2022.
[16] B. Yanto, L. Fimawahib, A. Supriyanto, B. H. Hayadi, and R. R. Pratama, "Klasifikasi Tekstur Kematangan Buah Jeruk Manis Berdasarkan Tingkat Kecerahan Warna dengan Metode Deep Learning Convolutional Neural Network," *INOVTEK Polbeng - Seri Inform.*, vol. 6, no. 2, p. 259, 2021.
[17] F. Sultana, A. Sufian, and P. Dutta, "Advancements in image classification using convolutional neural network," *Proc. - 2018 4th IEEE Int. Conf. Res. Comput. Intell. Commun. Networks, ICRCICN 2018*, pp. 122–129, 2018.
[18] I. Tabian, H. Fu, and Z. S. Khodaei, "A convolutional neural network for impact detection and characterization of complex composite structures," *Sensors (Switzerland)*, vol. 19, no. 22, pp. 1–25, 2019.
[19] M. E. Al Rivan and A. Setiawan, "Pengenalan Gestur Angka Pada Tangan Menggunakan Arsitektur AlexNet Dan LeNet Pada Metode Convolutional Neural Network," *Komputika J. Sist. Komput.*, vol. 11, no. 1, pp. 19–28, 2022.
[20] M. A. Pangestu and H. Bunyamin, "Analisis Performa dan Pengembangan Sistem Deteksi Ras Anjing pada Gambar dengan Menggunakan Pre-Trained CNN Model," *J. Tek. Inform. dan Sist. Inf.*, vol. 4, pp. 337–344, 2018.
[21] W. Setiawan, "Perbandingan Arsitektur Convolutional Neural Network Untuk Klasifikasi Fundus," *J. Simantec*, vol. 7, no. 2, pp. 48–53, 2020.
[22] F. F. Maulana and N. Rochmawati, "Klasifikasi Citra Buah Menggunakan Convolutional Neural Network," *J. Informatics Comput. Sci.*, vol. 1, no. 02, pp. 104–108, 2020.
[23] N. Bhargava, S. Dayma, A. Kumar, and P. Singh, "An approach for classification using simple CART algorithm in WEKA," *Proc. 2017 11th Int. Conf. Intell. Syst. Control. ISCO 2017*, pp. 212–216, 2017.
[24] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1578–1585, 2018.