Bulletin of Social Informatics Theory and Application
Vol. 6, No. 2, December 2022, pp. 158-167

Web-based system for medicinal plants identification using convolutional neural network

Franklin M.D. Mandagi [a,1], Frangky J. Paat [a,2], Dedie Tooy [a,3], Sandra E. Pakasi [a,4], Sofia Wantasen [a,5], Diane D. Pioh [a,6], Rinny Mamarimbing [a,7], Bobby J. V. Polii [a,8], Jantje Pongoh [a,9], Arthur G. Pinaria [a,10], Edwin Tenda [a,11], Luther Alexander Latumakulita [a,12,*], Noorul Islam [b,13]

[a] Sam Ratulangi University, Jalan Kampus, Manado 95115, Indonesia
[b] Kanpur Institute of Technology, A-1, UPSIDC, Rooma Industrial Area, Kanpur, 201008, India
[1] f.m.mandagi66@gmail.com; [2] frangkypaat@unsrat.ac.id; [3] dtooy@unsrat.ac.id; [4] sandrapakasi@unsrat.ac.id; [5] swantasen@unsrat.ac.id; [6] deibijdiane@unsrat.ac.id; [7] rinnyliwu@gmail.com; [8] bobbypolii@unsrat.ac.id; [9] jantjepongoh@unsrat.ac.id; [10] arthur.pinaria@unsrat.ac.id; [11] tenda.edwin@unsrat.ac.id; [12] latumakulitala@unsrat.ac.id; [13] noorul.islam3101@gmail.com
* corresponding author

ARTICLE INFO

Article history
Received October 30, 2022
Revised November 26, 2022
Accepted December 3, 2022

Keywords
medicinal plants
CNN
Identifications
Leaf images
Cross-validation

ABSTRACT

Indonesia has a variety of medicinal plants that are efficacious for preventing or treating various diseases. Each region has unique medicinal plants, such as in North Sulawesi, there are many medicinal plants with local names of "Jarak" (Jatropha curcas), "Jarak Merah" (Jatropha multifida), "Miana" (Coleus Scutellarioide), and "Sesewanua" (Clerodendron Squmatum Vahl). This research applies the Convolutional Neural Network (CNN) method to identify the types of medicinal plants of North Sulawesi based on leaf images. Data was collected directly by taking photos of medicinal plant leaves and then using the augmentation process to increase the data. The first stage is conducting training and validation processes using 10-fold cross-validation, resulting in 10 classification models. Evaluation results show that the lowest validation accuracy of 98.4% was obtained from fold-4, and the highest was 100% from fold-2. The third stage was to run the testing process using new data. The results showed that the worst model produced a test accuracy of 80.91% while the best model produced an accuracy of 87.73% which means that the identification model is quite good and stable in classifying types of medicinal plants based on its leaf images. The final stage is to develop a web-based system to deploy the best model so people can use it in real-time.

1. Introduction

Indonesia is known as a country rich in natural resources. About 30,000 plant species, and 7,000 of them have beneficial medicinal properties [1]. Medicinal plants are often used as an alternative to natural medicine, and around 80% of people still rely on traditional medicine for healing [2], [3]. However, due to the large number of medicinal plant species and their morphological similarity, the identification process can lead to errors that hurt the user and can even be fatal. The manual identification process takes a long time and requires assistance from experts [4]. Therefore, technology is needed to help identify the type of drug.

Over time, technologies such as Artificial Intelligence (AI) are increasingly popular in solving various problems. Several researchers have conducted studies on AI, such as that conducted by [5], who used the Backpropagation Neural Network (BPNN) method to identify local reef fish species in Bunaken National Park. They used three feature extraction processes, namely Geometric Invariant Moment (GIM), Gray Level Co-occurrence Matrix (GLCM), and Hue Saturation Value (HSV), with the lowest training accuracy of 75.00% and the highest of 88.73%, and the lowest validation accuracy

https:doi.org/10.31763/businta.v6i2.601

of 73.33%, and the highest 80.00%. This research was continued by research [6], which increased system performance with the lowest validation accuracy of 85% and the highest of 100%. Both studies used the K-Fold Cross Validation Technique to evaluate the model. Other studies have also used the BPNN method to predict and select prospective Bidikmisi scholarship recipients, considering the poverty level, and achieved an accuracy of 85.6% [7]. Meanwhile, researchers [8] identified 54 images of 3 types of ficus plants using the Support Vector Machine (SVM) and achieved an accuracy of 83.3%. Other researchers [9] used the SVM method to classify herbal leaves using the Scale Invariant Feature Transform (SIFT) technique as feature extraction. Research [10] classifies five different types of leaves using Law's Mask analysis and SVM as a classifier with an accuracy of 90.27%. [11] The Random Forest method was used with ten cross-validations to identify 32 images of 24 plant species on Mauritius Island, resulting in an accuracy of 90.1%.

Several studies use Convolutional Neural Networks (CNN) to identify types of medicinal plants, such as [12] analyzing the differences in leaf shape of 6 medicinal plants using an Android-based camera. The results show a training and validation accuracy of 100%. [13] authenticated herbal leaves using the CNN method on a Raspberry Pi for seven types of medicinal plants. The dataset is divided into 66.67% for training and 33.33% for data testing. The results show an accuracy of 93.62% for offline data testing and 91.04% for online data testing. In addition, other studies identify images of herbal plant leaves using CNN. [14] used 33 herbs with a dataset of 21,450 images, divided into 76.9% for training, 15.4% for validation, and 7.7% for testing. In the training and validation process, 150 epochs were carried out, with the highest accuracy of 94% and the lowest loss of 0.28%.

Based on the problems and research that has been done, this study aims to build a website-based system to classify four typical medicinal plants of North Sulawesi using the CNN method.

2. Method

2.1. Dataset

The data used in this study is in the form of leaf images of 4 types of medicinal plants: Jatropha curcas, Jatropha multifida, Coleus Scutellarioide, and Clerodendron Squmatum Vahl. The total original data is 300 photos and became 2320 through an augmented process. The distribution of training, validation, and testing data can be seen in Table 1, where Fig. 1 shows four types of medicinal plants as classification classes.

Table.1 Distribution of Data Training, Validation, and Testing

No	Medicinal plants	Data Group		
		Training	***Validation***	***Testing***
1	Jatropha curcas	423	47	110
2	Jatropha multifida	423	47	110
3	Coleus Scutellarioide	423	47	110
4	Clerodendron Squmatum Vahl	423	47	110

Jarak (*Jatropha curcas*) Jarak Merah (*Jatropha multifida*) Miana (*Coleus Scutellarioide*) Sesewanua (*Clerodendron Squmatum Vahl*)

Fig. 1. Four classification classes of medicinal plants

2.2. Research Stages

The research stages are shown in Fig. 2, which consists of 3 stages developing, testing, and deploying the model.

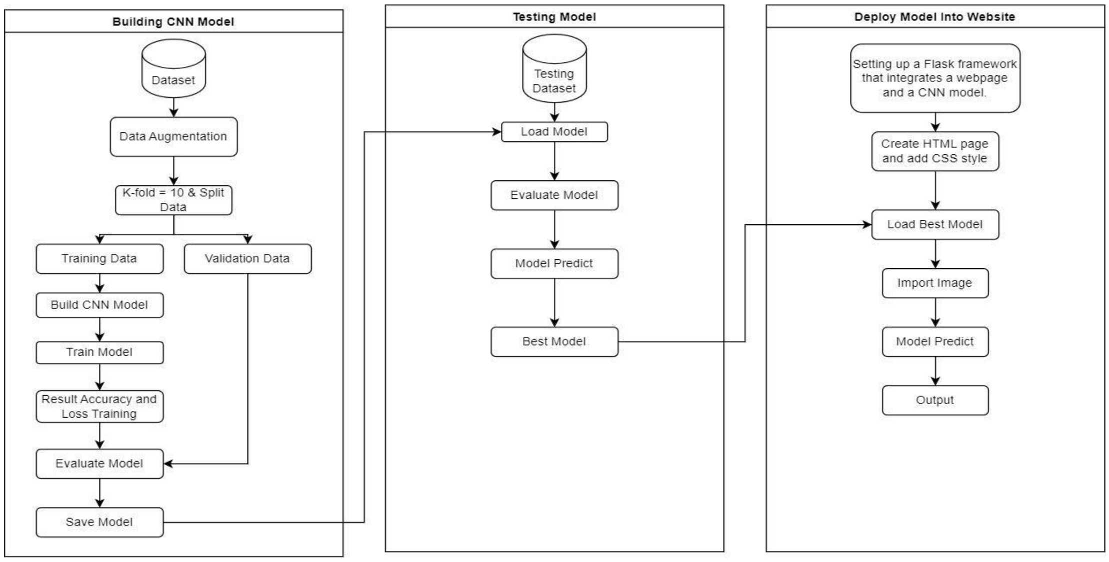

Fig. 2. Research Stages of Medicinal Plants Identification

In the first stage, the augmentation process was conducted to add more data and continued to training and validation processes. K-folds cross-validation technique with k was set up to 10, and then the evaluation process ran over ten built models. The testing process runs in the second stage by using new data never used in the previous stage. The best model will be selected to deploy in the third stage.

2.3. Cross Validation

Cross-validation is a standard evaluation technique in machine learning where data is divided into training and validation data according to the assigned K value [15]. Training and validation processes are repeated K times so that all subsets are used for evaluation. Thus, K-Fold Cross Validation produces K models tested on different data [16]. In this study using ten folds, this means that the data is divided 90% for training data and 10% for validation, as shown in Table 2.

Table.2 10-Fold Cross-Validation of Medicinal Plants Data

	Validation	**Training**	
Fold-1	Data 1-188	Data 189-1880	
Fold-2	1-188	189-376	377-1880
Fold-3	1-376	377-564	565-1880
Fold-4	1-564	565-752	753-1880
Fold-5	1-752	753-940	941-1880
Fold-6	1-940	941-1128	1129-1880
Fold-7	1-1129	1129-1316	1317-1880
Fold-8	1-1316	1317-1504	1505-1880
Fold-9	1-1504	1505-1692	1693-1880
Fold-10	1-1692	1693-1880	

2.4. Convulational Neural Network

Convolutional Neural Network (CNN) is an algorithm in Machine Learning that processes data in matrix forms, such as images or image data. CNN was developed from Artificial Neural Networks (ANN) with an architecture that processes spatial data. CNN is considered one of the best models for image object recognition problems. CNN consists of several layers that can be trained to extract the relevant features from the image.

Each layer in CNN receives input from the previous layer and produces output as a feature matrix. The processing at each layer consists of convolution, activation, and pooling operations. Convolution is used to extract features from the input data, activation is used to introduce non-linearity into the model, and pooling is used to reduce data dimensions [17], [18]. The architecture of CNN is shown in Fig. 3, where the input is medicinal plants leaf image and then processed at the feature extraction stage and ends with the classification stage.

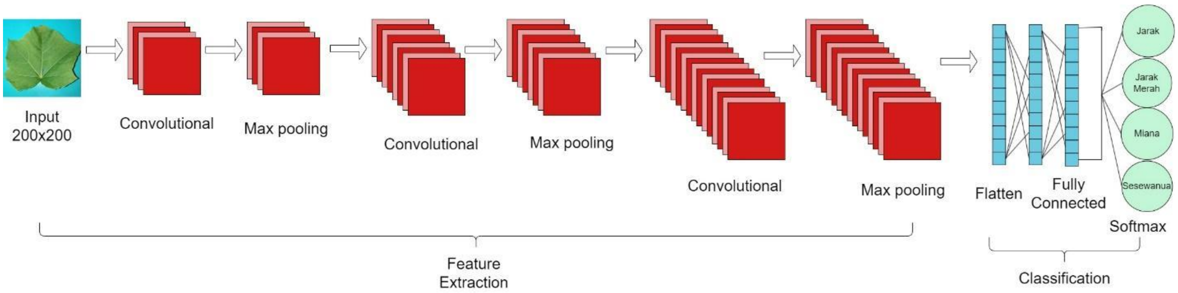

Fig. 3. Convolutional Neural Network Architecture

2.5. Confusion Matrix

Confusion Matrix is used to evaluate the performance of a model on classification problems [19], [20]. Table 3 shows the confusion matrix with four classification classes of Medicinal Plants Jatropha curcas, Jatropha multifida, Coleus Scutellarioide, and Clerodendron Squmatum Vahl.

Table.3 Confusion Matrix of Medicinal Plants With four classes

		Actual Values			
		Jatropha curcas	***Jatropha multifida***	***Coleus Scutellarioide***	***Clerodendron Squmatum Vahl***
Predicted Values	Jatropha curcas	T	F	F	F
	Jatropha multifida	F	T	F	F
	Coleus Scutellarioide	F	F	T	F
	Clerodendron Squmatum Vahl	F	F	F	T

Where T is True Prediction, and F is False Prediction. Accuracy can be calculated by using equation (1).

$$Accuracy = \left(\frac{Total\ of\ T}{(Total\ of\ T)+(Total\ of\ F)}\right) \quad (1)$$

3. Results and Discussion

3.1. Training and Validation Result

The training process is carried out with epoch = 5 parameters, batch size = 32, and learning rate = 0.001, 10 folds cross-validation is applied. Table 4 shows the loss and accuracy of validation result for ten folds, where Fig. 4 show the graphics of accuracy and loss for each fold.

Table.4 Loss and accuracy values of validation results

Fold	Loss	Accuracy
Fold-1	0.022688891738653183	0.9893617033958435
Fold-2	0.016021721065044403	1.0
Fold-3	0.04087042063474655	0.9946808218955994
Fold-4	0.07188611477613449	0.9840425252914429
Fold-5	0.0398603156208992	0.9840425252914429
Fold-6	0.03860936686396599	0.9840425252914429
Fold-7	0.02509928308427334	0.9840425252914429
Fold-8	0.04476083442568779	0.9840425252914429
Fold-9	0.025063488632440567	0.9893617033958435
Fold-10	0.04325980320572853	0.9840425252914429

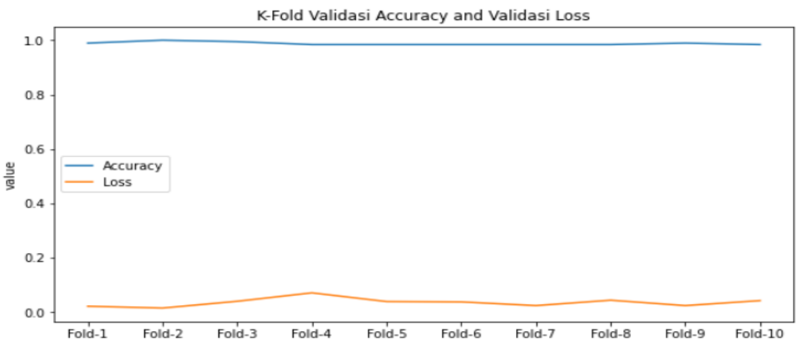

Fig. 4. Accuracy and Loss Per Fold Validation Graph

Fold-2 produces the most accurate model, whereas fold-4 yields the most miniature accurate model despite its sufficient accuracy at 98%. The best model is generated from fold-2. Fig. 5 and Fig. 6 compare the loss value and accuracy of the training and validation processes of the best model from fold-2. On the other hand, Fig. 7 and Fig. 8 present the same cases for the worst model from fold-4. This indicates that the loss values show the system performs well, and the accuracy is almost identical at the end of the epoch included for the worst model.

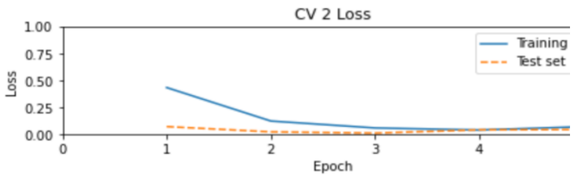

Fig. 5. Graphic of loss values of training and validation result of Fold-2

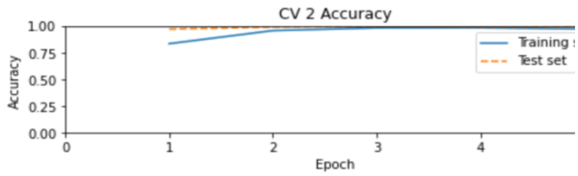

Fig. 6. Graphic of the accuracy of training and validation result of Fold-2

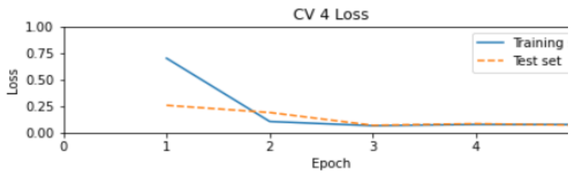

Fig. 7. Graphic of loss values of training and validation result of Fold-4

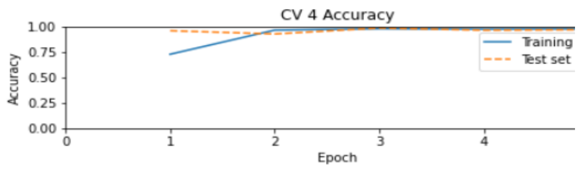

Fig. 8. Graphic of the accuracy of training and validation result of Fold-4

3.2. Testing Results

Based on the results of previous training and validation processes, the best model of fold-2 was selected. The testing process is conducted with new data to ensure that the model still provides excellent and stable performance. Fig. 9 shows the confusion matrix of the testing result of the best model fold-2, and Fig. 10 shows its classification reports.

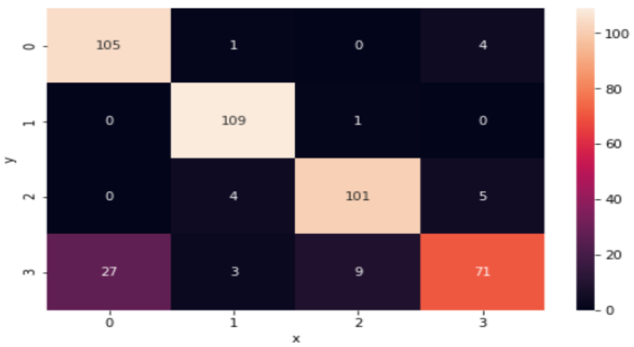

Fig. 9. Confusion Matrix of the best model of fold-2

```
              precision    recall  f1-score   support

       Jarak       0.80      0.95      0.87       110
 Jarak Merah       0.93      0.99      0.96       110
       Miana       0.91      0.92      0.91       110
   Sesewanua       0.89      0.65      0.75       110

    accuracy                           0.88       440
   macro avg       0.88      0.88      0.87       440
weighted avg       0.88      0.88      0.87       440

Accuracy on test data: 0.8773
Loss on test data: 0.3170
```

Fig. 10. Classification Report, the Best model of fold-2

As shown in Fig. 9 above that each class has 110 testing data. The result shows that 105 data from the "Jarak" class are correctly predicted, where 109 data from "Jarak Merah" is predicted as accurately, 101 from the "Miana" class are correctly detected, 71 data from the "Sesewanua" class are genuinely predicted. Table 5 shows detail of the testing results.

Table.5 Testing Result

Classification Classes	Prediction		Amount of testing data per class
	True	***False***	
"Jarak"	105	5	110
"Jarak Merah"	109	1	110
"Miana"	101	9	110
"Sesewanua"	71	39	110
	386		Total Testing Data = 440
	Accuracy = 368/440 * 100% = 87.73%		

Fig. 11 and Fig. 12 show the confusion matrix and classification report of the testing process of the worst model from fold-4. It can be seen that the worst model still gives a good accuracy of 80,91%, indicating that the proposed model for identifying the medicinal Plants discussed in this study is perfect and also stable. Based on all findings above, the model from fold-2 is selected to be deployed into a web-based system so it can be used online.

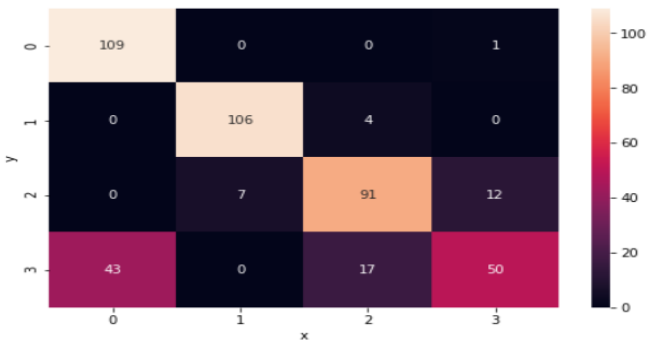

Fig. 11. The Confusion Matrix of the worst model

```
              precision    recall  f1-score   support

       Jarak       0.72      0.99      0.83       110
 Jarak Merah       0.94      0.96      0.95       110
       Miana       0.81      0.83      0.82       110
   Sesewanua       0.79      0.45      0.58       110

    accuracy                           0.81       440
   macro avg       0.82      0.81      0.80       440
weighted avg       0.82      0.81      0.80       440

Accuracy on test data: 0.8091
Loss on test data: 0.4901
```

Fig. 12. Classification Report of the Worst Model

3.3. Deployment

After several tests, the optimal value has been determined and will be integrated into the system. The system is designed as a web-based application that can detect medicinal plants. Fig. 13 illustrates the main menu header display of the application that contains various features, including the identification model for medicinal plants. The development of the application was based on the Python-Flask framework, a web application framework written in Python. This framework provides libraries, tools, and technologies that enable developers to build dynamic web applications. With the help of the Python-Flask framework, the application is designed to provide a user-friendly interface that can be easily accessed through a web browser. Overall, the system is expected to facilitate the identification of medicinal plants and support researchers in their study of herbal medicine.

Fig. 13. Main Menu of Web-Based Applications for Detect Medical Plants

In order to enable users to detect images of known medicinal plants, the website provides an image upload feature. This feature allows users to upload images of a medicinal plant they wish to identify. Fig. 14 depicts the form that enables users to browse an image of the medicinal plant they wish to detect.. The model uses deep learning algorithms to carry out the image recognition and detection process. This model has been trained on a large dataset of images of various medicinal plants, enabling it to accurately identify and classify the plant based on its unique features and characteristics. By providing this image upload feature, the website aims to provide a simple and effective way for users to identify medicinal plants and learn more about their properties and uses.

Fig. 14. Form Input image, which will detect

Fig. 15 shows the image the user has successfully uploaded and is ready to be detected. Upon clicking the "Detection" button, the website will start processing the image and display the detected medicinal plant with its name and other related information, such as its common name, scientific name, and medicinal properties, as in Fig. 16.

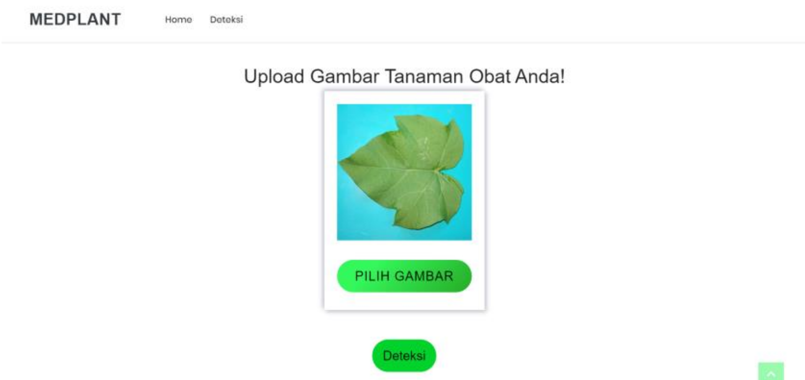

Fig. 15. Form to display the input image

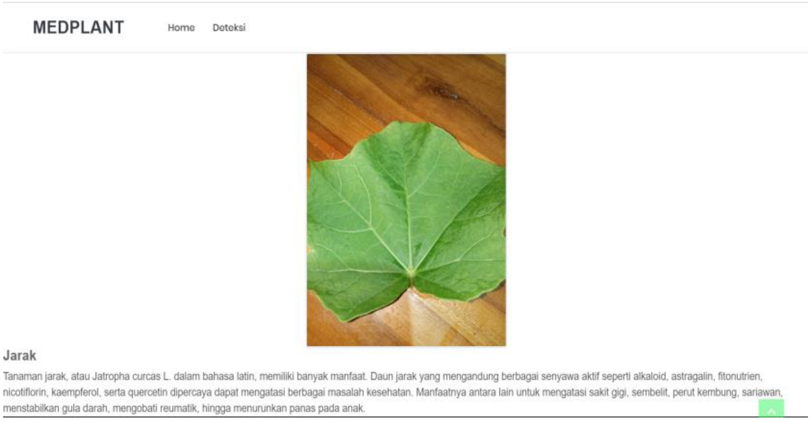

Fig. 16. Medicinal Plant Detection result

The web-based identification application created as part of this research presents a potential solution to the problem of assisting people in recognizing the native medicinal plants of the Sulawesi Utara Province and the advantages of these plants in preventing and treating illnesses. This program can provide an easy and readily available means for individuals to investigate the medicinal potential of local plants, which is particularly useful given the growing interest in alternative forms of treatment. This method can act as a basis for more research and development in the field of pharmacology based on traditional herbal medicine. Exploring the potential of natural treatments can bring new options for the development of drugs. This is especially important given the rising need for healthcare solutions that are both safe and effective.

In addition, the system may teach people how to combine herbal remedies from indigenous plants in North Sulawesi to prevent and treat ailments. This is one of the potential applications of the system. This method can potentially establish a new paradigm for health education based on local medicinal plants if it can capitalize on the knowledge of traditional medicine. The web-based identification application built throughout this research has a significant amount of potential to improve research in the field of pharmacology, advance the use of local medicinal plants, and give a new model for health education based on traditional medicine.

4. Conclusion

The identification model of medicinal plants using CNN proposed in this research provided excellent and stable performance. Ten identification models were created by using 10-fold cross-validation. Evaluation results show that even the worst model still gives a good validation accuracy of 80.91%, while the best model provides a perfect validation accuracy of 100%. The system also shows perfect accuracy in testing new data. The worst model provided a testing accuracy of 80,91%, the same as its validation accuracy, while the best model achieved an excellent testing accuracy of 87.73%. The website-based system has been successfully created to deploy the best model and runs properly.

Acknowledgment

We thank Kanpur Institute of Technology India and Sam Ratulangi University Manado Indonesia for this research collaboration.

References

[1] W. O. Jumiarni and O. Komalasari, “Inventory Of Medicines Plant As Utilized By Muna Tribe In Kota Wuna Settlement,” *Maj. Obat Tradis.*, vol. 22, no. 1, p. 45, Apr. 2017, doi: 10.22146/tradmedj.24314.

[2] X. Zhu, M. Zhu, and H. Ren, “Method of plant leaf recognition based on improved deep convolutional neural network,” *Cogn. Syst. Res.*, vol. 52, pp. 223–233, Dec. 2018, doi: 10.1016/j.cogsys.2018.06.008.

[3] H. A. Atabay, “A Convolutional Neural Network with a new architecture applied on leaf classification,” *IIOAB J.*, vol. 7, pp. 326–331, 2016, [Online]. Available at: https://www.iioab.org/articles/IIOABJ_7.S5_326-331.pdf.

[4] M. Musa, M. S. Arman, M. E. Hossain, A. H. Thusar, N. K. Nisat, and A. Islam, “Classification of

Immunity Booster Medicinal Plants Using CNN: A Deep Learning Approach," in *Communications in Computer and Information Science*, vol. 1440 CCIS, Springer Science and Business Media Deutschland GmbH, 2021, pp. 244–254, doi: 10.1007/978-3-030-81462-5_23.

[5] U. Andayani, A. Wijaya, R. F. Rahmat, B. Siregar, and M. F. Syahputra, "Fish Species Classification Using Probabilistic Neural Network," *J. Phys. Conf. Ser.*, vol. 1235, no. 1, p. 012094, Jun. 2019, doi: 10.1088/1742-6596/1235/1/012094.

[6] L. A. Latumakulita *et al.*, "Combination of Feature Extractions for Classification of Coral Reef Fish Types Using Backpropagation Neural Network," *JOIV Int. J. Informatics Vis.*, vol. 6, no. 3, p. 643, Sep. 2022, doi: 10.30630/joiv.6.3.1082.

[7] L. Latumakulita and T. Usagawa, "Indonesia Scholarship Selection Model Using a Combination of Back-Propagation Neural Network and Fuzzy Inference System Approaches," *Int. J. Intell. Eng. Syst.*, vol. 11, no. 3, pp. 79–90, Jun. 2018, doi: 10.22266/ijies2018.0630.09.

[8] S. J. Kho, S. Manickam, S. Malek, M. Mosleh, and S. K. Dhillon, "Automated plant identification using artificial neural network and support vector machine," *Front. Life Sci.*, vol. 10, no. 1, pp. 98–107, Jan. 2017, doi: 10.1080/21553769.2017.1412361.

[9] . P. P., "Robust Recognition And Classification Of Herbal Leaves," *Int. J. Res. Eng. Technol.*, vol. 05, no. 16, pp. 146–149, May 2016, doi: 10.15623/ijret.2016.0516031.

[10] D. Puri, A. Kumar, J. Virmani, and Kriti, "Classification of leaves of medicinal plants using laws' texture features," *Int. J. Inf. Technol.*, vol. 14, no. 2, pp. 931–942, Mar. 2022, doi: 10.1007/s41870-019-00353-3.

[11] A. Begue, V. Kowlessur, U. Singh, F. Mahomoodally, and S. Pudaruth, "Automatic Recognition of Medicinal Plants using Machine Learning Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 4, 2017, doi: 10.14569/IJACSA.2017.080424.

[12] L. P. D. S. Senevirathne, D. P. D. S. Pathirana, A. L. Silva, M. G. S. R. Dissanayaka, D. P. Nawinna, and D. Ganegoda, "Mobile-based Assistive Tool to Identify & Learn Medicinal Herbs," in *2020 2nd International Conference on Advancements in Computing (ICAC)*, Dec. 2020, pp. 97–102, doi: 10.1109/ICAC51239.2020.9357247.

[13] Haryono, K. Anam, and B. Sujanarko, "Herbal leaf authentication using convolutional neural network on Raspberry Pi 3," in *AIP Conference Proceedings*, Oct. 2020, vol. 2278, no. 1, p. 020047, doi: 10.1063/5.0014514.

[14] I. N. Purnama, "Herbal Plant Detection Based On Leaves Image Using Convolutional Neural Network With Mobile Net Architecture," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 6, no. 1, pp. 27–32, Jul. 2020, Accessed: Apr. 17, 2023. [Online]. Available at: https://ejournal.nusamandiri.ac.id/index.php/jitk/article/view/1400.

[15] Z. Xiong, Y. Cui, Z. Liu, Y. Zhao, M. Hu, and J. Hu, "Evaluating explorative prediction power of machine learning algorithms for materials discovery using k -fold forward cross-validation," *Comput. Mater. Sci.*, vol. 171, p. 109203, Jan. 2020, doi: 10.1016/j.commatsci.2019.109203.

[16] P. N. Andono, E. H. Rachmawanto, N. S. Herman, and K. Kondo, "Orchid types classification using supervised learning algorithm based on feature and color extraction," *Bull. Electr. Eng. Informatics*, vol. 10, no. 5, pp. 2530–2538, Oct. 2021, doi: 10.11591/eei.v10i5.3118.

[17] G. G. and A. P. J., "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Comput. Electr. Eng.*, vol. 76, pp. 323–338, Jun. 2019, doi: 10.1016/j.compeleceng.2019.04.011.

[18] V. Tiwari, R. C. Joshi, and M. K. Dutta, "Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images," *Ecol. Inform.*, vol. 63, p. 101289, Jul. 2021, doi: 10.1016/j.ecoinf.2021.101289.

[19] S. Ernawati, E. R. Yulia, Frieyadie, and Samudi, "Implementation of The Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, Aug. 2018, pp. 1–5, doi: 10.1109/CITSM.2018.8674286.

[20] A. F. Ab. Nasir *et al.*, "Text-based emotion prediction system using machine learning approach," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 769, no. 1, p. 012022, Feb. 2020, doi: 10.1088/1757-899X/769/1/012022.