# Identification of Potential Diabetes Mellitus Through Iris Images

*by* M Raafi' Joko

International Journal of Health and Information Technology
Volume. 1 No. 3 August 2024
e-ISSN: XXXX-XXXX; and p-ISSN: XXXX-XXXX, Pg. 01-11
DOI: ..............................................................

# Identification of Potential Diabetes Mellitus Through Iris Images

**M Raafi' Joko Wahyu Saputra**
Sekolah Tinggi Ilmu Kesehatan Semarang

**Mohammad Rofi'i**
Sekolah Tinggi Ilmu Kesehatan Semarang

**Diah Rahayu Ningtias**
Sekolah Tinggi Ilmu Kesehatan Semarang

Jl. Col. Warsito Sugiarto, km 2.5 Gunungpati, Semarang
mrofii@stikessemarang.co.id

***Abstract*** *. Diagnosis of a disease is usually carried out in a laboratory test, where this test requires a lot of money and time. As time goes by in the world of health, there is one method that can be used to detect the level of condition in the body by recognizing the patterns that form on the iris of the eye or better known as Iridology. By using digital image processing, the disease diagnosis process using the iridology method can be carried out. The aim of this program is to identify diabetes milletus by processing digital images in the form of iris images using Matlab . The research data used comes from the open source website, namely Github , which consists of diabetic iris image data and normal iris image data. Data is taken in the form of digital images. In this research there are several stages, including pre-processing, feature extraction, and classification. The feature extraction process is based on statistical characteristics (contrast, correlation, energy, homogeneity) from the iris image, then classified using the K-Nearest Neighbors (KNN) algorithm and Support Vector Machine (SVM) as a comparison algorithm. The results obtained from this research show that K-NN has better performance with accuracy values of 84%, sensitivity 72%, and specificity 96%.*

***Keywords*** *: Iridology, K-NN, Matlab, Image Processing, SVM.*

## INTRODUCTION

Diabetes Mellitus is a disease that be marked with hyperglycemia (enhancement rate sugar blood) Which continously And varied, especially after Eat. Diabetes mellitus is a condition of chronic hyperglycemia accompanied by several metabolic abnormalities due to disorders hormonal, which causes chronic complications in the eyes, kidneys and blood vessels, accompanied lesion on membrane basalis in examiner saan with a microscope electrons (Rahmawati & Amiruddin, 2017).

The results of basic health research in 2007,It was found that the proportion of causes of death due to diabetes mellitus on group age 45-54 year in urban areas are *ranked* 2nd, namely 14.7% while rural. Diabetes mellitus occupies the *king ranking* 6th, that is 5.8%. By Because That, important For identify And classify disease it's as fast as it gets Possible ( Rahmawati & Amiruddin , 201 7 ). Diagnosis of a disease is usually carried out in a laboratory test, where this test requires a lot of money and time. As time goes by in the world of health, there is one

*Received: May 31, 2024, Accepted : June 20, 2024, Published: August 31, 2024*
* M Raafi' Joko Wahyu Saputra, mrofii@stikessemarang.co.id

method that can be used to detect the level of condition in the body by recognizing the patterns that form on the iris of the eye or the better known as iridology (Sekar et al., 2021).

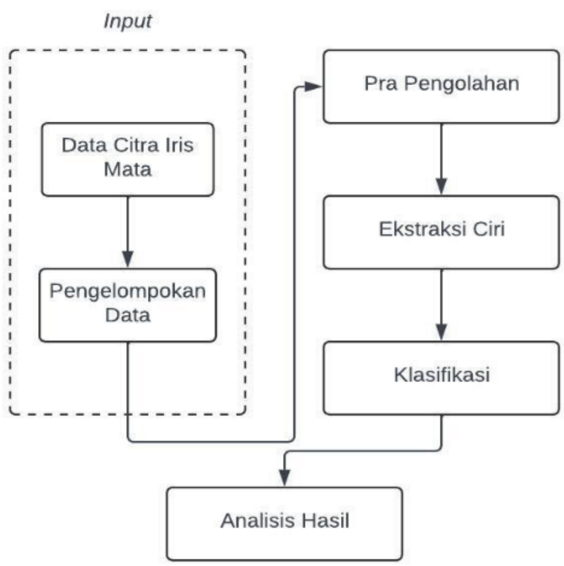
Iridology is a diagnostic approach that reads the iris, the colored part of the eye and determines it condition of any part of the body. Degree of intensity or irregularity The parts of the body affected by the disease are stored and recorded in the iris of the eye. This can used as a guide to diagnose various disorders. By using the symptoms shown by the organs in the iris, diagnosis can be made through the iris of the eye. Any changes and Imbalances found in the human body's organs are stored in the iris of the eye ( Divya et al., 20 21 ).

The iris or iris is the part of the eye that surrounds the pupil. The iris functions to give color to the eyes ( Rezika & Erlansari, 20 18 ). The iris is an internal organthe stable, externally visible, irises of two identical twins Although different and unrelated, irises have a rich physical structure and can provide a lot of data ( Diatry Indradewi , 20 18 ). Even though they have the same eye color different, the structure of the iris of the human eye is the same ( Sucipto & Riana, 20 13 ). These characteristics make iris identification known as a method Which safe and accurate in comparison other biometric methods ( Pavaloi et al., 20 19 ).

In development knowledge knowledge And technology, method for identification iris can done with help processing image. One way of image processing is by *Multi Thresholding* method for iris image segmentation, this process must be carried out so that the method is accurate and on target ( Bouaziz et al., 20 15 ). The image processing process carried out in this research uses *Gray Level Co-occurance Matrix* ( GLCM) feature extraction. This method extracts four texture statistics, namely *Contras, Correlation, Energy and Homogeneity*. (Sekar et al., 2021) . The KNN algorithm is considered to have simplicity in processing very large amounts of training data and testing data. The KNN algorithm is able to carry out training on diabetes datasets to see the negative impact of the loss of imputation values and solutions for healing (Saifur et al., 2021).

The aim of this research is use technique processing image iris using the *Multi Tresholding* segmentation method , then feature extraction is carried out using GLCM which extracts 4 features, namely *Contras, Correlation, Energy* and *Homogeneity* as well as identification image iris use algorithm *K-Nearest Neighbors* (KNN) and *Support Vector Machine* (SVM) as a comparison algorithm.

## RESEARCH METHODS

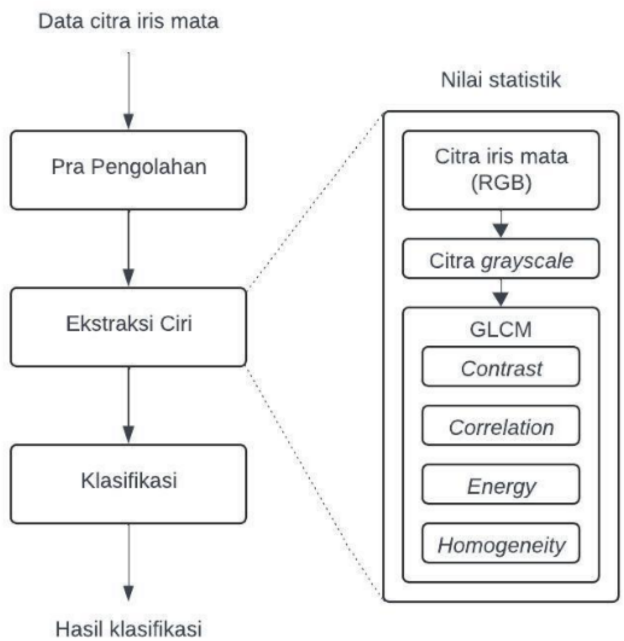



**Gambar 1. Tahapan Penelitian**

The material used in this research is 2-dimensional digital image data for two types of iris conditions taken from the open source website Github. The types of data taken were 50 normal iris data and 50 diabetic iris data, so the total amount of data was 100 data. Iris image processing in this research was carried out through several stages, including pre-processing, feature extraction and classification. These stages can be seen in Figure 1.

The pre-processing stage is carried out to get better data than the previous data. At this stage, segmentation of the iris image is carried out using the *Multi Thresholding method* , so that an iris image is obtained. Next, the feature extraction process is carried out from the segmented iris image. At the feature extraction stage, first the iris image is converted to *grayscale*.

Feature extraction is carried out to obtain characteristics of each type of iris image, namely normal iris images and diabetic iris images. At this stage, the feature extraction process is taken using the *Gray Level Co-occurance Matrix* ( GLCM) method which extracts four statistical values ( *Contras, Correlation, Energy and Homogeneity* ). In general, the stages of feature extraction can be seen in Figure 2 below. Statistical feature extraction is taken




**Gambar 2. Tahapan Ekstraksi Ciri**

from *grayscale images* . The statistical characteristics used are:

*a. Contrast*

*Contrast* ie measure intensity contrast in between neighboring pixels (equation 1)

$$\sum_{ij} |i - j| 2P(i,j) \quad (1)$$

*b. Correlation*

*Correlation* ie measure how much correlated pixels with their neighbors (equation 2)

$$\frac{\sum i \; \sum j (i - \mu i)(j - \mu j) p(i,j)}{\sigma i \sigma j} \quad (2)$$

*c. Energy*

*Energy* ie amount element square GLCM Which normalized (equation 3)

$$\sum_{ij} P(i,j) 2 \quad (3)$$

*d. Homogeneity*

*Homogeneity* is a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal, indicating a measure of the homogeneity of the image (equation 4)

$$\sum_{ij} \frac{p(i,j)2}{1 + |i - j|} \quad (4)$$

After obtaining these statistical characteristics, they are then classified using the *K-Nearest Neighbors* (KNN) method . In simple terms, the KNN algorithm is a technique based on an information system which is a technique for mapping (classifying) data into one or several classes that have been previously defined.

Testing of the classification method applied in this research was by calculating accuracy, sensitivity and specificity values. Each can be searched for using equations 5-7 below.

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

$$Sensitivitas = \frac{TP}{TP+FN} \quad (6)$$

$$Spesifitas = \frac{TN}{FP+TN} \quad (7)$$

Where TP = *True Positive*, FP = *False Positive*, TN = *True Negative*, FN = *False Negative*

Furthermore, as an evaluation material to obtain accuracy estimation results in this research, the total data amounting to 170 is divided into 2, namely 120 data as training data and 50 data as testing data, then the results of the testing data are presented in the form of a confusion matrix.

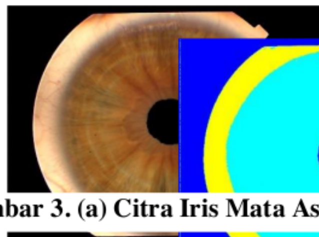
## RESULTS AND DISCUSSION

### 1.1 Pre-Processing

Pre-processing is carried out to obtain data that is better than the original data by manipulating the parameters contained in the data, so that a form is obtained that is more suitable to the desired values. The pre-processing stage begins with processing the iris image that has been obtained previously, so that an image is obtained that has the desired size and dimensions and color distribution (pixel values). The first step taken is segmentation.

(a) (b)

Segmentation is carried out to obtain an image that only consists of the desired object and



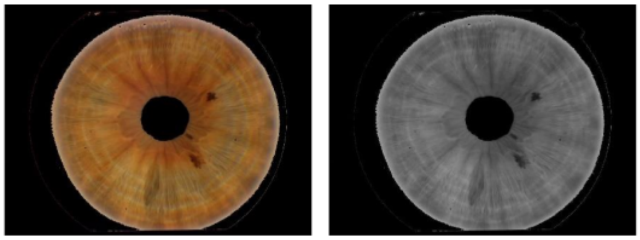
**Gambar 3. (a) Citra Iris Mata Asli (b) Hasil Segmentasi Citra Iris Mata**

**Gambar 4. Area Segmentasi Citra**

removes other objects that can interfere with the further process as in Figure 3. At this stage segmentation is carried out using the *Multi-Thresholding method, previously the image is converted into a gray or grayscale* image , then separation is carried out between regions in the iris image based on differences in pixel intensity values from *grayscale images* . In this image the iris is used two *threshold values* to separate *regions* . The *threshold* values used to segment this image are 6 and 190 which are included in region 2 and are marked in *cyan* in Figure 4.

### 1.2 Feature Extraction

The feature extraction process is carried out with the aim of obtaining characteristics from each type of iris image, namely normal iris images and diabetic iris images. The characteristics taken are statistical values from the *grayscale image* of the segmented iris image shown in Figure 5. The statistical values used are *Contras, Correlation, Energy and Homogeneity*.



(a) (b)

**Gambar 5. (a) Citra Hasil Segmentasi (b) Citra *Grayscale***

From the *grayscale image* above, statistical characteristics can be searched, so that each data is obtained as in table 1 below.

**Table 1. Statistical characteristics of *grayscale images***

| *Contrast* | *Correlation* | *Energy* | *Homogeneity* |
|---|---|---|---|
| *0.1759* | *0.9568* | *0.3546* | *0.9692* |

After obtaining the statistical characteristics, the next step is the classification process using the *K-Nearest Neighbors* (KNN) algorithm .

## 1.3 Classification

Classification is the process of finding models or functions that describe and differentiate data classes or concepts with the aim of being able to be used to predict the class of objects whose class labels are unknown. The classification method used in this research is *K-Nearest Neighbors* (KNN) and as a comparison algorithm *the Support Vector Machine* (SVM) is used.

The KNN algorithm is a *supervised learning method* because *the query instance results* need to be classified based on the majority of the KNN categories. The goal of this algorithm is to classify new objects based on attributes and *training samples* . *The classifier* does not use any model for fitting and is based only on memory. Given a *query point*, we will find a number of *K* objects or ( *training points* ) that are closest to the *query point* . Classification uses the most *votes* among the classifications of *K* objects. The KNN algorithm uses neighborhood classification as a prediction value for a new *query instance* .

*K–Nearest Neighbor* (KNN) works by classifying an object that has the closest similarity to other objects. KNN has an attribute which is initialized as k, namely the number of neighbor values which are used as a reference in the *k–nearest neighbor classification* to differentiate based on class. The reason the k-nearest neighbor method was chosen is because it can classify objects based on learning data that is closest to the object so that the results can be more accurate. In the learning phase, this algorithm only stores feature vectors and classifies the learning data. In the classification phase, the same features are calculated for test data ( data whose classification is unknown). The distance of this new vector to all learning data vectors is calculated, and the *K* closest ones are taken. The newly classified point is predicted to be among the most classified of these points. The best *K* value for this algorithm depends on the data. In general, a high *K value will reduce the effect of noise* on the classification, but make the boundaries between each classification more blurred. The special case where the classification is predicted based on the closest learning data (in other words, *K* = 1) is called the *nearest neighbor algorithm* . A good *K* value is selected by optimizing parameters by dividing all iris image data into two parts, namely training data and testing data, then the results are presented in the form of *a confusion matrix* which can be seen in Table 2 below.

**Table 2. *Confusion matrix* results of diabetes identification using KNN**

| | |
|---|---|
| TP = 18 | FN = 7 |
| FP = 1 | TN = 24 |

Based on Table 2 above, the results obtained from the classification performance using the KNN method based on statistical characteristics taken from *grayscale images* are an accuracy of 84% with a sensitivity of 72% and a specificity of 96%.

Meanwhile, *Support Vector Machine* (SVM) is a learning system whose classification uses a hypothesis space in the form of linear functions in a high-dimensional *feature space* , trained with a learning algorithm based on optimization theory by implementing *learning* bias derived from statistical learning theory. . In the SVM concept, we try to find the best separating function ( *hyperplane* ) among an unlimited number of functions. The best separating *hyperplane* between the two classes can be found by measuring *the margin of* the hyperplane and looking for the maximum point. Learning is carried out using pairs of *input data* and *output data* in the form of the desired target. Learning in this way is called directed learning *(supervised learning* ). With this directed learning, a function will be obtained that describes the form of dependence on input and output. Furthermore, it is hoped that the function obtained has good generalization capabilities. The results of the *Support Vector Machine (SVM)* classification method that have been applied in this research can be seen in table 3 below.

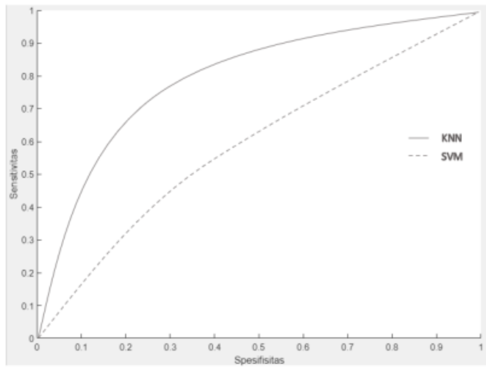
**Table 3. *Confusion matrix* results of diabetes identification using SVM**

| | |
|---|---|
| TP = 10 | FN = 15 |
| FP = 2 | TN = 23 |

Based on Table 3 above, the results obtained from the classification performance using the SVM method based on statistical characteristics taken from *grayscale images* are 66% accuracy, 40% sensitivity and 92% specificity.

So after obtaining the results of the classification of each algorithm, to see it more clearly, you can see the ROC ( *Receiver Operating Characteristic* ) curve which can describe the performance of each method, namely KNN and SVM as shown in Figure 6 below.

Based on the ROC curve in Figure 6 above, the AUC ( *Area Under ROC Curve* ) value for




**Gambar 6. Kurva ROC algoritma KNN dan SVM**

each algorithm has been obtained, namely KNN = 0.841 or 85.1%, and SVM = 0.642 or 64.2%. So from the AUC value it can be said that the KNN algorithm shows better performance than the comparison algorithm, namely SVM which has been used because it has the largest AUC value or is almost close to 100%. Because basically the AUC value has a range between 50% (0.5) to 100% (1). An AUC value of 50% is the worst AUC value, while an AUC value of 100% is the best value.

## CONCLUSION

After going through the various stages that have been carried out in this research, it shows that the extraction of statistical features taken from *grayscale images* can be used as a characteristic or *feature* for identifying diabetes iris images based on original iris image data. Meanwhile, for classification, the *K–Nearest Neighbor* (KNN) algorithm is used to identify diabetic iris images obtained better results by assessing the classification performance *index* resulting in an accuracy of 84.0%, sensitivity of 72%, and specificity of 96% with an AUC value = 0.841. Meanwhile, for SVM, it was obtained 66%, 40%, 92% in terms of accuracy, sensitivity and specificity with an AUC value of 0.642.

**REFERENCE LIST**

Bouaziz, Draa, A., & Chikhi, S. (2015). Artificial bees for multilevel thresholding of iris images. *Swarm Evol. Comput, 21* , 32–40. https://doi.org/10.1016/j.swevo.2014.12.002

Cahyanti, D., Rahmayani, A., & Ainy, S. (2020). Analysis of the performance of the Knn method on a dataset of patients with breast cancer. *Indonesia. J. Data Sci., 1* , 39–43.

Cholil, RS, Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementation of the K-Nearest Neighbor (KNN) Classification Algorithm for Scholarship Acceptance Selection Classification. *IJCIT Journal, 6* (2), 118-127. https://ejournal.bsi.ac.id/ejurnal/index.php/ijcit/article/view/10438

Diatri, IGAA (2018). Iris Identification with Snake Model-PSO and Gabor 2-D. *JST (Journal of Science and Technology, 7* (1), 25. https://doi.org/10.23887/jst-undiksha.v7i1.13010

Divya, D.C., Gururaj, L.H., Rohan, R., Bhagyalakshmi, V., Rashmi, A.H., Domnick, A., Flammini, F. (2021). An efficient machine learning approach to nephrology through iris recognition. *ResearchGate Journal, 1* (1), 10. https://www.researchgate.net/publication/355200132_An_efficient_machine_learning_approach_to_nephrology_through_iris_recognition

MM, R., Wijayanto, I., & Aulia, S. (2019). Biometric Iris Recognition Using Lbp With Classifiers Knn Biometrick Iris Recognition Using Dwt With Classifiers K-Nearest. *6* (1), 817–825.

Nugraha, ARD, Auliasari, K., & Agus Pranoto, Y. (2020). IMPLEMENTATION OF THE K-NEAREST NEIGHBOR (KNN) METHOD FOR SELECTION OF PROSPECTIVE NEW EMPLOYEES (Case Study: BFI Finance Surabaya). *JATI (Informatics Engineering Student Journal), 4* (2), 14–20. https://doi.org/10.36040/jati.v4i2.2656

Pavaloi, I., Nita, C.D., & Lazar, L.C. (2019). Novel matching method for automatic iris recognition using SIFT features. *ISSCS 2019 - Int. Symp. Signals, Circuits Syst* , 1–4. https://doi.org/10.1109/ISSCS.2019.8801797

Rahardja, A., Juardi, T., & Agung, H. (2019). Implementation of the K-Nearest Neighbor Algorithm on Laptop Recommendation Websites. *J. Buana Inform, 10* (1), 75. https://doi.org/10.24002/jbi.v10i1.1847

Rahmadya. (2019). Training, Validating, Testing and Corpus. Accessed from https://rahmadya.com/

Rahmawati & Amiruddin. (2017). Glycohemoglobin, Hypertension, IMT Against Vision Impairment in Elderly Diabetes Mellitus Patients. *Public Health Media Journal, 13* (1), 1528. https://journal.unhas.ac.id/index.php/mkmi/article/view/1582

Rosiva, AS, Zarlis, M., & Wanayumini. (2022). Leaf Image Classification with GLCM (Gray Level Co-Occurrence) and K-NN (K-Nearest Neighbor). *Journal of Matrics, 21* (2), 477-486. https://journal.universitasbumigora.ac.id/index.php/matrik/article/view/1572

Shidiq, F., Hidayat, WE, & Kurniati, IN (2021). Application of the K-Nearest Neighbor (KNN) Method to Determine Betta Fish by Extracting Shape and Canny Features. *Innovatics Journal, 3* (2), 39-46. https://jurnal.unsil.ac.id/index.php/innovatics/article/view/3093

Sucipto, DB, & Riana, D. (2013). Potential Glaucoma Diagnosis Application Through Iris Images Using Artificial Neural Networks Using Back Propagation Method. *1* (3), 19–27.

Suhendri & Rahayu, P. (2019). Gray Level Co-Occurrence (GLCM) Method for Classifying Water Guava Leaf Types Using the Neural Network Algorithm. *Joint Journal, 01* (01), 15-22. https://jurnal.stmik-amikbandung.ac.id/joint/article/view/4

Wahid, GSF, Purnamasari, R., & Saidah, S. (2019). Personal Identification Through the Iris of the Eye Using the Compound Local Binary Pattern Method and Support Vector Machine Classification Personal Identification Based on Compound Local Binary. *6* (2), 3959–3966.

Yahya, Y., & Puspita, WH (2020). Application of the K-Nearest Neighbor Algorithm to Classify the Effectiveness of Vape (Electric Cigarette) Sales in 'Lombok Vape On. *Infotek J. Inform. and Technol., 3* (2), 104–114. https://doi.org/10.29408/jit.v3i2.2279

Yani, S., Jumeilah, FS, & Kadafi, M. (2020). K-Nearest Neighbor Algorithm to Determine the Eligibility of Families Receiving Non-Cash Food Assistance (Case Study: Karya Jaya Village). *Journal of Information Technology Ampera, 1* (2), 75–87. https://doi.org/10.51519/journalita.volume1.isssue2.year2020.page75-87

Yustanti, W. (2012). K-Nearest Neighbor Algorithm to Predict Land Selling Prices. *JMSK: Journal of Mathematics, Statistics & Computing, 9* (1), 57-68. http://journal.unhas.ac.id/index.php/jmsk

# Identification of Potential Diabetes Mellitus Through Iris Images

ORIGINALITY REPORT

| 25% | 21% | 19% | 13% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | jurnal.itscience.org<br>Internet Source | 4% |
| 2 | aladzkiyajournal.com<br>Internet Source | 3% |
| 3 | C. D. Divya, H. L. Gururaj, R. Rohan, V. Bhagyalakshmi, H. A. Rashmi, A. Domnick, Francesco Flammini. "An efficient machine learning approach to nephrology through iris recognition", Discover Artificial Intelligence, 2021<br>Publication | 2% |
| 4 | adri27th.stkipsingkawang.ac.id<br>Internet Source | 2% |
| 5 | international.aripi.or.id<br>Internet Source | 2% |
| 6 | jurnal.fmipa.unmul.ac.id<br>Internet Source | 1% |
| 7 | Submitted to Associatie K.U.Leuven<br>Student Paper | 1% |

| | | |
|---|---|---|
| 8 | www.thescipub.com<br>Internet Source | 1 % |
| 9 | doctorpenguin.com<br>Internet Source | 1 % |
| 10 | Submitted to Universitas Diponegoro<br>Student Paper | 1 % |
| 11 | pji.ub.ac.id<br>Internet Source | 1 % |
| 12 | d.docksci.com<br>Internet Source | 1 % |
| 13 | Jung-Ho Choi, Hae-Yeoun Lee, Heung-Kyu Lee. "Color laser printer forensic based on noisy feature and support vector machine classifier", Multimedia Tools and Applications, 2011<br>Publication | 1 % |
| 14 | Ajrina Adistiabudi Khairani, Sarah Fairuz Imani, Rita Arbianti, Tania Surya Utami, Heri Hermansyah. "The effect of fermentation temperature and the crude extract concentration of bitter leaf (Andrographis Paniculata) on inhibition of α-glucosidase enzyme activity for antidiabetic drugs", AIP Publishing, 2020<br>Publication | 1 % |
| 15 | repositori.unsil.ac.id<br>Internet Source | |

| | | |
|---|---|---|
| | | 1% |
| 16 | Rohmat Indra Borman, Riduwan Napianto, Nurhasan Nugroho, Donaya Pasha, Yuri Rahmanto, Yohanes Egi Pratama Yudoutomo. "Implementation of PCA and KNN Algorithms in the Classification of Indonesian Medicinal Plants", 2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), 2021<br>Publication | 1% |
| 17 | Rafi Haidar Ramdlon, Entin Martiana Kusumaningtyas, Tita Karlita. "Brain Tumor Classification Using MRI Images with K-Nearest Neighbor Method", 2019 International Electronics Symposium (IES), 2019<br>Publication | 1% |
| 18 | dashboard-lldikti6.kemdikbud.go.id<br>Internet Source | 1% |
| 19 | www.dtic.mil<br>Internet Source | 1% |

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | < 1% |
| Exclude bibliography | On | | |