1570487555.pdf

by

Submission date: 21-Aug-2018 12:41PM (UTC+0700)

Submission ID: 991754898

File name: 1570487555.pdf (874.25K)

Word count: 3516

Character count: 19269

Comparison Analysis of the Artificial Neural Network Algorithm and K- Means Clustering in Gorontalo Herbal Plant Image Identification System

1st Mukhlisulfatih Latief Department of Informatics Engineering State University of Gorontalo Gorontalo, Indonesia mukhlis@ung.ac.id

2nd Novri Kandowangko Departement of Biology State University of Gorontalo Gorontalo, Indonesia novri1968@gmail.com

3rd Rampi Yusuf Department of Informatics Engineering State University of Gorontalo Gorontalo, Indones 4 rampirender 1@gmail.com

Abstract-The objective of this study was to analyze the comparison between artificial neural network algorithm and kmeans clustering to see the extent of the effectiveness of this algorithm on the identification of Gorontalo herbal plant image. This study uses a digital imaging processing method with segmentation and extraction techniques. Segmentation proses used thresholding method. The next process was extraction process of the characteristics of the image of the herbal plant using the shape and color characteristics to obtain the metric, eccentricity, hue, saturation, and value of the plant was carried out. These five parameters were used as parameters to identify the herbal plant image. This study used 91 images which consisted of 80 imagery training and 11 test images. The study revealed that k-means clustering accuracy was 27.27% whereas the artificial neural network algorithm accuracy was 54.54%.

Keywords— Artificial Neural Network, Image Processing

I. INTRODUCTION

Gorontalo is one of the regions with abundant herbal plants. To date, approximately 336 types of herbal plants have been identified [1]. People of Gorontalo often used these plants as traditional medicine as it is readily available, cheap, and can cure diseases.

Identification of herbal plants can be made by knowing the pattern of the leaves [2]. Identification of herbal plant is one of the common problems. This was due to the currently available method to identify the plant is by direct observation of the plant or asking the local traditional healers. This method was subjective. Thus, the identification results often inconsistent.

On the other hand, image processing has been widely applied to tropical fruit plants [3], identification of type and quality of tea [4], and the identification of herbal plants [5].

Digital image processing for identification of tropical fruit plants has been developed using the extraction method of the texture characteristics of Gray Level Co-Occurrence Matrix (GLCM). Types of fruit plants that have been successfully identified were starfruit, guava, jackfruit, mango, and sapodilla fruit. The identification algorithm uses the artificial neural network algorithm with back propagation which produces the highest accuracy by 90% [3].

In other research, identification of types and quality of the tea plant has been developed using thresholding segmentation method. Characteristic extraction uses the color characteristics extraction based on the value of red, green, blue, hue, saturation, and intensity. Identification algorithm used was Learning Vector Quantization (LVQ) artificial neural network. This study produces an accuracy of 62.7% during the training process and 42.31% accuracy on the testing process [4].

The next research [5] which designed the application to recognize the leaves pattern was developed by Indrawan. He designed the application to recognize the leaves pattern using the artificial neural network of Learning Vector Quantification (LVQ) to determine the herbal plant's type. The designed application was able to recognize the herbal plants accurately.

Further, implementation of the k-means algorithm can be found in the research of [6] on segmentation of lungs' image using the k-means which can visualize formulation of a more real color compared to the number of other clusters. In addition, research by [7] on Landsat digital image processing has produced the accuracy of 94.4%.

Based on the above description, this study was aimed at analyzing the comparison of the accuracy of artificial neural network and k-means clustering in the identification of Gorontalo herbal plants.

II. THEORETICAL REVIEW

A. Image Segmentation

Image segmentation is a process to obtain objects in an image or dividing the image into several objects or segments which have similar attributes [8].

Threshold is a basic technology that often used in image segmentation. This is one of the most straightforward methods to obtain clear image segmentation [8]. Threshold 121 duces a binary image, where the pixel of the object has the value of 1 and the background image has the value of 0. Therefore, the image will appear consistently clearer or darker than the background. The image segmentation stages are as follow [9]: Reading the original image

- 2. Converting the segment area for the image which was previously on RGB (Red, Green, Blue) into grayscale. The formula used to convert the RGB image into grayscale is Grayscale = 12989 * R + 0.5870 * G + 0.1140 * B
- 3. Converting the grayscale image into binary image through thresholding operation. The binary image from the

thresholding process is defined as:

$$g(x,y) = \begin{cases} 1 & \text{if } f(x,y) \ge T \\ 0 & \text{if } f(x,y) \end{cases} < 32$$

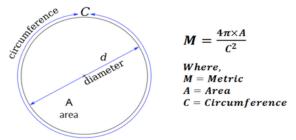
where g(x,y) is the binary image produced from thresholding, f(x, y) is the grayscale image, and T is the threshold value. 31

4. conducting the complement operation on the binary image produced from the thresholding, hence. Henceject is represented by I pixel1-pixel (white colored) and the background is represented by 0 pixel0-pixel (black colored).

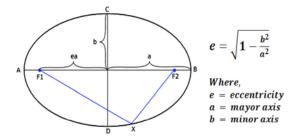
B. Characteristics Extraction

Following the separation of an object within the image with its background, the next process is the characteristics extraction. The extracted characteristics used to differentiate the one type of herbal plant from others. This characteristics extraction is conducted through analysis of shape and color. The analysis is carried out based on the calculation of the metric and eccentricity value. Whereas, the color analysis is carried out based on the hue, saturation, and value. These parameters are extracted from each image. The following formulas are used for this process [9]:

Metric Value



Eccentricity value



Hue, Saturation, and Value

$$r = \frac{R}{(R+G+B)}, g = \frac{G}{(R+G+B)}, b = \frac{B}{(R+G+B)}$$

$$V = \max(r, g, b)$$

$$S = \begin{cases} 0, & \text{jika } V = 0 \\ 1 - \frac{\min(r, g, b)}{V}, & V > 0 \end{cases}$$

$$H = \begin{cases} 0, & \text{jika } S = 0 \\ \frac{60*(g-b)}{S*V}, & \text{jika } V = r \end{cases}$$

$$60*\left[2 + \frac{b-r}{S*V}\right], & \text{jika } V = g$$

$$60*\left[4 + \frac{r-g}{S*V}\right], & \text{jika } V = b \end{cases}$$

$$H = H + 360 \text{ jika } H < 0$$

$$C. & \text{K-Means Clustering}$$

K-means is one of the methods in data mining to carry out unsupervised modeling by clustering the data into several clusters based on the similarity of characters. In brief, this kmeans algorithm is looking for the N object, based on the parameter into the k-cluster (k < N). This algorithm serves to cluster image data into several clusters based on the minimum distance [10]. K-means algorithm is widely known for its easiness and ability to quickly cluster the large data and outlier data. Based on partitioning clustering characteristics, each data should be grouped into a specific cluster and made it possible for each data in certain cluster during the specific process to move into another cluster [11].

In general, the K-Means algorithm is carried out with the following basic algorithm [12]:

- a. the data that would be clustered into k data cluster and the data values are randomized into cluster results which have similar data value numbers.
- b. each value of the data is calculated using the Euclidean distance for each cluster.
- c. if the value of the data is represented in the data cluster, let it be, and if the value of the data is unrepresented by that data cluster, then move the data into the cluster where the value of the data is represented.
- d. Repeat the step until all the data values are properly clustered.

D. Artificial Neural Network (ANN)

The artificial neural network is a large parallels distributed processor which tends to save knowledge which comes from perience and made it ready to be used. ANN is similar to the human brain in two ways: managing the knowledge obtained by the network through the learnin 7 process; the strength of relationship among neuron which known as synaptic values used to \$18e knowledge [13].

The artificial neural network is a computed mechanism which can acquire, represent, and calculate the mapping of a space with large variance of other information, and provided

with a 1st of data which represent that mapping. The objective of the artificial neural network is to develop a model of a data process; hence, a network can a cheralize and predict the output of previously unseen input. The learning process of the artificial neural network is carried out through adjustment of value between the neuron in responding toward the error between the actual output value and the output target value [14].

E. Learning Process of the Artificial Neural Network

The learning process of the attributed two learning systems: supervised learning and unsupervised learning. Supervised learning is a learning process which needs a teacher or something which have knowledge of the environment through the representation of input and output sample. The network parametral changes based on the training vector and the error signal between the ANN output and the expected output. This repeated process is carried out in order for the ANN to have knowledge similar to its teacher, or the reference input-output data. This explains that the ANN is trained to map a cluster of input-output samples with high accuracy. On the classification of a landslide area based on reference data previously obtained, the learning process is said to be supervised learning [13].

The supervised learning process has input data that will be trained by the network in the form of a pair of input and target. Thus, the desired value is obtained. The data pair will be treated as a teacher in the training process to train the network with the best form. In each training session, the network will produce output based on the input. The gap between output and the target output is the error that often happens. One of the networks that often used in supervised learning is the back propagation. [15].

F. Back propagation

This propagation training minimized the error between the targeted output and the actual output from the network. This algorithm is a multilayered neural network which consists of the input layer, hidden layer, and output layer. The process happened in this layer by inserting the input and multiplies it with the input and multiplies it with the input and multiplies it with the neuron system provides a model or reference that can be used to predict the value of the new target based on the input provided [14].

The more hidden layer used, the easier for the complex problem being inputted and the trained data target to be solved compared to the single screen. The output obtained from this training will have a lower level of error. However, the training process with may hidden layer will provide complex learning and often takes a long time with much iteration to obtain the expected result [15].

III. RESEARCH METHOD

The materials used in this study are 91 images consist of 80 image training and 11 test images. The image processing stage used in this stage is the image acquisition stage, processing, using the image segmentation, image extraction, characteristic extraction, and identification stage (training and testing), and analysis.

Preprocessing on the image uses thresholding segmentation and followed by the extraction of shape and color characteristics to obtain the metric, eccentricity, hue, saturation, and value.

The value obtained from the preprocessing then implemented on these two algorithms. Further, analysis based on the accuracy produced by these two algorithms.

The stages in this research can be described as follow:

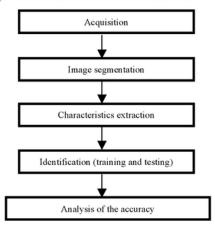


Figure 1. Research Stage

IV. RESULTS AND DISCUSSION

A. Image Segmentation Result

The result of this image segmentation through thresholding method in MATLAB programming the result is presented in Figure 2 below



Figure 2. Image Segmentation Process

(a) RGB Image; (b) grayscale Image;(c) thresholding result;

(d) complement result

B. The result of Characteristics Extraction

To obtain the metric and eccentricity value are extracted through shape characteristics extraction, whereas to obtain the hue, saturation, and value are carried through color characteristics extraction. The sample of several results of the extraction of characteristics of the herbal plant image is presented in Table I

TABLE I. SAMPLE OF THE CHARACTERISTICS EXTRACTION

No	Type of	Characteristics									
NO	Plant	Metric	eccentricity	Hue	saturation	value					
1	akar kucing	0.1754	0.9100	0.1502	0.2774	0.6826					
2	awar- awar	0.1673	0.5727	0.1435	0.1909	0.7900					
3	bambu kuning	0.1208	0.6800	0.1516	0.2997	0.7029					
4	bandotan	0.5688	0.8946	0.1176	0.2522	0.7725					
5	begonia	0.5942	0.8053	0.0820	0.2299	0.7835					

C. Chistering process based on k-means

All five of this extraction characteristics result is used as the input value in the k-means clustering algorithm. Out of 80 plant images available, the number of clusters used to identify the heal plants is 61. The clustering types of herbal plants using the k-means clustering algorithm are shown in Figure 3.



Figure 3. The interface of the herbal plant clustering using the k-means

In Figure 3, the sequence of different clusters in one type of herbal plant. Out of 80 images, only 39 images that were correctly identified, and the obtained accuracy value was 48.75%. This showed that the k-means clustering algorithm could not properly identify the plant image.

D. Testing process using the k-means

From the clustering above, the testing process was carried out using 11 testing images. The result of this test is presented in Table II below.

TABLE II. THE RESULT OF HERBAL PLANT IDENTIFICATION ON THE TESTING PROCESS USING THE K-MEANS CLUSTERING

No	Input cluster	Output cluster
1	Akar kucing	Tapa sambung nyawa
2	Awar-awar	Tapa lengkuas
3	Begonia	Begonia merah
4	Ceplukan	Temulawak
5	Jarak tintir	jarak tintir
6	Kananga	Labu kuning
7	Klayu garuga	Tutup merah
8	Labu kuning	Tapa meniran
9	Tayuman	Begonia
10	Tembelekan	tembelekan
11	Temulawak	temulawak

In table II above, it is clear that out of 11 tested images, there are only three images that are correctly identified. Hence the accuracy produced in the system in this testing process is

$$Accuract(\%) = \frac{number\ of\ correctly\ identified\ images}{Number\ of\ tested\ images} \times 100\%$$

$$=\frac{3}{11}\times 100\%$$

$$= 27.27\%$$

E. Training process with the artificial neural network

Artificial neural network algorithm is implemented to identify types of herbal plant based on five inputted values from the characteristics 23 raction. The identification process is conducted using the back propagation of artificial neural network algorithm with the architecture as shown in Figure 4 below:

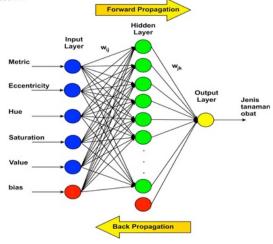


Figure 4. The architecture of back propagation artificial neural network for herbal plants

The artificial neural network used consists of three layers: an input layer, one hidden layer, and one output layer. The training process used Sigmoid Binary activation function on a hidden layer and Levenberg-Marquardt function process.

When the values from the result of the extracted image are obtained, the next process is the image training process. This training process serves to seek the best and most appropriate parameters and values for each layer that would be used in the testing process. The parameters trained are as follow

- Learning = 0.10
- Maximum iteration = 1000
- Error tolerance standard = 10-6
- Activation function = Sigmoid binary
- Number of trained data = 80

The artificial neural network training process is carried out by variating the number of the neuron on the hidden layer as shown in Figure 5 and several trial result of the training process is shown in Table III.

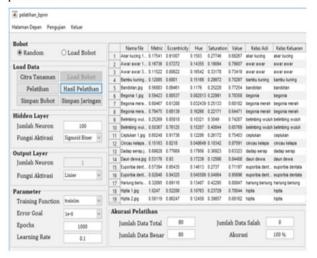


Figure 5. interface of the herbal plant training with the artificial neural network

TABLE III. THE RESULT OF THE NETWORK TRAINING USING PARAMETERS VARIATION

N	umber of neuro	Number of	accuracy		
Input later	Hidden Layer	Output Layer	epoch	(%)	
5	10	1	1000	14,28	
5	20	1	1000	42,85	
5	30	1	1000	48,35	
5	40	1	227	100	
5	50	1	219	100	
5	60	1	143	100	
5	70	1	327	100	

5	80	1	305	100
5	90	1	242	100
5	100	1	370	100

In Table III, it is clear the architecture that provides the highest accuracy and the least accuracy is the architecture with 60 neurons on the hidden layer. This architecture was then used to transfer the input value into the output value on system testing process. The output of this artificial neural network algorithm showed better clustering result.

F. Testing process using the artificial neural network

The testing is carried out using 11 testing images, and the result is shown in Table IV below:

TABLE IV. THE RESULT OF HERBAL PLANT IDENTIFICATION ON ARTIFICIAL NEURAL NETWORK TESTING SYSTEM

No	Input cluster	Output cluster
1	Akar kucing	Awar-awar
2	Awar-awar	Awar-awar
3	Begonia	Begonia
4	Ceplukan	Ceplukan
5	Jarak tintir	Jeruk nipis
6	Kananga	Kayu jawa
7	Klayu garuga	Klayu garuga
8	Labu kuning	Laping kubu
9	Tayuman	Tayuman
10	Tembelekan	Temulawak
11	Temulawak	Temulawak

In table IV above, it is clear that out of 11 images, there is six correctly identified image. Hence the accuracy of the system in this testing process is

$$Accuracy(\%) = \frac{number\ of\ correctly\ identified\ images}{number\ of\ tested\ images} \times 100$$

$$= \frac{6}{11} \times 100\%$$

= 54.54%

Based on the implementation of these two algorithms, artificial neural network algorithm showed better accuracy than the k-means clustering algorithm. The artificial neural network algorithm accuracy level was 54.54%, while the k-means accuracy was only 27.27%.

This high result by artificial neural network algorithm is due to this algorithm that works under supervision or through the training process. Hence, there is a learning process that should be done to recognize each type/cluster of the herbal plant. Whereas the k-means algorithm did not undergo this process (unsupervised).

Another thing to be highlighted in this research is that the accuracy for both of this algorithm was still very low (54.54%

and 27.27%). This was due to the number of images available for each plant is limited (mostly, one plant only has one image compared to a large number of plants type (61 types of plants).

V. CONCLUSION

The stage of image processing carried out is image segmentation, characteristics extraction, and identification. The segmentation process is carried out using the thresholding method to create a binary image which represents the object in white color and represent the background in black color. The characteristic extraction is based on the shape and color analysis. The shape analysis is based on the metric value and eccentricity, whereas the color analysis is based on the saturation, hue, and value. These five extracted characteristics is used as an input value in the identification process. Identification process is carried out using two algorithms, kmeans clustering algorithm and artificial neural network algorithm with back propagation. The clustering process uses 80 images whereas the testing process uses 11 images. The accuracy of training using the k-means clustering algorithm is 48.75% and during the testing process, the accuracy is only 27.27%. in artificial neural network algorithm, the accuracy during the training process is 100%, and the accuracy during testing process is 54.54%. This result showed that the clustering of the herbal plant using the artificial neural network system is better than the k-means clustering algorithm.

28

ACKNOWLEDGMENT

This research was funded by the Research Grant from the Ministry of Research, Technology and Higher Education in 2018 Fiscal Year

14 REFERENCES

- Novri K, et al, "Kajian etnobotani tanaman obat oleh masyarakat kabupaten Bone Bolango Propinsi Gorontalo [ethnobotanical study of the herbal plants by the community in Bone Bolango regency of Gorontalo Province]," Univ. Negeri Gorontalo, 2011.
- [2] Tjitrosoepomo, G., Taksonomi Umum, 3rd Edition, Yogyakarta, 25 nesia: Gadjah Mada University Press, 5 1-7, 50-54, 2005.
- [3] M.A. Agmalaro, A. Kustiyo, A.R. Akbar, "Identifikasi Tanaman Buah Tropika Berdasarkan Tekstur Permukaan Daun Menggunakan Jaringan Syaraf Tiruan [Identification of tropical fruit plants based on the su 24 texture of the plant using the artificial neural network]," Journal Ilmu Komputer Agri-Informatika 10 ol. 2 no.2, pp. 72-83, Nov. 2013.
- [4] M. Effendi, Fitriyah, U. Effendi, "Identifikasi Jenis dan Mutu Teh menggunakan Pengolahan Citra Digital dengan Metode Jaringan Syaraf Tiruan [identification of types and quality of tea using digital image processing with artificial neural network method]," Journal Teknotan, vo. 9, No. 2, Augustus. 2017.
- [5] F. Indrawan, "Aplikasi Pengenalan Pola Daun Menggunakan Jaringan Syaraf Learning Vector Quantification untuk Penentuan Tanaman Obat [application of leave pattern recognition using the neural learning vector quantification to determine herbal plant]," Proc. SemnasIF,2010, pp. 16-22
- [6] Atina, "Segmentasi Citra Paru Menggunakan 2 etode k-Means Clustering [segmentation of lung's image using k-means clustering method]," Journal Pendidikan Fisika dan Keilmuan, vol. 3 No. 2, 3 ptember. 2017.
- [7] Apriyanti, "Algoritma k-Means Clustering Dalam Pengolahan Citra Digital Landsat [k-means clustering algorithm in Landsat Digital Image

- Processing]," Journal Kumpulan jurnal Ilmu Komputer, vol. 2 No. 2, September. 2015.
- [8] Abdul kadir. Pengolahan Citra: Teori dan Aplikasi [image processing: 26 ry and application], Yogyakarta, Indonesia: ANDI, 2015.
- [9] Prasetyo, eko, Pengolahan Citra Digital dan Aplikasinya menggunakan mathlab [digital image processing and its application using mathlab],
 2) gyakarta, Indonesia: ANDI, 2011.
- [10] Tatiraju, S., & Mehta, A. Image Segmentation using k-means clustering, EM and Normalized Cuts. Department of EECS, University of California 11, 2008.
- [11] Noor m, Moch Hariadi, Image cluster berdasarkan warna untuk identifikasi kematangan buah tomat dengan metode valley tracing [color-based image clustering to identify tomato ripening with valley geing method]. Seminar nasional informatika 2009. Yogyakarta.
- [12] Widodo, Saptono; Hidayatno, Achmad; Isnanto, R Rizal, Segmentasi Citra Menggunakan Teknik Pemetaan Warna (Color Mapping) dengan Bahasa Pemrograman Delphi [image segmentation with color mapping technique with Delphi Programming Language]. Program Sarjana 23 Teknik Elektro Universitas Diponegoro, Semarang, 2011.
- [13] Suyanto, artificial intelligence Searching, Reasoning, Planning, Learning 2nd Revisio 4 Bandung, Indonesia: Informatika, 2014.
- [14] B. Pradhan, Saro I, "Landslide Risk Analysis Using Artificial Neuron Network Model Focusing on Different Training Sites," International 15 nal of Physical Sciences Vol. 4 (1), pp. 001-015, January, 2009.
- [15] Siang, J.J, Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab [Artificial Neural Network and Its Programming using the Mathlab], Yogyakarta, Indonesia: ANDI, 2005

OF							

SIMILARITY INDEX

INTERNET SOURCES

7%

PUBLICATIONS

STUDENT PAPERS

PRIMARY SOURCES

www.researchgate.net Internet Source

Atina Atina. "Segmentasi Citra Paru Menggunakan Metode k-Means Clustering", Jurnal Pendidikan Fisika dan Keilmuan (JPFK), 2017

Publication

klik.unlam.ac.id

Internet Source

1%

www.ejge.com

Internet Source

1%

journal.ipb.ac.id

Internet Source

1%

M. BINDER. "Application of an artificial neural 6 network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study", British Journal of Dermatology, 4/1994 Publication

Ali Taqi, Hayder M.A.A.Al-Assadi, Ahmed Azlan

Mat Is. "Chapter 6 An Improved Adaptive Kinematics Jacobian Trajectory Tracking of a Serial Robot Passing Through Singular Configurations", InTech, 2010

1%

Publication

8	www.jacn.net Internet Source	<1%
9	jurnal.upnyk.ac.id Internet Source	<1%
10	jurnal.unpad.ac.id Internet Source	<1%
11	repository.upnyk.ac.id Internet Source	<1%
12	Advances in Intelligent Systems and Computing, 2016. Publication	<1%
13	Submitted to Kennedy-Western University Student Paper	<1%
14	biologi.f-mipa.ung.ac.id Internet Source	<1%
15	prosiding-snastikomti.stth-medan.ac.id Internet Source	<1%
16	mdpi.com Internet Source	<1%

17	Submitted to University of Newcastle upon Tyne Student Paper	<1%
18	neural-network-matlab.doculib.org Internet Source	<1%
19	Submitted to Universiti Kebangsaan Malaysia Student Paper	<1%
20	kar.kent.ac.uk Internet Source	<1%
21	Lecture Notes in Computer Science, 2008. Publication	<1%
22	e-journal.unipma.ac.id Internet Source	<1%
23	ejournal.st3telkom.ac.id Internet Source	<1%
24	uad.portalgaruda.org Internet Source	<1%
25	journal.ugm.ac.id Internet Source	<1%
26	Agus Harjoko, Tri Ratnaningsih, Esti Suryani, Wiharto, Sarngadi Palgunadi, Nurcahya Pradana Taufik Prakisya. "Classification of acute myeloid leukemia subtypes M1, M2 and M3 using active contour without edge	<1%

segmentation and momentum backpropagation artificial neural network", MATEC Web of Conferences, 2018

Publication

- portalgaruda.ilkom.unsri.ac.id <1% Internet Source cmnt.lv 28 Internet Source Liu Hongwei, Li Li, Xuan Yiliang, Jiang Yongbo, 29 Qiang Yang. "An adaptive CBR model of call center systems", ICARCV 2004 8th Control, Automation, Robotics and Vision Conference, 2004., 2004 Publication www.ijaerd.co.in 30 Internet Source L.G. Shapiro. "Knowledge-based matching for 31 3D radiotherapy planning", [1992] Proceedings 11th IAPR International Conference on Pattern Recognition, 1992 Publication Russ, John, and F Neal. "Processing Binary <1% 32 Images", The Image Processing Handbook Seventh Edition, 2015. Publication
 - Amel, Feroui, Messadi Mohammed, and

Bessaid Abdelhafid. "Improvement of the Hard Exudates Detection Method Used For Computer- Aided Diagnosis of Diabetic Retinopathy", International Journal of Image Graphics and Signal Processing, 2012.

<1%

34

Xu, Baoyu Zhang, Hongjun Wang, Zhiteng W. "Model and algorithm of BP neural network based on expanded multichain quantum optimization.(Research", Mathematical Problems in Engineering, Annual 2015 Issue

<1%

Publication

Publication

Exclude quotes

Off Off Exclude matches

Off

Exclude bibliography