Jurnal Computech & Bisnis, Vol. 20, No. 1, June 2026, 50-57
ISSN (print): 1978-9629, ISSN (online): 2442-4943

Anthropometry: Palm Size Clustering Using the K-Means Method and Silhouette Coefficient

Cucut Hariz Pratomo[1], Anton Yudhana[2], Herman[3]
Universitas Ahmad Dahlan [1,2,3]
2537083012@webmail.uad.ac.id[1], eyudhana@s3difa.uad.ac.id[2], herman@s3difa.uad.ac.id[3]

Abstract

This study uses a palm image dataset to be used as training data and test data and the software used is the Python Programming Language and the OpenCV library which is quite light in processing on Google Collaboratory. The image data used is a palm image dataset from Kaggle which is open access to the public, the image data is then subjected to feature extraction using the Canny Edge Detection method with Euclidean distance calculations to see the edges of the object lines in the training image to clearly obtain the hand object that will be classified to see anthropometrically whether the palm is small, medium, and large. Grouping of 3 classes of palm size uses the K-Means method combined with the Silhouette Score to evaluate the results obtained. From a successful experiment, the Silhouette results show a score of 0.61, which means good..

Keyword: *Palm, Object Detection, Clustering, K-Means, Silhouette.*

INTRODUCTION

Anthropometry, the science of measuring the physical dimensions of the human body (Permana Ratumanan & Feinisa Khairani, 2023), is a crucial element in various industrial fields, from ergonomic design of work equipment and fashion product development to biometric security systems. One body part with high morphological variability yet highly functional is the palm of the hand (Oanta, 2024). For example, how the growth of finger length is influenced by prenatal hormones, where generally the ring finger (4D) will be longer than the index finger (2D) in men due to the influence of the testosterone hormone during fetal development, conversely in women, due to the influence of the estrogen hormone, the index finger (2D) will tend to be longer than the ring finger (4D) (South et al., 2023).

Previous research has shown that palm images can be used as an alternative biometric option besides fingerprints because they can be extracted using various methods such as KNN, ANN, Decision Tree, and even SVM with high accuracy results (Huang & Handhayani, 2025). The widespread use of palm biometrics has been studied in palm biometric-based CCTV movement detection using the Naive Bayes method, which obtained high accuracy results, (Yanuar Ilham et al., 2023) and can easily be recognized in research by using the CNN method (Wita & Liliana, 2022).

Clustering is a crucial technique in data analysis, aiming to group objects based on similarities in certain features(Firmansyah et al., 2025). The K-Means method is also known as the most popular clustering algorithm due to its simplicity and effectiveness. In the context of palm size measurement, K-Means is used to group palm size data into several clusters representing different size categories, facilitating analysis and practical applications such as ergonomics or product design, determining target markets, and so on. This method works by grouping data based on the closest distance to the cluster centroid. The optimal number of clusters (k) can be selected using the silhouette coefficient technique to improve cluster quality (Ruswanti et al., 2024; Yadav & Sharma, 2024).

Previous research has shown that optimizing K-Means, for example using purity methods or unsupervised approaches that automatically determine the number of clusters, can improve clustering efficiency and accuracy. Furthermore, K-Means has been successfully applied in various fields, such as clustering rice leaf diseases, clustering disaster-prone maps, and clustering customer behavior data (Gina Regiana et al., 2025; Hartama et al., 2024; Syahputri et al., 2024; Tan et al., 2025).

Based on our previous research, we ultimately propose this study, which aims to implement the K-Means algorithm for clustering palm sizes to generate more precise, data-driven size categories. The results of this clustering are expected to serve as a reference for standardizing more ergonomic product sizes or as a baseline for a hand anatomy-based identification system.

METHOD

To break this problem, a computational approach is needed that is able to automatically group palm size data based on their characteristic similarities (Humaira & Rasyidah, 2020). The K-Means Clustering algorithm is one of the unsupervised learning methods that is very effective in handling this kind of data grouping problem.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \|x_i^{(j)} - c_j\|^2$$

The K-Means method works by dividing data into k groups (clusters) based on the closest distance between the data and the cluster center point (centroid). By applying K-Means, researchers can find hidden patterns in the distribution of population hand sizes without requiring initial data labels (Khan et al., 2025; Zubair et al., 2024).

Systematic steps taken to process palm size data into groups (clusters) :

1. Data Acquisition
   The first step is to collect palm anthropometric data. This data can be obtained through manual measurements or through feature extraction from digital images. However, the researchers sourced the dataset from Kaggle: *Hands and Palm Images dataset*.
2. Pre-Processing Data
   Raw data typically has different scales in units (e.g., length in cm, area in pixels). Therefore, data normalization (such as Min-Max Scaling or Z-Score) is necessary to prevent any single variable from becoming dominant in distance calculations simply because it has a larger range of values.

In data pre-processing section, image feature extraction will be using canny edge detection method. The Canny edge detection algorithm is composed of 5 steps:

1) Noise reduction;
2) Gradient calculation;
3) Non-maximum suppression;
4) Double threshold;
5) Edge Tracking by Hysteresis.

Important thing to remember, is that the algorithm is based on grayscale pictures. Therefore, the pre-requisite is to convert the image to grayscale before following the above-mentioned steps.

3. Implementing K-Means
   i. Determine the Number of Clusters (k). Determine the desired number of size categories (e.g., 3 for S, M, L).
   ii. Initialize Centroids. Randomly select k points as the initial cluster centers (centroids).
   iii. Allocate Data to the Nearest Clusters. For each palm data item, the distance to each centroid is calculated using the Euclidean Distance formula:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

   Data is assigned to the cluster with the shortest distance.
   iv. Centroid Update: Recalculate the centroid position based on the average position of all data within the cluster.
   v. Iteration: Steps 3 and 4 are repeated continuously until the centroid position no longer changes (convergence) or reach the maximum iteration.
4. Evaluate
   To ensure that the clustering results are optimal, the Silhouette Coefficient evaluation method is used: Measuring how similar data is to its own cluster compared to another clusters (Punhani et al., 2022). A score close to 1 indicates good clustering, 0 indicates overlap, and -1 indicates that the data should probably be in another cluster. This includes silhouette plot visualization.

RESULTS AND DISCUSSION

When the drive is mounted, the image data will be read, and the training data will be read in RGB image format. Next, the image will be segmented by converting the RGB color format to grayscale. After the image data has been successfully converted to grayscale, the foreground and background will be separated. Object edges will be detected using the Canny Edge Detection method. and produces an image as shown in Figure 1.

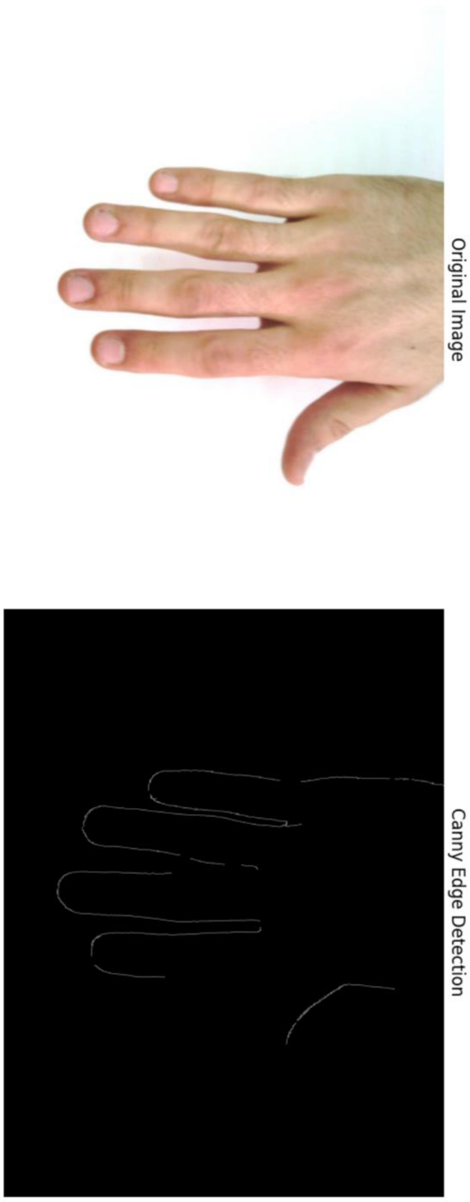

Figure 1. Feature Extraction

With the scikit-learn and pandas libraries, the object has been captured and then clustering will be carried out using the K-Means method, then to display the visualization results, the matplotlib library is used, which is capable of displaying visualization results in the form of scatter plots, as shown in Figure 2.

Figure 2 shows a scatterplot depicting the distribution of palm print image data into categories: Large (Red), Medium (Green), and Small (Violet).

The Large and Medium clusters are located close together, indicating a high degree of similarity between the two clusters. The Small cluster is located far away, indicating significant differences between the two classes.

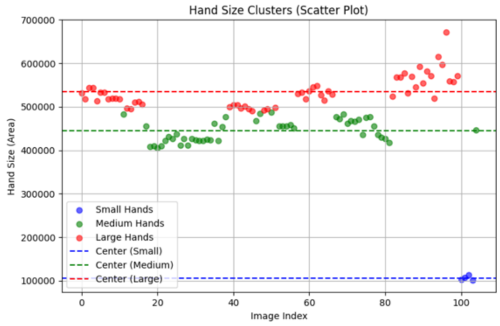

Figure 2. Cluster Plot

The details of figure 2 stands these cluster centers were mapped to descriptive categories: the cluster with a center of 106,744.75 was labeled 'Small' that captured at figure 3, 445,204.13 was labeled 'Medium' shown at figure 4, and 535,008.19 was labeled 'Large' at figure 5.

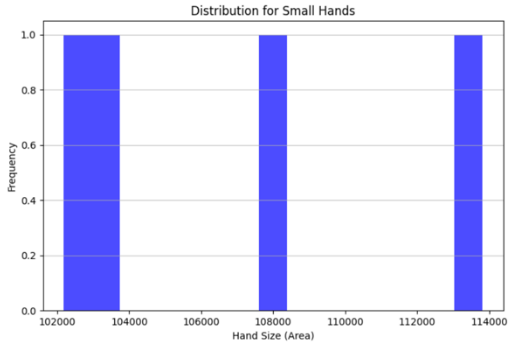

Figure 3. Small Plot

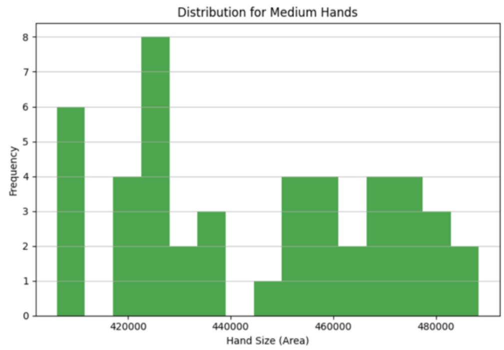

Figure 4. Medium Plot

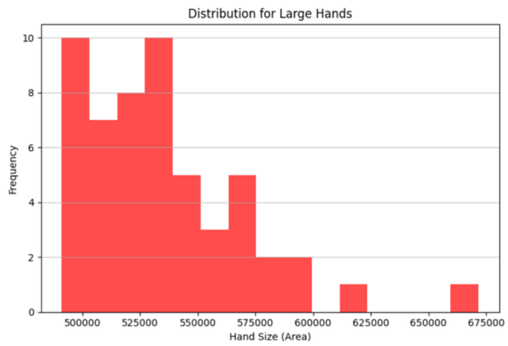

Figure 5. Large Plot

Figure 6. Image Cluster

Figure 6 shows an example of an original image that corresponds to the clusters formed using the K-Means method. Using the Silhouette Score evaluation method, the results obtained will appear as in Figure 7.

```
from sklearn.metrics import silhouette_score

# Calculate the Silhouette Score
silhouette_avg = silhouette_score(hand_sizes_reshaped, cluster_labels)

print(f"The average Silhouette score is: {silhouette_avg:.2f}")
```

```
The average Silhouette score is: 0.61
```

Figure 7. Silhouette Score

CONCLUSION

By using the Silhouette Score evaluation method, the results obtained will appear as in Figure 7. The average Silhouette score of 0.61 indicates that the clustering is of good quality. The clusters are well-separated, and the data points within each cluster are relatively compact and similar to each other. A score of 0.61 suggests a strong degree of separation and cohesion among the 'small', 'medium', and 'large' hand size categories.

For further research, it is recommended to collect primary images from respondents who truly have a distinct differentiation in terms of small, medium, and large hand sizes. Given consistent lighting and distance conditions, the dataset we used is sufficiently relevant.

REFERENCE

Firmansyah, M. I., Kustiyahningsih, Y., Rahmanita, E., Abidin, M. S., & Satoto, B. D. (2025). Optimization of MSMEs Clustering in Sampang District Using K-Medoids Method and Silhouette Coefficient Method. *Teknika*, *14*(1), 1–8. https://doi.org/10.34148/teknika.v14i1.1116

Gina Regiana, Irma Purnamasari, A., Bahtiar, A., & Tohidi, E. (2025). K-Means Algorithm to Improve Leaf Image Clustering Model for Rice Disease Early Detection. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, *4*(2), 1156–1160. https://doi.org/10.59934/jaiea.v4i2.840

Hartama, D., Wanayumini, W., & Damanik, I. S. (2024). Pengelompokan Algoritma K-Means dan K-Medoid Berdasarkan Lokasi Daerah Rawan Bencana di Indonesia dengan Optimasi Elbow, DBI, dan Silhouette. *Building of Informatics, Technology and Science (BITS)*, *6*(2). https://doi.org/10.47065/bits.v6i2.5851

Huang, J., & Handhayani, T. (2025). KLASIFIKASI TELAPAK TANGAN MENGGUNAKAN DEEP LEARNING DAN MACHINE LEARNING. *Jurnal Mnemonic*, *8*(2), 293–300. https://doi.org/10.36040/mnemonic.v8i2.15716

Humaira, H., & Rasyidah, R. (2020). Determining The Appropiate Cluster Number Using Elbow Method for K-Means Algorithm. *Proceedings of the Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia*. https://doi.org/10.4108/eai.24-1-2018.2292388

Khan, I., Daud, H. B. D., Zainuddin, N. binti Z., Sokkalingam, R. S., Azad, A. S. A., Samad, A., & Suleiman, A. A. S. (2025). Assessing the Efficiency and Accuracy of K-Means Clustering Compared to Other Clustering Techniques. *Data Science Insights*, *3*(2), 49–65. https://doi.org/10.63017/jdsi.v3i2.23

Oanta, E. M. (2024). K-means Clustering Solution for Image Compression. *2024 Advanced Topics on Measurement and Simulation (ATOMS)*, 248–251. https://doi.org/10.1109/ATOMS60779.2024.10921537

Permana Ratumanan, S., & Feinisa Khairani, A. (2023). *Metode Antropometri Untuk Menilai Status Gizi*. https://myjurnal.poltekkes-kdi.ac.id/index.php/hijp

Punhani, A., Faujdar, N., Mishra, K. K., & Subramanian, M. (2022). Binning-Based Silhouette Approach to Find the Optimal Cluster Using K-Means. *IEEE Access*, *10*, 115025–115032. https://doi.org/10.1109/ACCESS.2022.3215568

Ruswanti, D., Prayitno, I. J., & Al Haris, F. H. S. (2024). Optimizing Cluster Methods:

Combining K-Means with Hierarchical Techniques for Better Results. *2024 6th International Conference on Cybernetics and Intelligent System (ICORIS)*, 1–5. https://doi.org/10.1109/ICORIS63540.2024.10903822

South, A. J., Barkus, E., Walter, E. E., Mendonca, C., & Thomas, S. J. (2023). Dark Triad personality traits, second-to-forth digit ratio (2D:4D) and circulating testosterone and cortisol levels. *Biological Psychology*, *179*, 108567. https://doi.org/10.1016/j.biopsycho.2023.108567

Syahputri, Z., Sutarman, S., & Siregar, M. A. P. (2024). Determining The Optimal Number of K-Means Clusters Using The Calinski Harabasz Index and Krzanowski and Lai Index Methods for Groupsing Flood Prone Areas In North Sumatra. *Sinkron*, *9*(1), 571–580. https://doi.org/10.33395/sinkron.v9i1.13246

Tan, E. H. L., Hamed, Y., Daud, H., Abdul Wahab, M. A. F., Azhar, A. A. A., & Tan, S. Y. (2025). Profiling investor behavior in the Malaysian derivatives market using K-means clustering. *Frontiers in Artificial Intelligence*, *8*. https://doi.org/10.3389/frai.2025.1640776

Wita, D. S., & Liliana, D. Y. (2022). Klasifikasi Identitas Dengan Citra Telapak Tangan Menggunakan Convolutional Neural Network (CNN). *Jurnal Rekayasa Teknologi Informasi (JURTI)*, *6*(1), 1. https://doi.org/10.30872/jurti.v6i1.7100

Yadav, S., & Sharma, Dr. S. (2024). Study Of Existing Methods & Techniques Of K-Means Clustering. *Educational Administration: Theory and Practice*, 1806–1813. https://doi.org/10.53555/kuey.v30i4.1755

Yanuar Ilham, M., Wulanningrum, R., Nur Farida, I., & Ayu Dunia Widyadara, M. (2023). PENERAPAN SISTEM CERDAS KLASIFIKASI CITRA TELAPAK TANGAN DENGAN METODE NAÏVE BAYES. *Jurnal Simantec*, *11*(2), 139–146. https://doi.org/10.21107/simantec.v11i2.14708

Zubair, Md., Iqbal, MD. A., Shil, A., Chowdhury, M. J. M., Moni, M. A., & Sarker, I. H. (2024). An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling. *Annals of Data Science*, *11*(5), 1525–1544. https://doi.org/10.1007/s40745-022-00428-2