Plagiarism Checker X Originality Report

Similarity Found: 15%

Date: Tuesday, April 18, 2023
Statistics: 441 words Plagiarized / 2930 Total words
Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

1st Lekantara Annual Conference on Natural Science and Environment (LeNS 2021) IOP Conf. Series: Earth and Environmental Science 1097 (2022) 012001 IOP Publishing doi:10.1088/1755-1315/1097/1/012001 1 Machine Learning As Seed Image Identification Using Principal Component Analysis (PCA) Indra Laksmana 1*, Hendra 1, Jamaluddin 1, Trinovita Zuhara J 1, M Riza Nurtam1, Amrizal1, Rosda Syelly2 1Politeknik Pertanian Negeri Payakumbuh, West Sumatera, Indonesia 2Sekolah Tinggi Teknologi Payakumbuh, West Sumatera, Indonesia *indra.puskom@gmail.com Abstract. Plants can be identified using several variables, such as seeds' shapes, colors, and sizes.

However, several types of plants have close similarities to seed shapes. Therefore, additional characteristics are necessary to support the identification process. This study applied machine learning with the PCA method to identify plant species from seed shapes. The PCA simplifies the observed variables by reducing data dimensions and storing 75% of the informatio n.

The procedure did not eliminate too much important information while reducing data size and processing time. We collected 100 images of plant seeds similar to one another, such as sapodilla seeds, soursop, cucumber, star fruit, grape, melon, apple, lime, watermelon, and chili. A measurement system was designed using the K-Fold Cross Validation, and 10 tables of experimental results discovered a good level of accuracy of 83%.

The Omission error occurred in the seeds of soursop, starfruit, grape, apple, lime, and watermelon while the most commission errors occurred in apple seeds (8 times). K eywo rd s : Seed Image Identification; Machine Learning; Principal Component Analysis 1. Introduction Fruit plants are one of the plantation crops that are widely planted by farmers in Indonesia.

These plants are usually produced to meet self-sufficiency and local market needs [1]. In general, fruit plants are planted around the house or in pots as gard en decorations. [2]. Farmers commonly propagate fruit plants by planting seeds. Planting with seeds provides a strong root system, a simple technique, and a long period of fruiting; however, the nature of the offspring is not the same as that of the mother plant [3]. The shapes and structures of the seeds vary depending on the types of species and environmental conditions in which the plant lives [4].

Therefore, the seeds can be used as a marker to differentiate plant species [5]. However, the wide diversity of forms sometimes complicates the identification of a particular plant. In general, there are four ways to identify plants [6]. However, these ways complicate the manual classification because they require a l ong time and special understanding.

Therefore, we need a system that can simplify the data when identifying plants. One of the techniques to simplify the data as an image classifier is the principal component analysis (PCA). The PCA is used to reduce the dimensions of data without significantly reducing the characteristics of the data [7].

According to Santosa [8], the PCA is a reliable technique for extracting the structure of a data set with mass dimensions. The PCA projects images in to eigenspace planes by finding the eigenvectors of each image; then projects them into the obtained eigenspace. Research by [9] entitled "Plant classification based on leaf recognition" revealed that the PCA method could classify plants with 1st Lekantara Annual Conference on Natural Science and Environment (LeNS 2021) IOP Conf. Series: Earth and Environmental Science 1097 (2022) 012001 IOP Publishing doi:10.1088/1755-1315/1097/1/012001 2 an accuracy rate of more than 90 %. Moreover, [10] discovered medicinal plant leaf images with an accuracy of more than 90%.

[11] used the PCA to identify typical compounds from various Eucalyptus plant species and obtained that the compound groupings could clearly describe the specifications of each species.This study applied the Principal Component Analysis (PCA) to identify plant species based on seed images. The PCA simplifies images of seeds by reducing the dimensions and not losing a lot of crucial information of the images. 2.

Methodology The stages of this research were collecting or acquiring image data, converting the image data into vectors, sharing the data (training and testing), conducting the PCA, and analyzing the accuracy. These stages are shown in Figure1.

Acquiring Image Data Changing Images to Vectors Image Testing Saved as Matrix Data Testing Saved as Matrix Data Training PCA Transformation of 1 Dimension Matrix Transformation Recognizing and Characterizing Extraction Analyzing Accuracy PCA Process Yes No Figure 1.

Research Method Image Data This study employed 100 images, consisting of 10 types of plants, each of which used 10 images. Seed images were taken under the same lighting conditions and distances; the quality of the data and test results were not significantly affected [12] [13]. Ten types of plants in this research had morphologically similar seeds, such as sapodilla (Manilkara zapota), soursop (Annona muricata), cucumber (Cucumis sativus), star fruit (Averrhoa carambola), grape (Vitis vinifera), melon (Cucumis melo), apple (Malus domestica), lime (Citrus aurantifolia), watermelon (Manilkara zapota) and chili (Capsicum annum L.). All images were saved in JPG formats and RGB modes with image dimensions of 100x75 pixels.

Reading Image and Splitting Data From the RGB mode, the image data were converted to greyscales. The matrix of an image was saved as a 1x7500 row vector. This process was repeated for the entire image . Then, the image data were combined into matrix X and matrix O. Matrix X referred to a collection of row vectors for training images, while matrix O referred to a collection of row vectors for testing images.

The image-sharing applied the k-fold cross-validation with a value of k=10, also known as 10- fold cross-validation. The data were divided into two parts: a training set of 9 images and a testing set of 1 image. Therefore, there were 9 x 10 = 90 images for the training data. Meanwhile, the testing image consisted of 1x10 = 10 images. The first experiment used the first image as a testing image, and so forth.

Thus, the tenth experiment used the tenth image as a testing image. In one experiment, the process of reading the training image was repeated nine times for ten seed images to obtain a combined matrix measuring 90x7500. The testing image in an experiment was an exception or other than the data from the matrix X because one image from each class (seed) would serve as the training image or would not be processed when reading the matrix X.

Meanwhile, ten images would 1st Lekantara Annual Conference on Natural Science and Environment (LeNS 2021) IOP Conf. Series: Earth and Environmental Science 1097 (2022) 012001 IOP Publishing doi:10.1088/1755-1315/1097/1/012001 3 become row vectors. The next stage was dividing the matrix O; thus, the matrix O served as a composite matrix of testing images with a size of 10x7500.

The Process of Principal Component Analysis (PCA) The PCA procedure aims to simplify the observed variables by reducing their dimensions [14]. A digital image consists of the smallest elements usually called pixels. The pixels store information in the form of the color intensity of images at these coordinates. The images can be translated as a matrix while pixels can be translated as elements of the matrix [15].

Artificial intelligence uses many images as inputs which are processed in clustering and classification. The image is defined as a two-dimensional function f(x,y) where x and y are the spatial coordinates; and the amplitude f, at either x or y, is called the intensity or degree of gray. If x, y, and f are finite, the image is considered a digital image [16].

In this study, seed image was detected in several stages: reading the images, determining the covariance matrix, determining the feature vector and roots that correspond to the covariance matrix, selecting the largest feature root to contribute to the desired information, determining the matrix transformation using the characteristic vectors according to the root of the feature, determining the resulted matrix from the transformation of the seed image, determining the average vector of the transformation results for each class (plant species) , determining the Euclidean distance between the testing image and each class average vector, and recognizing the images based on the minimum Euclidean distance. When the minimum distance of the image was obtained from the same class average vector, the image was recognized.

After matrix X had been obtained, the covariance was calculated to obtain the sigma matrix. The calculation of this covariance used the equation of ______ÿ___

. To calculate the covariance, the previous unit type matrix data must have been changed to the double type; thus, the value range of the matrix elements changed from 0-255 to 0.0-1. The next step after obtaining the sigma covarian ce matrix was determining the feature vectors and roots of the covariance matrix.

The resulted output was a matrix v and d; each showed a feature vector and root. This step also determin ed the diagonal of the matrix d. The output of the diagonal matrix d was lambda, referring to the characteristic root value of the sigma covariance matrix. Furthermore, after obtaining the values of the feature roots and vectors, the features were extracted from the image.

Several characteristic roots were taken, and they were large enough to represent the desired information. This study took 75% of the initial information. 3. Results and Discussion The computer program developed to run the PCA algorithm and the 10-fold cross-validation revealed a matrix from the first to the tenth seed of the training and detection samples. The results are summarized in Table 1. Table 1.

Results of Image Recognition for 10 Experiments Experiment Original Images 1 2 3 4 5 6 7 8 9 10 Detected as 1 1 2 3 4 7 6 5 8 7 10 2 1 2 3 4 4 6 7 8 7 10 3 1 2 3 4 5 6 7 8 7 10 4 1 2 3 4 5 6 4 8 7 10 5 1 2 3 8 9 6 7 8 9 10 6 1 5 3 4 7 6 7 6 7 10 7 1 2 3 4 7 6 7 8 9 10 8 1 2 3 9 5 6 7 8 9 10 9 1 2 3 4 5 6 9 8 9 10 10 1 2 3 4 5 6 9 8 9 10 Table 1 shows that the images of seeds 1 (sapodilla), 3 (cucumber), 6 (melon), and 10 (chili) in each experiment did not have errors when detecting the image. However, image 5 (grape) was detected 1st Lekantara Annual Conference on Natural Science and Environment (LeNS 2021) IOP Conf.

Series: Earth and Environmental Science 1097 (2022) 012001 IOP Publishing doi:10.1088/1755-1315/1097/1/012001 4 as an image of soursop seeds in the sixth experiment and as an image of apple seeds in the first experiment. In the fifth experiment, the image of lime seeds was detected as an image of star fruit seeds.

The image of star fruit seeds was detected as an image of grape seeds in the second experiment and as an image of apples in the fourth experiment. The image of watermelon seeds was detected as an image of star fruit seeds and apples. Meanwhile, the image of apple seeds was detected as the images of grape seeds and watermelons.

The shape of the seed images and the patterns of image data taken from the same angle influenced this error detection. We iterated the experiment ten times to get information on how many times the system would accurately detect the plant seeds from their images. The sum of each element from the experimental results shows the accuracy of the image de tection system. Table 2 summarizes the sum of all experiments.

The main diagonal shows the number of correct image detections in 10 trials. Table 2. Result of the Sum of All Trials Label Detected as 1 2 3 4 5 6 7 8 9 10 1 10 0 0 0 0 0 0 0 0 0 2 0 9 0 0 1 0 0 0 0 0 3 0 0 10 0 0 0 0 0 0 0 4 0 0 0 8 0 0 0 1 1 0 5 0 0 0 0 6 0 3 0 1 0 6 0 0 0 0 0 10 0 0 0 0 7 0 0 0 1 1 0 6 0 2 0 8 0 0 0 0 0 1 0 9 0 0 9 0 0 0 0 0 0 5 0 5 0 10 0 0 0 0 0 0 0 0 0 10 Accuracy Analysis The accuracy of the seed image detection system can be calculated by counting the number of diagonals in the overall result table, dividing them by a lot of data, and multiplying them by 100%. The built system has an accuracy of 83%. Table 3.

Omission Errors Plants Seed Number Omission Errors Sapodilla (Manilkara zapota) 1 0 Soursop (Annona muricata) 2 1 Cucumber (Cucumis sativus) 3 0 Starfruit (Averrhoa carambola) 4 2 Grape (Vitis vinifera) 5 4 Melon (Cucumis melo) 6 0 Apple (Malus domestica) 7 4 Lime (Citrus aurantifolia) 8 1 Watermelon (Citrullus lanatus) 9 5 Chili (Capsicum annum L.) 10 0 Omission and Commission Error The omission constitutes the number of testing images that do not match the classification category.

Meanwhile, the commission is the number of detected testing images that disagree with the actual situation/classification in the class. The omission error can be calculated by adding up every row of the resulted matrix, except the elements in the main diagonal. The output of the omission error is shown in Table 3. Most omission errors occurred in watermelon seeds (see Table 3).

Watermelon seeds were detected as seeds of other plants in five of ten experiments. Meanwhile, the seeds of sapodilla, cucumber, melon, and chili did not occur omission. This phenomenon indicated that the system had successfully 1st Lekantara Annual Conference on Natural Science and Environment (LeNS 2021) IOP Conf. Series: Earth and Environmental Science 1097 (2022) 012001 IOP Publishing doi:10.1088/1755-1315/1097/1/012001 5 and correctly detected the seeds.

The commission errors can be calculated by adding up every column of the resulted matrix, except the elements on the main diagonal. The output of the commission errors is shown in Table 4. Table 4 shows that most commission errors of ten exp eriments occurred in apple seeds. This phenomenon denoted that the system had detected other seeds as themselves or detected errors from other plant seeds eight times.

Meanwhile, sapodilla, soursop, cucumber seeds, and chili seeds never received errors when detected from other seeds. Table 4. Commission Errors Plants Seed number

Commission Errors Sapodilla (Manilkara zapota) 1 0 Soursop (Annona muricata) 2 0 Cucumber (Cucumis sativus) 3 0 Starfruit (Averrhoa carambola) 4 1 Grape (Vitis vinifera) 5 2 Melon (Cucumis melo) 6 1 Apple (Malus domestica) 7 8 Lime (Citrus aurantifolia) 8 1 Watermelon (Citrullus lanatus) 9 4 Chili (Capsicum annum L.) 10 0 4. Conclusion The results of this study concluded that the PCA technique could be used to reduce dimensions.

The information taken was 75% of the 100 processed plant seed images. This study obtained good results because the images were successfully and correctly recognized with a percentage rate of 83%. Image retrieval techniques are urgently needed so that the images taken as data have similar lighting, distances, and dimensions.

Moreover, all available research evidence is crucially identified, assessed, and interpreted to answer specific research questions. References [1] N. L. M. Pradnyawathi, I. K. Sardiana, and N. N. Darmiati, “Pengembangan Bibit Buah Lokal Unggul Wani Bali Tanpa Biji,” Bul. Udayana Mengabdi , vol. 19, no. 1, Feb. 2020, doi: 10.24843/BUM.2020.v19.i01.p06. [2] M. R. Suhartanto and E. Gunawan, Untung besar dari bisnis bibit tanaman buah . Jakarta: AgroMedia Pustaka, 2012.

[3] A. Barus and Syukri, Agroteknologi Tanaman Buah-buahan. Medan: USU Press, 2008. [4] A. R. Muhammad, “Hubungan Bentuk Biji Dengan Karakteristik Morfologi Bibit Tanaman Pala (Myristica fragrans H.),” Andalas University, 2016. [5] I. Laksmana, Y. Herdiyeni, and E. A. M. Zuhud, “Genetic Programming for Medicinal Plant Family Identification System,” J. ICT Res. Appl., vol. 7, no. 3, pp. 217–234, Dec.

2013, doi: 10.5614/itbj.ict.res.appl.2013.7.3.4. [6] I. P. B. Laboratorium Ekologi Hutan, Fakultas Kehutanan, Diktat Dendrologi . Bogor: Laboratorium Ekologi Hutan, Fakultas Kehutanan, Insitut Pertanian Bogor, 2004. [7] L. I. Smith, A tutorial on Principal Components Analysis. John Wiley & Sons Inc, 2002. [8] S. B, Data Mining (Teori dan Aplikasi). Yogyakarta: Graha Ilmu, 2007. [9] A.

Ehsanirad, “Plant Classification Based on Leaf Recognition,” 2010. [10] I. Laksmana et al. , “Identification of Tropical Plants Leaves Image Base on Principal Component Analysis,” J. Appl. Agric. Sci. Technol. , vol. 4, no. 1, pp. 64–75, Feb. 2020, doi: 10.32530/jaast.v4i1.156. [11] P. J. Dunlop, C. M. Bignell, J. F. Jackson, and D. B.

Hibbert, “Chemometric analysis of gas chromatographic data of oils from Eucalyptus species,” Chemom. Intell. Lab. Syst., vol. 30, no. 1, pp. 59–67, Nov. 1995, doi: 10.1016/0169-7439(95)00036-4.

(2022) 012001 IOP Publishing doi:10.1088/1755-1315/1097/1/012001 6 [12] K. Delac, M. Grgic, and T.

Kos, “Sub-image homomorphic filtering technique for improving facial identification under difficult illumination conditions.” [13] V. V. Starovoitov, D. I. Samal, and D. V. Briliuk, “Image Enhancement for Face Recognition,” 2003. [14] H. Abdi and L. J. Williams, “Principal component analysis,” Wiley Interdiscip. Rev. Comput. Stat., vol. 2, no. 4, pp. 433–459, Jul. 2010, doi: 10.1002/wics.101. [15] A.

Rahmad, “Perbandingan Metode Ekst raksi Ciri FFT, PCA, dan FPE dalam Pengenalan Karakter Tulisan Tangan,” INSTITUT PERTANIAN BOGOR, 2012. [16] R. C. Gonzalez and R. E. Woods, Digital image processing. Upper Saddle River, N.J.: Prentice Hall, 2008.

INTERNET SOURCES:

1% -
https://researchportal.northumbria.ac.uk/files/65771863/_2022_IOP_Conf._Ser._Earth_Environ._Sci._966_011001.pdf
<1% -
https://www.semanticscholar.org/paper/Machine-Learning-As-Seed-Image-Identification-Using-Laksmana-Hendra/2bac3c3d8dc06d387d39038b53ac9a182bb9ea7b
<1% - https://leverageedu.com/blog/plants/
<1% -
https://www.section.io/engineering-education/how-to-implement-k-fold-cross-validation/
<1% - https://pubmed.ncbi.nlm.nih.gov/33477975/
<1% - https://greenupside.com/what-are-grafted-tomato-plants-plus-5-rootstocks/
<1% - https://brainly.com/question/13849781
<1% - https://ieeexplore.ieee.org/document/1558909
<1% -
https://stackoverflow.com/questions/22557883/scikit-learn-pca-dimension-reduction-data-lot-of-features-and-few-samples
<1% -
https://opus.lib.uts.edu.au/bitstream/10453/151753/2/Septarini_2021_IOP_Conf._Ser.__Earth_Environ._Sci._776_012004.pdf
<1% - https://www.nature.com/articles/nmeth.4346
<1% - https://www.nature.com/articles/s41598-021-01712-6
1% - http://repository.ppnp.ac.id/1086/
<1% -
https://towardsdatascience.com/k-fold-cross-validation-are-you-doing-it-right-e98cdf3

e6690
1% -
https://www.lekantara.com/conference/1supst-sup-lekantara-annual-conference-on-nat
ural-science-and-environment-lens
<1% - https://www.javatpoint.com/principal-component-analysis
<1% - https://theailearner.com/2018/10/06/what-is-a-digital-image/
1% - https://www.geeksforgeeks.org/digital-image-processing-basics/
<1% - http://gvpcew.ac.in/Material/ECE/DIP-4%20ECE-1%20&%202.pdf
<1% -
https://www.analyticsvidhya.com/blog/2020/02/4-types-of-distance-metrics-in-machine
-learning/
<1% - https://www.sciencedirect.com/topics/engineering/minimum-euclidean-distance
<1% -
https://math.libretexts.org/Bookshelves/Linear_Algebra/Understanding_Linear_Algebra_(
Austin)/04%3A_Eigenvalues_and_eigenvectors/4.03%3A_Diagonalization_similarity_and_
powers_of_a_matrix
<1% - https://faculty.uml.edu/darcus/47.375/materials/resultsmpl_02.htm
<1% - https://www.youtube.com/watch?v=xVqlH22CRy4
<1% - https://pubmed.ncbi.nlm.nih.gov/34063422/
<1% - https://pubmed.ncbi.nlm.nih.gov/27190684/
<1% -
https://www.geeksforgeeks.org/how-to-find-the-sum-of-all-elements-of-a-given-array-i
n-javascript/
<1% -
https://stackoverflow.com/questions/14952818/display-1-2-3-4-5-6-8-10-11-as-1-6-8-1
0-11
<1% - https://www.netspotapp.com/blog/ip-addresses/10-0-0-0-1.html
<1% -
https://stackoverflow.com/questions/66235113/runtime-error-when-counting-the-num
ber-of-diagonals-queen-can-attack-from-a-spec
<1% -
https://www.eajournals.org/wp-content/uploads/The-Influence-of-Oil-Palm-Plantation-t
oward-the-Insect-Ecology-on-Chilli-Plant-in-Desa-Bangun-Rejo-Labuhanbatu-Utara-Ind
onesia.pdf
<1% -
https://www.geeksforgeeks.org/program-to-find-the-sum-of-each-row-and-each-colum
n-of-a-matrix/
<1% -
https://scholar.its.ac.id/en/publications/acc1-and-acc2-genes-expression-microalgae-skl
etonema-costatum-on-

<1% - https://scholar.google.com/citations?user=cJwDa1MAAAAJ
<1% - https://www.tandfonline.com/doi/full/10.1080/26895293.2022.2037472
<1% -
https://simdos.unud.ac.id/uploads/file_penelitian_1_dir/7f69c520d3dee7151418f00e004a3576.pdf
<1% -
https://garuda.kemdikbud.go.id/author/view/2927337?jid=943&jname=Buletin%20Udayana%20Mengabdi
<1% - http://repository.unimus.ac.id/2565/8/DAFTAR%20PUSTAKA.pdf
<1% - http://scholar.unand.ac.id/3243/2/BAB%20I.pdf
<1% -
https://journals.itb.ac.id/index.php/jictra/citationstylelanguage/get/modern-language-association?submissionId=842&publicationId=588
<1% - http://journal.itb.ac.id/index.php?li=article_detail&id=1940
<1% - https://scholar.google.com/citations?user=w2T_IbkAAAAJ
<1% - https://en.wikipedia.org/wiki/Wiley_(publisher)
<1% - https://www.eurekaselect.com/185583/article
<1% -
https://kinfopolitani.com/files/journals/1/articles/156/submission/query/156-1-843-1-18-20200305.pdf
<1% - https://sinta.kemdikbud.go.id/authors/profile/5992329/?view=garuda
<1% - https://link.springer.com/chapter/10.1007/978-3-662-03331-9_6
<1% -
https://www.semanticscholar.org/paper/Disturbance-detection-and-isolation-by-dynamic-Ku-Storer/0189961bf20b2c9dc33ef6162e561a7ec432c6ce
<1% -
https://www.academia.edu/es/53653570/Sub_image_homomorphic_filtering_technique_for_improving_facial_identification_under_difficult_illumination_conditions
<1% - https://scholar.google.com/citations?user=mQ3xI48AAAAJ
<1% - https://personal.utdallas.edu/~herve/abdi-awPCA2010.pdf
<1% -
https://adoc.pub/perbandingan-metode-ekstraksi-ciri-histogram-dan-pca-dalam-m.html
<1% -
https://www.scirp.org/(S(25svyev5vqjboq45r4ni4f55))/reference/referencespapers.aspx?referenceid=667966