

Research Article

Faults Detection for Photovoltaic Field Based on K-Means, Elbow, and Average Silhouette Techniques through the Segmentation of a Thermal Image

Abdelilah Et-taleby , Mohammed Boussetta , and Mohamed Benslimane 

Innovative Technologies Laboratory Université Sidi Mohamed Ben Abdellah Fez, Morocco

Correspondence should be addressed to Mohammed Boussetta; mohammed.boussetta@usmba.ac.ma

Received 29 October 2020; Accepted 21 November 2020; Published 4 December 2020

Academic Editor: Dhruva B. Khadka

Copyright © 2020 Abdelilah Et-taleby et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Clustering or grouping is among the most important image processing methods that aim to split an image into different groups. Examining the literature, many clustering algorithms have been carried out, where the K-means algorithm is considered among the simplest and most used to classify an image into many regions. In this context, the main objective of this work is to detect and locate precisely the damaged area in photovoltaic (PV) fields based on the clustering of a thermal image through the K-means algorithm. The clustering quality depends on the number of clusters chosen; hence, the elbow, the average silhouette, and NbClust R package methods are used to find the optimal number K. The simulations carried out show that the use of the K-means algorithm allows detecting precisely the faults in PV panels. The excellent result is given with three clusters that is suggested by the elbow method.

1. Introduction

Solar energy source is considered one of the most important energy sources, which has attracted considerable attention worldwide because it provides clean, reliable, and unlimited power. Furthermore, solar energy receives significant investments to develop and improve the productivity of the solar panels, which was evaluated for \$131.1 billion in 2019 [1].

Solar energy is captured using photovoltaic panels; these latter present several faults and anomalies that influence the production of the PV systems. On this way, several techniques have been proposed in many works in the literature to ensure reliable and efficient PV operation; these techniques are mainly split into two categories: electrical methods and nonelectrical methods.

Electrical methods are mostly based on I-V characteristics analysis [2], power losses analysis (PLA) [3], statistical and signal processing approaches (SSPA) [4], etc. On the other hand, the nonelectrical techniques are based on infrared thermography, which has known increasing use to con-

trol the renewable energy system over recent years [5–7]. These techniques are based on the use of thermal cameras and drones in order to detect the hottest areas in the PV field, as shown in Figure 1.

In this context, the applications of the machine learning algorithms to detect and locate the damaged area precisely through the segmentation of a thermal image are investigated.

Segmentation, clustering, or grouping of data is an unsupervised machine learning technique that aims for the partition of data into many groups based on the similarity of properties (e.g., color, size, and shapes). This technique has been widely studied in the literature in a wide range of fields over recent years [8–11]. Therefore, several algorithms have been proposed, such as the K-means algorithm (KMA) [12], the Fuzzy C Means algorithm (FCM) [13], and the Mean Shift algorithm (MSA) [14]. Among them, the K-means method is acknowledged as the simplest, fastest, and most popular clustering method. In spite of having these features, K-means has a disadvantage related to the number of

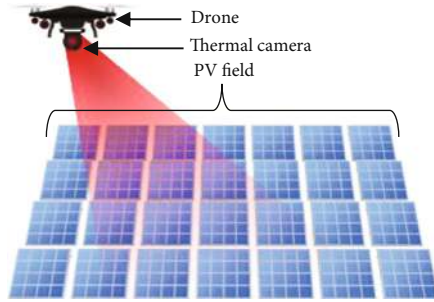


FIGURE 1: Control photovoltaics field using thermal camera on drone.

clusters K that must be optimal to get excellent results. To address this issue, numerous methods are proposed, such as elbow method (EM) [15], average silhouette method (ASM) [16], and gap statistic method (GSM) [17].

In this paper, the k -means algorithm's application to detect the damaged solar panels in the PV field is investigated. More precisely, the K -means algorithm is employed to cluster the thermal image into many regions and identify the damaged area. Furthermore, Elbow and gap methods are used to select the optimal number of clusters by setting the system with various values for K .

The remainder of this work is structured as follows: Section 2 provides the clustering techniques and K -means algorithm. Elbow, average silhouette, and NbClust methods are explained in Section 3. Simulation results and discussion are presented in Section 4, whereas Section 5 summarizes the main conclusions of the present work.

2. Clustering Techniques and K-Means Algorithm

2.1. Clustering Techniques. Clustering or grouping is one of the most interesting parts of unsupervised machine learning, which has been extensively employed for image processing, and has many applications whether image retrieval or annotation [18]. This technique aims to classify data into many regions (clusters) depending on the similarity of characteristics.

As shown in Figure 2, data clustering can be divided into two types: hierarchical [19, 20] and partitional [21] clustering. The partitional clustering is the most straightforward technique for grouping data that provides nonoverlapping clusters, where each object belongs to one group, whereas hierarchical clustering permits clusters to have subclusters; therefore, each object can belong to many groups.

Figure 3 exhibits the different algorithms clustering for each category, where partitional includes, for instance, K -mean [12] and fuzzy c -mean [13] algorithms. In contrast, hierarchical clustering is split into two types of algorithms: agglomerative algorithms and divisive algorithms [22].

Image segmentation is a process of clustering techniques and the most important key in image processing that aims to divide an image into different segments based on their intensity. In this context, this paper focuses on utilizing image processing to detect the fault in the PV system. In other words,

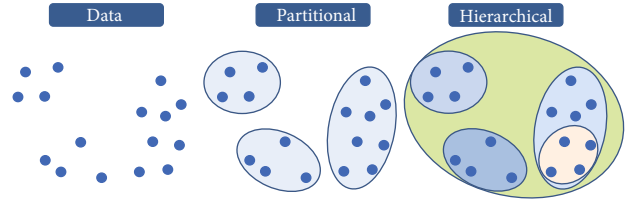


FIGURE 2: Type of clustering.

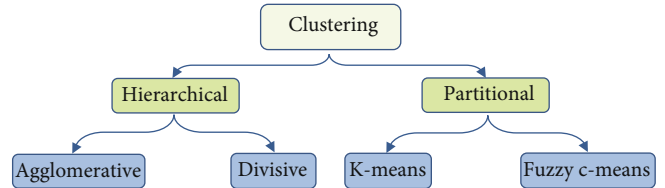


FIGURE 3: Type of FDD.

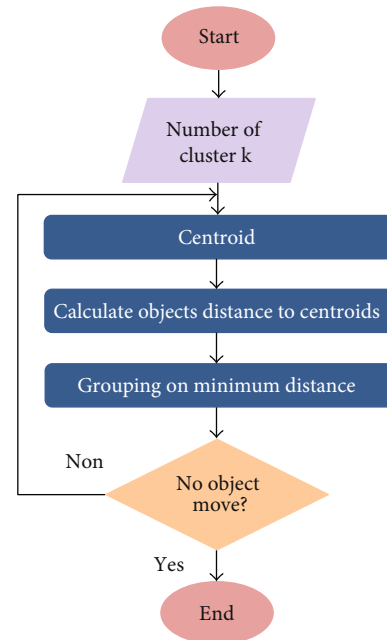


FIGURE 4: Flowchart of K -means algorithm.

the application of the unsupervised machine learning algorithm is used to locate the damage in solar panels.

2.2. The K -Means Algorithm. The K -means is the simplest unsupervised machine learning algorithm used for clustering, proposed by Mac Queen in 1967 [23]. This algorithm is based on two main steps: the first step is to define the k centroid, and the second step is to attach each point to the nearest cluster. The flowchart diagram of the K -means algorithm is shown in Figure 4.

In this paper, the K -means algorithm is used as a clustering technique to detect and locate precisely the faults areas in a thermal image of the solar field. Otherwise, the thermal image is classified into K groups, and each group contains pixels with similar properties (intensity). The K -means

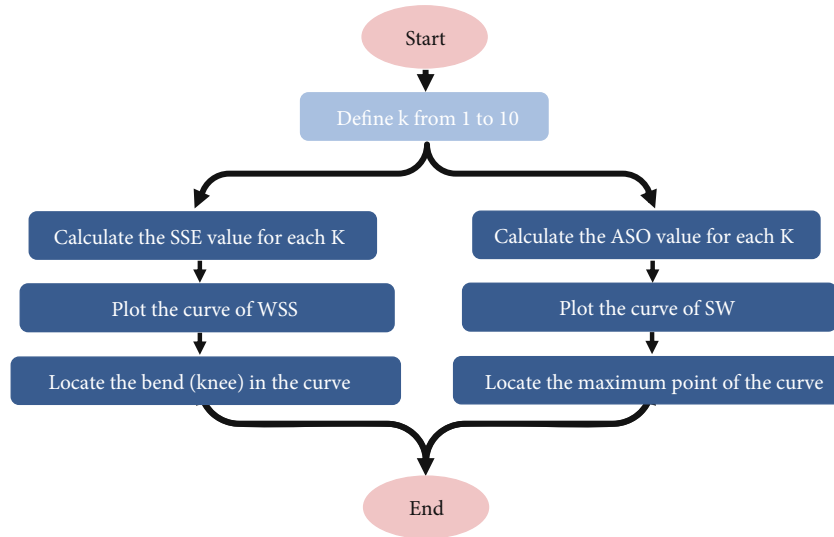


FIGURE 5: Flowchart of the elbow and the average methods.

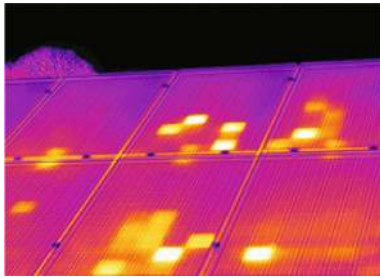


FIGURE 6: Thermal image of PV panels.

algorithm procedure to cluster an image includes the following steps, as noted in [24]:

We consider an image with a resolution of (x, y) , $p(x, y)$ is the input, and C_k be the cluster centers.

Step 1. Initialization of the cluster k and center

Step 2. Calculation of the distance d between the center and each pixel of the image using the following equation:

$$d = \|p(x, y) - C_k\| \quad (1)$$

Step 3. Attach all pixels to the nearest center using the calculated distance

Step 4. Calculation of the new center position using the following equation:

$$C_k = \frac{1}{k} \sum_{y \in C_k} \sum_{x \in C_k} p(x, y) \quad (2)$$

Step 5. Repeat steps 2, 3, and 4 until the center C_k no longer moves.

Step 6. Reconstruction of the image by reshaping the pixels of the cluster.

3. Optimal Number of Clusters

3.1. Elbow Method. In literature, many research works proposed several methods to determine the optimal number of clusters K . The elbow method has been considered among the excellent techniques. It is based on the square distance between the centroid of the cluster and each cluster's sample points. The sum of squared errors (SSE) is the performance indicator, which is calculated for each value K using the following equation [15]:

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|X_i - C_k\|_2^2, \quad (3)$$

where x is the data present in each cluster and C_k is the K th cluster.

The optimal value of K is found when the SSE value drops on the curve drastically and forms a smaller angle.

3.2. Average Silhouette Method. The average silhouette method has the same purpose as the elbow method that was proposed by Liu and Sarkar [25]. It is based on computing the average silhouette of observations for different values of k . Otherwise, the difference between distances an object has to other objects in the same cluster and the distance it has to other objects in other clusters. For object i in a cluster C_i , the silhouette width (SW) is defined using the following equation [16]:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)}, \quad (4)$$

where $b_i = \min(\sum_j d(i, j) / |C_j|)$, $C_i \neq C_j$ and $a_i = \min(\sum_j d(i, j) / |C_i|)$, $C_i = C_j$.

The SW is considered as a performance indicator varying from -1 to +1, where the optimal number of clusters is corresponding to the highest value.

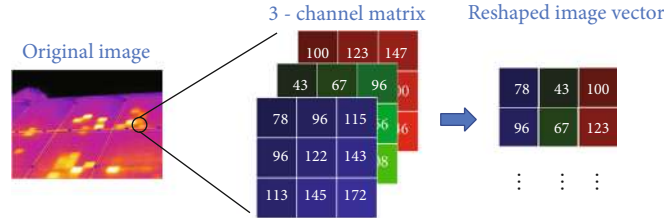


FIGURE 7: Converting Image to Vector.

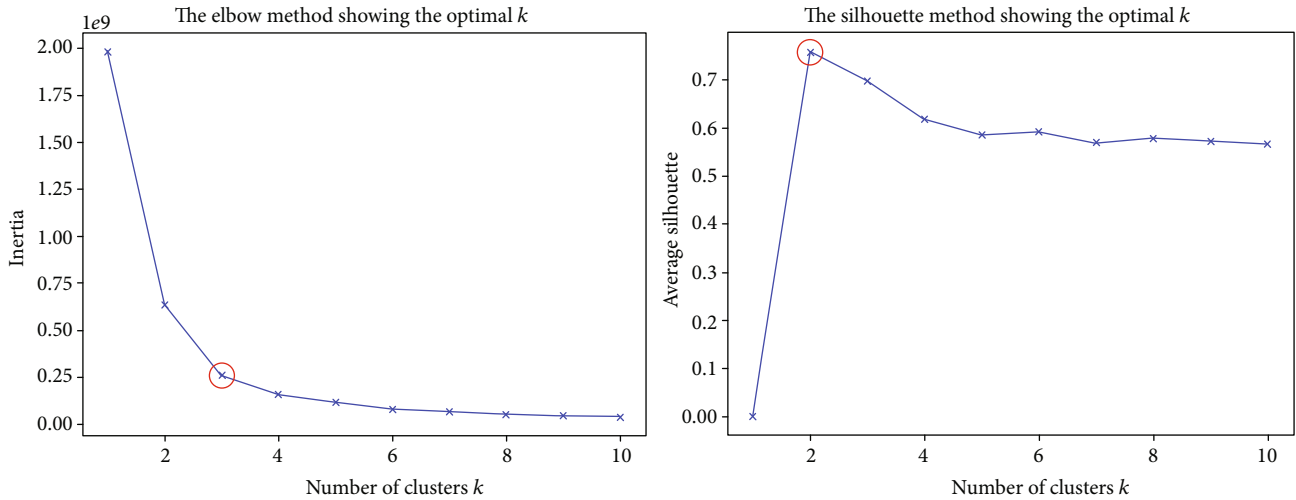


FIGURE 8: Optimal number of clusters suggested by the elbow and the average silhouette methods.

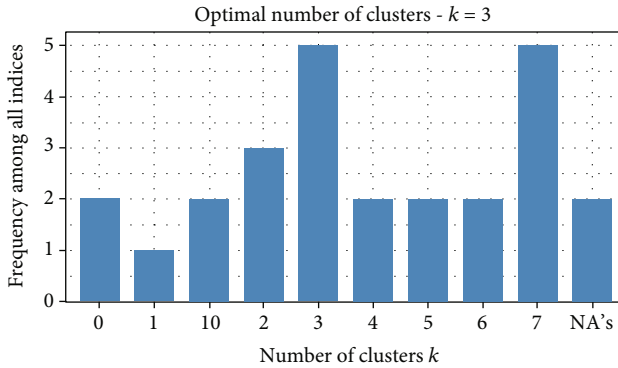


FIGURE 9: Optimal number of clusters using NbClust R package.

The flowchart diagram of the elbow method and the average silhouette method are illustrated in Figure 5.

3.3. NbClust Package. To ensure the clustering quality, it is crucial to select the best number of clusters suggested by the most methods cited in the literature. Hence, NbClust has been developed for this purpose, which is a function of the NbClust R package [26] that allows selecting the optimal number of clusters in a dataset by varying all combinations of the number of clusters, distance measures, and clustering methods. In other words, NbClust is based on thirteen indices, as presented in [26], and it provides the indices number suggesting for each number of clusters. The optimal number will be chosen according to the majority rule.

4. Results and Discussion

In this section, we are going to present the results obtained using the K-means algorithm to detect and locate the faults in PV panels. Firstly, before applying this algorithm, the optimal number of clusters K has to be defined; hence, the Elbow and the average silhouette methods are used.

All implementations are carried out on the thermal image presented in Figure 6; after getting permission from the online website [27], the simulation was performed using python 3.7.0 and an i5 8th gen 1.8 GHz machine with 8 GB RAM platform (Windows10 64 bit).

Before implementing the methods, the image must be converted to vector in the 3-D space of RGB; hence, the input image is converted from RGB (red, green, and blue) colors space to HSV (hue, saturation, and value). Therefore, the $L \times H \times 3$ image is transformed into a $K \times 3$ matrix with $K = L \times H$, as shown in Figure 7.

The results of applying the Elbow and the average silhouette methods are presented in Figure 8. For the elbow method, it is clear that the elbow point occurs at $K = 3$; hence, it is the optimal value suggested. On the other side, according to the curve of silhouette width, the highest value corresponding to $K = 2$, which is the optimal number of clusters proposed by the average silhouette method.

In addition, the NbClust function [28] was also used to find the optimal number of clusters based on 26 indexes, as shown in Figure 9. According to the majority rule, the best number of clusters is 3.

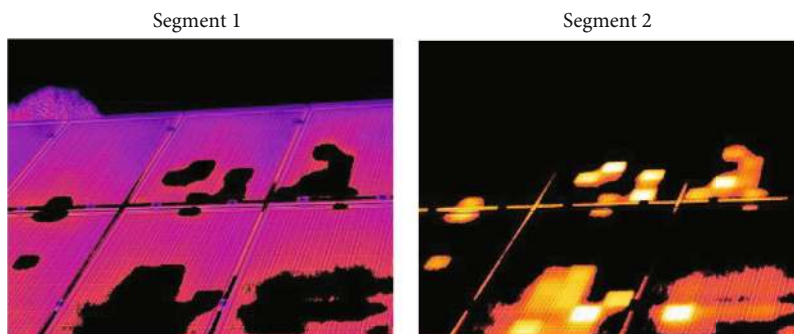


FIGURE 10: Output results using K-means algorithm for $k = 2$.

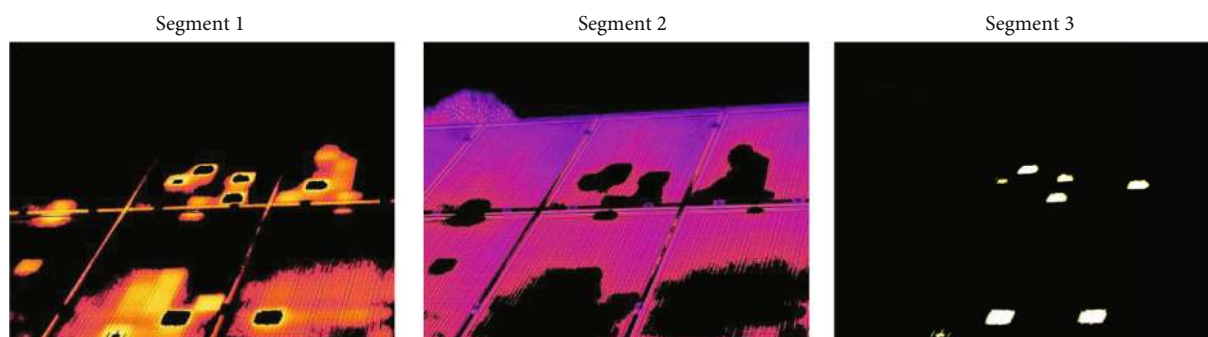


FIGURE 11: Output results using K-means algorithm for $k = 3$.

Thermal image				
Faults detected				

FIGURE 12: The application results of the K-means algorithm.

Figures 10 and 11 present the outputs results of applying the K-means algorithm on a thermal image of solar panels, using the optimal values of clusters suggested, respectively, by the Elbow and the average silhouette methods.

The output results are fragmented into segments depending on the number of clusters. It is evident that Figure 11 pre-

sents excellent results that allow to detect and identify the damaged areas that are presented in segment 3. Hence, the elbow method has given the most optimal number of clusters than the average silhouette method.

Four results were carried out in Figure 12 in order to validate the effectiveness of the K-means algorithm for detecting

the faults in PV panels. As can be noticed from this table, the K-means algorithm has successively detected all faults presented in the four thermal images of PV panels.

Finally, the K-means algorithm with the optimal number of clusters has demonstrated the excellent performance that can be integrated to the drone and the thermal system to identify precisely the damaged solar panels in the PV field.

5. Conclusion

The energy demand has increased quite fast, which requires the improvement of all energy sources. Especially, solar energy providing clean, reliable, and unlimited energy. Accordingly, many works have been studied to diagnose and control the production of PV panels, among them, the thermal drone system, which allows controlling the vast farms of PV panels.

In this paper, the application of the K-means algorithm is investigated to cluster a thermal image of PV panels and automatically detect the damaged areas. The quality of clustering depends on selecting an optimal number of clusters; hence, the elbow and the average silhouette methods were used.

The K-means algorithm and elbow method together provide excellent results to find and detect precisely the faulty panels. Therefore, this algorithm can be integrated into the thermal drone system to find the faults in the PV system easily and in real-time, especially for large PV fields.

Abbreviations

ASM: Average silhouette method
 EM: Elbow method
 FCM: Fuzzy C means algorithm
 GSM: Gap statistic method
 HSV: Hue, saturation, and value
 KMA: K-means algorithm
 MSA: Mean shift algorithm
 SSE: Sum of squared errors
 SSPA: Statistical and signal processing approaches
 SW: Silhouette width
 PLA: Power losses analysis
 PV: Photovoltaic
 RGB: Red, green, and blue.

Data Availability

The data used to support the findings of this study have not been made available because it is confidential.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

A. Et-taleby carried out the simulations. The results were discussed and evaluated by M. Boussetta and M. Benslimane. A. Et-taleby wrote the first version of the paper, M. Boussetta corrected and modified it afterwards, and after supervision

of the article by all the team, M. Boussetta submitted the article to the journal. All authors of this research paper have directly participated in the planning, execution, or analysis of this study. All authors read and approved the final manuscript.

References

- [1] T. Ajadi, V. Cuming, R. Boyle et al., *Global Trends in Renewable Energy Investment 2020*, 2020.
- [2] A. Mellit, G. M. Tina, and S. A. Kalogirou, "Fault detection and diagnosis methods for photovoltaic systems: a review," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 1–17, 2018.
- [3] J. Solórzano and M. A. Egido, "Automatic fault diagnosis in PV systems with distributed MPPT," *Energy Conversion and Management*, vol. 76, pp. 925–934, 2013.
- [4] F. Harrou, Y. Sun, B. Taghezouit, A. Saidi, and M.-E. Hamlati, "Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches," *Renewable Energy*, vol. 116, pp. 22–37, 2018.
- [5] B. Du, Y. He, Y. He, and C. Zhang, "Progress and trends in fault diagnosis for renewable and sustainable energy system based on infrared thermography: a review," *Infrared Physics & Technology*, vol. 109, p. 103383, 2020.
- [6] Z. A. Jaffery and A. Haque, "Temperature measurement of solar module in outdoor operating conditions using thermal imaging," *Infrared Physics & Technology*, vol. 92, pp. 134–138, 2018.
- [7] M. Cubukcu and A. Akanalci, "Real-time inspection and determination methods of faults on photovoltaic power systems by thermal imaging in Turkey," *Renewable Energy*, vol. 147, pp. 1231–1238, 2020.
- [8] A. Khan, Z. ur Rehman, M. A. Jaffar et al., "Color image segmentation using genetic algorithm with aggregation-based clustering validity index (CVI)," *Signal, Image and Video Processing*, vol. 13, no. 5, pp. 833–841, 2019.
- [9] M. Ali, M. Khan, and N. T. Tung, "Segmentation of dental X-ray images in medical imaging using neutrosophic orthogonal matrices," *Expert Systems with Applications*, vol. 91, pp. 434–441, 2018.
- [10] T. Ren, H. Wang, H. Feng, C. Xu, G. Liu, and P. Ding, "Study on the improved fuzzy clustering algorithm and its application in brain image segmentation," *Applied Soft Computing*, vol. 81, p. 105503, 2019.
- [11] X. Ji, J. F. Henriques, and A. Vedaldi, "Invariant information clustering for unsupervised image classification and segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9865–9874, Seoul, Korea (South), 2019.
- [12] D. Li, S. Wang, N. Gao, Q. He, and Y. Yang, "Cutting the Unnecessary Long Tail: Cost-Effective Big Data Clustering in the Cloud," *IEEE Transactions on Cloud Computing*, p. 1, 2019.
- [13] J. C. Bezdek, *Fuzzy Mathematics in Pattern Classification*, [Ph. D. thesis], Applied Mathematics, Cornell University, 1973.
- [14] K. Wisaeng and W. Sa-Ngiamvibool, "Exudates detection using morphology mean shift algorithm in retinal images," *IEEE Access*, vol. 7, pp. 11946–11958, 2019.
- [15] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile

- Cluster,” in *IOP Conference Series: Materials Science and Engineering*, p. 12017, Surabaya, Indonesia, 2018, IOP Publishing.
- [16] H. Yahyaoui and H. S. Own, “Unsupervised clustering of service performance behaviors,” *Information Sciences*, vol. 422, pp. 558–571, 2018.
- [17] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic,” *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, vol. 63, no. 2, pp. 411–423, 2001.
- [18] Y. Ren, N. Wang, M. Li, and Z. Xu, “Deep density-based image clustering,” *Knowledge-Based Systems*, vol. 197, p. 105841, 2020.
- [19] K. Li, Z. Ma, D. Robinson, and J. Ma, “Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering,” *Applied Energy*, vol. 231, pp. 331–342, 2018.
- [20] C. Li, J. Bai, Z. Wenjun, and Y. Xihao, “Community detection using hierarchical clustering based on edge-weighted similarity in cloud environment,” *Information Processing and Management*, vol. 56, no. 1, pp. 91–109, 2019.
- [21] S. Ben Salem, S. Naouali, and Z. Chtourou, “A fast and effective partitionial clustering algorithm for large categorical datasets using a k-means based approach,” *Computers and Electrical Engineering*, vol. 68, pp. 463–483, 2018.
- [22] P. Govender and V. Sivakumar, “Application of k-means and hierarchical clustering techniques for analysis of air pollution: a review (1980–2019),” *Atmospheric Pollution Research*, vol. 11, no. 1, pp. 40–56, 2020.
- [23] J. MacQueen, “Some Methods for Classification and Analysis of Multivariate Observations,” *Proc. Fifth Berkeley Symp. Math. Stat. Probab.*, 1967, pp. 281–297, Oakland, CA, USA, 1967.
- [24] N. Dhanachandra, K. Manglem, and Y. J. Chanu, “Image segmentation using K-means clustering algorithm and subtractive clustering algorithm,” *Procedia Computer Science*, vol. 54, pp. 764–771, 2015.
- [25] Z. Liu and S. Sarkar, “Simplest representation yet for gait recognition: Averaged silhouette,” in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, pp. 211–214, Cambridge, UK, 2004, IEEE.
- [26] C. Malika, N. Ghazzali, V. Boiteau, and A. Niknafs, “NbClust: an R package for determining the relevant number of clusters in a data set,” *Journal of Statistical Software*, vol. 61, pp. 1–36, 2014.
- [27] “Infrared solar panels,” May 2020, <https://thermodrone-inspection.com/panneaux-solaires-et-fermes-solaires/>.
- [28] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, and M. M. Charrad, “NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set,” *Journal of Statistical Software*, vol. 61, pp. 1–36, 2014.