applied sciences

MDPI

Article

Artificial Afterimage Algorithm: A New Bio-Inspired Metaheuristic Algorithm and Its Clustering Application

Murat Demir

Department of Software Engineering, Faculty of Engineering and Architecture, Muş Alparslan University, 49250 Muş, Türkiye; m.demir@alparslan.edu.tr

Abstract: Metaheuristic methods are optimization methods that look for different ways to converge to a solution to a problem where it is difficult to find a solution analytically. Their difference from known optimization methods is that they imitate living things or systems in nature. Each metaheuristic method has its equations, and the solution is found using these equations. In this study, a new, metaheuristic method called the afterimage algorithm is proposed. The proposed method was developed inspired by the fact that when we close our eyes after looking at a luminous image for a while, the vision still occurs in our minds. This is called an afterimage. The proposed method first pre-processes with the operator called afterimage and calculates the best and worst solution values. The visual angle value is then calculated, and new solutions are produced around this value. Three different datasets were used in experimental studies on data clustering. Accuracies of 96.66% for the iris plant dataset, 92% for the Wisconsin breast cancer dataset, and 95% for the occupancy detection dataset were obtained.

Keywords: metaheuristic methods; optimization; visual angle; perceptual size; afterimage; clustering

Academic Editor: Juan A. Gómez-Pulido

Received: 30 December 2024
Revised: 23 January 2025
Accepted: 24 January 2025
Published: 28 January 2025

Citation: Demir, M. Artificial Afterimage Algorithm: A New Bio-Inspired Metaheuristic Algorithm and Its Clustering Application. *Appl. Sci.* **2025**, *15*, 1359. https://doi.org/10.3390/app15031359

1. Introduction

The most effective and definitive way to solve problems is classical mathematical solutions. However, in some cases, obtaining an analytical solution is either very difficult or disadvantageous in terms of time.

There are three main techniques for solving problems: experimental, analytical, and numerical. Experimental methods are costly and have disadvantages in terms of time. They are also not very useful in changing parameters. Analytical methods are precise in terms of solution but difficult to use in complex problems. Numerical methods have become popular with the advancement of computational techniques. They produce approximate solutions to the problems. Their accuracy rates are also considered sufficient for engineering studies [1].

Data environments are growing day by day due to increasing technological devices, sensors, and resources, etc. The huge increase in the volume of data makes it difficult to analyze such large data. For this reason, it can sometimes be very difficult or impossible to obtain a solution using analytical methods. In places where analytical solutions are insufficient, artificial intelligence and several different methods that support artificial intelligence come into play instead of classical analytical methods.

Today's computers can process large amounts of data very quickly. Thanks to this capacity, they contribute greatly to problem solving in every field. Artificial intelligence and optimization are important fields that bring a perspective other than classical solutions

to problems. In this way, we can shorten the problem-solving processes or reach alternative solutions [2].

Real-world problems require significant computational costs and many complex solution methods. Therefore, mathematical programming approaches or optimization techniques, called metaheuristics, have been proposed many times to solve these problems [3]. Metaheuristic methods are used in cases where developing analytical solutions are not possible or very difficult. Metaheuristic methods are algorithms inspired by nature. Every method has a living thing or event from which it draws inspiration [4].

The motivation of this study is from metaheuristic algorithms and clustering. The proposed artificial afterimage algorithm (AAIA) is an algorithm that has not been proposed before in the literature. Therefore, a list of different metaheuristic algorithms on clustering is presented below.

Chunzhong Li, C., et al. [5] propose a metaheuristic algorithm based on cognitive feature integration in their study. The proposed algorithm uses no parameter density estimation and maximum likelihood estimation to integrate cognitive features and finally produces clustering results. Pattanayak, P., K., et al. [6] propose a new metaheuristic method for a graph-based topic modeling technique in their study. Latent Dirichlet analysis (LDA) and the Gibbs sampling Dirichlet mixture model (GSDMM) are popularly used methods for topic modeling. Shaikh, M. S., et al. [7] propose the improved gray wolf clustering algorithm (IGWOA). It improves the gray wolf algorithm's convergence rate and reduces the risk of getting stuck in the local optimum in their study. There has been a proposed study to cluster students' stress levels. Tsai, C-W, et al. [8] present a hyper-heuristic clustering algorithm based on diversity detection and optimization detection operators in their study, in order to determine when to switch from one metaheuristic algorithm to another in order to improve the clustering result. Christou, I., T., et al. [9] propose a new and effective metaheuristic algorithm for minimizing the distance of the weighted sum of squares appearing on the Euclidean surface in many biomass supply-chain management applications in their study. Lamari [10] presents new and effective parameter-free metaheuristic methods for clustering large datasets in her thesis. The first proposed method is a fast heuristic method called the transitive heuristic method. Alam and Sadaf [11] present an approach to cluster the results of an information retrieval system in their work. The work is a combination of metaheuristic methods using cosine similarity and k-means technique. Al-betar et al. [12] propose an approach known as the bare-bones salp swarm algorithm (BBSSA) which generates new individuals and directs the population to solve the text document clustering problem using Gaussian search equations, inverse hyperbolic cosine control strategies and greedy selection techniques. Hugo Schnoering et al. [13] carry out a study to determine the groups of addresses belonging to the same entity on Bitcoin blockchains using metaheuristic methods in their study. Thus, clustering operations of addresses belonging to blockchains were performed. Puri and Gupta [14] propose Heuristic-mrk-medoids clustering in their work to perform linear time clustering of big data. In the work, the modified squirrel search algorithm (M-SSA) was designed to align the mrk-medoids value optimally. Kumar, G., K. et al. [15] focus on wireless sensor network security issues in their work. For this purpose, they propose an optimized metaheuristic method. The method aims to follow the clustering-based privacy key agreement route. Gao, C. et al. [16], in their work, propose a self-adaptive logarithmic spiral path black hole algorithm. The algorithm uses a parameter to control randomness. It defines a self-adaptive parameter to control the switching mechanism and maintain the balance of the algorithm between exploration and exploitation.

The above-mentioned studies in literature are all studies carried out for the purpose of data clustering. However, the afterimage infrastructure presented in this study has not

been encountered in any previous study. In this respect, this study is presented for the first time in the literature. It also has very good results in terms of success metrics. The main goal in metaheuristic algorithms is to provide methods that can solve problems by imitating the systems and living things that exist in nature. This study achieves this in this respect.

1.1. Metaheuristic Methods

The non-linearity of the data makes it difficult to solve with experimental, analytical, and numerical methods. For this reason, metaheuristic methods, a new field of study that provides solutions with approximative methods, have come to the fore as an alternative to support artificial intelligence.

These methods are offered as an alternative in cases where we cannot obtain a solution with known classical methods or when we have time problems with very complex solutions.

We suggest the following as examples of metaheuristic algorithms developed in recent years: snooker-based optimization algorithm (SBOA) [17], a new heuristic algorithm based on minimum spanning tree [18], GPU-based efficient parallel heuristic algorithm for high-utility itemset mining [19], learning search algorithm (LSA) [20], stadium spectators optimizer (SSO) [21], piranha foraging optimization algorithm (PFOA) [22], dynamic population optimization (DPO) [23], the heuristic simulated annealing algorithm [24], elite opposition-based bare-bones mayfly algorithm (EOBBMA) [25], black-winged kite algorithm (BWKA) [26], hippopotamus optimization algorithm [27], special forces algorithm (SFA) [28], walrus optimization algorithm (WaOA) [29], equilibrium slime mold algorithm (EOSMA) [30].

Each metaheuristic method has its own mathematical equations. It tries to improve the solution space, which starts with random solutions, by using these mathematical equations. While making these improvements, it uses problem-specific functions called fitness functions. The fitness function of each problem is determined by the designer specifically for the problem. In each iteration, the solution population is changed using the metaheuristic method's own equations. The best solution is searched up to a stopping criterion. The result is the ideal possible solution to the problem.

1.2. Afterimage and Visual Angle

1.2.1. Afterimage

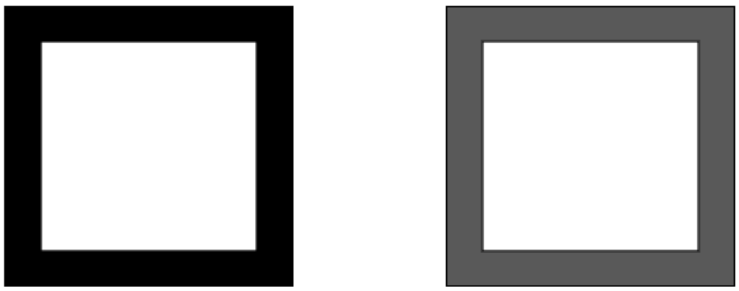
Almost all of us have looked at a luminous object or an object passing around or through it for a while and then continued to see the image of the object when we closed our eyes. This is called afterimage in scientific literature.

After looking at an original image for a while, even if the image disappears, it continues to be visible in the eyes. This is called afterimage. After the afterimage is formed, even if the eyes do not see the original image, the photochemical effect on the retina of the eye continues for a while. Therefore, we continue to see the image for a while. Afterimages are perceptual phenomena that can be seen frequently in daily life. They are ghost images that appear when moving away from a bright light source. If the color contrast is significant, the afterimage can also be colored. It also disappears within a certain time interval. [31–33].

If the eye is closed after looking at any luminous object, we will continue to see this object for a short time. This is called positive afterimage. After looking at a colored object for a long time if we turn our eyes to a neutral background, we see that the shape of that object is formed in a complementary color. This is called negative afterimage [34].

Negative afterimages are formed in the retina of the eye and can be modified by adaptation of the retinal ganglion cells. Positive afterimages appear in colors identical to the original image. The reason is not clear, but it is probably due to the retinal photoreceptor cells continuing to send signals to the occipital lobe [35].

Figure 1 is an example of a positive afterimage. (a) In the picture on the left is the original image. The image on the right (b) is its positive afterimage reflection. The image has similar tones to the original picture, but the sharpness of the colors will not be the same.



(**a**) (**b**)

Figure 1. Positive afterimage instance: (**a**) original image, (**b**) reflection image.

Figure 2 shows negative afterimage examples for example 1 and example 2. In example 1, while the ground color is white and the object color is black, the reflected part appears as black and white, which are the complementary colors. In example 2, the ground color is blue and red; in the reflected image, the ground color appears as yellow, and the color of the object appears as yellow and green/turquoise.

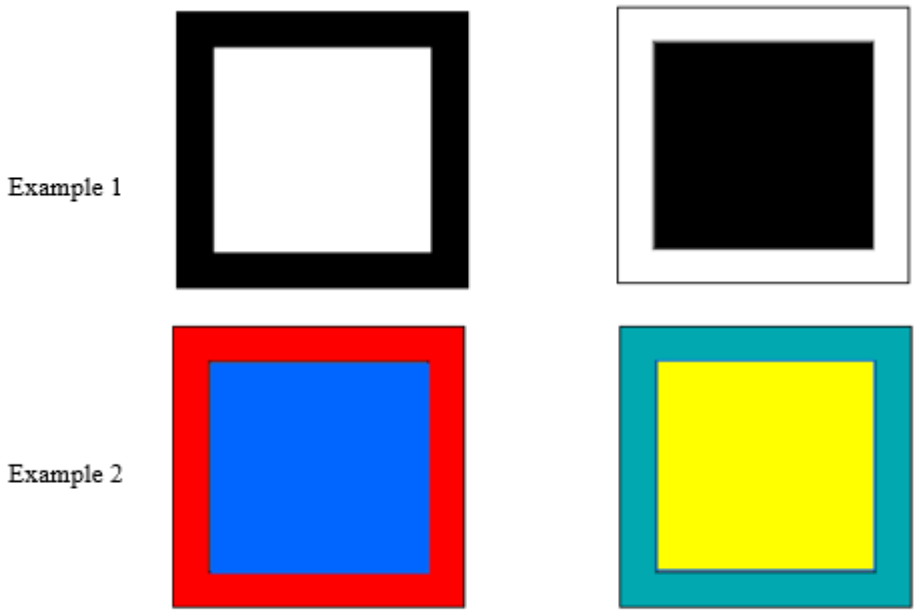


Figure 2. Negative afterimage instance, (**a**) original image, (**b**) reflection image.

1.2.2. Visual Angle

In the visual angle, an object is visible when observed. It is expressed in degrees of arc [36]. Visual angle is formed by the rays reflected from the eye above and below an object. It is used to indicate the size of the retinal image of an object. The larger the angle of view is, the larger the retinal image size will be. Figure 3 shows how the visual angle is formed.

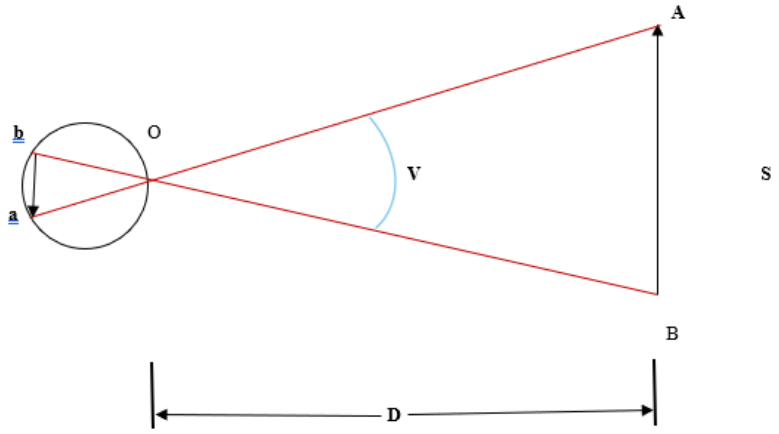


Figure 3. Visual angle representation.

The angle of view is calculated by two values: the size of the object and the distance of the object from the eye. Equation (1) gives the mathematical expression of the visual angle [36].

$$Visual\ Angle\ (V) = 2 \cdot \arctan\left(\frac{object\ size}{2 * \ object\ distance}\right) \quad (1)$$

Object size (S) is the distance between points A and B. It is the real size of the object. Object distance (D) is the distance of the object from the eye. It is the distance between O and B.

The perceptual size of the observed object (S) is expressed by Equation (2) [36].

$$S = V \cdot D \quad (2)$$

As can be seen, after finding the V visual angle, the perceptual size of the observed object can be found.

2. Data Clustering

Machine learning is a field of artificial intelligence that allows the development of complex techniques. It is very suitable for data analysis. According to the observed data, it is divided into four classes: supervised, unsupervised, semi-supervised, and reinforcement [37].

Clustering is a technique used to organize related data points into different groups. It is a technique in which the structure is determined, so that the objects with the most similarities are in a group [38]. Clustering is an unsupervised learning technique. Unsupervised learning uses unlabeled data. Grouping is carried out based on the distance relationship between the attributes [39]. Various distance measure equations exist to measure distances between attributes. The most well-known of these are Euclidean, Manhattan, and Minkowski distance measures [40]. In this study, the Manhattan distance measure was used as the distance measure. The formula for the Manhattan distance measure (MAND) is given in Equation (3) [40].

$$MAND\ (x, y) = \sum_{i=1}^{n} |x_i - y_i| \quad (3)$$

x_i and y_i represent the representation of two different attributes.

Figure 4 represents the basic representation of the clustering process.

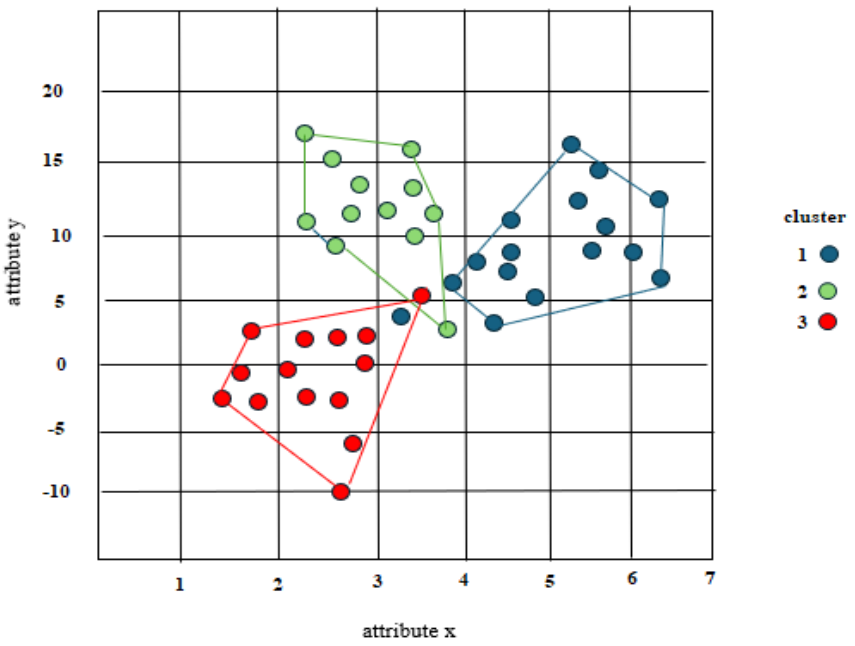


Figure 4. Clustering process representation.

In Figure 4, the points of different colors in different clusters are examples of incorrectly clustered data. These will be considered as errors.

Confusion Matrix and Evaluation Metrics

The confusion matrix shows the success of the model. Each cell represents an evaluation criterion. The criteria are the numerical equivalents of the actual values and the values to be predicted. Four different values are presented as true positive (TP), true negative, false positive (FP) and false negative (FN).

If a positive result is obtained for an actually positive situation, it is called TP; if a negative result is obtained for an actually positive situation, it is called FN; if a negative result is obtained for an actually negative situation, it is called TN; and if a positive result is obtained for an actually negative situation, it is called FP [41].

The performance metrics calculated with these values obtained from the confusion matrix are given in Equations (4)–(7) [41].

Accuracy is the ratio of correctly predicted values to the total dataset.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

Precision is the ratio of true positives to total positives.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

Recall is the ratio of true positives to the sum of true positives and false negatives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

F1 score is the harmonic meaning value of recall and precision values.

$$\text{F1 score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \tag{7}$$

In the Materials and Methods Section, the method of the proposed AAIA algorithm is presented. For this purpose, the calculations related to visual angle and perceptual size are given in the adapted versions of the AAIA algorithm. It is explained how the concepts of visual angle and perceptual size are used in the proposed method. The pseudocodes of the AAIA algorithm and the general working principle of the method are presented.

In the Experimental Results Section, after the data properties of the three datasets are presented, the confusion matrices are obtained according to the clustering results produced by the AAIA algorithm that are given. Then, the obtained clustering intervals are presented, and their evaluations are made. The clustering results are visualized with the help of the presented graphics. In addition, in this section, the results of other studies conducted in the literature with three datasets are given. Thus, the advantages of the AAIA algorithm compared to the literature are also presented.

In the Comparison of AAIA with Methods in Literature Section, the AAIA algorithm is compared with other heuristic methods in the literature in terms of clustering. The studies compared in the literature are selected especially from the studies using the three datasets in this study.

In the Discussion Section, the distributions of the datasets are analyzed and the effects of the distribution on clustering in the distance-based clustering method are discussed. The related graphics are presented.

3. Materials and Methods

3.1. Innovative Aspect of the Manuscript

The proposed algorithm AAIA is an original work in this respect. Because there is nothing similar in the literature, this makes the study original. In the literature, studies on afterimage have generally focused on painting, photography, and digital imaging. No previous studies have been found as a metaheuristic method. In this respect, this study offers a new field of study. In addition, clustering with different metaheuristic methods was attempted on the datasets on which the experimental studies were conducted in the article. Although the proposed method is new and still open to development, it has achieved very good success compared to known distance-based methods. This shows that the presented method is a successful approach. Furthermore, as mentioned before, the aim of metaheuristic methods is to solve problems by imitating the events and systems that exist in nature. The presented AAIA method has achieved this very successfully. In many metaheuristic methods, new solutions are obtained from the best values. In this proposed method, calculations are made from both the best and worst values. In this respect, it is expected to prevent becoming bogged down on local solutions.

3.2. The Proposed Method: Artificial Afterimage Algorithm (AAIA)

In this section of the study, the structure of the artificial afterimage algorithm is explained. The afterimage infrastructure was presented in Section 1.2. The developed method, artificial afterimage algorithm, is an original new method that has not been studied in the literature before. It is a method inspired by the light reflections that occur during a visual event in the eye. It has an infrastructure based on afterimage and visual angle mentioned in Section 1.2. The pseudo code of the method is presented in Algorithm 1.

This study was carried out on data clustering. Here, the concept called object size represents the current value of the candidate solution, and object distance represents the distance from the best solution. These values are used in Equations (1) and (2) to calculate the visual angle and perceptual size of the object. Then, based on these two values, the new values of the candidate solutions of the population are updated around the best solution.

After this process is applied for a certain number of steps, the best solution available gives us the best value for the clustering process.

Algorithm 1: Pseudo code of AAIA

```
Begin
Produce initial population of candidate solution
Determine initial best solution
while (count ≤ iteration)
    Determine Object size (value of current solution)
      Determine Object distance (distance of the current solution from the best solution)
      Calculate Visual angle (V)
      Calculate Perceptual size of object (S)
      local best solution ← Determine the current best values
      if  fitness value of local best solution < fitness of best solution then
        best ← local best values
      end if
    Calculate New solution for population
  end while
  Present best solution
end begin
```

Equations (8)–(10) show the calculations of visual angle, perceptual size of object, and update solutions.

$$V = 2 * \text{atan}\left(\frac{pop_{i,j}}{2 * pop_{i,j} - best_{1,j}}\right) \tag{8}$$

$$S_{i,j} = V * \left(pop_{i,j} - best_{1,j}\right) \tag{9}$$

$$New\ pop_{i,j} = S_{i,j} + \left(\left| S_{i,j-}(best_{1,j} - \left|\left(worst_{i,j} - pop_{i,j}\right)\right|\right) * rand \tag{10}$$

$pop_{i,j}$, specified in Equations (8)–(10), is used instead of object size. It is the size of the candidate solution as the current value. $pop_{i,j} - best_{i,j}$ used instead of object distance is the distance of the cim from the best value, best. The worst value used in Equation (10) represents the worst result. Best and worst values represent positive and negative afterimage.

Here, the worst means the worst solution; it expresses the opposite of the color of the real image in the event of vision. The distance of the candidate solution from the worst value actually represents the colors we see inverted when we close our eyes. In other words, it is like looking at a white object and seeing black when we close our eyes. Subtracting this value from the best value represents the difference from the original values. By adding this to the perceptual size value, we calculate the new solution value.

The algorithm aims to produce the new pop values in Equation (10) by using the visual angle and perceptual size values calculated in Equations (8) and (9). New pop values are the new population values obtained in each generation. Obtaining them by using the best and worst values allows better values to be obtained each time. In this way, the values obtained in the end can give very close results to the expected ones.

This situation is expressed in Figures 2 and 3 in Sections 1.2.1 and 1.2.2 In light of this information, the pseudo code of the method is as in Algorithm 2.

Algorithm 2: Final state of Pseudo code of AAIA

Begin
Produce initial population of candidate solution
Determine initial best solution
while (count $\leq$ iteration)
Determine Object size
    Determine Object distance
    $V \leftarrow 2 * \text{atan}(\frac{pop_{i,j}}{2 * pop_{i,j} - best_{1,j}})$
$S \leftarrow V * \left(pop_{i,j} - best_{1,j}\right)$
Local best solution $\leftarrow$ determine current best values
    if fitness value of local best solution < fitness value of best solution **then**
        best $\leftarrow$ local best
    end if
Calculate $New\ pop_{i,j} = S_{i,j} + \left(\left| S_{i,j-}(best_{1,j} - \left|\left(worst_{i,j} - pop_{i,j}\right)\right|\right) * rand\right)$
End while
Present best solution values
end begin

Figure 5 shows the general workflow of this study. This study used the AAIA method to cluster three different datasets.

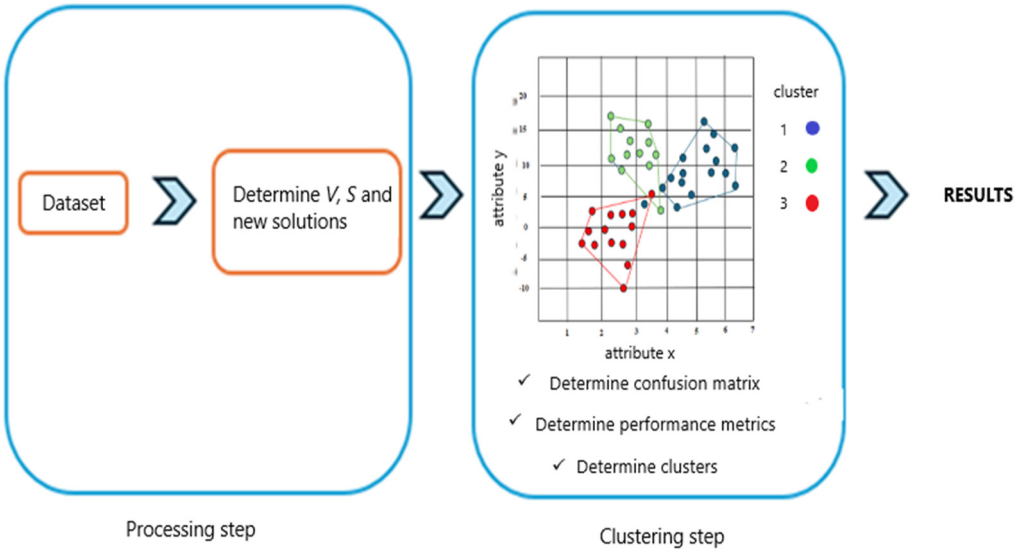


Figure 5. General workflow study.

4. Experimental Results

Results of Datasets

The iris plant dataset, Wisconsin breast cancer dataset, and occupancy detection dataset [42–44] used in this study are datasets frequently used in data clustering studies. There are 150 samples in total in the iris plant dataset. Each sample contains four data. In the study, 30 samples from each cluster were used for training the algorithm and 20 samples were used for testing. There are 685 samples in total in the Wisconsin breast cancer dataset. Each sample contains nine data. In the study, 585 samples from each cluster were used for training the algorithm and 100 samples were used for testing. There are 8143 samples in total in the Wisconsin breast cancer dataset. Each sample contains five data. In the study,

7943 samples from each cluster were used for training the algorithm and 200 samples were used for testing. Table 1 shows the V, best and worst values obtained for datasets.

Table 1. Experimental results of datasets.

Obtained Values for Iris Plant Dataset					
Best	2.9800	3.3094	1.0343	0.0543	
Worst	2.2846	1.6018×10^{306}	4.0205×10^{305}	0.0440	
V			0.8707		
Obtained values for Wisconsin Breast Cancer Dataset					
Best	3.2452	0.4773	0.3837	0.6072	0.7001
	0.5530	1.4196	0.3803	0.4494	
Worst	2.4528×10^{304}	0.2938	0.2169	0.4762	0.4199
	0.3523	2.5194×10^{305}	0.3	4.4899×10^{305}	
V			0.9685		
Obtained values for Occupancy Detection Dataset					
Best	0.5102	0.5625	0.4983	2.9965	1.6872
Worst	0.3088	0.3840	2.6569×10^{304}	1.6602	1.4436
V			0.9947		

The matrix indicated as best in Table 1 is the value that can perform the clustering process best. The best and worst matrices represent the equivalents of the positive afterimage and negative afterimage values in the afterimage algorithm. They are used to calculate the new values in Equation (10). Figure 6 shows the confusion matrix for datasets.

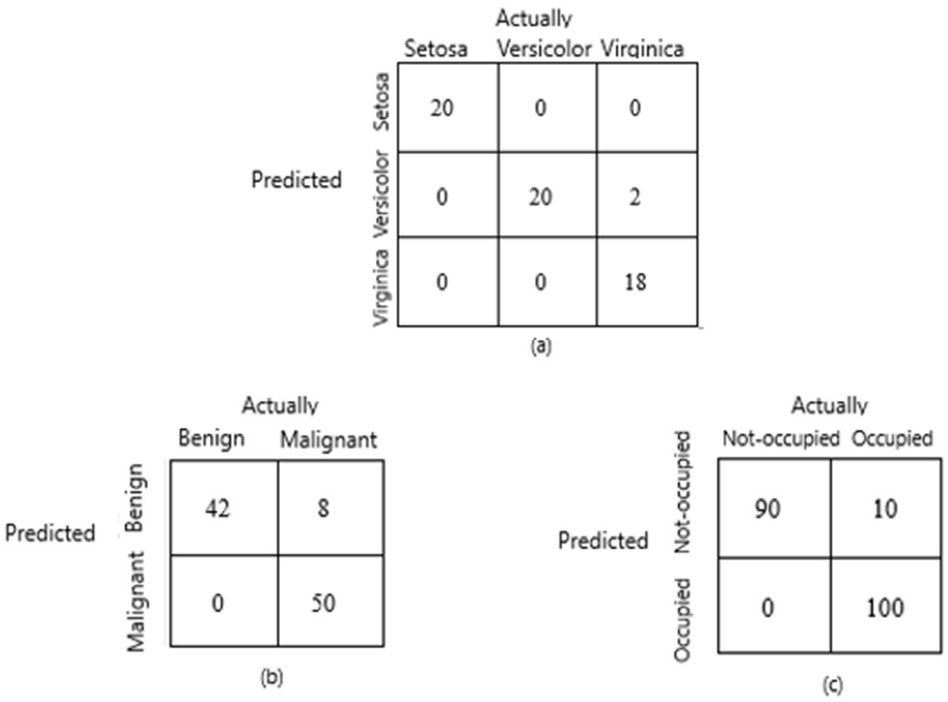


Figure 6. (**a**) Confusion matrix results of iris plant dataset, (**b**) confusion matrix results of Wisconsin breast cancer dataset, (**c**) confusion matrix results of occupancy detection dataset.

Figure 6 shows the confusion matrices of all three datasets. As can be seen from the figure, there are three clusters for the iris plant dataset. Only for one cluster, 2 out of 20 values are clustered incorrectly. There are two clusters for the Wisconsin breast cancer original dataset. Only for one cluster, 8 out of 50 values are clustered incorrectly. There are two clusters for the occupancy detection dataset. Only for one cluster, 10 out of 100 values are clustered incorrectly.

Tables 2–4 show the cluster interval values of the test data of the datasets studied. These intervals are the values obtained according to the distances of the best matrix obtained as a result of the algorithm's operation to the test data.

Table 2. Cluster interval values of iris plant dataset.

Number	Setosa	Versi Color	Virginica	Number	Setosa	Versi Color	Virginica
1	2.1410	6.2408	**8.8408 (error)**	11	2.9220	8.3408	10.4408
2	2.2410	7.0408	**9.3408 (error)**	12	2.9220	8.3408	10.4408
3	2.3220	7.2408	9.5408	13	3.0220	8.5408	10.5408
4	2.5410	7.2408	9.6408	14	3.0220	8.6408	10.9408
5	2.6410	7.5408	9.8408	15	3.3220	8.8408	11.0408
6	2.6410	7.5408	9.9408	16	3.3220	9.0408	11.7408
7	2.7220	7.5408	10.0408	17	3.3220	9.2408	12.5408
8	2.7220	7.8408	10.1408	18	3.4220	9.4408	12.7220
9	2.7410	7.8408	10.3220	19	4.0220	9.4408	12.8408
10	2.8220	7.9408	10.4408	20	4.1220	9.4408	13.0220

Table 3. Cluster interval values of Wisconsin breast cancer original dataset.

Number	Benign	Malignant	Number	Benign	Malignant
1	5.2746	19.6232	26	10.1137	45.7841
2	5.6232	19.7841	27	10.2746	46.7841
3	5.7841	23.7841	28	10.2746	46.7841
4	6.1137	25.7841	29	10.7841	47.7841
5	6.2746	26.2746	30	11.1137	47.7841
6	6.7841	27.7841	31	11.2746	48.7841
7	7.1137	30.7841	32	11.7841	49.7841
8	7.1137	32.7841	33	12.7841	50.7841
9	7.2746	35.7841	34	12.7841	51.7841
10	7.2746	35.7841	35	13.1137	51.7841
11	7.2746	35.7841	36	13.6232	51.7841
12	7.7841	35.7841	37	13.7841	52.2746
13	7.7841	36.7841	38	13.7841	52.7841
14	7.7841	37.7841	39	13.7841	54.7841
15	8.1137	37.7841	40	14.7841	54.7841
16	8.2746	38.7841	41	16.7841	54.7841
17	8.7841	38.7841	42	17.2746	55.7841
18	8.7841	39.7841	43	**24.7841 (error)**	56.2746
19	8.7841	40.7841	44	**28.2746 (error)**	57.7841
20	8.7841	40.7841	45	**29.7841 (error)**	61.7841
21	8.7841	42.2746	46	**32.7841 (error)**	61.7841
22	9.6232	42.7841	47	**32.7841 (error)**	65.7841
23	9.7841	42.7841	48	**32.7841 (error)**	66.7841
24	9.7841	42.7841	49	**38.7841 (error)**	72.7841
25	9.7841	44.7841	50	**40.7841 (error)**	75.7841

If Table 2 is examined carefully, the clustering range for the iris setosa cluster is between 2.1410 and 4.1220. The clustering range for the iris versi color cluster is between 6.2408 and 9.4408. The clustering range for the iris virginica cluster is between 9.5408 and 13.0220. In this case, two values are determined as error for the iris versi color cluster.

If Table 3 is examined carefully, the clustering range for the benign cluster is between 5.2746 and 17.2746. The clustering range for the malignant cluster is between 19.6232 and 75.7841. In this case, the value of eight for the benign cluster is clustered incorrectly.

Table 4. Cluster intervals of occupancy detection dataset.

Number	Not Occupied	Occupied	Number	Not Occupied	Occupied	Number	Not Occupied	Occupied	Number	Not Occupied	Occupied
1	466.8480	926.3340	26	523.2060	1027.5710	51	596.6010	1243.2140	76	623.5500	1627.5490
2	467.8430	943.3040	27	526.7020	1034.6440	52	598.6010	1247.8140	77	623.9110	1637.8820
3	468.6930	946.6980	28	541.2010	1039.9820	53	600.1010	1250.0890	78	624.5110	1657.3050
4	472.3480	950.2870	29	576.4060	1072.1770	54	600.1010	1258.3080	79	625.3610	1658.1600
5	472.8930	950.3380	30	576.9710	1075.8490	55	602.1010	1267.4640	80	627.4110	1662.2300
6	472.9530	950.4340	31	577.9510	1101.9940	56	602.1380	1273.8660	81	627.4110	1686.9500
7	476.4960	951.9040	32	579.4110	1116.8770	57	603.2110	1274.5120	82	627.5110	1693.2050
8	477.0030	952.4770	33	580.1750	1121.7310	58	603.6290	1276.6590	83	629.4360	1701.6720
9	477.5030	954.3370	34	581.4560	1122.8990	59	604.2110	1285.8090	84	632.1110	1745.8050
10	477.7580	954.5040	35	583.0010	1149.3040	60	604.6560	1306.8570	85	672.5570	1854.2280
11	485.8130	958.0990	36	584.5010	1151.4010	61	605.2110	1307.8070	86	703.1560	1958.2440
12	486.4700	960.3380	37	584.6680	1155.8240	62	607.2110	1322.4140	87	734.06890	1960.0500
13	488.4790	961.9040	38	585.0010	1165.7570	63	609.3110	1324.8730	88	837.8160	2055.0030
14	489.1120	963.0710	39	585.0010	1171.0710	64	609.3110	1332.4160	89	840.1180	2102.2000
15	490.4020	963.2410	40	587.0010	1189.4820	65	610.4110	1333.3150	90	841.6300	2116.3260
16	491.7920	964.4540	41	589.0010	1191.0320	66	610.8460	1334.2330	91	**936.6120 (error)**	2214.3370
17	492.9580	966.0770	42	589.0010	1192.9990	67	611.7450	1351.6910	92	**1115.2950 (error)**	2277.2730
18	495.5420	972.6140	43	589.0560	1194.8170	68	611.8610	1352.0060	93	**1268.7300 (error)**	2284.7810
19	495.8130	984.8370	44	593.5060	1199.3820	69	612.7110	1367.2500	94	**1276.5020 (error)**	2286.3630
20	496.9500	995.2390	45	594.5510	1205.7640	70	613.6780	1510.4600	95	**1541.3150 (error)**	2309.3360
21	499.4130	996.3590	46	594.6010	1211.5070	71	614.4110	1517.4150	96	**1548.8480 (error)**	2347.6400
22	501.1030	998.5140	47	595.1010	1217.0640	72	615.4110	1561.8650	97	**1623.3480 (error)**	2454.3030
23	511.1030	1002.9100	48	595.6010	1219.6570	73	617.2110	1582.7950	98	**1633.5000 (error)**	2501.5470
24	515.1760	1005.2400	49	595.6010	1225.8740	74	619.9110	1592.6050	99	**1644.2050 (error)**	2503.3430
25	518.2020	1019.1110	50	595.7350	1230.7320	75	621.9110	1599.0380	100	**1754.6550 (error)**	2518.8030

If Table 4 is examined carefully, the clustering range for the not occupied cluster is between 466.8480 and 841.6300. The clustering range for the occupied cluster is between 926.3340 and 2518.8030. In this case, 10 values are clustered incorrectly for the not occupied cluster.

Figures 7–9 show graphical representations of clustering intervals for the iris plant dataset, Wisconsin breast cancer original dataset, and occupancy detection dataset.

If the figures are examined carefully, it will be seen that the values detected as errors for the clustering intervals are marked. Two points are marked in Figure 2, eight points in Figure 3, and ten points in Figure 4. These points are the points where there are error values for three datasets. The clustering interval for these remained within the range of another cluster. Therefore, an error occurred at these points.

Table 5 shows the performance metrics of the datasets. The values of the performance metrics formulated in Equations (4)–(7) were calculated separately for each of the three datasets and are presented in Table 5.

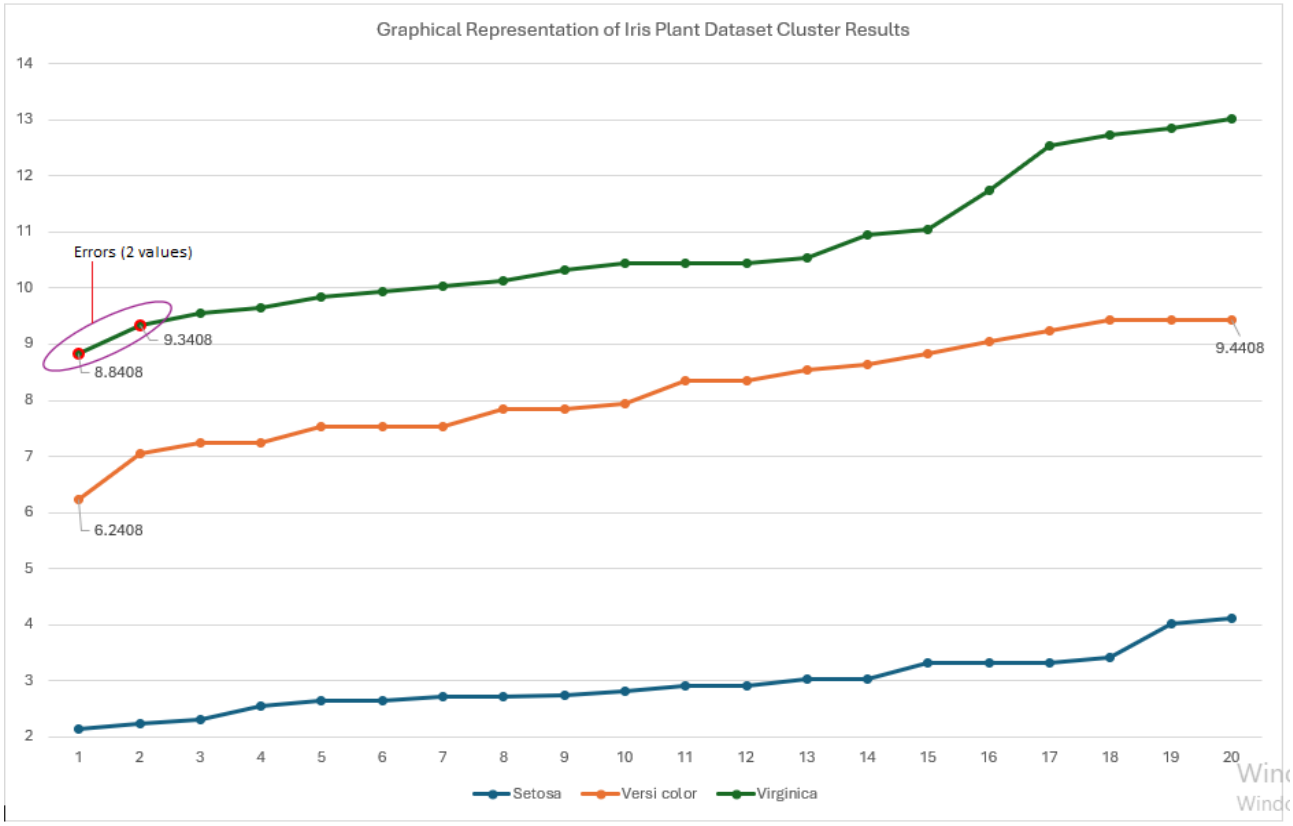


Figure 7. Graphical representation of iris plant dataset cluster results.

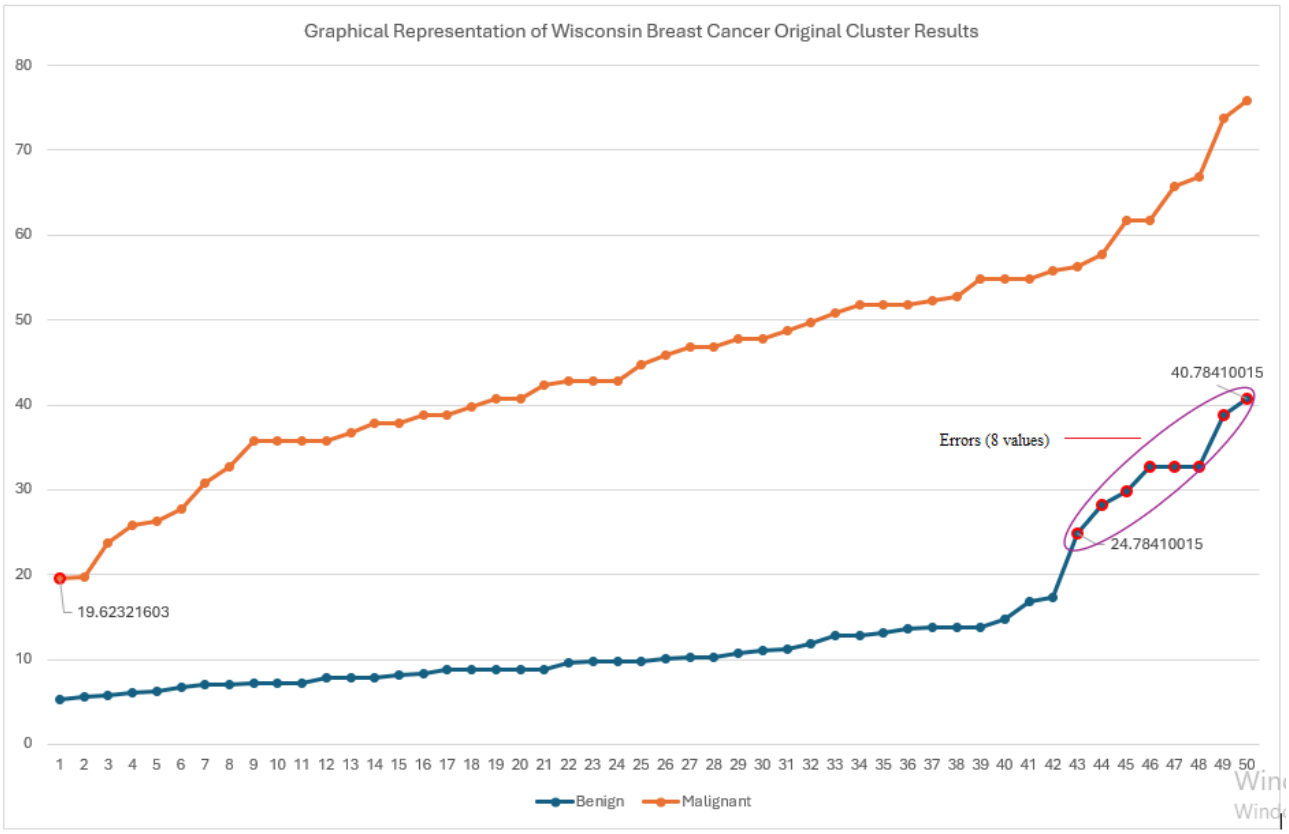


Figure 8. Graphical representation of Wisconsin breast cancer original cluster results.

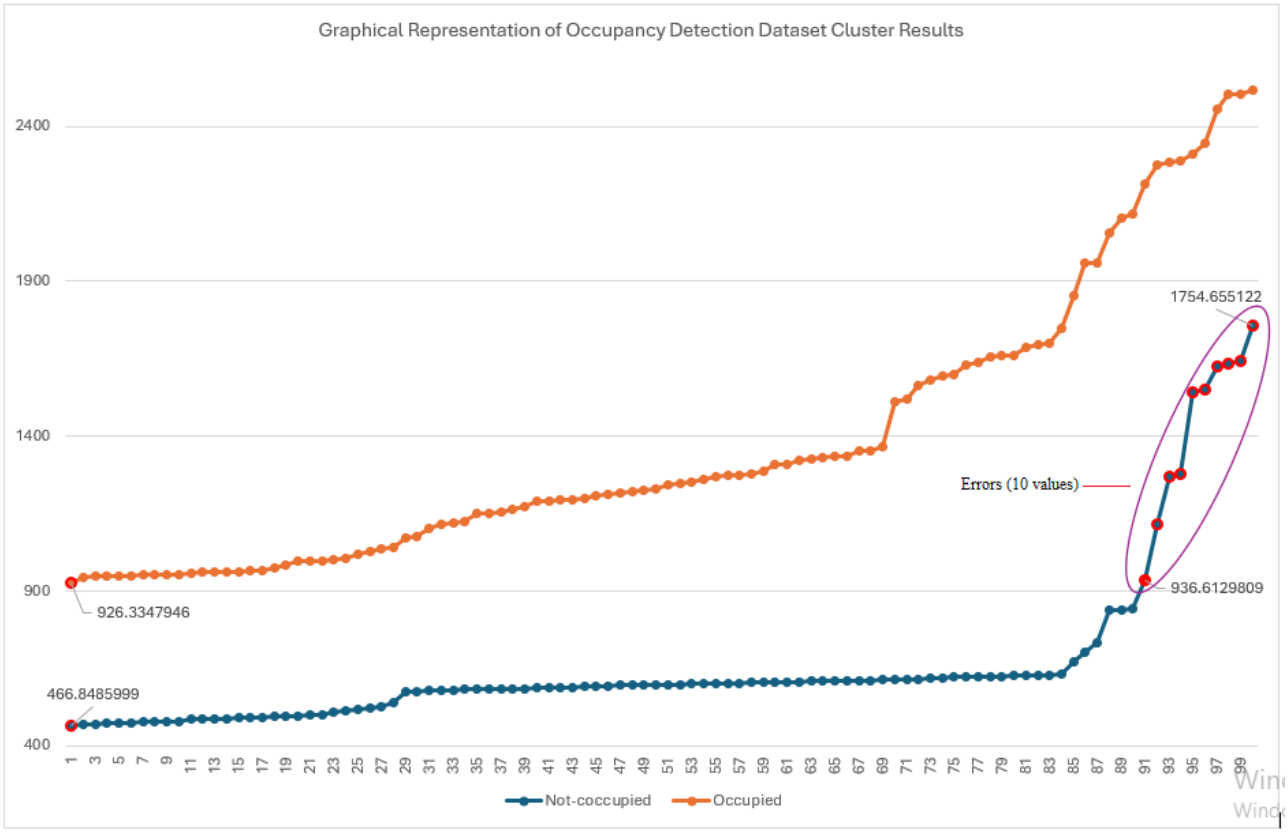


Figure 9. Graphical representation of occupancy detection dataset cluster results.

Figure 10 shows the graphical representation of the performance metrics calculated for all datasets.

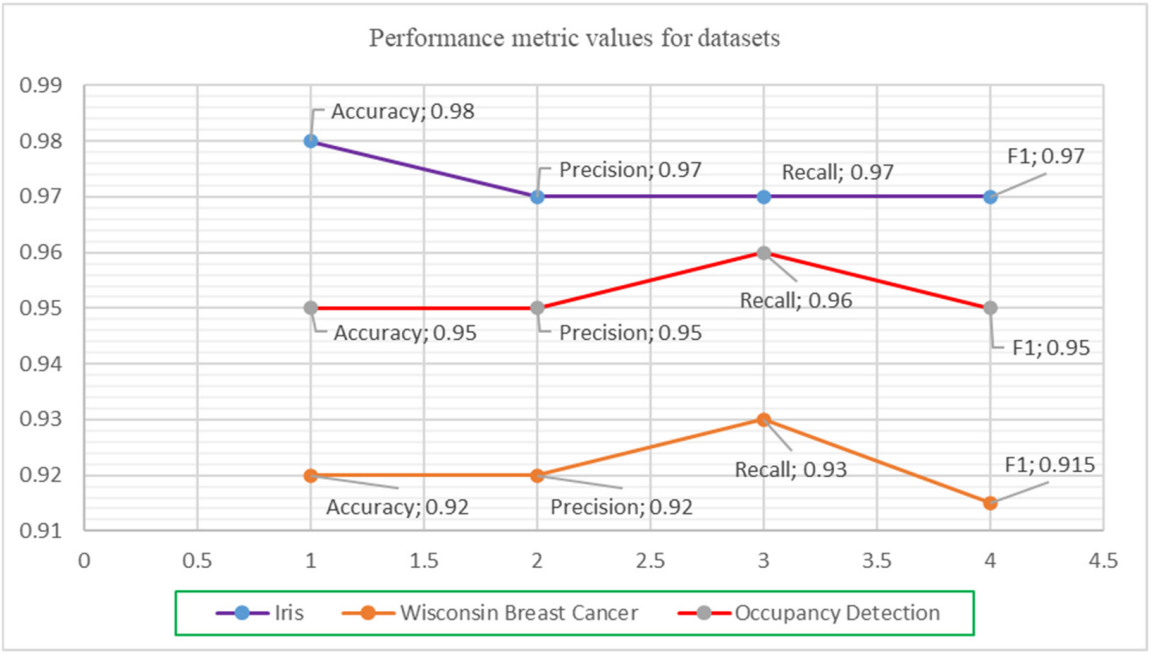


Figure 10. Performance metrics values for datasets.

Table 5. Performance metrics of datasets.

	Iris Plant Dataset			
	Accuracy	**Precision**	**Recall**	**F1**
Iris Setosa	1	1	1	1
Iris Versicolor	0.97	0.91	1	0.96
Iris Virginica	0.97	1	0.9	0.95
Average	**0.98**	**0.97**	**0.97**	**0.97**
	Wisconsin Breast Cancer Dataset			
Benign	0.92	0.84	1	0.91
Malignant	0.92	1	086	0.92
Average	**0.92**	**0.92**	**0.93**	**0.915**
	Occupancy Detection Dataset			
Not Occupied	0.95	0.9	1	0.95
Occupied	0.95	1	0.91	0.95
Average	**0.95**	**0.95**	**0.96**	**0.95**

5. Comparison of AAIA with Methods in the Literature

A comparison of the datasets studied in this publication, and the clustering studies using metaheuristic methods is given in Table 3. There are many methods in the literature such as clustering methods. However, the methods given in Table 3 are examples of clustering studies, especially those conducted with metaheuristic methods.

In addition, in order to make comparisons with studies in the literature, methods that evaluate with accuracy, precision, recall, and F1 metrics should be selected, because if your metrics are different, it will not be correct to compare with other methods.

If Table 6 is examined carefully, although the proposed AAIA algorithm is a newly proposed algorithm, its experimental results are quite successful when compared with other metaheuristic methods.

Table 6. Comparison of AAIA with methods in literature.

	Iris Dataset			
	Accuracy	Precision	Recall	F1
AAIA	0.98	0.97	0.97	0.97
Elfarra et al. [45]	0.90	-	-	-
Shial et al. [46]	0.97	-	-	0.97
Prakash et al. [47]	-	-	-	0.82
Leela et al. [48]	0.85	-	-	-
Huang and Gel [49]	0.78	-	-	-
	Wisconsin Breast Cancer Original Dataset			
AAIA	0.92	0.92	0.93	0.915
Elfarra et al. [45]	0.70	-	-	-
Ayoob [50]	0.97	-	-	-
Dubey et al. [51]	0.92	-	-	-
Pantazi et al. [52]	0.95	-	-	-
	Occupancy Detection Dataset			
AAIA	0.98	0.98	0.98	0.97
Fährmann et al. [53]	0,96	-	-	-
Prabhakaran et al. [54]	-	0.78	0.62	0.85
Huang and Gel [49]	0.77	-	-	-

6. Discussion

In order to perform clustering analysis on data more accurately, a number of normalization operations are performed. The Z-score method is one of them [55]. Q-Q plot graphs are plotted using the Z-score values of an attribute in the dataset. When the expected and actual values overlap, a line appears on the graph at a 45-degree angle. In a normal distribution, all expected values are on or very close to the 45-degree line [56].

Figures 11–13 show the Q-Q plot distributions of the datasets.

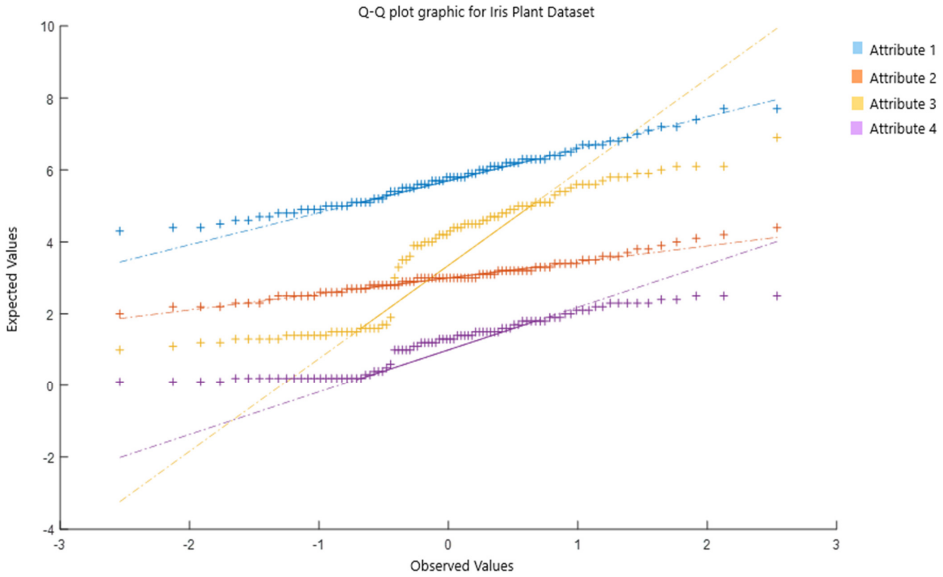


Figure 11. Q-Q plot graphic for iris plant dataset.

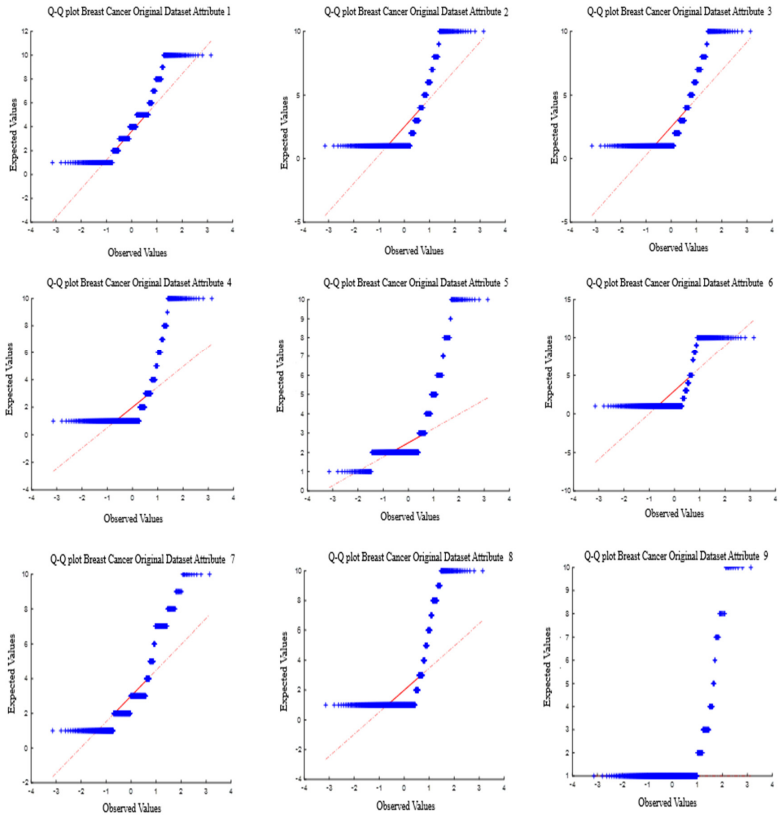


Figure 12. Q-Q plot graphics for breast cancer original dataset.

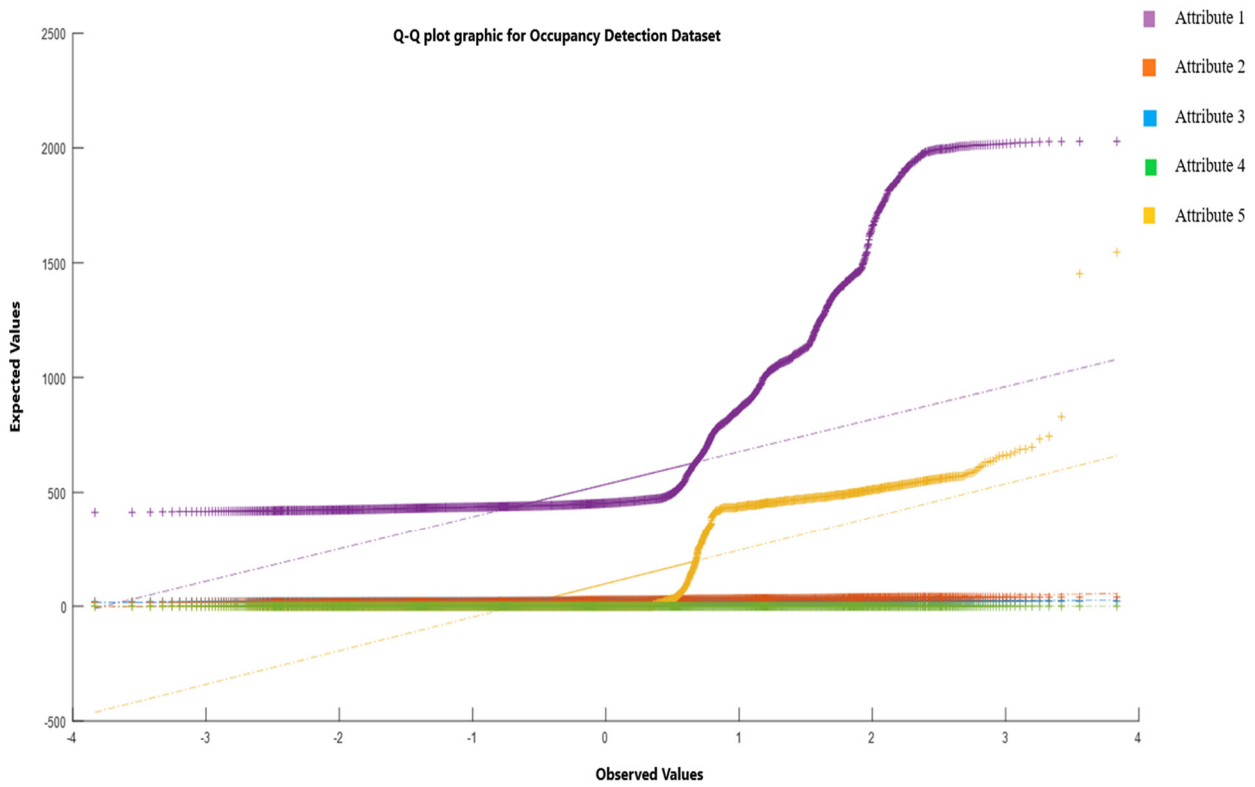


Figure 13. Q-Q plot graphic for occupancy detection dataset.

Figure 12 is the Q-Q plot graphs of the breast cancer original dataset. Separate graphs are given for each attribute. Because there are too many discrete values, when plotted on the same graph, the points overlap. Therefore, the graph looks meaningless. Therefore, it is plotted separately for each attribute. If Figures 11–13 are examined carefully, it is seen that there are many data points separated from the 45-degree axis line for each dataset. These points are meaningful points to separate the clusters from each other. If all the points are distributed around the axis, the effects of the attributes will not emerge. So, it is seen that there are attribute distributions that can make this distinction in all three datasets. It will be advantageous to take this issue into consideration in algorithms that use the distance-based clustering approach. Otherwise, when choosing the clustering method, grid-based or density-based clustering methods can be selected instead of distance-based. This clustering will affect performance.

7. Conclusions

With the rapid advancement of technology, the use of digital technologies in every field is increasing. Today, digital technologies are used in all fields such as engineering, medicine, finance, entertainment, health, etc. The data used is also increasing with these developments. The increase in data is both in terms of abundance and size. This also means that the volume of data to be processed is growing every day. It is also becoming very difficult to solve the relationships between data and parameters in very large volumes of data.

Metaheuristic methods provide great advantages in such problems that are very difficult to solve with classical mathematical methods, or they cannot be solved. If we have consistent and sufficient data, metaheuristic methods can approach the result. In cases where a solution cannot be obtained with classical methods, metaheuristic methods can reach the result through training from the existing data. Metaheuristic methods act with the approach of resembling the best. For this reason, they try to resemble the best solution in

each iteration. This approach is a method that can be applied to every discipline. Therefore, solutions can be easily developed for the solution of problems in every discipline where data can be obtained. This makes them more widely used and popular.

In this article, the artificial afterimage algorithm method inspired by the phenomenon of vision has been developed. Experimental studies have been conducted on the clustering problem, which is applied very frequently in literature. The datasets used are consistent datasets that are used very frequently. The results obtained are very successful. The amount of data in the first and second datasets is smaller than the third. However, the algorithm has achieved a good success rate in the third dataset, which contains more data.

In future studies, experimental studies can be conducted on datasets that contain more data and are larger in size. The fact that the artificial afterimage algorithm has not been studied in the literature before makes the algorithm open to development in many ways.

Funding: This research received no external funding.

Data Availability Statement: No new data was created in this study. The data studied were obtained from open access databases on websites.

Conflicts of Interest: The authors state that they have no known competing financial interests or personal ties that could have influenced the research presented in this study.

References

1. Creosteanu, A.; Gavrila, G.; Creosteanu, L. Comparison between an analytical method and two numerical methods on a given electrostatic potential determination problem. In Proceedings of the 15 International Symposium on Antenna Technology and Applied Electromagnetics, Toulouse, France, 25–28 June 2012; pp. 1–6. [CrossRef]
2. Köse, B.; Ansay, S.; Pektaş, M. Veri, Enformatik, Yapay Zeka ve Optimizasyon. *Kuantum Teknol. Enformatik Araştirmalari* **2023**, *1*, 35–40. [CrossRef]
3. Dirik, M. Comparison of Recent Meta-Heuristic Optimization Algorithms Using Different Benchmark Functions. *J. Math. Sci. Model.* **2022**, *5*, 113–124. [CrossRef]
4. Rajwar, K.; Deep, K.; Das, S. An exhaustive review of the metaheuristic algorithms for search and optimization: Taxonomy, applications, and open challenges. *Artif. Intell. Rev.* **2023**, *56*, 13187–13257. [CrossRef] [PubMed]
5. Li, C.; Zhang, Y.; Chen, X. Heuristic Clustering Based on Centroid Learning and Cognitive Feature Capturing. *Math. Probl. Eng.* **2019**, *2019*, 1530618. [CrossRef]
6. Pattanayak, P.K.; Trıpathy, R.M.; Padhy, S. A Novel Heuristic for Graph-Based Topic Modeling Using Spectral Clustering. *J. Theor. Appl. Inf. Technol.* **2024**, *102*, 664–672.
7. Shaikh, M.S.; Dong, X.; Zheng, G.; Wang, C.; Lin, Y. An Improved Expeditious Meta-Heuristic Clustering Method for Classifying Student Psychological Issues with Homogeneous Characteristics. *Mathematics* **2024**, *12*, 1620. [CrossRef]
8. Tsai, C.-W.; Song, H.-J.; Chiang, M.-C. A hyper-heuristic clustering algorithm. In Proceedings of the 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Republic of Korea, 14–17 October 2012; pp. 2839–2844. [CrossRef]
9. Christou, I.T.; Psathas, F.; Rentizelas, A.; Papadakis, A.; Georgiou, P.; Anastasopoulos, D.; Lappas, P.Z. A heuristic for improving clustering in biomass supply chains. *Int. J. Syst. Sci. Oper. Logist.* **2024**, *11*, 2378859. [CrossRef]
10. Lamari, Y. New Heuristics for Clustering Big Data. Ph.D. Thesis, Faculty of Scineces, University Ibn Zohr, Agadir, Morocco, 2018.
11. Alam, M.; Sadaf, K. Web Search Result Clustering based on Heuristic Search and k-means. *arXiv* **2015**, arXiv:abs/1508.02552. [CrossRef]
12. Al-Betar, M.A.; Abasi, A.K.; Al-Naymat, G.; Arshad, K.; Makhadmeh, S.N. Bare-Bones Based Salp Swarm Algorithm for Text Document Clustering. *IEEE Access* **2023**, *11*, 100010–100028. [CrossRef]
13. Schnoering, H.; Porthaux, P.; Vazirgiannis, M. Assessing the Efficacy of Heuristic-Based Address Clustering for Bitcoin. *arXiv* **2024**, arXiv:abs/2403.00523. [CrossRef]
14. Puri, D.; Gupta, D. A novel linear time clustering using heuristically improved mrk-medoids based on modified squirrel search algorithm. *Aust. J. Electr. Electron. Eng.* **2024**, *21*, 358–373. [CrossRef]
15. Kumar, G.K.; Prashanth, S.K.; Padmalatha, E.; Reddy, M.V.K.; Devi, N.R.; Abualigah, L.; Chithaluru, P.; Kumar, M. An optimized meta-heuristic clustering-based routing scheme for secured wireless sensor networks. *Int. J. Commun. Syst.* **2024**, *37*, e5791. [CrossRef]
16. Gao, C.; Yong, M.; Gao, Y.; Li, T. An improved black hole algorithm designed for K-means clustering method. *Complex Intell. Syst.* **2024**, *10*, 5083–5106. [CrossRef]

17. Diao, Q.; Junaidi, A.; Chan, W.; Zain, A.M.; Yang, H. SBOA: A Novel Heuristic Optimization Algorithm. *Baghdad Sci. J.* **2024**, *21*, 764–779. [CrossRef]
18. Masoumi, M.; Behnamian, J. A New Heuristic Algorithm Based on Minimum Spanning Tree for Solving Metric Traveling Salesman Problem. *Int. J. Ind. Eng. Prod. Res.* **2024**, *35*, 1–13. [CrossRef]
19. Fang, W.; Jiang, H.; Lu, H.; Sun, J.; Wu, X.; Lin, J. GPU-Based Efficient Parallel Heuristic Algorithm for High-Utility Itemset Mining in Large Transaction Datasets. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 652–667. [CrossRef]
20. Qu, C.; Peng, X.; Zeng, Q. Learning search algorithm: Framework and comprehensive performance for solving optimization problems. *Artif. Intell. Rev.* **2024**, *57*, 139. [CrossRef]
21. Nemati, M.; Zandi, Y.; Agdas, A.S. Application of a novel metaheuristic algorithm inspired by stadium spectators in global optimization problems. *Sci. Rep.* **2024**, *14*, 3078. [CrossRef]
22. Cao, S.; Qian, Q.; Cao, Y.; Li, W.; Huang, W.; Liang, J. A Novel Meta-Heuristic Algorithm for Numerical and Engineering Optimization Problems: Piranha Foraging Optimization Algorithm (PFOA). *IEEE Access* **2023**, *11*, 92505–92522. [CrossRef]
23. Mirkhan, A.; Çelebi, N. A hybrid model of heuristic algorithm and gradient descent to optimize neural networks. *Bull. Pol. Acad. Sci. Tech. Sci.* **2023**, *71*, e147924. [CrossRef]
24. Sarkheil, H.; Hosseini, M.H. Heuristic Simulated Annealing Modeling to Optimum Target Audience Identification in Digital Marketing: A Case Study of a Mining Industry Training Service Company. *Int. J. Supply Oper. Manag.* **2024**, *11*, 188–202. [CrossRef]
25. Zhou, G.; Zhang, T.; Zhou, Y. Elite Opposition-Based Bare Bones Mayfly Algorithm for Optimization Wireless Sensor Networks Coverage Problem. *Arab. J. Sci. Eng.* **2024**, *50*, 719–739. [CrossRef]
26. Wang, J.; Wang, W.; Hu, X.; Qiu, L.; Zang, H. Black-winged kite algorithm: A nature-inspired meta-heuristic for solving benchmark functions and engineering problems. *Artif. Intell. Rev.* **2024**, *57*, 98. [CrossRef]
27. Amiri, M.H.; Hashjin, N.M.; Montazeri, M.; Mirjalili, S.; Khodadadi, N. Hippopotamus optimization algorithm: A novel nature-inspired optimization algorithm. *Sci. Rep.* **2024**, *14*, 5032. [CrossRef]
28. Zhang, W.; Pan, K.; Li, S.; Wang, Y. Special Forces Algorithm: A novel meta-heuristic method for global optimization. *Math. Comput. Simul.* **2023**, *213*, 394–417. [CrossRef]
29. Trojovský, P.; Dehghani, M. A new bio-inspired metaheuristic algorithm for solving optimization problems based on walruses behavior. *Sci. Rep.* **2023**, *13*, 8775. [CrossRef]
30. Yin, S.; Luo, Y.; Zhou, Y. EOSMA: An equilibrium optimizer slime mould algorithm for engineering design problems. *Arab. J. Sci. Eng.* **2022**, *47*, 2. [CrossRef]
31. Bender, M.B.; Feldman, M.; Sobın, A.J. Palinopsia. *Brain J. Neurol.* **1968**, *91*, 321–338. [CrossRef]
32. Gersztenkorn, D.; Lee, A.G. Palinopsia revamped: A systematic review of the literatüre. *Surv. Ophthalmol.* **2014**, *60*, 1–35. [CrossRef]
33. Lőrinc, G.; András, H. Analyzing afterimage strength and duration test results with k-means clustering. In *Szemelvények a BGE Kutatásaiból II. Kötet*; Budapesti Gazdasági Egyetem: Budapest, Hungary, 2023; pp. 103–112. ISBN 978-615-6342-76-8.
34. Lee, H.C. *Introduction to Color Imaging Science*; Cambridge University Press: Cambridge, UK, 2005.
35. Zaidi, Q.; Ennis, R.; Cao, D.; Lee, B. Neural locus of color afterimages. *Curr. Biol.* **2012**, *22*, 220–224. [CrossRef]
36. Visual Angle-Wikipedia. Available online: https://en.wikipedia.org/wiki/Visual_angle (accessed on 20 November 2024).
37. Yakovyna, V.; Shakhovska, N.; Szpakowska, A. A novel hybrid supervised and unsupervised hierarchical ensemble for COVID-19 cases and mortality prediction. *Sci. Rep.* **2024**, *14*, 9782. [CrossRef]
38. Velunachiyar, S.; Sivakumar, K. Some Clustering Methods, Algorithms and their Applications. *Int. J. Recent Innov. Trends Comput. Commun.* **2023**, *11*, 401–410. [CrossRef]
39. Çatak, F.Ö. Development of Data Mining Software Framework by Using Map/Reduce Method in Cloud Computing Systems. Ph.D. Thesis, İstanbul University Institute of Science, Department of Electronics, Informatics Program, İstanbul, Turkey, 2014.
40. Blanco-Maloo, E.; Morán-Fernández, L.; Remeseiro, B.; Bolón-Canedo, V. Do all roads lead to Rome? Studying distance measures in the context of machine learning. *Pattern Recognit. Vol.* **2023**, *141*, 109646. [CrossRef]
41. Göde, A.; Kalkan, A. Performance comparison machine learning algorithms in diabetes disease prediction. *Eur. Mech. Sci.* **2023**, *7*, 178–183. [CrossRef]
42. Iris–UCI Machine Learning Repository. Available online: https://archive.ics.uci.edu/dataset/53/iris (accessed on 25 November 2024).
43. Occupancy Detection-UCI Machine Learning Repository. Available online: https://archive.ics.uci.edu/dataset/357/occupancy+detection (accessed on 25 November 2024).
44. Breast Cancer Wisconsin (Original)-UCI Machine Learning Repository. Available online: https://archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original (accessed on 25 November 2024).
45. Elfarra, B.K.; Salaha, M.A.A.; Ashour, W.M. Black Hole Clustering: Gravity-Based Approach with No Predetermined Parameters. *Data Sci. J.* **2024**, *23*, 27. [CrossRef]

46. Shial, G.; Sahoo, S.; Panigrahi, S. A Nature-Inspired Hybrid Partitional Clustering Method Based on Grey Wolf Optimization and Jaya Algorithm. *Comput. Sci.* **2023**, *24*, 361–405. [CrossRef]
47. Prakash, K.L.; Suryanarayana, G.; Swapna, N.; Bhaskar, T.; Kiran, A. Optimizing K-Means Clustering using the Artificial Firefly Algorithm. *Int. J. Intell. Syst. Appl. Eng.* **2023**, *11*, 461–468.
48. Leela, V.; Sakthipriya, K.; Manikandan, R. Comparative Study of Clustering Techniques in Iris Data Sets. *World Appl. Sci. J.* **2014**, *29*, 24–29.
49. Huang, X.; Gel, Y.R. CRAD: Clustering with Robust Autocuts and Depth. In Proceedings of the IEEE International Conference on Data Mining, New Orleans, LA, USA, 18–21 November 2017; pp. 925–930. [CrossRef]
50. Ayoob, N.K. Breast Cancer Diagnosis Using K-means Methodology. *J. Babylon Univ./Pure Appl. Sci.* **2018**, *26*, 9–16. [CrossRef]
51. Dubey, A.K.; Gupta, U.; Jain, S. Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. *Int. J. Comput. Assist. Radiol. Surg.* **2016**, *11*, 2033–2047. [CrossRef]
52. Pantazi, S.; Kagolovsky, Y.; Moehr, J.R. Cluster Analysis of Wisconsin Breast Cancer Dataset Using Self-Organizing Maps. *Stud. Health Technol. Inform.* **2002**, *90*, 431–436. [PubMed]
53. Fährmann, D.; Boutros, F.; Kubon, P.; Kirchbuchner, F.; Kuijper, A.; Damer, N. Ubiquitous multi-occupant detection in smart environments. *Neural Comput. Appl.* **2023**, *36*, 2941–2960. [CrossRef]
54. Prabhakaran, K.; Dridi, J.; Amayri, M.; Bouguila, N. Explainable K-Means Clustering for Occupancy Estimation. *Procedia Comput. Sci.* **2022**, *203*, 326–333. [CrossRef]
55. Şen, A.; Gökgöz, T. Kümelemede Normalleştirmenin Etkisi. In Proceedings of the TMMOB Harita ve Kadastro Mühendisleri Odası, 14. Türkiye Harita Bilimsel ve Teknik Kurultayı, Ankara, Turkey, 14–17 May 2013.
56. Can, A. *SPSS ile Bilimsel Araştırma Sürecinde Nicel Veri Analizi*, 7th ed.; Pegem Akademi: Ankara, Turkey, 2019.