Indonesian Journal of Applied Technology, Computer and Science
Vol. 2, No. 2, August 2025, pp. 46-56
ISSN: 3063-8305

Application of Support Vector Machine and Image Segmentation Methods to Classify Types of Medicinal Leaves in Betel Leaves

Dani Damara Simamora[1], Mhd Basri[2], Andi Zulherry[3]
[1,2]Department of InformationTechnology, Universitas Muhammadiyah Sumatera Utara, Indonesia
[3]Department of Data Science, Universitas Muhammadiyah Sumatera Utara, Indonesia

ABSTRACT

Betel leaves (Piper betle) are widely recognized for their medicinal properties and are commonly used in traditional medicine. Accurate classification of betel leaf types is essential for quality control, proper utilization, and preservation of their medicinal value. This study proposes a classification approach combining image segmentation techniques with the Support Vector Machine (SVM) algorithm to identify and categorize different types of medicinal betel leaves. The methodology involves image acquisition, preprocessing, and segmentation to isolate leaf regions from the background, followed by feature extraction based on color, texture, and shape descriptors. These features are then used to train the SVM classifier for accurate leaf type identification. Experimental results demonstrate that the proposed method achieves high accuracy in classifying betel leaf types, with performance metrics including precision, recall, and F1-score confirming the robustness of the approach. This research contributes to the development of automated, non-destructive, and efficient systems for medicinal plant classification, supporting applications in agriculture, herbal medicine, and botanical research. Indonesia has a rich biodiversity, including medicinal plants. Leaves are one of the plant parts widely used as traditional medicine. However, people often find it difficult to identify types of medicinal leaves. This study aims to classify types of medicinal leaves using the Support Vector Machine (SVM) method and digital image segmentation. The data used are images of betel leaves, which are processed through several stages: color transformation from RGB to HSV, image segmentation using the thresholding method, and texture and color feature extraction. The results show that the SVM method and image segmentation can classify leaf types with high accuracy. Therefore, this method can help people identify medicinal leaves more easily and accurately.

Keyword : Digital Image Processing; Image Segmentation; Support Vector Machine (SVM); Feature Extraction; Betel Leaves

Corresponding Author:
Dani Damara Simamora
Department of Information Technology
Universitas Muhammadiyah Sumatera Utara
Jl. Kapten Mukhtar Basri No 3 Medan, 20238, Indonesia.
Email : danidamarasimamora@gmail.com

Article history:
Received April, 2024
Revised April, 2024
Accepted Mei, 2024

1. INTRODUCTION

Indonesia is a country rich in biodiversity, including medicinal plants. Medicinal plants are plants that have properties to cure diseases. Leaves are one of the parts of the plant that are most widely used as traditional medicine ingredients (Prayoga et al., 2023). Classification of types of medicinal leaves is one effort to preserve and utilize medicinal plants optimally. It is difficult for people to determine whether leaves are medicinal or not in today's era, and this classification can make it easier for people to determine whether leaves are medicinal or not. This classification can be done in various ways, one of which is by using digital image processing methods.

Digital image processing is a technique used to manipulate digital images. One of the digital image processing techniques that can be used to classify types of medicinal leaves is image segmentation. Image segmentation is the process of separating objects in an image into smaller parts. Image segmentation can be done using various methods, one of which is the thresholding method. The thresholding method is a method that divides an image into two parts, namely the foreground and background. The foreground is the part that contains the objects to be classified, while the background is the part that does not contain objects. After the objects in the image have been segmented, the next step is feature extraction. Feature extraction is the process of identifying the characteristics of an object that can be used to distinguish it from other objects.

Features that can be used to classify types of medicinal leaves include shape, texture, and color features. Shape features are features that describe the shape of an object, such as size, shape, and direction. Texture features are features that describe the surface texture of an object, such as particle size, regularity, and smoothness. Color features are features that describe the color of an object, such as hue, saturation, and brightness. After the features have been extracted, the next step is classification. Classification is the process of determining the class of an object based on the features that have been extracted.

Classification methods that can be used to classify types of medicinal leaves include the support vector machine (SVM) method. The SVM method is a classification method that uses a kernel function to separate object classes. Based on the background above, the author conducted a study entitled the application of support vector machine methods and image segmentation to classify types of medicinal leaves as an effort to preserve and optimally utilize medicinal plants.

2. RESEARCH METHOD/MATERIAL AND METHOD/LETERATURE REVIEW

A. Research Stages

To carry out research so that it can be carried out properly and in accordance with the objectives, it is necessary to compile structured and planned research stages. The steps in conducting research to classify medicinal images by applying the Support Vector Machine (SVM) algorithm are presented in Figure 1.

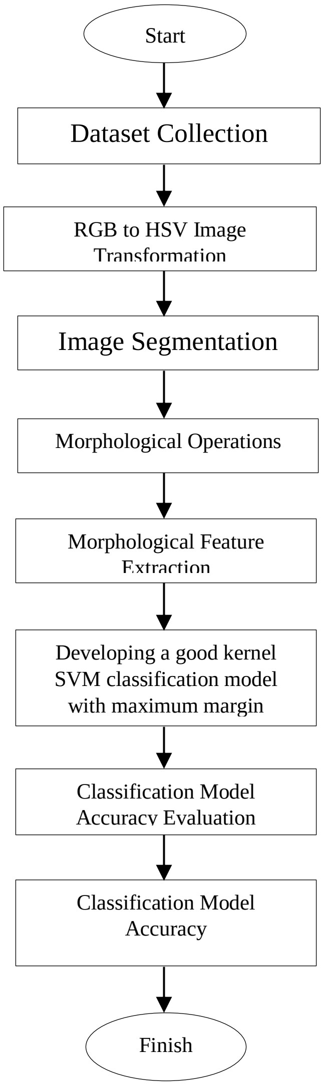

Fig 1. Research Stages Flowchart

B. Dataset Collection

Before classifying the image, the initial step taken is to collect image data of medicinal leaf types which will later become a data set or what is called a dataset. Datasets are an important aspect in image processing, this is because the availability of datasets determines the performance of a model (A. Mulyanto, 2021). The types of medicinal leaves used are based on the book "Various Medicinal Plants" (M. Megawati, 2021), then selected based on medicinal leaves that are easy to find, namely there are 5 (five) types of medicinal leaves, including: Bay Leaves, Betel Leaves, Guava Leaves, Turmeric Leaves and Soursop Leaves. The process of obtaining datasets through image capture is taken independently using a camera with the same light intensity. For dataset distribution based on the division of 70% as data for training and 30% as data for testing. The total image data that has been collected is as many as images of medicinal plant types with each class of 200 images. By dividing the distribution of training data and testing data that has been determined so that the number of training data used is 120 images and the testing data used is 75 images.

C. RGB to HSV Image Transformation

To facilitate image processing, generally images with RGB color space will be converted first to Hue, Saturation and Value or HSV color space. The HSV color model provides an overview of coloring that is the same as human intuition in capturing color (F. Sun, 2020, 2021). The process of converting RGB images to HSV images is useful for facilitating the segmentation process because this process will produce images with HSV colors that are visible on the objects to be segmented.

D. Image Segmentation

At this stage, the image is useful for distinguishing between objects and their backgrounds. In image segmentation, objects will be separated based on their regional boundaries so that the object and its background will be visible. So, at this stage, the image will be converted into a binary image, where the desired object gets a value of 1, while the background will have a value of 0 (R. I. Borman I. A., 2022). The image segmentation approach applied in this study is Thresholding. In this technique, the appropriate threshold value will be sought, to make it easier to distinguish objects in the image. (Priyopradono, 2018) The stages in Thresholding are the same as the image quantization process. To perform image quantization through equations (1) and (2).

$$x = b * int\left(\frac{w}{b}\right) \qquad (1)$$

$$b = int\left(\frac{256}{a}\right) \qquad (2)$$

Based on equations (1) and (2), w shows the gray level value before Thresholding is performed. Then, x is the gray level value after Thresholding.

E. Morphological Operations

At this stage, the aim is to improve the results of image segmentation so that the segmented objects can be seen clearly. Generally, morphological operations are implemented on binary images to improve segmentation results (Perangin-angin, 2018). There are many types of morphological operations, in this study the morphological operations used are filling holes, opening and closing operations. The morphological operation of filling holes is used to fill the entire area with a value of 1 and in this operation with a reference based on pixel values. The morphological operation of opening is used to smooth the lines that form objects, remove narrow parts and remove thin protrusions. While the morphological operation of closing rejects fragments, removes holes and fills gaps in lines. To perform filling holes, opening and closing operations, equations (3), (4) and (5) can be used.

$$X_k = (X_{k-1} \oplus B) \cap A^c \qquad (3)$$

$$A \circ B = (A \ominus B) \oplus B \qquad (4)$$

$$A \cdot B = (A \oplus B) \ominus B \qquad (5)$$

Where X is a point inside the boundary, A indicates the initial matrix in the site before the morphological operation is performed, while B indicates the operator matrix used to perform the morphological operation.

F. Morphological Feature Extraction

Feature extraction can be interpreted as a stage to obtain features that are used to distinguish between one object and another (R. I. Borman F. R., 2021). The features obtained are then used as information that can characterize objects to be managed by the classification algorithm. In the leaf, the features that can be extracted are features based on their shape. The feature extraction used is morphological feature extraction with parameters area, perimeter, eccentricity, major axis length and minor axis length. Area can be obtained from the number of pixels contained in an object in an image. Perimeter can be obtained from the number of pixels surrounding the object. Then the major axis length is the length of the diameter of an area and the minor axis length is the shortest diameter of an area. Based on the major and minor axes, other morphological features can be calculated, namely eccentricity. Eccentricity is the ratio of the length between the major axis and the minor axis. The eccentricity parameter can be calculated using the following equation (6).

$$e = \sqrt{1 - \frac{b^2}{a^2}} \qquad (6)$$

G. Classification with Support Vector Machine

In 1992, the support vector machine (SVM) was introduced by Vladimir Vapnik, Bernhard Boser, and Isabelle Guyon. Before 1992, precisely in the 1960s, the basis for SVM already existed (including early work by Vapnik and Alexei Chervonenkis on statistical learning theory). Support Vector Machine (SVM) is a supervised type of classification because when the training process requires a specific learning target. SVM is an algorithm that works to solve classification problems by finding the maximum marginal hyperplane (MMH). The concept of SVM can be simply stated as an effort to find the optimal separator function (classifier/hyperplane) that can separate two data sets from two different classes. (Safik Rionga Alven, 2018).

H. Evaluation of Classification Result Accuracy

At this stage, an accuracy evaluation is carried out by matching the classification prediction data using the support vector machine (SVM) method with the best kernel so that in the confusion matrix stage, namely visualizing the performance of the prediction model in classification. In the example below, table 1 can be seen as the solution to the confusion matrix with multi-classes. The equation for precision, recall, and accuracy of the original data with the following equation. (Khalimi, 2021).

$$Presisi = \frac{TP}{TP+FP}\ x\ 100\% \qquad (7)$$

$$Recall = \frac{TP}{TP+FN}\ \text{x}\ 100\% \qquad (8)$$

$$Akurasi = \frac{TP+TN}{TP+FN+FP+FN}\ \text{x}\ 100\% \qquad (9)$$

$$F1\ Score = 2 * \frac{Precision*Recall}{Precision+Recall} \qquad (10)$$

3. RESULTS AND DISCUSSION

In the results carried out in this research, there are processes that must be carried out, so that the desired research results are achieved, including:

A. RGB to HSV Image Transformation

The following is an image of the result of RGB to HSV Image Transformation, along with hue channels, saturation channels, and value channels:

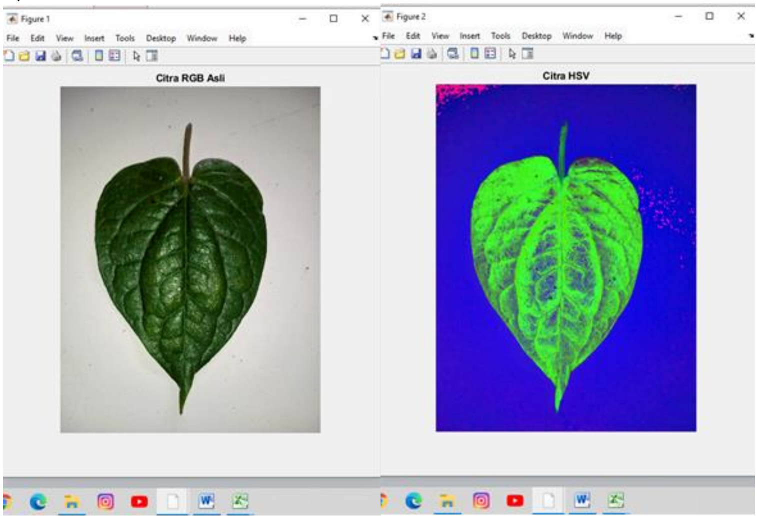

Fig 2. Original RGB Image & HSV Image

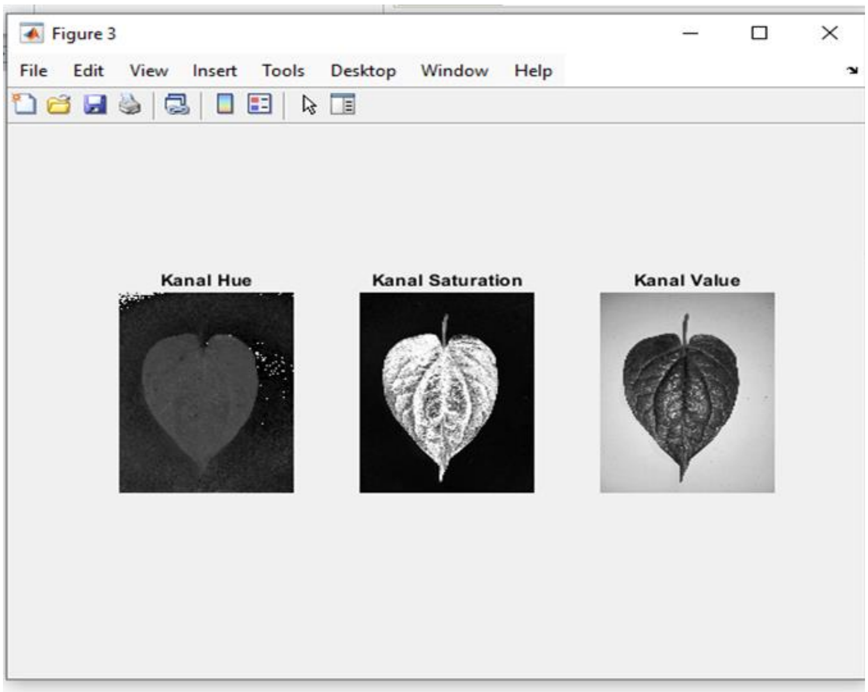

Fig 3. Kanal Hue, Kanal Saturation, & Kanal Value

B. Image Segmentation

The following is an image of the results of image segmentation which includes, among others:

Fig 4. Original Image

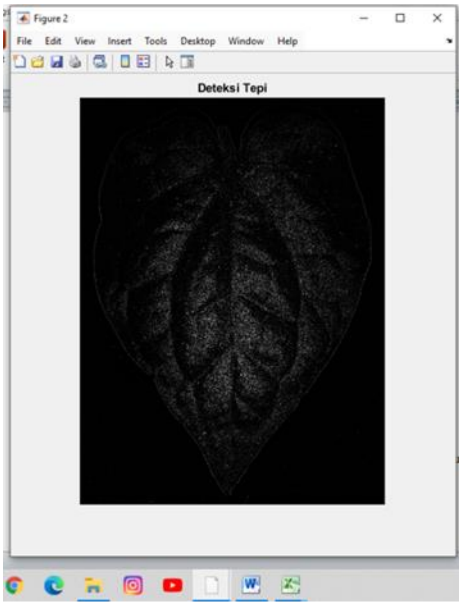

Fig 5. Edge Detection

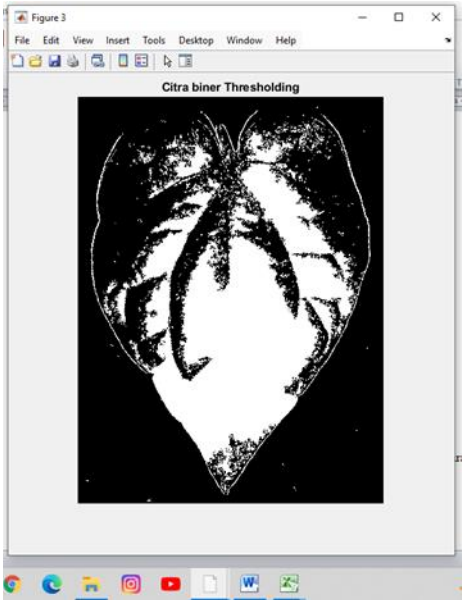

Fig 6. Thresholding Binary Image

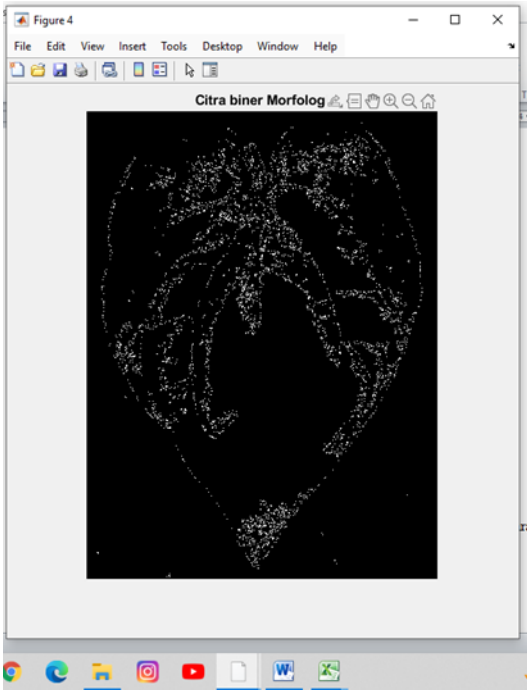

Fig 7. Morphological Binary Image

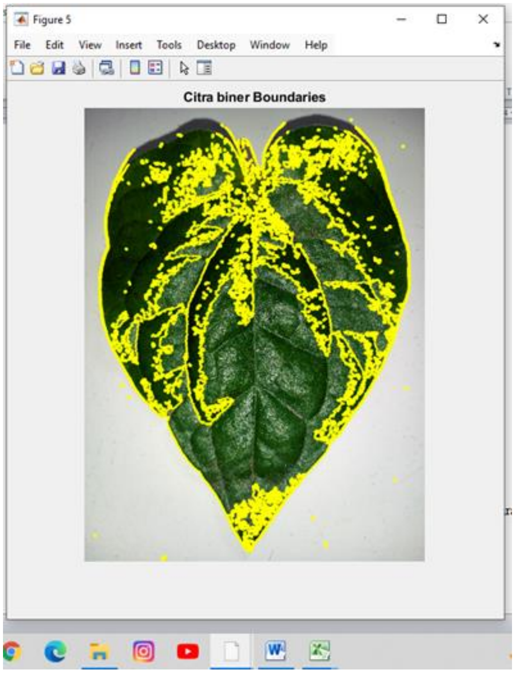

Fig 8. Binary Image Boundaries

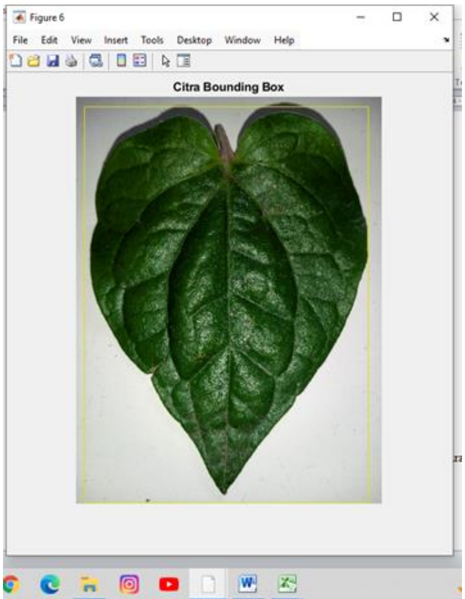

Fig 9. Image Bounding Box

Fig 10. Cropping Segmentation Results

C. Classification with Support Vector Machine

The following is the calculation of the confusion matrix for the average feature extraction section of the table above:

Table 1. Confusion Matrix for Average Feature Extraction

	Predcited Positive	Predicted Negative
Actual Positive	3(TP)	1(FN)
Actual Negative	1(FP)	5(TN)

Akurasi:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{3+5}{3+5+1+1} = \frac{8}{10} = 0{,}8$$

Presisi

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{3}{3+1} = \frac{3}{4} = 0{,}75$$

Recall

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{3}{3+1} = \frac{3}{4} = 0{,}75$$

Spesifisitas

$$\text{Spesifitas} = \frac{TN}{TN+FP} = \frac{5}{5+1} = \frac{5}{6} = 0{,}83$$

F1-Score

$$\text{F1-Score} = \frac{2 \text{ x Presisi x Recall}}{\text{Presisi + Recall}} = \frac{2 \text{ x } 0.75 \text{ x } 0,75}{0,75 + 0,75} = \frac{1,125}{1.5} = 0,75$$

4. CONCLUSION

Based on the results and discussions above, there are several conclusions concluded by the author: Transformation from RGB to HSV facilitates color analysis and processing in various applications, such as color segmentation, object tracking, and image enhancement. By separating color information from intensity and saturation, the HSV model is more intuitive and closer to the way humans perceive color. Image segmentation is an important process in image processing that includes various techniques for dividing images into meaningful parts. Edge detection helps find object boundaries, thresholding produces binary images, morphological operations improve binary images, boundary determination finds object contours, bounding boxes surround objects with bounding boxes, and cropping cuts off parts of the image that contain objects. The combination of these techniques allows for more effective and efficient image processing and analysis. In the classification of betel leaf images using SVM, the combination of feature extraction shows varying levels of accuracy. Histogram Texture gives the best results with only 1 leaf being misclassified from 1000 test data. GLCM also shows good performance with only 3 errors. The Contrast feature has the lowest performance with 9 errors out of 1000 test data. Overall, the SVM method with proper feature extraction can effectively classify betel leaf images, and proper feature selection is essential to improve classification accuracy.

REFERENCES

Sari, I.P., Hariani, P.P., Al-Khowarizmi, A., Ramadhani, F., Sulaiman, O.K., Satria, A, & Manurung, A.A. (2024). CLUSTERING HIV/AIDS DISEASE USING K-MEANS CLUSTERING ALGORITHM. Proceeding International Seminar on Islamic Studies 5 (1), 1668-1676

Sari, I.P., Ramadhani, F., Satria, A., & Sulaiman, O.K. Leukocoria Identification: A 5-Fold Cross Validation CNN and Adaboost Hybrid Approach. 2023 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 486-491

Manurung, A.A., Nasution, M.D., & Sari, I.P. (2023). Implementation of Fuzzy K-Nearest Neighbor Method in Dengue Disease Classification. 2023 11th International Conference on Cyber and IT Service Management (CITSM), 1-4

Sari, I.P., Ramadhani, F., Satria, A., & Apdilah, D. (2023). Implementasi Pengolahan Citra Digital dalam Pengenalan Wajah menggunakan Algoritma PCA dan Viola Jones. Hello World Jurnal Ilmu Komputer 2 (3), 146-157

Sari, I.P., Al-Khowarizmi, A, Sulaiman, O.K., & Apdilah, D. (2023). Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases. Journal of Computer Science, Information Technology and Telecommunication Engineering 4 (2), 402-412

Sulaiman, O.K & Batubara, I.H. (2021). Implementation Data Mining For Level Analysis Traffic Violation By Algorithm Association Rule. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal 2 (2), 128-135

Sari, I.P., Batubara, I.H., & Al-Khowarizmi, A. (2021). Sensitivity Of Obtaining Errors In The Combination Of Fuzzy And Neural Networks For Conducting Student Assessment On E-Learning. International Journal of Economic, Technology and Social Sciences (Injects) 2 (1), 331-338

Sari, I.P., Al-Khowarizmi, A., & Batubara, I.H. (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering For Grouping Students' Abilities In Online Learning Process. Journal of Computer Science, Information Technology and Telecommunication Engineering 2 (1), 139-144

Apdilah, D., & Sari, I.P. (2021). Optimization Of The Fuzzy C-Means Cluster Center For Credit Data Grouping Using Genetic Algorithms. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal 2 (2), 156-163

Indah Purnama Sari. Algoritma dan Pemrograman. Medan: UMSU Press, 2023, pp. 290.

Indah Purnama Sari. Buku Ajar Pemrograman Internet Dasar. Medan: UMSU Press, 2022, pp. 300.

Indah Purnama Sari. Buku Ajar Rekayasa Perangkat Lunak. Medan: UMSU Press, 2021, pp. 228.

Janner Simarmata Arsan Kumala Jaya, Syarifah Fitrah Ramadhani, Niel Ananto, Abdul Karim, Betrisandi, Muhammad Ilham Alhari, Cucut Susanto, Suardinata, Indah Purnama Sari, Edson Yahuda Putra. Komputer dan Masyarakat. Medan: Yayasan Kita Menulis, 2024, pp.162.

Mahdianta Pandia, Indah Purnama Sari, Alexander Wirapraja Fergie Joanda Kaunang, Syarifah Fitrah Ramadhani Stenly Richard Pungus, Sudirman, Suardinata Jimmy Herawan Moedjahedy, Elly Warni, Debby Erce Sondakh. Pengantar Bahasa Pemrograman Python. Medan : Yayasan Kita Menulis, 2024, pp.180

Zelvi Gustiana Arif Dwinanto, Indah Purnama Sari, Janner Simarmata Mahdianta Pandia, Supriadi Syam, Semmy Wellem Taju Fitrah Eka Susilawati, Asmah Akhriana, Rolly Junius Lontaan Fergie Joanda Kaunang. Perkembangan Teknologi Informatika. Medan: Yayasan Kita Menulis, 2024, pp.158