

CLASSIFICATION OF EDELWEISS FLOWERS USING DATA AUGMENTATION AND LINEAR DISCRIMINANT ANALYSIS METHODS

Fransiscus Rolanda Malau^{1*}, Dadang Iskandar Mulyana²

Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, DKI Jakarta, Indonesia^{1,2}

fransiscus.rolanda.malau@gmail.com

Received : 24 August 2022, Revised: 10 September 2022, Accepted : 10 September 2022

*Corresponding Author

ABSTRACT

Edelweiss is a plant that grows at a height, and is known as a perennial flower because it has beautiful petals and does not wilt easily. Although edelweiss in Indonesia is still in the same family as Leontopodium Alpinum, it turns out that the type of edelweiss found in the mountains of Indonesia is different from edelweiss found abroad. Therefore, in this study, an image processing system was developed that can classify the types of edelweiss flowers based on their image using Linear Discriminant Analysis to classify data into several classes based on the boundary line (straight line) obtained from linear equations. In this study, the types of edelweiss flowers used in this study were Anaphalis Javanica and Leontopodium Alpinum, the two types of edelweiss flowers were distinguished based on their color characteristics using hue and saturation values. The images used are 1500 images for training data and 450 test data images with a training and test data ratio of 70:30, so that the accuracy produced in the testing process is 99.77% in the Linear Discriminant Analysis method.

Keywords : Edelweiss Flowers; Data Augmentation; Linear Discriminant Analysis

1. Introduction

The edelweiss flower or perennial flower is actually a Leontopodium Alpinum flower found in the highlands of the Alps. In Indonesia, edelweiss flowers were first discovered in 1819 by a naturalist from Germany named Georg Carl Reinwardt on the slopes of Mount Gede. However, the name edelweiss comes from German, which consists of the words edel (noble) and white (white). Translated into Indonesian, it means "high white flower" (Kiswanto and Susanto, 2021; Martin & Susandi, 2022).

The splendor of this white flower is then used as a sign or sign of eternal love. Not only that, the beauty of this eternal flower can also be a magnet for climbers to use as selfies. Although edelweiss flowers in Indonesia are still in the same family as Leontopodium Alpinum, it turns out that the types of edelweiss found in the mountains of Indonesia are different from the types of edelweiss found abroad. Types of Edelweiss Flowers Anaphalis javanica is one type of edelweiss flower that is often found by Indonesian mountain climbers (Kiswanto and Susanto, 2021; Hanafi, et al., 2019).

The crown of Javanese edelweiss consists of hundreds of small, round white flower buds that are not pointed. In the center is a yellow "flower head". Edelweiss colors such as brown, blue and pink are the result of artificial dyes. In contrast to edelweiss anaphalis javanica, European edelweiss is an edelweiss flower that is easily found in the Alps. This plant is widespread in Alpine countries such as Austria, Germany, Italy, France, and Switzerland. This perennial flower has a different shape from the Javanese edelweiss. In one flower of Edelweiss Leontopodium Alpinum there are 500 to 1,000 flower buds with 2 to 10 "flower heads" surrounded by spiky velvety white leaves (Kiswanto and Susanto, 2021).

In this study, the color feature extraction used is feature extraction based on hue and saturation values. Color feature extraction with HSV is used to obtain various information from the color in the image to facilitate the identification process. HSV (Hue, Saturation, Value) is a type of perceptual color space. HSV has cylindrical coordinates, which consists of three color channels namely Hue, Saturation and Value. In addition, LDA will achieve optimal projections to be able to enter spaces with smaller dimensions and look for patterns that can be separated to group them from contour lines obtained from linear equations (Kiswanto and Susanto, 2021).

Classification is the process of finding a model that can divide the data by class, and is divided into two phases: a training phase (learning) and testing to understand how the categories of data are known to improve. The model exposure assessment phase is the result of the training phase using the new data as test data. The result of this phase is the level of accuracy/performance of the model when predicting unknown class data, especially test data (Antoko, et al., 2021; Hanafi et.al, 2019; Kartika Wisnudhanti, 2020).

Digital Image Processing Image is a two-dimensional matrix resulting from a continuous two-dimensional analog image into a discrete image through a sampling process. An image can be represented in the form of a two-dimensional matrix (with two variables x and y), where x and y are spatial coordinates and $f(x, y)$ is the image at those coordinates. The smallest signal unit of the matrix is called a pixel (Hashari et.al, 2018).

Machine learning is a computational algorithm or computer process that works based on historical data to improve performance in creating predictors. In machine learning, there are three learning methods, namely unsupervised learning, supervised learning, and reinforcement learning (Pradika, et al., 2020). In unsupervised learning, the training data used does not yet have a class, so the data are grouped based on the same characteristics. Supervised learning is a learning method for training data that already has classes. Furthermore, the reinforcement learning will look for the right steps in order to obtain the right predictions and in accordance with the existing conditions (Pratama et.al, 2018).

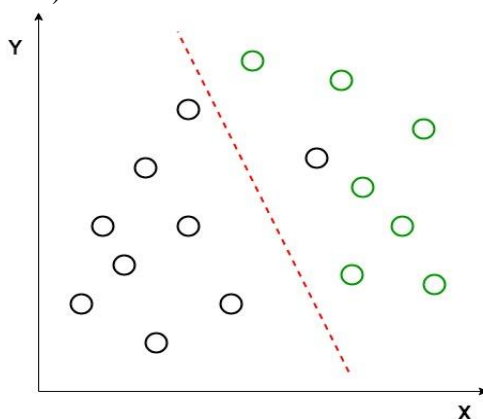


Fig 1, Linear Discriminant Analysis (Source: <https://media.geeksforgeeks.org>)

Linear Discriminant Analysis (LDA) was first applied to the facial recognition process by Etemad and Chellapa. Linear Discriminant Analysis works based on scatter matrix analysis which aims to find an optimal projection that can maximize the spread between classes and minimize the spread within the face data classes. The LDA algorithm has almost the same matrix calculation characteristics as PCA (Prasetyanto, et al., 2022; Putra, 2017). The basic difference is that in LDA, there is a minimum difference between the images in the class. The difference between classes is represented by the S_b matrix (scatter between class) and the difference within the class is represented by the S_w matrix (scatter within class). The covariance matrix is obtained from the two matrices. To maximize the distance between classes and minimize the distance within the class, a discriminant power is used (Ramdhani, 2015).

Data Augmentation is the process of enriching training data which aims to avoid the appearance of overfitting. The data augmentation process consists of several stages, namely horizontal flip, shear range, and zoom range. Shear range and zoom range itself has a value of 0.2. The horizontal flip stage works to increase the amount of training data by rotating the image or image horizontally by 90 degrees. The Shear range stage applies the shear transformation method, which is to add variations to the image by rotating the image to a certain degree, and the zoom range step is to enlarge the image to a certain scale from the original image (Naufal and Kusuma, 2021; Fadillah et al., 2021; Harahap & Muslim, 2018; Nana, et al., 2022).

Based on the background of the problem that has been described, this research will build a classification system for the types of edelweiss flowers. The system built can distinguish the types of edelweiss flowers with a digital image processing approach. Where the color feature extraction

will be carried out based on the hue and saturation values and then the extraction results will be classified using Linear Discriminant Analysis (LDA)(Nurhalimah, et al., 2020).

2. Research Methods

Research Stages

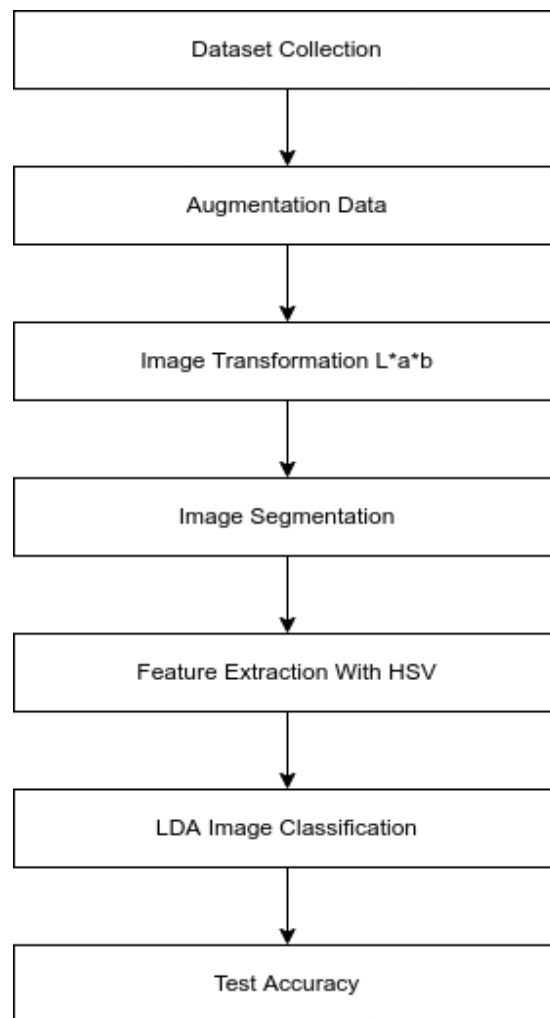


Fig 2. Research Stages

Systematic and structured research must be carried out throughout the research stages so that the research is right on target and in accordance with the research objectives. This research consists of several stages to conduct research so that it can be carried out properly which can be seen from Figure 2. At the initial stage is to collect datasets from two types of edelweiss flowers which will be used as training and test data. This phase is very important because dataset availability is an important factor for image processing performance. The quality and number of records affect the classification results, so preparation is needed at the time of collection. From the data set that has been collected, then the data augmentation process is carried out to significantly increase the diversity of the data set that has been obtained without losing the essence or essence of the data(Sanjaya & Ayub, 2020). So in this study the dataset used consisted of 1950 images, then the dataset was divided into 70% . 30% training and testing to determine the structure of the model. Therefore, the training data used are 1500 training data and 450 test data. After collecting the dataset and augmentation data, the next step is to perform an image transformation using the L*a*b color space, which is intended to digitally identify color content. The steps taken are to change and change the image color space from RGB to XYZ. In addition, the resulting RGB color value is used as a value to calculate the L*a*b* value. After the image transformation is done, the next step is to perform image segmentation which serves to separate one object from another. Separation is carried out based on regional boundaries that have the same shape or layout.

The output of this process is a binary image with a value of 1 (white) for the desired object and a value of 0 (black) for the background. Image segmentation in this paper uses thresholding technique. The process of transforming an image into binary format so that the feature extraction process can be easily carried out. The next step is to increase the information for feature extraction by using HSV color features based on hue and saturation values(Sentosa, et al., 2022).

Color feature extraction with HSV is used to obtain different information from colors in an image to facilitate the identification process. After HSV feature extraction, color feature information is obtained for the identification process in the LDA algorithm. LDA maintains the information index area, but includes more classes. Classes are separated with the aim that this condition increases the distance between classes and reduces the distance of information processing in the classroom. The number of features produced by LDA depends on the number of classes and the number of poses performed. After all the stages are carried out in the last stage, a test will be carried out to see how well the built model works. In this phase, the validity of the developed algorithm or model is tested(Sinulingga, et al., 2016).

LDA Methods

This edelweiss flower classification system uses the LDA method. The purpose of Linear Discriminant Analysis is to classify objects into one of two or more groups based on various features that describe the class or group(Husein & Harahap, 2017). The edelweiss flower classification process consists of a training process and a classification process. To carry out the training process, the covariance matrix in the SW class is first searched, and the covariance matrix between the SB classes is defined as follows:

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i) (x_k - \mu_i)^T$$

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)^T (\mu_i - \mu)$$

- Xk : image k,
- Ni : number of samples in class Xi,
- C : number of classes,
- μi : average image of class and average image of class-i.

Furthermore, the search for the eigenvectors of the multiplication matrix between SB and the inverse SW is carried out using Eq.

$$COV = S_B S_W^{-1}$$

The eigenvector is selected based on the largest eigenvalue. The eigenvector values are used to make projections for each training data using Eq. (Bimantoro, 2020).

$$FP_T = x_i W$$

Where :

- FPT : data projection value,
- Xi : input data
- W : the vector eigenvalues are selected based on the largest eigenvalues.

The results of the training process in the form of projections from each training data are then stored for comparison with test data. The classification process is carried out by projecting the test data (Bimantoro, 2020).

The test data projection is done by multiplying the test data with the eigenvector used in the training process. The classification stage is carried out by comparing the projected data from the training results with the projected results from the test data. Furthermore, to find out the class of the test data, the distance between the projected training data and the projected test data is carried out using Eq

Where :

d_{ij} : distance between projected training data and test data

x_{ik} : training data projection

x_{ij} : test data projection.

The results of the search for the distance between the projected training data and the test data are then sorted from the largest to the smallest. The result of the largest distance search is the classification result of the test data. These results are then verified by comparing the actual class with the class classification results. The software used in the analysis of edelweiss flower image data is Matlab software.

Data Collection Process

In this study, there were 5 images per each type of dataset obtained from image searches on google search which were disseminated from various sources.

Datasets Creation

Most of the image data in this study were downloaded from the internet and of course have different sizes. Therefore, then the pixels of each image used are changed to 300x300 pixels and then the background of each image is changed to red in order to facilitate the feature extraction process because the edelweiss flower itself is white.

From the image, the data augmentation process is then carried out to increase the diversity of data available for the training model, the process of augmenting data in this study uses a library of the python programming language(Solihin, et al., 2022).

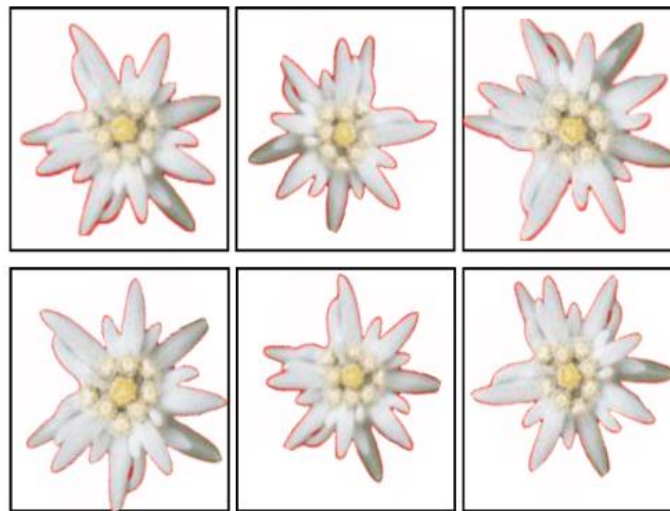


Fig 3. Process Data Augmentation

Augmentation techniques such as cropping, padding, and horizontal flipping are commonly used to train large neural networks. The sample in this study took 2 types of edelweiss flowers with a sample of 1,500 images for training data and 450 images for test data so that the data generated from each type became as follows:

Table 1 - Research Datasets

Class	Training Data	Test Data
<i>Anaphalis Javanica</i>	750	225
<i>Leontopodium Alpinum</i>	750	225
Total	1500	450

Resize the used image to fit the image for training or testing. Feature extraction is an object recognition technique that looks at certain features of an object that aims to perform calculations and comparisons to classify an image. Two separate data sets were used in this study, each containing training data and test data.

Test Design

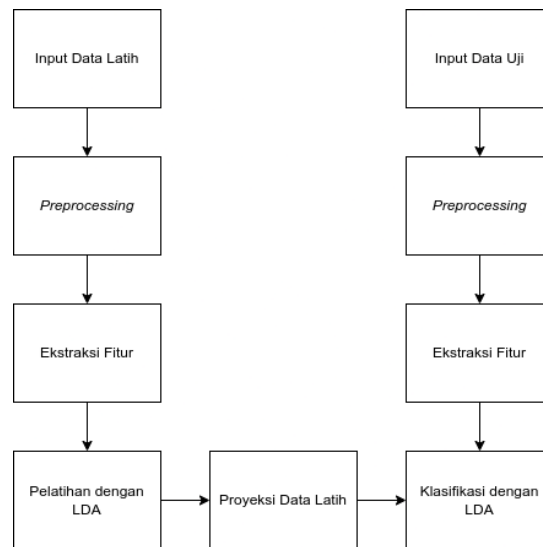


Fig 4 Test Design

Figure 4 shows the test of the design. The first stage is an image that has been cropped and resized outside the system and whose background has been manually changed to input for training. Next is the preprocessing stage. In this phase, the image is converted from the RGB color space to L^*a^*b and the image segmentation process is carried out using the thresholding method. These features are trained in LDA. The results of the research are data predictions, stored and used in the classification process.

3. Results and Discussions

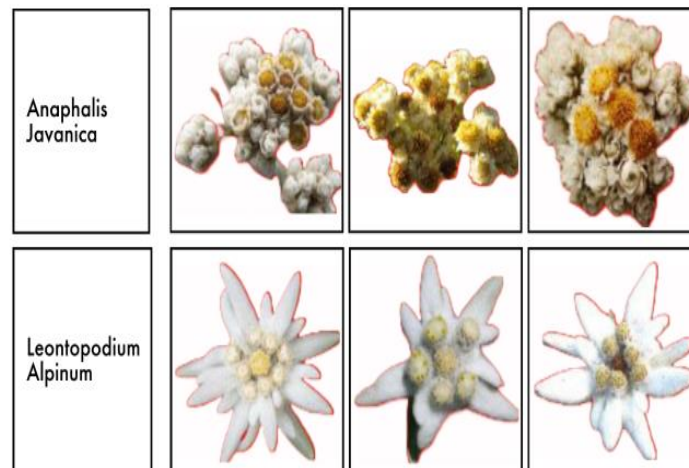


Fig 5. Examples of Flower Images in Each Class

The image data used in this study is data on 2 types of edelweiss flower images, totaling 1500 training data, with test data of 450 edelweiss flower images. From the results of the training carried out to get very high accuracy results, namely with 100% accuracy. After doing training on the image.

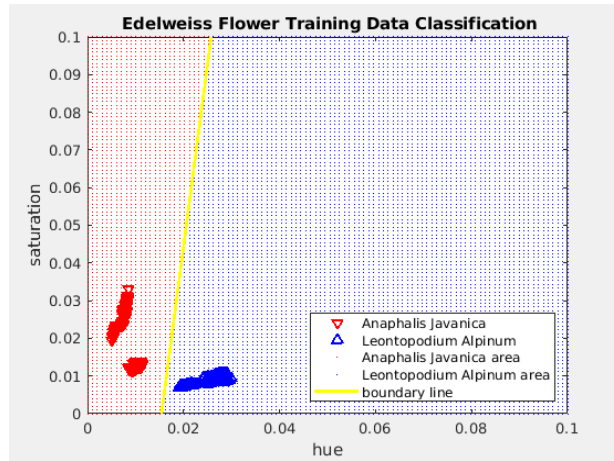


Fig 6. Training Data Distribution Graph

Based on the results of the distribution of data in Figure 6 which shows that there is a clear difference between edelweiss flowers of Anaphalis Javanica and Leontopodium Alpinum, this causes a high level of accuracy

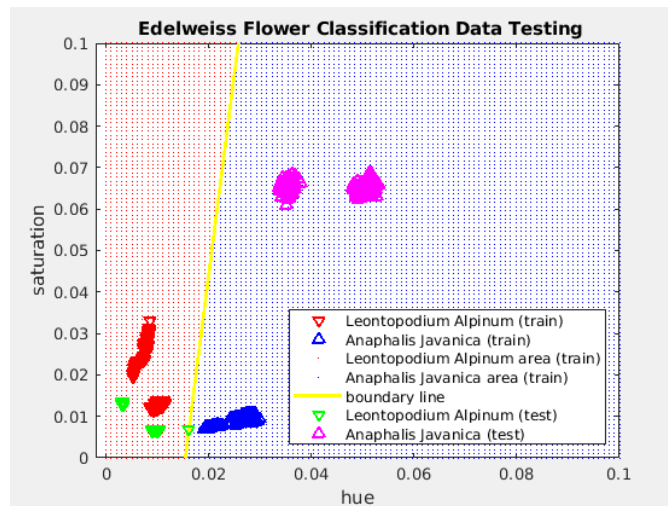


Fig 7. Test Data Distribution Graph

Also based on the results of the distribution of data in each class along with the resulting boundary line in Figure 7. Thus the magnitude of the resulting accuracy value indicates that of the 450 test data images carried out there is only 1 wrong image.

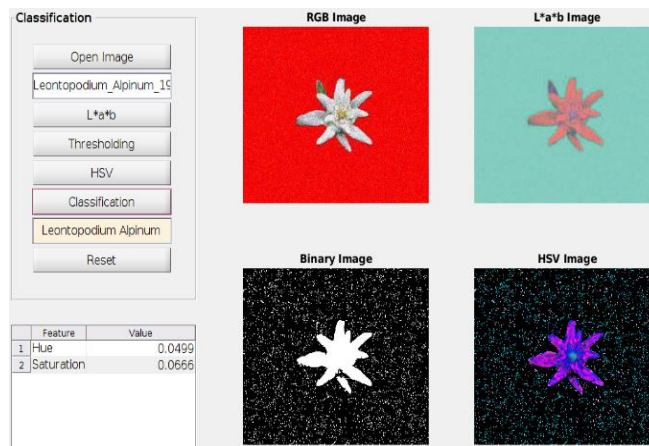
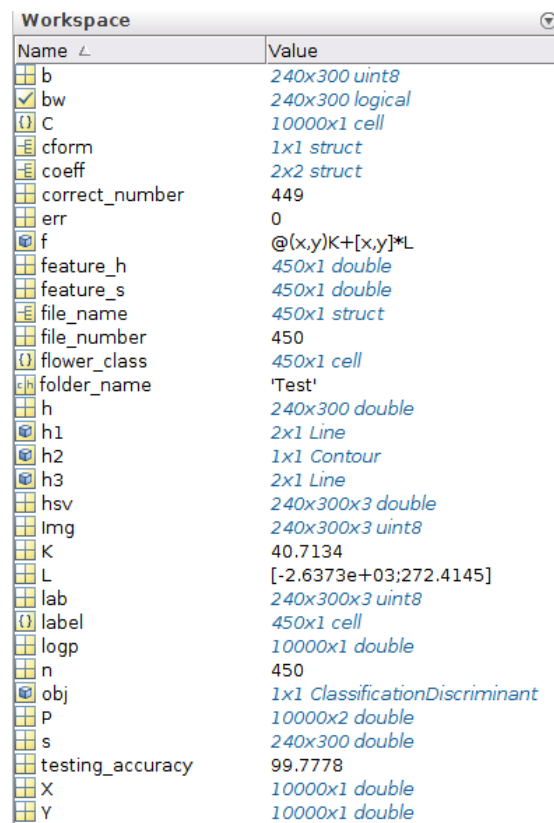


Fig 8. GUI Classification of Flower Types Leontopodium Alpinum

Similarly, the test result from the GUI program in Figure 8 show that the result of the image input process up to classifying with very good result.



Name	Value
b	240x300 uint8
bw	240x300 logical
C	10000x1 cell
cform	1x1 struct
coeff	2x2 struct
correct_number	449
err	0
f	@(x,y)K+[x,y]*L
feature_h	450x1 double
feature_s	450x1 double
file_name	450x1 struct
file_number	450
flower_class	450x1 cell
folder_name	'Test'
h	240x300 double
h1	2x1 Line
h2	1x1 Contour
h3	2x1 Line
hsv	240x300x3 double
Img	240x300x3 uint8
K	40.7134
L	[-2.6373e+03;272.4145]
lab	240x300x3 uint8
label	450x1 cell
logp	10000x1 double
n	450
obj	1x1 ClassificationDiscriminant
p	10000x2 double
s	240x300 double
testing_accuracy	99.7778
X	10000x1 double
Y	10000x1 double

Fig 9. Value Accuracy

From Figure 9 it can be seen that the testing accuracy value for 450 edelweiss flower images is 99.77%. These results can be categorized as very good. These results are influenced by several factors, including the classification can be optimal when feature extraction can provide the best information, and color-based feature extraction is very easy to recognize when the tested images have different colors. Based on the results of the author's test, the author also knows that there are several factors that cause misclassification, including: (1) The amount of training and test data is too small and must be added because the more models are trained, the more models are trained, the more models. resulting from. (2) If the picture is not clear, it will be difficult to classify the model, so that there are still errors in classification.

4. Conclusion

This study uses the Linear Discriminant Analysis (LDA) method to classify the types of edelweiss flower images based on their color characteristics. This study uses a programming platform that uses the Matlab matrix-based language. Based on the results of the tests that have been carried out, it can be concluded that the classification system for the type of edelweiss flower image using the Linear Discriminant Analysis method was successfully built with the resulting accuracy in determining the flower image of 99.77%. The use of color feature extraction with the HSV algorithm helps to obtain various information from the colors in the image to assist in the recognition process. LDA can then obtain an optimal projection that can enter a smaller dimension space by looking for patterns that can be separated so that they can be grouped according to the boundaries obtained from the linear equation so that it can be determined. LDA retains the information index area, but includes more classes. Classes are separated with the aim that this condition increases the distance between classes while reducing the distance for preparing information in the classroom. The number of features produced by LDA depends on the number of classes and the number of poses performed. There are two determining classes in this study, namely *Anaphalis Javanica* and *Leontopodium Alpinum*. Suggestions for further research can add some other classes or use more complex categories for further investigation. In addition, improvements are needed to maximize feature extraction and identification and increase the number of datasets used for both training and testing. In addition, to achieve better feature

extraction and identification, deep learning algorithms should be used to add feature extraction algorithms based not only on color but also on texture and shape.

References

- Antoko, T. D., Ridani, M. A., & Minarno, A. E. (2021). Klasifikasi Buah Zaitun Menggunakan Convolution Neural Network. *Komputika: Jurnal Sistem Komputer*, 10(2), 119-126.5.
- Fadillah, R. Z., Irawan, A., & Susanty, M. (2021). Data Augmentasi Untuk Mengatasi Keterbatasan Data Pada Model Penerjemah Bahasa Isyarat Indonesia (BISINDO). *Jurnal Informatika*, 8(2), 208-214.
- Hanafi, M. H., Fadillah, N., & Insan, A. (2019). Optimasi Algoritma K-Nearest Neighbor untuk Klasifikasi Tingkat Kematangan Buah Alpukat Berdasarkan Warna. *IT Journal Research and Development*, 4(1), 10-18.
- Harahap, R. N., & Muslim, K. (2018). Peningkatan Akurasi Pada Prediksi Kepribadian MBTI Pengguna Twitter Menggunakan Augmentasi Data. *Teknologi Informasi dan Ilmu Komputer*, 815-822.
- Hashari, I., Hidayat, B., & Arif, J. (2018). Identifikasi Fosil Gigi Geraham Manusia Berbasis Pengolahan Citra Digital Dengan Metode Gabor Wavelet Dan Klasifikasi Linier Dicurian Analysis (LDA). *eProceedings of Engineering*, 5(2).
- Husein, A. M., & Harahap, M. (2017). Penerapan Metode Distance Transform Pada Kernel Discriminant Analysis Untuk Pengenalan Pola Tulisan Tangan Angka Berbasis Principal Component Analysis: Penerapan Metode Distance Transform Pada Kernel Discriminant Analysis Untuk Pengenalan Pola Tulisan Tangan Angka Berbasis Principal Component Analysis. *Sinkron: jurnal dan penelitian teknik informatika*, 2(2), 31-36.
- Kartika Wisnudhanti, F. C. (2020). Metode Convolutional Neural Network Dalam Klasifikasi Citra Tiga Tokoh Wayang Pandawa. 7(2018), 1-5.
- Kiswanto, A., & Susanto, D. R. (2021). Strategi Pengembangan Desa Wonokriti Sebagai Desa Wisata Edelweis Di Kawasan Taman Nasional Bromo Tengger Semeru. *Journal of Tourism and Economic*, 4(2), 119-134.
- Martin, K., & Susandi, D. (2022). Perancangan dan Implementasi Sistem Irigasi Kabut Otomatis Tanaman Edelweis Menggunakan Mikrokontroler Arduino Uno. *ikraith-informatika*, 6(1), 57-66.
- Nana, N., Mulyana, D. I., Akbar, A., & Zikri, M. (2022). Optimasi Klasifikasi Buah Anggur Menggunakan Data Augmentasi dan Convolutional Neural Network. *Smart Comp: Jurnalnya Orang Pintar Komputer*, 11(2), 148-161.
- Naufal, M. F., & Kusuma, S. F. (2021). Pendeteksi Citra Masker Wajah Menggunakan CNN dan Transfer Learning. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 8(6), 1293-1300.
- Nurhalimah, N., Wijaya, I. G. P. S., & Bimantoro, F. (2020). Klasifikasi Kain Songket Lombok Berdasarkan Fitur GLCM dan Moment Invariant Dengan Teknik Pengklasifikasian Linear Discriminant Analysis (LDA). *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTika)*, 2(2), 173-183.
- Pratama, A. S. S., Wibawa, A. P., & Handayani, A. N. (2022). Convolutional Neural Network (Cnn) Untuk Menentukan Gagrak Wayang KULIT. *Mnemonic: Jurnal Teknik Informatika*, 5(2), 98-102.
- Pradika, S. I., Nugroho, B., & Puspaningrum, E. Y. (2020, November). Pengenalan Tulisan Tangan Huruf Hijaiyah Menggunakan Convolution Neural Network Dengan Augmentasi Data. In *Prosiding Seminar Nasional Informatika Bela Negara* (Vol. 1, pp. 129-136).
- Prasetyanto, A. E., Kusri, K., & Hartanto, A. D. (2022, February). Analisis Review Siswa Selama Pembelajaran pada Masa Pandemi Menggunakan Metode Topic Modelling LDA. In *STAINS (SEMINAR NASIONAL TEKNOLOGI & SAINS)* (Vol. 1, No. 1, pp. 241-246).
- Putra, I. M. K. B. (2017). *Analisis topik informasi publik media sosial di surabaya menggunakan pemodelan latent dirichlet allocation (LDA)* (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

- Ramdhani, Y. (2015). Komparasi Algoritma LDA Dan Naïve Bayes Dengan Optimasi Fitur Untuk Klasifikasi Citra Tunggal Pap Smear. *Jurnal Informatika*, 2(2)..
- Sanjaya, J., & Ayub, M. (2020). Augmentasi Data Pengenalan Citra Mobil Menggunakan Pendekatan Random Crop, Rotate, dan Mixup. *Jurnal Teknik Informatika dan Sistem Informasi*, 6(2).
- Sentosa, E., Mulyana, D. I., Cahyana, A. F., & Pramuditasari, N. G. (2022). Implementasi Image Classification pada Batik Motif Bali dengan Data Augmentation dan Convolutional Neural Network. *Jurnal Pendidikan Tambusai*, 6(1), 1451-1463.
- Sinulingga, S., Faticah, C., & Yuniarti, A. (2016). Pengenalan Wajah Menggunakan Two Dimensional Linear Discriminant Analysis Berbasis Optimasi Feature Fusion Strategy. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 3(1), 1-11.
- Solihin, A., Mulyana, D. I., & Yel, M. B. (2022). Klasifikasi Jenis Alat Musik Tradisional Papua menggunakan Metode Transfer Learning dan Data Augmentasi. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, 5(2), 36-44.