Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi
Vol. 12, No. 2, Desember 2023
ISSN 0216 – 0544
e-ISSN 2301– 6914

OPTIMIZING LANTANA CLASSIFICATION: HIGH-ACCURACY MODEL UTILIZING FEATURE EXTRACTION

[a]Adri Gabriel Sooai, [b]Sisilia Daeng Bakka Mau, [c]Donatus Joseph Manehat, [d]Yovinia Carmeneja Hoar Siki, [e]Shine Crossifixio Sianturi, [f]Alicia Herlin Mondolang

[a a-d]Ilmu Komputer Universitas Katolik Widya Mandira Kupang
[e]Universitas Sanata Dharma Yogyakarta
[f]Universitas Negeri Malang
E-mail: [a]adrigabriel@unwira.ac.id, [b]sisilia901@gmail.com, [c]dmanehat@gmail.com, [d]yoviniacarmeneja@gmail.com, [e]185314091@student.usd.ac.id ,[f]aliciaherlinm@gmail.com

Abstract

As an invasive and poisonous plant, Lantana has become a pest in the agricultural world. Still, on the other hand, it becomes an ornamental plant with different positive potentials. Lantana flower datasets are not yet widely available for open image classification research, given that the research needs are still broad in remote sensing. This study aims to provide a model with classifier accuracy that outperforms similar studies and Lantana datasets for classification needs using several algorithms that can be run on small source computers. This study used five types of lantana colors, red, white, yellow, purple, and orange, as the primary dataset, which had 411 instances. VGG16 assisted feature extraction in preparing datasets for the data training using three classifiers: decision tree, AdaBoost, and k-NN. 2-fold cross-validation, 5-fold cross-validation, and a self-organizing map are used to help validate each process. The experiment to measure the classifier's performance resulted in a good figure of 99.8% accuracy for 2-fold cross-validation, 100% for 5-fold cross-validation, and a primary dataset of lantana interest that can be accessed freely on the IEEE Data port. This study outperformed other related studies in terms of classifier accuracy.

Key words: classification, feature extraction, image processing, lantana, machine learning.

INTRODUCTION

Lantana flowers are initially shrubs that can be considered disruptors of agricultural and livestock areas [1]. Some studies report that if it is edible by livestock, it will cause livestock to become sick and die because of the toxin content in the lantana flower in question [2]. However, over time, ornamental plant enthusiasts used it to become an artistic decoration of gardens. Having a variety of unique colors such as Red, Purple, Yellow, White, and Orange, this flower can be pretty attractive for garden decoration, even developed to raise butterflies [3]. Research involving lantana flowers has been carried out to monitor and prevent the expansion of lantana wild plants [4]. From small to large scale, remote sensing technology with a massive coverage area has been carried out [5], even making lantana profit for various fields in agriculture [6]. The process is to record the area of lantana wild plants and save them into an image format for further processing. Image processing can undoubtedly be done with the help of various image classifications, both traditional ones [7] [8] and assisted by deep learning. Each classifier has its performance that depends on the quality of the dataset used. Measuring the performance of a classifier is usually done by testing the dataset. The better the quality of the dataset used, the higher the quality of the model, in this case, the resulting prediction accuracy [9]–[12]. Various ways are done to get a dataset with good quality, including by doing the correct ground truth [13]–[15]. Improved dataset quality can also be enhanced using feature selection and outlier detection. This will cut features and observations that do not represent the data well. If the dataset has been validated and undergoes good preprocessing, the next step is to test through a performance comparison against the dataset. This performance comparison research will overview data sets considered and tested classifiers.

This study aims to make modeling using a primary dataset of lantana flowers with reasonable accuracy; the contribution offered is an effort to increase the accuracy of image classification using initial processing, namely feature extraction. This is done considering that other related research has not used that process, so it requires heavy classifiers such as deep learning. In contrast, in this study, the classification process uses light or traditional classifiers such as k-NN[16]. This study starts from the data acquisition process, extraction of VGG16-assisted features, and validation to testing performance comparisons between classifiers. The selection of VGG16 was based on more features because of the feature extraction process compared to some other extractors. By comparison, VGG16 produces 4096, four times more features than SqueezNet 1000 and InceptionV3 2048 features. The many features create an advantage, allowing for increased accuracy through feature selection and other preprocessing if needed. This is one of the reasons why this research is essential, namely how an experiment is equipped with the right choice of tools according to the characteristics of the data. The limitation of this study is that the dataset used is lantana flowers with flower color classes, namely Red, White, Yellow, Purple, and Orange. The total imagery used was 411, divided into five color categories. The classifiers used are Decision Tree[17], [18], AdaBoost, and k-NN, Validation using 2-fold cross-validation, and Self-Organizing Map visualization. This research was carried out on machines with limited resources, namely 8 Giga Byte RAM and 1.8 Gigahertz CPU, without GPU. This research contributes to the availability of lantana flower classification models and their datasets for free for various research reproducible needs. The dataset can be accessed using the lantana interest keyword on the IEEE Data port. Knowledge of the importance of preparing datasets that have good quality will significantly help the variety of classifying research in the future. The systematics of writing this research report begins with preliminary research followed by related research that discusses what the researcher has done before, which explains the research gap to be answered with solutions, followed by the method used, the results of the experiment, and closed with conclusions.

Invasive Plants Mappings

Wild plants that are considered pests can be used as ornamental plants, among others, are Lantana [19]. The introduction of a variety of wild plants with the potential for ornamental plants has been carried out previously by Omeer & Desmuk. Extensive data acquisition is used, assisted by remote sensing technology with advanced equipment such as the Field Spec 4 Hi-Res: High-Resolution Spectroradiometer ASD Scanner, and results in an introduction accuracy of 99.3% assisted by the Convolution Neural Network algorithm. However, the research that has been carried out does not provide an open dataset for other researchers to use to advance science.

Artificial Intelligence for Data Classification

Various studies on modeling a dataset using a classifier are scattered for multiple purposes. Measurement of classifier performance is carried out for plant and animal datasets. For example, the classification of betta fish uses a combination of neural networks with three color combinations in four types of betta fish. The study was conducted using Gabor filters with a combination of CMYK, HSV, and RBG colors to classify the four types of betta fish and assess the highest classification accuracy in the combination of neural networks with Gabor filters and RBG colors at an accuracy figure of 78.81% with a comparison of training and testing data of 90:10. The research that was carried out gave examples of relatively many combinations [20]. However, the classification accuracy is low using neural networks and a combination of Gabor filters and CMYK, RGB, and HSV color options. Low accuracy may be due to the dataset quality [21].

Subsequent research is still classifying living creatures carried out to group Starling Birds. It uses a 90:10 pattern for training and testing data on neural network classifiers and is assisted by manual segmentation to improve classifier accuracy. The affirmation of textures, shapes, and colors is also prepared in the feature extraction stage to enhance the classifier's performance [22]. However, this study only achieved an accuracy figure of 93%. The quality of the dataset likely used affects the performance of the classifier [23], [24].

Furthermore, classifier performance test research has been conducted on the primary dataset, specifically the flower dataset. Three cutting-edge classifiers from the deep learning family, namely ResNet18, RestNet50, and DenseNet121, are used for datasets with relatively small sizes. The results of the accuracy of the classifiers were 91.88%, 97.34%, and 99.82%. The study also used the help of joint supervision in the form of center loss and L2-Softmax loss [25]. They were related to image classification research conducted using the lantana flower dataset incorporated in the DeepWeed dataset collected from various locations in Northern Australia. Researchers use a deep learning classifier packaged in the YOLO framework. The study successfully classified lantana flowers (Hi and Wibowo, 2022). However, the resulting model only achieved an accuracy figure of 90.52%.

Regarding using various images of plants, flowers, and similar multiple classification research, another study focused on modeling herbal plants as substitute drugs in health. The researchers classified 25 types of herbal plants for human and animal health needs. The image used is the image of the leaves of herbal plants. A comparison of the classifier's accuracy performance was carried out between deep learning using original data and augmentation data; then, classification was also carried out using multi-layer perceptron [26]. However, the study only managed to achieve accuracy figures of 97.68%, 98.08%, and 82.51% for both classifiers and two types of datasets.

Efforts to improve the accuracy of the classifier are also carried out with the help of several initial stages using augmentation and segmentation. The dataset used is a variety of agricultural products such as fruits, flowers, and vegetables. The classifier used to conduct the research in question is a convolution neural network that utilizes the help of augmentation and a generative adversarial network. The study used secondary datasets: PlantDoc, Plants, Fruits-360, and PlantVillage [27]. However, after using various variations of preliminary processing and classifier combinations, the resulting model accuracy was only 99.57% for all datasets. The overall related research that has accuracy under this study is summarized and can be seen in Table 1.

Table 1. Related research in classifying image dataset

Author	Model	Accuracy
Hidayat	Neural Network with Gabor Filter	78.1%
Rahman	Neural Network	93%
Zhang	ResNet18, ResNet50, DensNet121	91.88%, 97.34%, and 99.82%.
Hi and Wibowo	Deep Learning YOLO Framework	90.52%
Kumar and Kumar	Deep Learning, Multi-Layer Perceptron	98.08%, 82.51%
Batchuluun	Convolution Neural Network	99.57%
This study	VGG16 + kNN	99.8%, 100%

MATERIAL AND METHODS

The entire experiment stage is arranged in five steps based on data acquisition to become the primary dataset. It explains how the data is obtained, followed by sequential stages of the entire modeling process, starting from data acquisition, feature extraction, and classifying using three classifiers, namely the Decision Tree, AdaBoost, and k-NN. Measure results by comparing classification results using the self-organizing map method to see where and what data are successfully identified and failed, namely those identified as the wrong class. While testing and evaluating, use confusion matrix tools to see predicted and actual data. In order the entire process picture can be seen in Fig 1.

Data Acquisition

The data acquisition process is carried out to obtain the primary dataset of lantana flowers. A total of Five types that have different colors are prepared for shooting. Each flower is shot with a pixel size of 1600 x 1200, a vertical and horizontal resolution of 72dpi, a bit depth 24, and RGB color mode. The model's camera is Xiaomi's M2101K7AG, with focus f/2.4, exposure 1/33sec, ISO speed 169, focal length 2mm, shooting does not use a flashlight, and illuminance when shooting is 77. The total imagery produced is 411, divided into five categories: Red, White, Purple, Orange, and Yellow.

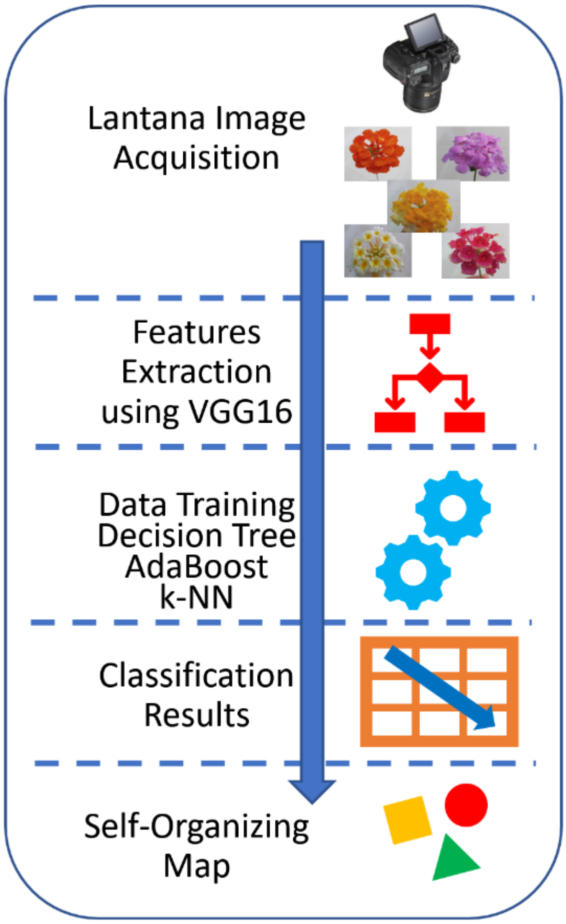

Fig 1. Experiment flow.

Feature Extraction

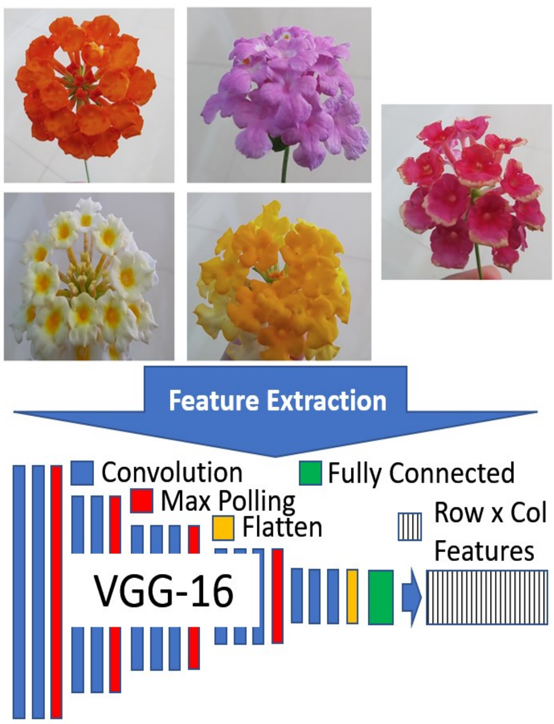

Fig 2. Illustration of feature extraction process using VGG16.

After the data acquisition process, the next stage is to perform feature extraction using the VGG16 algorithm. A dataset in the form of a numeric table is generated with a size of 411 instances and 4096 features. Briefly, the dataset is passed on the 41 layers of the VGG16 algorithm to be decomposed into numerical data to facilitate the data training. The reason for choosing VGG-16 to carry out the feature extraction process is that the features produced are relatively more than other algorithms, and this will facilitate the feature selection process if the accuracy results are low or if needed because feature selection will allow increasing the accuracy of the resulting model [28]. The illustration is seen in Figure 2.

The next classifier is decision tree 2-fold cross-validation. The way it works is to check whether the node representing the data class is pure or not by using the Gini Impurity Index; if the node is not mixed with other class data, then the value of impurity = 0; if the node is combined with the data of different classes, then the value of impurity = 0.5, as for the equation is as follows:

$$impurity = \sum_{i=1}^{n} p(i) \times (1 - p(i)) \qquad (1)$$

The last classifier used was Nearest Neighbor; the k-NN classifier used the Manhattan distance metric with the following equation:

$$d_1(p, q) = \|p - q\|_1 = \sum_{i=1}^{n} |p_i - q_i| \qquad (2)$$

From the equation (6), d is the distance, where p and q are points measured in distance.

Visualization using Self-Organizing Map

The classification process can be seen using visualization tools to determine the extent to which group data distribution is class-appropriate or intersects with other class data. This visualization can use the Self-organizing Map, which uses a coherent network formulation. The equation is as follows:

$$W_v(s+1) = W_v(s) + \theta(u, v, s).\alpha(s).(D(t) - W_v(s)) \qquad (3)$$

Evaluation

The evaluation was carried out by looking at the results of the classifiers from the three algorithms, namely kNN, Decision Tree, and AdaBoost. Each classifier works using 2-fold cross-validation, which also serves as proof of successful training and data testing. In 2-fold cross-validation, the data is divided into two parts: the first tests the second and vice versa. Each classifier will display the results of its prediction in the confusion matrix. Each confusion matrix will contain predictive and actual data to describe the performance of each classifier.

RESULT AND DISCUSSION

The results of image classification obtained in this study consist of several forms. The following sequence of processes gives their respective outputs, chronologically starting from the data source and ending with the accuracy value. Using smartphone cameras to acquire datasets of lantana flowers shows a practical and relevant approach to everyday use situations. The data acquisition results obtained through smartphone cameras provide several images with a size of 1600 x 1200 pixels, density of 72dpi, color depth of 24bit, and RGB color combinations. The total number of images acquired was 411 images. Classification Based on Color, classifying lantana flowers based on five color classes shows the model's sophistication in recognizing differences in color substances. In the data acquisition process, five colors of lantana flowers were obtained: Orange, Purple, Red, White, and Yellow. Using VGG-16, all 411 acquired images undergo a feature extraction process, breaking each image into rows and columns containing vector values and into 4098 features. In VGG-16, VGG-16 consists of a series of convolution layers and a fully connected (FC) layer. The architecture consists of several convolution blocks, each with several successive layers, followed by max-pooling. These blocks are designed to capture different features in the image of lantana flowers, from low features to high-level features. The feature extraction process for lantana flowers occurs through a convolution layer and a max-pooling layer in the VGG-16 architecture. Each layer of convolution works to identify specific features in the image of the lantana flower, such as lines, edges, or more complex objects. Max-pooling helps reduce the spatial dimensionality of lantana flower imagery from 1600 x 1200 to the smaller 224 x 224 while retaining the most significant feature information. After several convolution and max-pooling layers, the processed lantana flower image is flattened into a one-dimensional vector. This is done to convert the structure of the resulting matrix from convolution into a vector representation that can be fed to the next fully connected layer. After flattening the lantana flower image, the vector is arranged through several fully connected (FC) or dense layers. FC layers are responsible for connecting each element of the feature vector with all neurons in this layer. These neurons combine information from the features found in the previous step. In the context of feature extraction, this layer output produces a final vector representation of size 4098, representing the features extracted from the image. Each image is fed through the network and pulled into a feature vector that has dimensions of 4098. This extracted data set will have 411 rows (number of images processed) and 4098 columns (vector feature dimensions). Thus, after going through the above steps, VGG-16 will generate 4098 vector features for each treated image, resulting in 411 rows of data (images) in the feature representation matrix. Furthermore, all data transformation results from the image into a collection of numbers are arranged in a table with dimensions of 411 rows and 4098 columns in five color classes.

Division of datasets into 60% for training data and 40% for testing is standard practice. The results of this dataset division form two tables with a composition of 246 rows for training data and 154 rows for testing. After going through the dataset division, the following process is continued by training the data using the Decision Tree classifiers, AdaBoost and kNN. Each classifier produces 87.1%, 88.6%, and 99.8%, respectively. This result is obtained by setting 2-fold cross-validation. Furthermore, for the 5-fold cross-validation setting sequentially, the resulting accuracy figures are 91.2%, 90%, and 100%

The experiments that have been carried out provide three results from each classifier supported by comparisons between classifiers and visualizations that utilize self-organizing maps. The performance of each classifier is displayed in the form of a confusion matrix. Meanwhile, the comparison between classifiers is shown by comparing several values such as Area Under Curve, Classification Accuracy, F1, Precision, and Recall. Next, a data grouping distribution diagram is displayed in a self-organizing map format.

Decision Tree

Actual \ Predicted	Orange	Purple	Red	White	Yellow	Σ
Orange	75	0	6	1	7	89
Purple	3	86	3	0	7	99
Red	0	6	79	0	0	85
White	0	3	2	50	5	60
Yellow	6	0	3	1	68	78
Σ	84	95	93	52	87	411

Fig 3. Result from decision tree on confusion matrix.

The second experiment in the classification process was carried out using AdaBoost with the help of 2-fold cross-validation. The overall results of the classification can be seen in Figure 3, where the progress of the classifier's performance is displayed on the diagonal confusion matrix, while the right and left sides of the diagonal are failed predictions. As seen for the Orange Lantana class, it was predicted as much as 75 out of 89 instances or about 84.2%. Some instances fail to be expected and are spread across red class number 6, White number 1, and Yellow number 7. In the next class, purple lantana, it was successfully predicted as many as 86 instances out of 99, or about 86.6%. Then, for the Red Lantana class, as many as 79 of the total 85 instances, or about 92.9%, were successfully predicted appropriately. At the same time, the remaining six instances are incorrectly predicted as purple lantana. In the White Lantana class, out of a total of 60, only 50 instances were successfully expected precisely, or about 83%. The rest are scattered on the Purple, Red, and Yellow classes, with 3, 2, and 5 instances, respectively. The last class predicted by decision tree classifiers is Yellow Lantana. Sixty-eight of the total 78 instances were expected precisely, or about 87.1%. The rest are scattered in the Orange, Red, and White classes, with 6, 3, and 1 instance, respectively.

AdaBoost

Actual \ Predicted	Orange	Purple	Red	White	Yellow	Σ
Orange	84	1	1	0	3	89
Purple	2	91	4	0	2	99
Red	4	1	73	0	7	85
White	0	0	0	58	2	60
Yellow	8	5	4	3	58	78
Σ	98	98	82	61	72	411

Fig 4. Result from AdaBoost on confusion matrix.

The first experiment in the classification process was carried out using the Decision Tree with the help of 2-fold cross-validation. The overall results of the classification can be seen in Figure 4, where the progress of the classifier's performance is displayed on the diagonal confusion matrix, while the right and left sides of the diagonal are failed predictions. It was predicted that 84 out of 89 instances, or about 94.3%, were successfully expected for the Orange-colored Lantana class. At the same time, some instances fail to be anticipated and are spread out on the Purple, Red, and Yellow classes, with the numbers being 1, 1, and 3 instances, respectively. In the next class, purple lantana, it was predicted that there would be as many as 91 instances out of 99, or about 91.9%. The rest are scattered in the Orange, Red, and Yellow classes, with numbers 2, 4, and 2 instances, respectively.

The following result for the Red Lantana class is that as many as 73 of the total 85 instances were successfully predicted appropriately, or about 85.8%. At the same time, the remaining six instances are incorrectly predicted as purple lantana. In the White Lantana class, out of 60, as many as 58 instances, or about 96.6%, were successfully and appropriately expected. The rest is as many instances as falsely indicated as Yellow Lantana. The last class predicted by decision tree classifiers is Yellow Lantana. Fifty-eight of the total 78 instances were expected precisely, or about 74.3%. The remaining are scattered in the Orange, Purple, Red, and White classes, with 8, 5, 4, and 3 instances.

k-NN

Actual \ Predicted	Orange	Purple	Red	White	Yellow	Σ
Orange	89	0	0	0	0	89
Purple	0	99	0	0	0	99
Red	0	0	85	0	0	85
White	0	0	0	60	0	60
Yellow	1	0	0	0	77	78
Σ	90	99	85	60	77	411

Fig 5. Result from k-NN classifier on confusion matrix.

The third experiment conducted on the Lantana dataset using the k-NN classifier with 2-fold cross-validation can be seen in Figure 5. Unlike the previous two classifiers, there was a considerable jump in the resulting accuracy of the resulting classifier. The four classes of lantana flowers, namely Orange, Purple, Red, and White, were predicted precisely with a figure of 100%. Only one class, namely

Yellow, experienced a slight margin; only one instance failed to be expected and was declared a member of the Yellow lantana class. The overall accuracy for this third experiment is 99.7569%, so if the figure is rounded upwards to 99.8%.

Classifier Comparison

Table 2. Result in uusing 2-Fold cross validation.

Model	AUC	CA	F1	Pre cision	Re call
kNN	0.998	0.998	0.998	0.998	0.998
Tree	0.928	0.886	0.885	0.885	0.886
Ada Boost	0.924	0.871	0.872	0.876	0.871

The entire experiment can also be measured using several other magnitudes such as Area Under Curve (AUC), F1, Precision, and Recall. Seen in Table 2, a Comparison of the achievements of each classifier. As previously explained using visualizations in the confusion matrix, the decision tree classifier ranks at the bottom with an AUC figure of 92.4%, then CA and Recall of 87.1%, as opposed to F1 and Precision at 87.2% and 87.6%. The AdaBoost classifier is followed by second place, with AUC at 92.8%, then CA and Recall at 88.6% as far as F1 and Precision at 88.5%. Furthermore, for the k-NN classifier, all the values of AUC, CA, F1, Precision, and Recall obtained a figure of 98.8%.

Table 3. Result in using 5-Fold cross validation.

Model	AUC	CA	F1	Precision	Recall
kNN	1.000	1.000	1.000	1.000	1.000
Tree	0.942	0.900	0.900	0.901	0.900
AdaBoost	0.944	0.912	0.912	0.913	0.912

Based on the results obtained from the classification process using 2-fold cross-validation, it has exceeded several previous studies. However, another experiment was carried out by increasing the fold number to 5-fold cross-validation, as seen in Table 3. This resulted in a significant increase where each classifier gave higher accuracy figures of 91.2%, 90%, and 100%. In the case of lantana flowers, 411 image data is available. In 2-fold cross-validation, the lantana flower dataset is divided into two equal parts. The first part contains 205 data, and the second part includes 206 data. The model will be trained with one part and evaluated with the other, and vice versa. However, this approach may be less practical in describing variations in the dataset because there are only two divisional combinations. In a 5-fold cross-validation, the lantana flower dataset is divided into five equally large parts, each containing about 82 data points.

The model is trained with 4 out of 5 parts and evaluated with the remaining parts. This process will be repeated five times with different combinations of divisions each time. This approach will better show how well lantana flower models perform on different data variations. 5-fold cross-validation can improve classifier accuracy compared to 2-fold cross-validation because dividing the dataset into more folds assumes that each fold will represent a better variation in the original dataset, so the model will better cope with different data variations.

Self-Organizing Map

A means of seeing the distribution of data groupings that are successfully classified according to their class and some that are mixed can use the help of Self-Organizing Maps as seen in Figure 6. At first glance, three groups of data mixing between classes framed by hexagons containing 3 to 4 cells. The remaining three cells are stand-alone each and have a mixture of two classes of flowers.

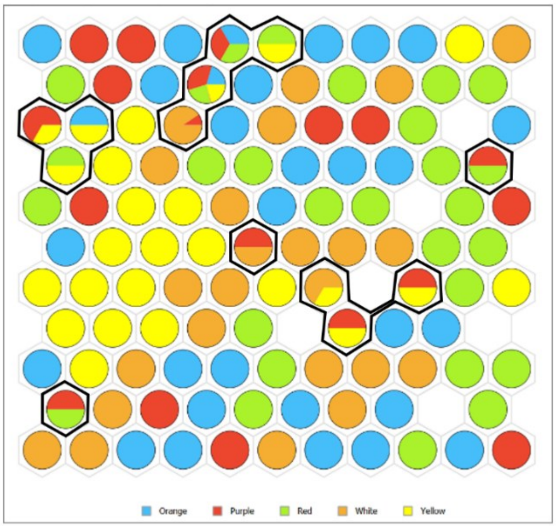

Fig 6. Self-Organizing map based on lantana dataset.

CONCLUSION

This research has done some experiments to measure the classifier's performance against the primary dataset of lantana flowers, which consists of five classes in Red, White, Yellow, Purple, and Orange. The classifiers that measured performance were decision trees, AdaBoost, and k-NN. Assisted by VGG16 for the feature extraction process and 2-fold cross-validation and 5-fold cross-validation, resulting in the highest classifier accuracy performance of 99.8% and 100% for k-NN. Supported by visualization, the self-organizing map displays which data overlaps in hexagon cells, thereby reducing prediction accuracy. Using feature extraction before the modeling process helps increase the prediction accuracy rate and produce models with better accuracy than previous studies. In addition to the accuracy figures obtained, the study provided a primary dataset that is open and available in the IEEE Data port for reuse to reproduce this study. The suggestion for further research is that it is still possible to improve accuracy even using low n-fold cross-validation by utilizing a feature selection process. This is possible considering that the relatively enormous number of image extraction features, namely 4096 features, can still be reduced to obtain features that have more impact in improving accuracy.

REFERENCES

[1] [1] T. Samarajeewa et al., “Identification of Lantana Camara Distribution Using Convolutional Neural Networks,” 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer), pp. 221–228, 2018, doi: 10.1109/ICTER.2018.8615559.

[2] M. Ntalo, K. E. Ravhuhali, B. Moyo, O. Hawu, and N. H. Msiza, “Lantana camara: Poisonous Species and a Potential Browse Species for Goats in Southern Africa—A Review,” Sustainability, vol. 14, no. 2, p. 751, Jan. 2022, doi: 10.3390/su14020751.

[3] S. S. Mukherjee, M. T. Ahmed, and A. Hossain, “Role of a Global Invasive Species (GIS), Lantana camara in conservation and sustenance of local butterfly community,” Acta Ecologica Sinica, Mar. 2021, doi: 10.1016/j.chnaes.2021.02.008.

[4] R. Gupta, D. Niyogi, R. Nayan, S. Singh, A. Mishra, and V. Varun, “Clinico-pathological study of Lantana camara toxicity in a sheep farm,” Journal of Pharmacognosy and Phytochemistry, vol. 8, no. 4, p. 3, 2019.

[5] A. Tarugara, B. W. Clegg, and R. Matuvhunye, “Mapping the spatial distribution and canopy cover of Lantana camara in the Zaka district of Zimbabwe,” Scientific African, vol. 17, p. e01339, Sep. 2022, doi: 10.1016/j.sciaf.2022.e01339.

[6] G. C. S. Negi et al., “Ecology and Use of Lantana camara in India,” Bot. Rev., vol. 85, no. 2, pp. 109–130, Jun. 2019, doi: 10.1007/s12229-019-09209-8.

[7] N. Z. Fanani et al., “Two Stages Outlier Removal as Pre-processing Digitizer Data on Fine Motor Skills (FMS) Classification Using Covariance Estimator and Isolation Forest,” IJIES, vol. 14, no. 4, pp. 571–582, Aug. 2021, doi: 10.22266/ ijies2021. 0831.50.

[8] N. Z. Fanani, A. G. Sooai, S. Sumpeno, and M. H. Purnomo, “Penentuan Kemampuan Motorik Halus Anak dari Proses Menulis Hanacaraka Menggunakan Random Forest,” JNTETI, vol. 9, no. 2, Art. no. 2, May 2020, doi: 10.22146/jnteti.v9i2.153.

[9] A. J. Lado et al., “Comparison of Neural Network and Random Forest Classifier Performance on Dragon Fruit Disease,” in 2021 International Electronics Symposium (IES), Sep. 2021, pp. 287–291. doi: 10.1109/IES53407.2021.9593992.

[10] A. G. Sooai, K. Yoshimoto, H. Takahashi, S. Sumpeno, and M. H. Purnomo, “Dynamic Hand Gesture Recognition on 3D Virtual Cultural Heritage Ancient Collection Objects Using k-Nearest Neighbor,” Engineering Letters, vol. 26, no. 3, pp. 356–363, 2018.

[11] A. G. Sooai, A. N. Rumaksari, K. Khamid, N. Z. Fanani, S. Sumpeno, and M. H. Purnomo, “Deteksi Gestur Lengan Dinamis pada Lingkungan Virtual Tiga Dimensi Koleksi Warisan Budaya,” Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI), vol. 7, no. 4, 2018, doi: 10.22146/jnteti.v7i4.457.

[12] A. G. Sooai et al., “Comparison of Recognition Accuracy on Dynamic Hand Gesture Using Feature Selection,” in 2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM), Surabaya: Institut Teknologi Sepuluh Nopember, Nov. 2018, pp. 270–274. doi: 10.1109 /CENIM. 2018.8711397.

[13] A. B. Gumelar, E. M. Yuniarno, D. P. Adi, A. G. Sooai, I. Sugiarto, and M. H. Purnomo, “BiLSTM - CNN Hyperparameter Optimization for Speech Emotion and Stress Recognition,” in Wireless Technologies and Intelligent Systems for Better Human Lives, Surabaya, Indonesia: IEEE, Sep. 2021, pp. 220–225. doi: 10.1109 /IES53407.2021. 9594024.

[14] A. B. Gumelar et al., “Human Voice Emotion Identification Using Prosodic and Spectral Feature Extraction Based on Deep Neural Networks,” in 2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH), Kyoto, Japan: IEEE, Aug. 2019, pp. 1–8. doi: 10.1109/SeGAH. 2019. 8882461.

[15] E. Talahaturuson et al., “Exploring Indonesian Netizen’s Emotional Behavior Through Investment Sentiment Analysis Using TextBlob-NLTK (Natural Language Toolkit),” in Technology 4.0 for Smart Ecosystem: A New Way of Doing DIgital Business, Semarang, Indonesia: IEEE, Sep. 2022.

[16] H. A. Roysid, A. Maulana, and U. Pujianto, “CAN K-NEAREST NEIGHBOR METHOD BE USED TO PREDICT SUCCESS IN INDONESIA STATE UNIVERSITY STUDENT SELECTION,” Jurnal Ilmiah Kursor, vol. 9, no. 4, Art. no. 4, Dec. 2018, doi: 10.28961 /kursor. v9i4.186.

[17] A. Z. Nur, H. Suyono, and M. Aswin, “APPLICATION OF HYBRID GA-PSO TO IMPROVE THE PERFORMANCE OF DECISION TREE C5.0,” Jurnal Ilmiah Kursor, vol. 10, no. 4, Art. no. 4, Dec. 2020, doi: 10.21107/kursor.v10i4.248.

[18] M. R. Siki, N. H. Harani, and C. Prianto, “Electronic Data Interchange (EDI) Applications Use the Decision Tree Method to Determine Vendor Recommendations,” Jurnal Ilmiah Kursor, vol. 10, no. 2, Art. no. 2, Dec. 2019, doi: 10.21107/kursor.v10i2.217.

[19] A. A. Omeer and R. R. Deshmukh, “Deep Learning-Based Models for Classification of Invasive Plant Species from Hyperspectral Remotely Sensed Data,” in Proceedings of the International Conference on Data Science, Machine Learning and Artificial Intelligence, in DSMLAI “21.” New York, NY, USA: Association for Computing Machinery, Aug. 2021, pp. 222–230. doi: 10.1145/3484824.3484884.

[20] S. Hidayat, A. Y. Rahman, and Istiadi, “Betta Fish Image Classification Using Artificial Neural Networks with Gabor Extraction Features,” in 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), Malang, Indonesia: IEEE, Jun. 2022, pp. 270–273. doi: 10.1109/CyberneticsCom55287.2022.9865509.

[21] O. Loyola-González, M. A. Medina-Pérez, and K.-K. R. Choo, “A Review of Supervised Classification based on Contrast Patterns: Applications, Trends, and Challenges,” J Grid Computing, vol. 18, no. 4, pp. 797–845, Dec. 2020, doi: 10.1007/s10723-020-09526-y.

[22] A. Y. Rahman, “Image Classification of Starlings Using Artificial Neural Network and Decision Tree,” in 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), Malang, Indonesia: IEEE, Jun. 2022, pp. 286–291. doi: 10.1109/CyberneticsCom55287.2022.9865465.

[23] M. Piernik and T. Morzy, “A study on using data clustering for feature extraction to improve the quality of classification,” Knowl Inf Syst, vol. 63, no. 7, pp. 1771–1805, Jul. 2021, doi: 10.1007/s10115-021-01572-6.

[24] M. Rzanny, P. Mäder, A. Deggelmann, M. Chen, and J. Wäldchen, “Flowers, leaves or both? How to obtain suitable images for automated plant identification,” Plant Methods, vol. 15,

no. 1, p. 77, Jul. 2019, doi: 10.1186/s13007-019-0462-4.

[25] R. Zhang et al., "Metric learning for image-based flower cultivars identification," Plant Methods, vol. 17, no. 1, p. 65, Jun. 2021, doi: 10.1186/s13007-021-00767-w.

[26] G. Kumar and V. Kumar, "Herbal Plants Leaf Image Classification Using Deep Learning Models Based On Augmentation Approach." Rochester, NY, Jul. 31, 2022. doi: 10.2139/ssrn.4292344.

[27] G. Batchuluun, S. H. Nam, and K. R. Park, "Deep Learning-Based Plant-Image Classification Using a Small Training Dataset," Mathematics, vol. 10, no. 17, Art. no. 17, Jan. 2022, doi: 10.3390/math10173091.

[28] A. G. Sooai, "Seleksi Fitur untuk Pengenalan Gestur Tangan 3D secara Dinamis pada Interaksi Benda Virtual Menggunakan Hybrid GMM," doctoral, Institut Teknologi Sepuluh Nopember, Surabaya, 2020. Accessed: Jul. 04, 2023. [Online]. Available: https://repository.its.ac.id/81786/

.