# Image Clustering Optimization: A Comparison of Single vs Hybrid Feature Extraction Technique

1st Akhiar Wista Arum
*Dept. of Computer Engineering*
*Universitas Sriwijaya*
Palembang
akhiarwistaarum@unsri.ac.id

2nd Dian Palupi Rini*
*Dept. of Informatics Engineering*
*Universitas Sriwijaya*
Palembang
dprini@unsri.ac.id

3rd Dewa Sheva Dzaky
*Dept. of Informatics Engineering*
*Universitas Sriwijaya*
Palembang
09021182126005@student.unsri.ac.id

4th Iman Carrazzi Syamsidi
*Dept. of Informatics Engineering*
*Universitas Sriwijaya*
Palembang
09021182126009@student.unsri.ac.id

5th Affandi Arrizal
*Dept. of Informatics Engineering*
*Universitas Sriwijaya*
Palembang
09021182126027@student.unsri.ac.id

6th Anharul Zikri
*Dept. of Informatics Engineering*
*Universitas Sriwijaya*
Palembang
0902182126034@student.unsri.ac.id

7th Wahyu Nugraha
*Dept. of Informatics Engineering*
*Universitas Sriwijaya*
Palembang
09021182126070student.unsri.ac.id

***Abstract*—In image processing, clustering techniques aim to group images into distinct classes. These methods utilize feature extraction, the first step toward pattern recognition, and analyze unprocessed data to derive valuable information. The impact of various feature extraction methods on the clustering outcomes is the focus of this research study. In this case, the issue to answer is: does the use of different feature extraction methods affect the effectiveness of clustering? The featured methods of extraction in this study are Histogram of Oriented Gradients, Local Binary Patterns, and HOG-LBP. These methods were used in conjunction with the Self-Organizing Map and K-means. The results show that LBP with K-Means gives exceptionally effective results with silhouette value 0.3615, a Davies-Bouldin index of 0.8128, and a Calinski–Harabasz index of 51940.5105. These results confirm the effectiveness of the extraction method used and the clustering algorithm. Like the other methods of extraction, the dependability of effectiveness centers on the capability of the feature selection algorithm's ability to differentiate the dataset. This is one of the feature extraction challenges for image clustering in large datasets – the focus is on feature extraction and the need for precise definition is critical.**

***Keywords—Feature Extraction, HOG, LBP, K-Means, Self-Organizing Maps***

## I. Introduction

Clustering is a sophisticated technique of data analysis with a myriad of uses such as in image processing. It is very helpful in image analysis as well as pattern recognition [1]. The image analysis process includes the image features extraction phase where the image features are detected and organized through pixels or objects within an image having shared characteristics.

The process of image analysis entails the identification and grouping of pixels or objects within an image that align in terms of characteristics. The use of specific processing techniques helps to improve understanding of the visual data received. Data being clustered is heavily reliant upon the level of detail which is extracted from an image. In order to yield meaningful representations, sharp focus features extracted from an image have to be of high quality and meaningfully represent the data's underlying structure. In this discipline, developing effective techniques for image feature extraction remains dominant as it has strong impact on the accuracy and precision of the clustering results [2], [3].

Defining feature extraction (FE) in images refers to extracting relevant information from images for further analysis, which may include grouping. This critical step involves utilizing Fourier transforms, intensity histograms, and texture descriptors like Local Binary Patterns (LBP). The Fourier transforms are particularly adept at analyzing frequency components [4], while intensity histograms offer insight into the distribution of pixel values. Additionally, LBP has demonstrated effectiveness in capturing the details of texture [5]. Typically, a single approach to feature extraction is inadequate, particularly in complex, heterogeneous image datasets. The use of a singular method, in this case, could be too blunt to appreciate the rich structure of the dataset, resulting in poor, even nonsensical analysis of data. Substantiated evidence provided suggests that many feature extraction techniques must be incorporated to improve image data analysis, thus making clustering the images more meaningful and consequential, statistically validating the outcomes [6].

To enhance the performance of clustering, it is necessary to implement a more advanced approach to extracting image features. Hybrid feature extraction methods which combine several techniques to overcome the strengths and weaknesses of each method seem to be a viable approach to solve this problem. The merging of various techniques yields more informative and relevant data regarding images [7]. This feature integration captures a more complete set of image features and improves the accuracy of clustering. This means that hybrid FE approaches can improve the effectiveness of clustering methods on complex and heterogeneous image data sets [8]. Nevertheless, the performance results, or whether other factors must be considered to optimize grouping results, especially on image data, are unclear.

## II. Literature Review

Eisha et al. [9] investigated a feature extraction method for multiple object detection using Histogram of Oriented Gradients (HOG) combined with Local Ternary Patterns (LTP). By integrating the HOG and LTP models, this feature extraction process aims to identify significant regions within images, thereby enhancing the classification accuracy by capturing both oriented and texture features. The proposed method achieved an accuracy of 92.48%, significantly outperforming the existing Multi-Object Detection and Tracking (MODT) method, which recorded an accuracy of 76.23% in detecting multiple objects. Marlen et al. utilize only Principal Component Analysis (PCA) to reduce data dimensionality while retaining the most critical variables from each category. The principal components represent linear combinations of the variables in the tuberculosis clustering data. Sujan et al. [10]. They integrated two feature extraction techniques, principal component analysis and linear discriminant analysis. The resulting features are then utilized by both linear and kernel Support Vector Machines (SVM) to classify attention patterns. The classification results are compared against those achieved using linear and kernel SVM. The results show that the hybrid approach with both techniques integrated outperform others with the greatest scores in accuracy and other metrics precision, recall, F1, and kappa.

This study builds upon earlier research because, based on the results, there appears to be a lack of investigation into how combined multi-feature extraction is implemented in clustering analysis, especially with images. The current study examines the effect of using single and hybrid feature extraction using LBP and HOG on clustering performance. The hypothesis is that the synergy between LBP's local texture detail resolution and HOG's shape and edge capturing ability provides a positive contribution to the accurate clustering results [11].

Both the Self-Organizing Map (SOM) and K-Means algorithms will be utilized to evaluate the effectiveness of a clustering algorithm through the use of image dataset. K-Means is a widely-used clustering algorithm that partitions data into K clusters based on centroid distances, optimized iteratively [12], [13]. This algorithm is known for its simplicity and efficiency in handling large datasets. In contrast, SOM is a type of artificial neural network that uses unsupervised learning to create a two-dimensional representation of high-dimensional input data, facilitating easier visualization and analysis [14], [15]. These two algorithms will be assessed with multiple tests based on the metrics silhouette score, Davies-Bouldin index, and Calinski-Harabasz index [13]. The use of these metrics allows quantifiable comparisons regarding the performance of SOM and K-Means concerning data partitioning, aiding in understanding the data's structure and each algorithm's advantage in differing situations.

## III. Research Method

This study is multi-phased with the first phase being a pre-processing stage which incorporates image rescaling, converting to grayscale, histogram equalization, and normalization. Following the completion of data processing, feature extraction is achieved through multiple methods such as HOG, LBP, and HOG with LBP. The data is then trained with classical techniques such as K-Means and Self-Organizing Map (SOM). The last step is analyzing the model performance and its metrics involving the silhouette score, Davies-Bouldin index, and Calinski-Harabasz index. The research workflow is shown in Fig. 1.

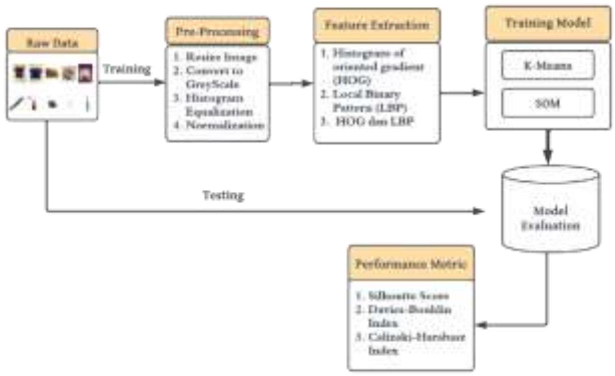



Fig. 1. Research Workflow.

### A. Data Acquisition

This information was gathered from Kaggle under the "garbage—classification—v2" dataset. This dataset contains over 23642 images classified into batteries, biological waste, cardboard, clothing, glass, metal, paper, plastic, shoes, and general waste. This dataset is described in detail in the overview Table 1, which shows the distribution and number of images per category.The Amount of Data in Each Class of the Dataset

TABLE I. DATASET DETAIL

| Garbage Category | Amount |
|---|---|
| Metal | 1869 |
| Glass | 4097 |
| Biological | 985 |
| Paper | 2727 |
| Battery | 945 |
| Trash | 834 |
| Cardboard | 2341 |
| Shoes | 1977 |
| Clothes | 5325 |

The dataset "garbage-classification-v2" has a wide range of image sizes, varying from small to large. The differences in resolution and size reflect the dataset's variation of images. This makes it challenging to process the data, and thus a pre-processing step is required. The pre-processing steps involve changing the size of the images, changing them to grayscale, performing histogram equalization, and normalizing the data. These steps are crucial in creating a dataset that meets the requirements for analysis. Such uniformity ensures that accuracy in feature extraction and further machine learning processes is enhanced.

### B. Data Pre-Processing

As part of every image analysis workflow, data preprocessing and the corresponding activities must be executed to augment an image's quality before presenting it for feature extraction. The steps involve resizing, grayscale conversion, histogram equalization, the addition of Gaussian noise, and normalization. These steps help prepare the image for accurate and efficient analysis during the other more advanced stages of the processing workflow. Image resizing is changing the dimensions and proportions of an image by either increasing or decreasing the height and width. In simpler terms, standardizing the dimensions of images to be processed for easy analysis [16]. During this process step, the image proportions are set to ensure that the image will not be distorted. In this study, all images were resized to a lower resolution of 128 by 128 pixels to be easily processed during the next steps.

This change replaces the color details of the image with brightness and intensity levels. The process transforms the picture into a black and white image (greyscale) which consists of pixel values from zero for black to 255 for white on an 8-bit scale. [17]Removal of a color component lowers an image's complexity while preserving its information. For example, grayscale images are easier to analyze than colored images in edge detection, image segmentation, and feature extraction tasks.

Normalization is a method in image processing that deals with rescaling specific values to enhance the image between ranges of 0-1 or 0-255. The primary purpose is to improve the contrast of the image, along with normalizing it for further analysis or processing. Normalization has many benefits, including eliminating or reducing the impacts caused by differences in illumination, thus making matching or blending images from diverse sources much more reliable. Within the field of image processing, blending images with varying degrees of intensity owing to illumination differences poses a challenge. Standardizing pixel intensities across images solves this challenge.

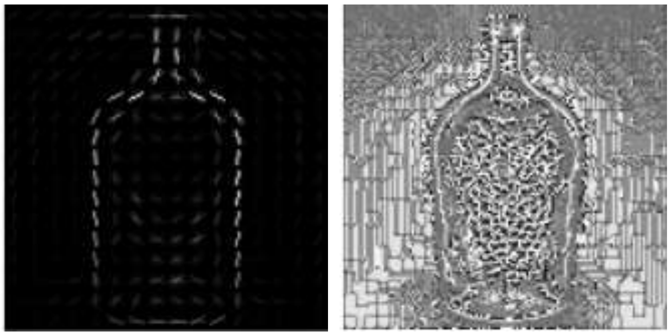
The resizing of the image follows soon after and is arguably the finest step to take in the analysis. The step reduces the number of features contained in the data set while keeping the information intact. The raised dimension decreased the model’s complexity; therefore, the efficiency and effectiveness of the analysis algorithm were improved. In this work, the reduction of dimensions was performed using the principal component analysis (PCA) method [18]. This allows the normalized image to be stored in a lower feature space while keeping the image data's structure. This improves the effectiveness of the further analysis, which includes estimating, clustering, and classification.

### C. *Feature Extraction*

Image processing involves extracting salient components of an image serving a specific purpose which may involve edges, shapes, color and texture. These features can be useful in segmentation, pattern recognition, clustering, classification, and many more domains. With the extraction of features from an image, algorithms can deal with the most significant portions of the image efficiently and perform tasks on them which, in tern, increases the accuracy of operations performed on images [19].

Histograms of oriented gradients (HOG) is a method of feature extraction used in image processing. It finds applications in pattern recognition and object detection within images. The HOG technique computes the measure of an object’s features by determining the gradient distribution i.e., measuring the change in pixel intensity in different parts of the picture. The image makes it possible to determine the direction and the magnitude of the gradient. Thus the HOG technique acquires data that is related to the shape and the structure of various objects in an image [20].

Local Binary Patterns (LBP) are a feature of image processing used for analysis of textures of images. This method is based on the evaluation of texture patterns through the analysis of the intensity values of pixels relative to the intensity values of the surrounding pixels. The comparison generates a corresponding binary code which illustrates the local structures in an image [5]. Fig. 2 shows the result of the feature extraction of the sample image in the glass class.



(a) (b)

Fig. 2. The image extraction feature in the glass class. (a) HOG feature extraction, (b) LBP feature extraction.

A blend of HOG and LBO techniques after the extraction of image features creates a unified representation, which enhances the depiction of images visually by using the strength of both extraction methods. The incorporation of shape and structural information captured by HOG alongside the fine detail textures recognized by LBP increases image representation. The integration of these features leads to abundant representation, which is useful in clustering, segmentation, and object classification. The results of this integration are demonstrated in Fig. 3.

```
[[0.02855194 0.         0.26826727 ... 0.05957031 0.22619629 0.06567383]
 [0.03375177 0.04482515 0.29275048 ... 0.08233643 0.13183594 0.10235596]
 [0.         0.         0.         ... 0.0269165  0.70318604 0.03808594]
 ...
 [0.12143355 0.1170457  0.14769375 ... 0.08575439 0.13708496 0.16412354]
 [0.         0.         0.         ... 0.06329346 0.32226562 0.05792236]
 [0.         0.         0.         ... 0.03094482 0.71026611 0.02539062]]
```

Fig. 3. The array of results obtained from the hybrid feature.

### D. *Clustering*

Clustering has always been an important concept in machine data analysis. The aim is to divide a set of objects such that all items within each group are more or less similar. This research applied K-Means and Self Organizing Maps (SOM) clustering techniques. K-Means is one of the methods that classifies information into a predetermined number of clusters by measuring distances, and subsequently attempts to optimize the division to minimize variance within each cluster [21], [22]. Self-organizing maps or SOMs is an approach based on artificial neural networks that project data with high dimensionality into lower dimensionality while retaining the topological features of the data [23]. The basis for combining these two techniques is the expectation of obtaining a more in-depth clustering analysis.

### E. *Performance Metric*

The metrics that will be used in benchmarking the clustering performance from K-Means and Self-Organizing Maps (SOM) techniques are the Silhouette Score, Davies Bouldin Index, and Calinski Harabasz score [13], [24]. This score measures the degree to which an object is similar to its cluster as opposed to other clusters which indicates the quality of clustering. Calculating the distance to every single object within a cluster and the distance between cluster components yields the ratio of the Davies Bouldin Index. Even higher values of the Silhouette score are given when more members are included within a cluster and the Calinski Harabasz index measures the totality of variability that exists within the cluster as opposed to the totality of the variability of the clusters forming which makes the process of clustering more efficient.

## IV. RESULT AND DISCUSSION

The data outlined in Table 2 confirms that the Local Binary Pattern (LBP) feature extraction technique yields better clustering outcomes for both K-Means and Self-Organizing Map (SOM) methodologies. This achievement is apparent in the analysis's use of various evaluation metrics:

- Silhouette Score - Silhouette coefficient is a statistical measure that captures intra-cluster cohesion and inter-cluster separation. A high Silhouette coefficient suggests that LBP-generated clusters have strong intra-cluster cohesion and low inter-cluster separation.
- Davies-Bouldin Index - The highest internal similarity and substantial inter-cluster distance is characteristic of LBP-generated clusters which explains the lower Davies-Bouldin index value.
- Calinski-Harabasz Index - Compact and well-separated clusters produced by LBP are validated by the highest Calinski-Harabasz index.

The evaluation from different aspects and conditions confirms that the LBP feature extraction method, versus competing methods, consistently outperforms in improving the clustering process. This confirms the effectiveness of LBP, especially when used with K-Means and SOM algorithms.

This research pinpointed the best model as the K-means algorithm integrated with LBP feature extraction. This choice was based on the Calinski-Harabasz index value which was notably greater than that of other models. This value is highly suggestive of well-formed clusters with enough internal density and separability, qualities denoting high-performance clustering. The K-Means algorithm with the LBP feature extraction method was particularly high in terms of visibility and cluster formation compared to other combinations in this research study.

TABLE II. PERFORMANCE METRICS OF THE MODELS

| Algorithm | Feature Extraction Method | Silhouette Score | Davies-Bouldin Index | Calinski-Harabarz Index |
|---|---|---|---|---|
| K-Means | HOG | 0.3326 | 0.852 | 20428.9489 |
| K-Means | LBP | 0.3615 | 0.8128 | 51940.5105 |
| K-Means | HOG+LBP | 0.3344 | 0.8532 | 20452.4623 |
| SOM | HOG | 0.2972 | 0.9497 | 17597.5445 |
| SOM | LBP | 0.3687 | 0.8191 | 46465.6623 |
| SOM | HOG+LBP | 0.3177 | 0.9034 | 17228.1758 |

Further analysis of these results raises some issues that help explain why this combination works well. First, K-Means clustering is one of the well-known clustering algorithms designed for the data that do not have any prior classification. K-Means clustering divides multi-dimensional data into a certain number of clusters and minimizes the variance for each cluster. For that reason, K-Means algorithm remarkably partitions the data into its salient features and relationships which is often very difficult using other clustering methods. Second, local binary pattern LBP texture extraction algorithm is well known for its capability capturing area of interest within the image. This technique calls for transforming every pixel in the image into a binary value based on the values of the neighboring pixels to create a powerful descriptor that is invariant to changes in lighting.. LBP extracts the features that distinguish between different classes of rubbish when applied to the “garbage-classification-v2” dataset. The incorporation of K-Means and LBP leads to a greater combined effect than using either of the methods individually. The combined effect is exhibited in the visualization results of K-Means + LBP as shown in Figure 4 where the approach results in distinct and well-separated clusters.

Fig 4. shows the results of clustering using the K-Means algorithm which is visualized in two-dimensional space via Principal Component Analysis (PCA). PCA is used to reduce the dimensions of data so that it can be visualized in two dimensions. This helps in seeing the structure and separation of the clusters formed. Although PCA only preserves some of the variability of the original data, the results still provide a fairly good picture of how the data is clustered. These clustering results show the effectiveness of K-Means in finding data structures that are not directly visible in high-dimensional data. K-Means can separate data into several clusters well, although some data points from different clusters appear close together, which may indicate K-Means' limitations in handling overlap between clusters.


Fig. 1. The Clustering of the "garbage-classification-v2" dataset Using the K-Means Algorithm and Feature Extraction LBP.

From the illustration, it can be seen that the K-Means algorithm accurately divides the data into 9 distinct clusters. The distinct presence of these clusters proves that the algorithm has achieved its purpose on the patterns and structural components of LBP. It further indicates that the application of K-Means in conjunction with LBP as a feature extraction technique has produced well-defined clusters as confirmed by the color distribution within the plot. Figure 4. illustrates the importance of employing methods like PCA for the visualization of results obtained from clustering high-dimensional data, aiding in a better understanding of the clustering results.

This study raises a point on the need to select a feature extraction technique for computer clustering which is an enhancement in quality. The feature extraction method implemented influences the outcome and the reliability of the clustering results. The results obtained using the Local Binary Pattern (LBP) feature extraction method in this study enabled salient features of the data to be extracted, and therefore resulted in more recognizable and well separated clusters. The application of a multitude of features with the use of hybrid methods tends to create redundancy and noise that can overlap with the cluster boundaries and reduce the accuracy of the clustering. This research helps the domain by showing the need to invest attention and resources in the selection and analysis of feature extraction methods for improving image processing analysis results.

## V. Result and Discussion

In this study, the single-feature extraction method was found to perform better in clustering images than the hybrid approach. This finding indicates that the integration of Local Binary Patterns (LBP) and the K-Means clustering algorithm is the most effective approach among those evaluated. This conclusion was determined by calculating a silhouette score of 0.3615, a Davies-Bouldin index of 0.8128, and a Calinski-Harabasz index of 51,940.5105. These values indicate that the combination of LBP and K-Means has a higher intra-cluster similarity and inter-cluster distance value than the features obtained by the hybrid feature extraction approach. This observation lends support to the notion that LBP execution is characterized by its ability to encapsulate predominant structural texture features, which are indispensable for ensuring the reliability of the clustering process. Additionally, the findings underscore the necessity for caution when employing feature extraction methods intended for a particular clustering algorithm, as these methods can significantly influence the outcome of the clustering process. Further research could apply single feature extraction techniques or hybrids tailored to specific datasets to validate or optimize the results of the research.

## References


[1] Dr. N. Gomathi and A. Geetha, 'An Efficient Hybrid Clustering and Feature Extraction Techniques for Brain Tumor Classification', Webology, vol. 18, no. 2, pp. 556–581, Dec. 2021, doi: 10.14704/web/v18i2/web18339.

[2] A. Stephan et al., 'Text-Guided Image Clustering', Long Papers, 2024. [Online]. Available: https://github.com/AndSt/

[3] Y. Li, P. Hu, D. Peng, J. Lv, J. Fan, and X. Peng, 'Image Clustering with External Guidance', Oct. 2023, [Online]. Available: http://arxiv.org/abs/2310.11989

[4] M. M.ansal, 'A Review Of Research Papers On Fourier Transforms & Statistical Fourier Analysis', 2018. [Online]. Available: www.jetir.org

[5] S. Alkhazali and M. El-Bashir, 'Local Binary Pattern Method (LBP) and Principal Component Analysis (PCA) for Periocular Recognition', IJACSA) International Journal of Advanced Computer Science and Applications, vol. 11, no. 8, pp. 70–76, 2020, doi: 10.14569/IJACSA.2020.0110810.

[6] W. K. Mutlag, S. K. Ali, Z. M. Aydam, and B. H. Taher, 'Feature Extraction Methods: A Review', in Journal of Physics: Conference Series, IOP Publishing Ltd, Aug. 2020. doi: 10.1088/1742-6596/1591/1/012028.

[7] M. Azzeh and A. B. Nassif, 'A hybrid model for estimating software project effort from Use Case Points', Applied Soft Computing Journal, vol. 49, pp. 981–989, 2016, doi: 10.1016/j.asoc.2016.05.008.

[8] J.-S. Dessureault and D. Massicotte, 'Feature selection or extraction decision process for clustering using PCA and FRSD', Nov. 2021, [Online]. Available: http://arxiv.org/abs/2111.10492

[9] E. Akanksha and P. R. K. Rao, 'A Feature Extraction Approach for Multi-Object Detection Using HoG and LTP', International Journal of Intelligent Engineering and Systems, vol. 14, no. 5, pp. 259–268, 2021, doi: 10.22266/ijies2021.1031.24.

[10] S. Poudyal, M. J. Mohammadi-Aragh, and J. E. Ball, 'Hybrid Feature Extraction Model to Categorize Student Attention Pattern and Its Relationship with Learning', Electronics (Switzerland), vol. 11, no. 9, pp. 1–23, May 2022, doi: 10.3390/electronics11091476.

[11] N. Sabri, H. N. A. Hamed, Z. Ibrahim, K. Ibrahim, M. A. Isa, and N. M. Diah, 'The hybrid feature extraction method for classification of adolescence idiopathic scoliosis using Evolving Spiking Neural Network', Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 10, pp. 8899–8908, Nov. 2022, doi: 10.1016/j.jksuci.2022.08.019.

[12] S. Mulyaningsih and J. Heikal, 'K-Means Clustering Using Principal Component Analysis (PCA) Indonesia Multi-Finance Industry Performance Before and During Covid-19', Asia Pacific Management and Business Application, vol. 011, no. 02, pp. 131–142, Dec. 2022, doi: 10.21776/ub.apmba.2022.011.02.1.

[13] I. Firman Ashari, E. Dwi Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, 'Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta', 2023. doi: https://doi.org/10.30871/jaic.v7i1.4947.

[14] P. Bashivan, M. Schrimpf, R. Ajemian, I. Rish, M. Riemer, and Y. Tu, 'Continual Learning with Self-Organizing Maps', Apr. 2019, [Online]. Available: http://arxiv.org/abs/1904.09330

[15] M. Cottrell, M. Olteanu, F. Rossi, and N. Villa-Vialaneix, 'SELF-ORGANIZING MAPS, THEORY AND APPLICATIONS', 2018.

[16] H. Talebi, P. Milanfar, and G. Research, 'Learning to Resize Images for Computer Vision Tasks', in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 487–496. doi: 10.1109/ICCV48922.2021.00055.

[17] I. Zeger, S. Grgic, J. Vukovic, and G. Sisul, 'Grayscale Image Colorization Methods: Overview and Evaluation', IEEE Access, vol. 9, no. 1, pp. 113326–113346, 2021, doi: 10.1109/ACCESS.2021.3104515.

[18] S. Mishra et al., 'Principal Component Analysis', International Journal of Livestock Research, p. 1, 2017, doi: 10.5455/ijlr.20170415115235.

[19] J. Prasad and M. Sreelatha, 'Feature Extraction And Clustering Techniques on Remote Sensing Images-A Survey', International Journal of Recent Technology and Engineering (IJRTE), vol. 7, no. 6S, 2019.

[20] M. Aminul Islam, M. Sayeed Iftekhar Yousuf, and M. M. Billah, 'Automatic Plant Detection Using HOG and LBP Features With SVM', International Journal of Computer, 2019, [Online]. Available: http://ijcjournal.org/

[21] Y. Sopyan, A. D. Lesmana, and C. Juliane, 'Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan', Building of Informatics, Technology and Science (BITS), vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2697.

[22] K. Balabaeva and S. Kovalchuk, 'Post-hoc Interpretation of Clinical Pathways Clustering using Bayesian Inference', Procedia Comput Sci, vol. 178, no. 2019, pp. 264–273, 2020, doi: 10.1016/j.procs.2020.11.028.

[23] A. H. Primandari, A. Kesumawati, and T. Purwaningsih, 'Supporting of Waste Management in Indonesia Using Self Organizing Map for Clustering Analysis', in Journal of Physics: Conference Series, IOP Publishing Ltd, Apr. 2021. doi: 10.1088/1742-6596/1863/1/012072.

[24] M. Gaido, 'Distributed Silhouette Algorithm: Evaluating Clustering on Big Data', Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.14102