Turkish Journal of Science & Technology
20(1), 299-307, 2025

Research Paper
https://doi.org/10.55525/tjst.1630246

Risk Detection in Medical Data Using Artificial Afterimage Algorithm

Murat DEMİR[1*]
[1] Department of Software Engineering, Faculty of Engineering-Architecture, Muş Alparslan University, Muş, Türkiye
[*1] m.demir@alparslan.edu.tr

(Geliş/Received: 30/01/2025; Kabul/Accepted: 06/03/2025)

Abstract: Metaheuristic algorithms have become a frequently used methodology in many fields such as genetics, bioinformatics, microbiology, etc. related to human health. Metaheuristic methods provide efficient solutions when classical approaches fail or are computationally expensive. In this study, Artificial Afterimage Algorithm was applied to 4 different medical data sets obtained from Kaggle. There is no previous study in the literature that models the afterimage algorithm as a heuristic method. Its mathematical infrastructure is simpler than many other methods. Using the Artificial Afterimage Algorithm, clusters of test samples taken from healthy individuals and patients were tried to be detected. Accuracy, precision, recall and F1 values of the clusters were calculated. The highest Accuracy value was obtained as 0.85, Precision value as 0.9, Recall value as 1 and F1 score value as 0.86. The study shows that the method can perform a good rate of risk detection in medical data.

Key words: Metaheuristic, Artificial Afterimage Algorithm, Clustering.

Yapay Afterimage Algoritması Kullanarak Medikal Verilerde Risk Tespiti

Öz: Metaheuristic algoritmalar, insan sağlığını ilgilendiren; genetik, biyoinformatik, mikrobiyoloji vb. birçok alanda sıkça kullanılan bir metodoloji olmuştur. Metaheuristic yöntemler, klasik yaklaşımlar başarısız olduğunda veya hesaplama açısından pahalı olduğunda verimli çözümler sunar. Bu çalışmada Yapay Afterimage Algoritması, Kaggle'dan elde edilen 4 ayrı medikal veri seti üzerine uygulanmıştır. Literatürde daha önce Afterimage Algoritmasını sezgisel bir yöntem olarak modelleyen bir çalışma yoktur. Matematiksel alt yapısı diğer birçok yönteme göre daha basittir. Yapay Afterimage Algoritması kullanılarak, sağlıklı kişiler ve hastalardan alınan test örneklerinin kümeleri tespit edilmeye çalışılmıştır. Kümelere ait Accuracy, Precision, Recall ve F1 değerleri hesaplanmıştır. En yüksek Accuracy değeri 0,85, Precision değeri 0,9, Recall değeri 1 ve F1 skor değeri 0,86 olarak elde edilmiştir. Çalışma göstermektedir ki, yöntem medikal verilerde iyi bir oranda risk tespitini gerçekleştirebilmektedir.

Anahtar kelimeler: Metasezgisel, Yapay Afterimage Algoritması, Kümeleme.

1. Introduction

The development of technology has increased the use of digital devices in all areas. The spread of digital technologies has made this technology a driving factor in innovation and transformation in the scientific world. The digital tools that information technologies provide to the digital world form the basis for this situation [1]. In parallel with these developments, a very critical increase in data density has emerged. Big data is being recorded in many areas day by day.

One of the most important areas that has made significant progress in the field of digital technologies is artificial intelligence technology. Artificial intelligence, which affects every area of life in daily life, is gaining more and more importance every day. Apart from daily activities and business life activities, it also significantly affects our perspective on problems, our understanding of entertainment and our lifestyles [2].

Artificial intelligence was first presented as a proposal letter by John McCarthy, Marvin L. Minsky, Nathaniel Rochester and Claude E. Shannon at the Dortmund Conference in 1956 [3]. However, John McCarthy is considered the inventor of artificial intelligence. McCarthy [4] defined artificial intelligence as "the science and engineering of creating human-like intelligent machines, especially intelligent computer programs."

As in every field, very critical volumes of data are generated in studies concerning the field of health. A digital health system that integrates with traditional methods is being tried to be put forward in health technologies. Since this development is very rapid, the volume of investment in digital technologies in the health sector is also increasing [5]. Artificial intelligence accelerates many processes such as diagnosis, treatment and preventive health services in the field of health [6].

[*] Corresponding author: m.demir@alparslan.edu.tr ORCID Number of author: [1] 0000-0001-7362-0401

1.1. Metaheuristic Methods

This field of study, known as metaheuristic methods, is actually artificial learning studies. They try to product solutions to problems by imitating systems, events and living things in nature. It is expressed as inspired by nature [7]. Although there are many types; evolutionary algorithms, physics-based algorithms, swarm-based algorithms and human-based algorithms are the four most studied types of metaheuristic algorithms [8].

Metaheuristic algorithms act on the principle of creating a solution set and continuing on their way by selecting the best solution candidate from among them. When creating new solution values, the values of the best solution are acted upon. Each metaheuristic method has mathematical formulas that it uses to produce new values. The method uses a mathematical infrastructure in such a way that it imitates the living being/system that is inherent in the method and imitates it. In general, the solution space is initially generated randomly. However, it is later generated in a way that approaches the best solutions using these equations.

1.2. Afterimage and Artificial Afterimage Algorithm

The Artificial Afterimage Algorithm [9] is an algorithm that has recently entered the literature of heuristic methods. Even if the eyes are closed after looking at light emitting objects for a certain period of time, negative or positive afterimage reflections continue to be created by our brain for a while. The structure on which the algorithm is based is the mathematical substructure in the creation of this image.

1.2.1. Afterimage

When we look at a lighted object for a certain period of time, the brain continues to create an image for a while even if our eyes are closed. This phenomenon is called afterimage. Even if our eyes do not see the original object, the chemical effect caused by the light in the eye continues for a while. These images are also called ghost images. Afterimages can also be colored. After a while, this image becomes clear and disappears [10-12]. If we continue to see the same color but less clearly even when our eyes are not open, this is positive afterimage. When the object we are looking at is colored, if we turn our eyes to a neutral surface, we continue to see it with complementary colors. This is called negative afterimage [13]. Figure 1 is an example of a positive afterimage and Figure 2 is an example of a positive afterimage.

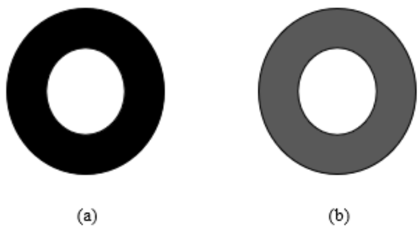

Figure 1. Example of positive afterimage: (a) original image, (b) reflection image.

Figure 1 (a) represents the original image, and (b) represents the reflection image in the eye. Although the reflection image will be the same color as the original, its brightness will be less.

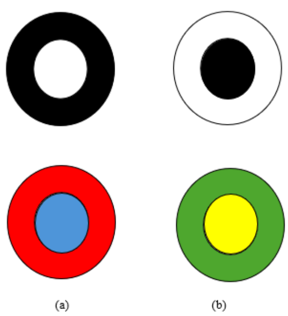

Figure 2. Example of negative afterimage: (a) original image, (b) reflection image.

Figure 2 the negative reflections of the images in (a) will appear as in (b). The colors will appear different, as the reflection colors.

1.2.2. Visual Angle and Perceptual Size

The rays reflected from our eyes pass through the top and bottom of an object to create a viewing angle. This is called the Visual Angle. The size of the image formed on the retina is determined by this angle. The image size is proportional to the size of this Visual Angle. Figure 3 shows the representation of the viewing angle.

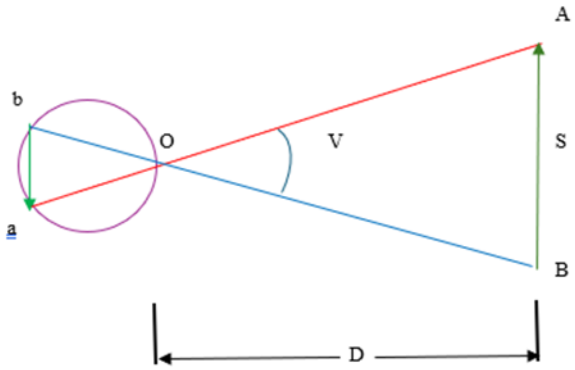

Figure 3. Visual Angle.

Visual Angle is calculated by the expression in Equation (1) [14].

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

When calculating the visual angle (V), 2 values are used. The first is the size of the object. Its real size is the distance between A and B. The second is the D value. It is the distance between the object and the eye. This value also shows the distance between O and B.

Perceptual size is the size perceived by the eye. It is expressed by Equation (2) [14].

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

1.3. Artificial Afterimage Algorithm

The algorithm that uses the concepts of afterimage, visual angle and perceptual size is a new algorithm in literature. Since this study is applied to data clustering, the value of the candidate solution is used instead of the concept of object size, and the distance of the object from the eye is used instead of the distance from the best solution. Equations (3) and (4) [9] show the visual angle (V), perceptual size (S) equations of the Artificial Afterimage Algorithm adapted to data clustering. Equation (5) [9] shows the formula to be used when generating new solution values. V is the AAIA adapted version of the visual angle. In fact, it is the result of the ratio of the values of the solution population to the difference between them and the best values. $S_{i,j}$ is the value obtained by multiplying the distance of a solution from the best by the visual angle after obtaining the visual angle value. $New\ pop_{i,j}$ is the step in which new candidate solutions are obtained using the obtained V and S(i,j) values.

$$V = 2 * \text{atan}\left(\frac{pop_{i,j}}{2 * pop_{i,j} - best_{1,j}}\right) \quad (3)$$

$$S_{i,j} = V * (pop_{i,j} - best_{1,j}) \quad (4)$$

$$New\ pop_{i,j} = S_{i,j} + (|\, S_{i,j-}(best_{1,j} - |(worst_{i,j} - pop_{i,j})|) * rand) \quad (5)$$

The matrix expressed as best represents the solution with the best fitness value and worst represents the solution with the worst fitness value. After these updates are made, the resulting pseudo code is as in Algorithm 1 [9]. Figure 4 shows the flow chart of the Artificial Afterimage Algorithm.

Algorithm 1: Pseudo code of Artifical Afterimage Algorithm

Begin
Create initial population
Create initial best solution
while(count⩽iteration)
  Specify Object size
  Specify Object distance
  V←$2 * \text{atan}\left(\frac{pop_{i,j}}{2*pop_{i,j} - best_{1,j}}\right)$
  S←$V * (pop_{i,j} - best_{1,j})$
  Local best values←determine current best values
 if fitness value of local best value < fitness value of best value **then**
   best←local best
  end if
 Compute $new\ pop_{i,j} = S_{i,j} + (|\, S_{i,j-}(best_{1,j} - |(worst_{i,j} - pop_{i,j})|) * rand)$
End while
Print best values
end begin

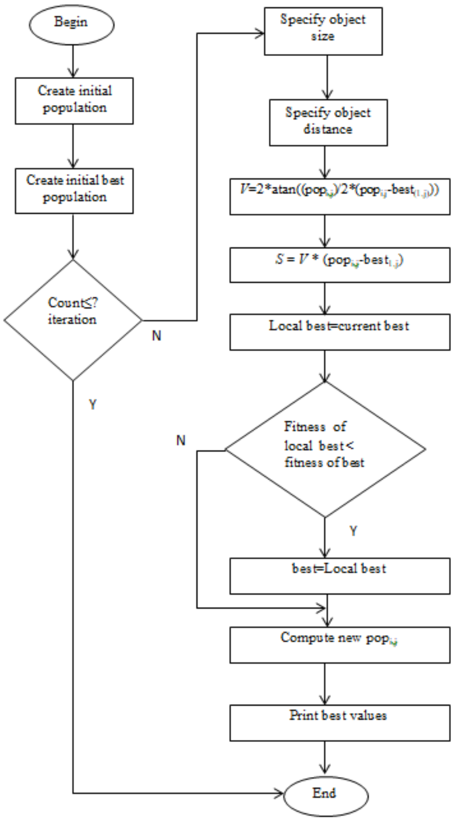

Figure 4. Flowchart of the Artificial Afterimage Algorithm.

2. Data Clustering

The name of the analysis technique that tries to group a group of data according to the similarities of their attributes is clustering. This grouping process is carried out according to the criteria we have determined beforehand [15]. Since the similarities of the existing data are tried to be grouped, a similarity measurement criterion is needed. Distance and similarity measurements are used for this task [16]. There are some distance measurement methods in the literature. In this study, the Manhattan distance measurement method is used.

Machine learning algorithms are basically divided into three different classes: supervised learning, unsupervised learning and reinforcement learning. In supervised learning, a labeled training set is used for output. The data in the training set consists of binary input objects and output values in a vector format. Unsupervised learning uses unlabeled data. Grouping is done according to the distance relationship between the attributes [17].

Clustering is unsupervised learning. Thanks to this feature, it manages to find hidden patterns in the data. In this respect, it has been the subject of important research in many fields such as pattern recognition, signal processing, bioinformatics, image processing and data mining [16]. Figure 5 is given as a representative example of the clustering process.

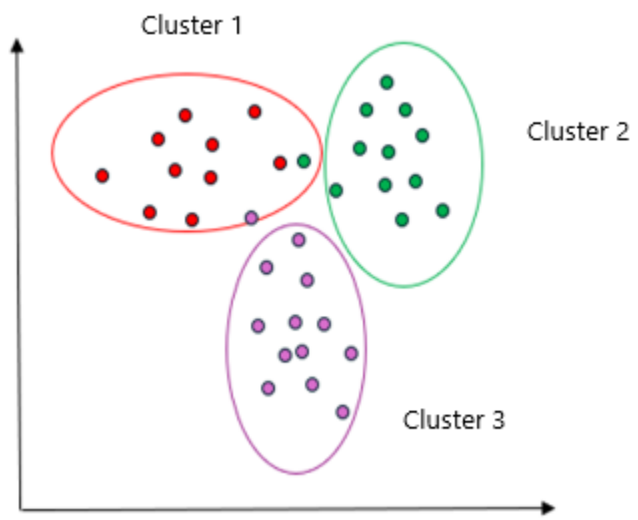

Figure 5. Visual representation of the clustering process.

In Figure 5, some data are not in the cluster they should be in. These data are detected as errors. In other words, they are not detected in the cluster they should be in. Points other than this are considered to be detected correctly.

2.1. Confusion Matrix and Performance Metrics

Confusion matrix is an important concept in the evaluation of machine learning classification and clustering. For performance evaluation, accuracy calculation is the basis. Confusion matrix is used when evaluating accuracy. In fact, confusion matrix is a summary of the prediction results in a classification problem [18]. We have 4 different values in the confusion matrix [19].

TP (True Positive): Cases where the predicted positive class overlaps with the true value
TN (True Negative): Cases where the predicted negative class overlaps with the true value
FP (False Positive): Cases where the predicted positive class does not overlap with the true value
FN (False Negative): Cases where the predicted negative class does not overlap with the true value

Figure 6 shows a representation of the confusion matrix.

	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

Figure 6. Confusion matrix representation.

Performance evaluation metrics are obtained from this matrix using the formulas in Equations 6-9 [20].

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (6)$$

Accuracy is obtained by dividing the correctly predicted values by the total value in the data set.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (7)$$

Precision is obtained by dividing the true positives by the total positives.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (8)$$

Recall is obtained by dividing the true positives by the sum of the true positives and false negatives.

$$\mathrm{F1\ score} = 2 * \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad (9)$$

The F1 score is the harmonic mean value of Precision and Recall.

3. Experimental Results

In this study, clustering operations of Artificial Afterimage Algorithm were performed on disease data obtained from Kaggle database [21-24]. Thus, risk factor assessment of diseases could be successfully presented with Artificial Afterimage Algorithm which is newly introduced to the literature.

Risk detection in medical data is of critical importance for the diagnosis of the disease. Because it can be very difficult to detect risk, especially in large-scale disease data or in data where the number of parameters affecting the disease is high. Although the Artificial Afterimage Algorithm has recently entered the literature; it was able to successfully perform risk detection on the data in this study. Thanks to the simple mathematical infrastructure it uses, the Artificial Afterimage Algorithm can reach a result without causing too much complexity. In this respect, it has an advantageous infrastructure.

Table 1 shows the best, worst and V values obtained by the algorithm for clustering process. V value was obtained with Equation (3) and perceptual angle was calculated with Equation (4). Finally, new candidate solutions of the solution set were calculated with Equation (5). While some of the data sets were used for train process, some were separated for testing.

Table 1. Experimental results of Datasets.

			Colon Cancer Dataset			
best	0.424	0.789	0.168	0.318	0.09	0.524
	0.72	1.198E+14	0.310	0.397	0.362	0.524
worst	2.445E+260	0.669	1.704E+259	6.730E+260	5.87E+260	0.417
	8.603E+260	0.815	5.937E+260	1.182E+260	6.148E+260	8.864E+260
V	1.358					
			Coronary Heart Disease Dataset			
best	2.053E+14	1.149E+14	3.456E+14	0.710	0.171	
	0.422	0.439	4.635E+14	2.094E+14		
worst	0.916	4.343E+251	3.126E+14	0.505	3.179E+252	
	0.310	2.315E+252	1.419E+252	3.688E+252		
V	0.999					
			Indian Liver Patient Records Dataset			
best	3.940E+14	1.140E+14	6.804E+14	0.1623	2.320E+14	
	0.168	3.754E+14	0.0464	0.178	0.270	
worst	3.017E+14	1.673E+257	1.086E+257	1.993E+257	1.018E+14	
	8.249E+256	1.745E+14	4.669E+256	3.053E+257	4.986E+257	
V	3.123					
			Prostate Cancer Dataset			
best	0.458	3.284E+14	0.536	0.681	1.339E+14	
	0.385	0.071	0.199			
worst	3.330E+251	1.852E+14	0.222	0.513	3.58E+250	
	2.328E+251	2.188E+251	4.011E+251			
V	1.497					

According to the results obtained, the confusion matrices of each data set are as shown in Figure 7.

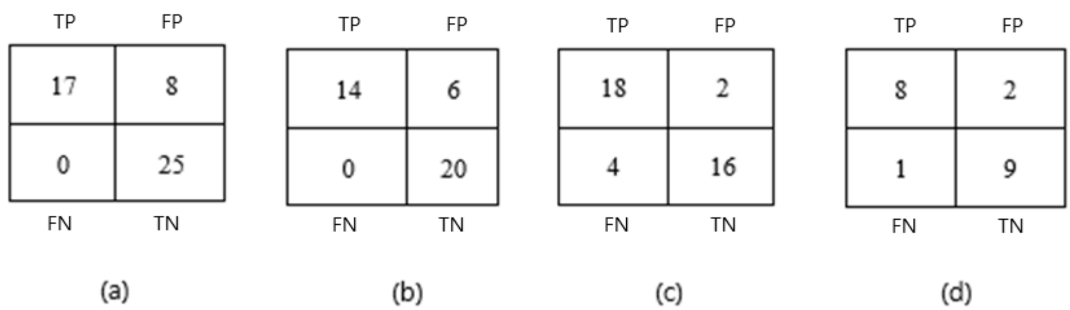

Figure 7. Confusion matrices for the evaluated datasets: (a) Colon Cancer Dataset, (b) Coronary Heart Disease Dataset, (c) Indian Liver Patient Records Dataset, (d) Prostate Cancer Dataset

Table 2 shows the performance metrics of the experimental results.

Table 2. Performance metrics of AAIA on four medical datasets

	Accuracy	Precision	Recall	F1
Colon Cancer Dataset	0.84	0.68	**1**	0.81
Coronary Heart Disease Dataset	**0.85**	0.7	**1**	0.82
Indian Liver Patient Records Dataset	**0.85**	**0.9**	0.82	**0.86**
Prostate Cancer Dataset	**0.85**	0.8	0.89	0.84

Figure 8 shows the graphical representation of performance metrics.

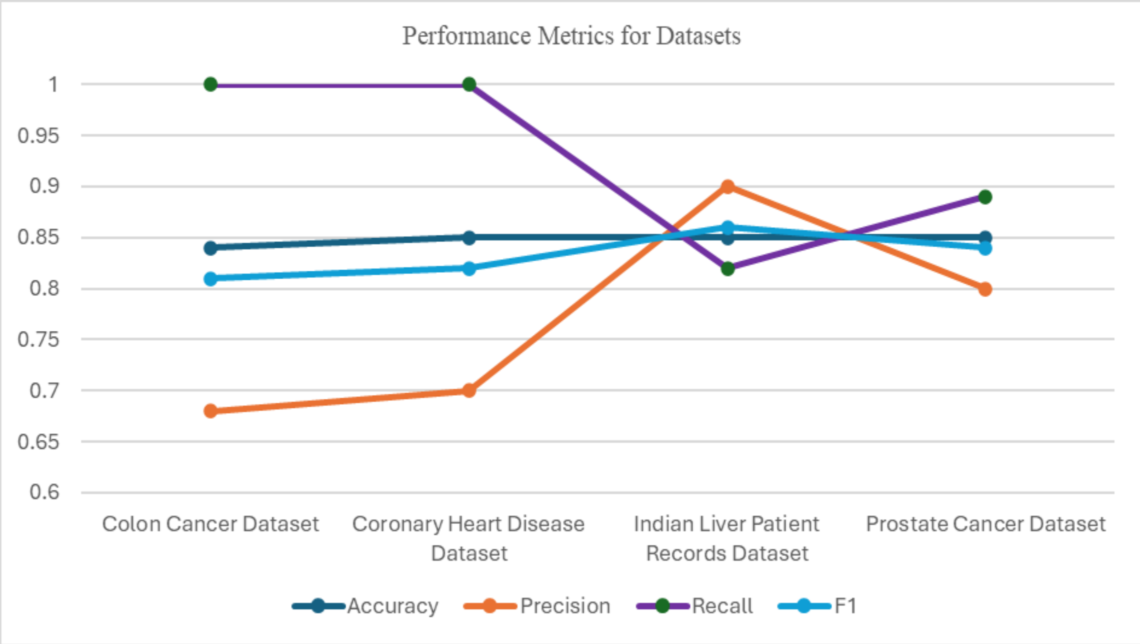

Figure 8. Graphical representation of performance metrics.

4. Discussion and Conclusion

Metaheuristic algorithms are frequently preferred methods, especially in problems where a solution cannot be found with analytical methods or where the solution time may be a problem even if it is found. In this study, the Artificial Afterimage Algorithm was applied to medical data. The Artificial Afterimage Algorithm clustering metrics were examined for disease risk analysis detection.

Although the Artificial Afterimage Algorithm is a new algorithm in the literature; It has obtained significant results in terms of Accuracy, Precision, Recall and F1 metrics. A distance-based evaluation measure was used in the clustering method. In this methodology, the distribution of values in the datasets greatly affects the clustering performance. Therefore, this situation should be taken into account when selecting the data to be worked on. The distribution of the data should be examined and then worked on. In addition, disadvantageous situations in terms of time may occur in very large datasets. The Artificial Afterimage Algorithm has a distance-based infrastructure. Grid-based and hierarchical-based versions can be developed; how they will behave on data sets with different distributions may be the aspects waiting to be studied.

References

[1] Ersöz B, Özmen M. Dijitalleşme ve Bilişim Teknolojilerinin Çalışanlar Üzerindeki Etkileri. AJIT-e: 2020; 11(42): 170-179.

[2] Aydın A. Devlet erkinin yönetim paradigmasının yapay zeka bağlamında dönüşümü. G. Telli (Ed.), Yapay Zeka ve Gelecek içinde İstanbul: Doğu Kitapevi; 2019.

[3] McCarthy J, Minsky ML, Rochester N, Shannon CE. “A proposal for the Dartmouth summer research project on artificial intelligence” https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html(accessed Nov 15, 2024).

[4] McCarthy J. “What is artificial intelligence?” http://www-formal.stanford.edu/jmc/whatisai/(accessed Nov 18, 2024).

[5] Şimşir İ, Mete B. Sağlık Hizmetlerinin Geleceği: Dijital Sağlık Teknolojileri. JOINIHP 2021; 2(1): 33-39.

[6] Yorgancıoğlu Tarcan G, Yalçın Balçık, P, Sebik, NB. Türkiye ve Dünyada Sağlık Hizmetlerinde Yapay Zekâ. MEÜ Tip Fakültesi Tıp Tarihi ve Folklorik Tıp Dergisi 2024; 14(1):50-60.

[7] 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.

[8] Faramarzi A, Heidarinejad M, Stephens B, Mirjalili S. Equilibrium optimizer: A novel optimization algorithm. Knowl Based Syst 2020; 191.
[9] Demir M. Artificial Afterimage Algorithm: A New Bio-Inspired Metaheuristic Algorithm and Its Clustering Application. Appl Sci (Basel) 2025; 15(3): 1359.
[10] Bender, MB, Feldman M, Sobin AJ. Palinopsia. Brain 1968; 91(2): 321–338.
[11] Gersztenkorn D, Lee AG. Palinopsia revamped: A systematic review of the literatüre. Surv Ophthalmol 2014; 60: 1–35.
[12] L˝orinc G, András H. Analyzing afterimage strength and duration test results with k-means clustering. In Szemelvények a BGE Kutatásaiból (II. Kötet) Budapest; 2023.
[13] Lee HC. Introduction to Color Imaging Science Cambridge University Press: Cambridge; 2005.
[14] Visual Angle-Wikipedia. Available online: https://en.wikipedia.org/wiki/Visual_angle(accessed Nov 20, 2024).
[15] Taşkın Ç, Emel, GG. Clustering Approaches in Data Mining and an Application with Kohonen Networks in Retailing Sector. Süleyman Demirel University the Journal of Faculty of Economics and Administrative Sciences 2010; 15(3): 395-409.
[16] Avşar İİ. Clustering of Türkiye and European Union Countries by Length of Railroad Lines. Journal of The Faculty of Applied Sciences of Tarsus University 2023; 3(1): 13-25.
[17] Çatak FÖ. Development of data mining software framework by using map/reduce method in cloud computing systems. Phd. Thesis, İstanbul University Institute of Science, Department of Electronics, Informatics Program, İstanbul, 2014.
[18] Chen RC, Dewi C, Huang SW, Caraka RE. Selecting critical features for data classification based on machine learning methods. J Big Data 2020; 7:52.
[19] Yeşildal G. Diagnosing COVID-19 Disease through Medical Images. MSc, Ankara University Institute of Science, Department of Computer Engineering, Ankara, 2022.
[20] Aslanyürek M, Mesut A. A New Method to Measure Clustering Performance and its Evaluation for Text Clustering. Eur J Sci Technol 2021; 27:, 53-65.
[21] Colon Cancer, https://www.kaggle.com/datasets/lakshmi25npathi/colon-cancer(accessed Dec 10, 2024).
[22] Prostate Cancer, https://www.kaggle.com/datasets/sajidsaifi/prostate-cancer(accessed Dec 10, 2024).
[23] Indian Liver Patient Records, https://www.kaggle.com/datasets/uciml/indian-liver-patient-records(accessed Dec 10, 2024).
[24] Coronary Heart Disease, https://www.kaggle.com/datasets/billbasener/coronary-heart-disease?resource=download (accessed Dec 10, 2024).