# Analysis of the development of fruit trees diseases using modified analytical model of fuzzy c-means method

**Ali Abdulkarem Habib Alrammahi[1], Farah Abbas Obaid Sari[1], Haidar Abdulwahab Habeeb Shamsuldeen[2]**
[1]Department of Computer Sciences, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq
[2]Department of Information Research and Training Center in training, University of Kufa, Najaf, Iraq

**Article Info**

***Article history:***

Received Jul 22, 2022
Revised Sep 11, 2022
Accepted Sep 28, 2022

***Keywords:***

Big data
Distance scale
False negative
False positive
Priority fuzzy C-means
Similarity coefficient
Smart garden

**ABSTRACT**

The use of digital technologies in agriculture has become very important to ensure the protection of trees from disease and limit their development, which leads to increased production, so the paper proposes a modified analytical model to analyze the data and graphical parts of the leaves of fruit trees using priority fuzzy C-means (PFCM). Based on the proposed distance scale to obtain a clustering with a less error rate and fairly close to accuracy for the purpose of monitoring the development of diseases of fruit trees, by classifying the diseases and medications needed for each disease, a database was created containing large samples of data and images, where the results of Analysis of previous studies that analyzes of large amounts of data give accurate results. The proposed method was used in smart gardens with large areas and we got the desired results.

*This is an open access article under the CC BY-SA license.*

***Corresponding Author:***

Ali Abdulkarem Habib Alrammahi
Department of Computer Sciences, Faculty of Computer Science and Mathematics, University of Kufa
54003 Kufa, Najaf, Iraq
Email: alia.alramahi@uokufa.edu.iq

## 1. INTRODUCTION

Today, the agro-industrial complex lags far behind the world leaders in the use of digital technologies, as these technologies are used in agriculture in proportions of less than 2%, while in the countries of the European Union and the USA, the use of digital solutions is 80% and 60%, respectively. The results of studies conducted at the present time allow us to draw a disappointing conclusion that the population of our country is not sufficiently provided with horticultural products, both fresh and processed, due to the relevance of the issues of organizing import substitution in order to supply the population with competitive horticultural products in various forms, it is necessary to provide widespread use of digital solutions by agricultural producers, in particular, in horticulture. One of the key areas in the organization of digital agriculture is the development of horticulture, based on new principles of control, optimization and management decision-making.

In industrial horticulture, very effective techniques for growing fruit crops are currently being used. However, it must be admitted that insufficient attention is paid to the issues of systemic intensification, which can provide a multiple increase in the economic efficiency of the use of lands occupied by orchards, difficult climatic conditions, as a rule, bring great risks to the organization of horticulture. In this regard, one of the fundamental elements of new technologies for regulating horticulture is the regulation of irrigation systems, the use of fertilizers and digital monitoring and control systems [1], [2]. The issues of irrigation and organization of drip irrigation are considered in the works [3]-[7].

The widespread use of the concept of creating a smart garden provides the ability to conduct an operational analysis of soil and climatic conditions based on the processing of big data [8]-[11] obtained from sensors installed both in the vicinity of the root system of a single tree, and directly on its trunk, to carry out the application of fertilizers of various types depending on the characteristics of a particular tree, carry out preventive measures to combat pests and diseases of fruit plants in a smart garden, based on the analysis of graphic information, which is represented by images of tree leaves. At the same time, issues related to the organization and storage of large volumes of both data and graphical choice of processing graphic information are of particular importance, the analysis of which allows you to take prompt measures in the process of fruit growth, taking into account the state of fruit trees. To implement an intelligent system and predict the water regime of the soil, it is necessary to use mathematical modeling of moisture transfer processes. However, this system lacks the ability to process large amounts of data and graphic information, which makes it possible to identify characteristic classes [12]-[14], which is necessary for solving problems of analysis, forecasting and provides an increase in storage efficiency and increases the efficiency of managerial decision-making. Currently, there are a large number of methods that allow processing data used in information systems to implement the concept of a smart garden [15]-[17], however the resulting images from camera may have defects: defocusing, blurring, distortion, violation of brightness and contrast. In addition, the analysis of the soil data that we obtained from the sensors leads us to determine the percentage of moisture and drought to support decision-making in smart garden control systems. In this regard, the paper is aimed at building analytical models of information processing in information systems used within the concept of a smart garden. The modified model uses a different distance scale than the usual one that is used with segmentation methods with contribution criteria, all of which together lead to increased accuracy in the diagnosis of diseases and the organization of large amounts of data.

The following is the outline of the paper. Section 2 of this paper presents the basic information and differences between measures of the distance between the points of the data set or image and the center of clusters in order to have a clear picture of the advantages and disadvantages of each measure. Section 3 presents the proposed method for analyzing the data collected from the sensors and segmentation of the fruit tree leafes images. Section 4 describes the analysis of results related to the use of the modified analysis model. As for the last section, it includes the details of the conclusion of the work which presented in this paper.

## 2. DISTANCE MEASURES

The purpose of cluster analysis is to determine the stratification of the initial observations into clearly defined clusters-clusters that lie at a certain distance from each other, but do not break into parts that are equally distant from each other, which, in fact, is a grouping with the identification of natural stratification. Cluster analysis can be done in several ways. Each method is characterized by three features: a measure of the proximity of any two objects, a measure of the proximity of two groups of objects, and a rule for choosing the final version of the classification. The basis for clustering is the matrix of distances between objects. There are several ways to determine the distance between every two objects.

### 2.1. Euclidean distance

This seems to be the most common type of distance. It is simply a geometric distance in multidimensional space and is calculated as (1).

$$dist\,(\mathrm{x,y}) = \sqrt{\sum_i (X_i - Y_i)^2} \quad (1)$$

Note that the Euclidean distance (and its square) is calculated from the original data, not from the standardized data. This is the usual way of calculating it, which has certain advantages (for example, the distance between two objects does not change when a new object is introduced into the analysis, which may turn out to be an outlier). However, distances can be greatly affected by differences between the axes from which the distances are calculated. For example, if one of the axes is measured in centimeters, and then you convert it to millimeters (by multiplying the values by 10), then the final Euclidean distance (or the square of the Euclidean distance) calculated from the coordinates will change dramatically, and, as a result, the results of the cluster analysis can be very different from the previous ones.

### 2.2. The square of the Euclidean distance

Sometimes you may want to square the standard Euclidean distance to give more weight to more distant objects. This distance is calculated as (2).

$$dist\ (\mathrm{x,y}) = \sum_i (X_i - Y_i)^2 \quad (2)$$

### 2.3. City block distance (Manhattan distance)

This distance is simply the average of the differences over the coordinates. In most cases, this measure of distance leads to the same results as for the usual Euclid distance. However, note that for this measure the influence of individual large differences (outliers) decreases (because they are not squared). Manhattan distance is calculated using:

$$dist\ (\mathrm{x,y}) = \sum_i |X_i - Y_i| \quad (3)$$

### 2.4. Minkowski distance and Chebyshev distance

A parametric metric on Euclidean space that can be thought of as a generalization of Euclidean distance and city block distance (Manhattan).
p⩾1 Minkowski distance
p<1 is not a metric
When p=∞ If P=0 {1/P= (infinity) will get the Chebyshev distance, this distance between two n-dimensional points or vectors is the maximum modulus of the difference in the coordinates of these points. The Chebyshev distance can be useful when one wishes to define two objects as "different" if they differ in any one coordinate (any one dimension). The Chebyshev distance is calculated by (4).

$$dist\ (\mathrm{x,y}) = \mathrm{Max}|X_i - Y_i| \quad (4)$$

## 3. MODIFIED ANALYTICAL MODEL OF PRIORITY C-MEANS METHODS FOR PROCESSING DATA AND FRAGMENTS OF GRAPHIC INFORMATION

Consider an analytical model that allows you to perform the process of fuzzy clustering, the construction of which takes into account both local and non-local information and their relative contribution [18]-[21]. In contrast to the standard approach based on the use of FCM, for fuzzy clustering in the proposed fuzzy clustering model, the definition of distance is changed in order to take into account the mutual influence of pixels during segmentation and reduce the noise of others.

The optimization criterion, on the basis of which the clustering is carried out, has the form:

$$O_{fcm}(U,C) = \sum_{k=1}^{n} \sum_{i=1}^{v} (u_{ik}^{s})^m d^g(x_k, c_i)\,, \quad (5)$$

$$g = \{0,1\}.$$

if $g=0$, то $d$ – Chebyshev distance.
if $g=1$, то $d$ – Minkowski distance.
Spatial membership function $u_{ik}^{s}$
Where $P_{ik}$– a priori probability that the *k* pixel is in the *i* cluster

$$P_{ik} = \frac{NN_i(k)}{N_k} \quad (6)$$

where $NN_i(k)$ is the number of pixels in the vicinity of the *k* pixel, which are located in the *i* cluster after removing the fuzziness. $N_k$ is the total number of pixels in the neighborhood. $d_{iz}$ is the distance between the *i* cluster and its *z* neighborhood. This makes it possible to calculate the center of each cluster $c_i^z$ using (7).

$$c_i^s = \frac{\sum_{k=1}^{n}\left((u_{ik}^{s})^m x_k\right)}{\sum_{k=1}^{n}(u_{ik}^{s})^m} \quad (7)$$

The proposed fuzzy clustering model has a distinctive feature, which is that the distance measure [22] -[25] which should take into account both local and non-local information, is formalized in:

$$d_{ki}(x_j, v_i) = (1-\lambda_j) d_l^g(x_i, v_i) + \lambda_j d_{nl}^g(x_j, v_i) \quad (8)$$

where $d_l(x_j, v_i)$ - distances, taking into account the influence of local information, and $d_{nl}^g(x_j, v_i)$ - distances, taking into account the influence of non-local information; $\lambda_j$ is a parameter ranging from 0 to 1, a weight factor which is determined by the user when solving the problem and can be corrected when an unsatisfactory solution is obtained. A larger value is taken to make the contribution of the distance to the local information larger, and a smaller value is taken to reduce the contribution of the non-local information distance. The measure, for local information, is defined as:

$$d_l^g(x_j, v_i) = \frac{d_{nl}^g(x_j, x_i)}{\sum_{x_k \in N_i} \omega_l(x_k, x_j)} \tag{9}$$

where $d_l^g(x_j, v_i)$ is a measure determined by local information;
$\omega_l(x_k, x_j)$- pixel weight $N_j$

The measure $d_{nl}$ determined by non-local information is defined as a weighted sum over image pixels:

$$d_{nl}^g(x_j, v_i) = \sum_{x_k \in N_j} \omega_{nl}(x_k, x_j) d^g(x_k, v_i) \tag{10}$$

where $\omega_l(x_k, x_j)$ is the pixel weight in $N_i$

Membership for representative points should be as high as possible, and for non-representative points, as low as possible. An optimization criterion that satisfies these requirements is defined as shown in and looks like:

$$min\{J_m(x, \mu, c) = \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{ij}^m d_{ij}^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{N} (1 - \mu_{ij})^m\} \tag{11}$$

$d_{ij}$ - is the distance between the *j* point and the *i* center of the cluster;
$\mu_{ij}$ - membership degree;
$m$ – degree of fuzziness;
$\eta_i$ - positive number;
$c$ – number of clusters;

$$\mu_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^g}{\eta_i}\right)^{\frac{1}{m-1}}} \tag{12}$$

The value of $\eta_i$ corresponds to the distance at which the value of the membership function of a point to a cluster is 0.5. The constructed analytical model allows you to fix $\eta_i$ or change it at each iteration by changing $d_{ij}$ and $\mu_{ij}$., which increases the resistance to noise when searching for valid clusters and determining their centers. Figure 1 show the working method of the proposed method for processing data and images of fruit tree leaves.

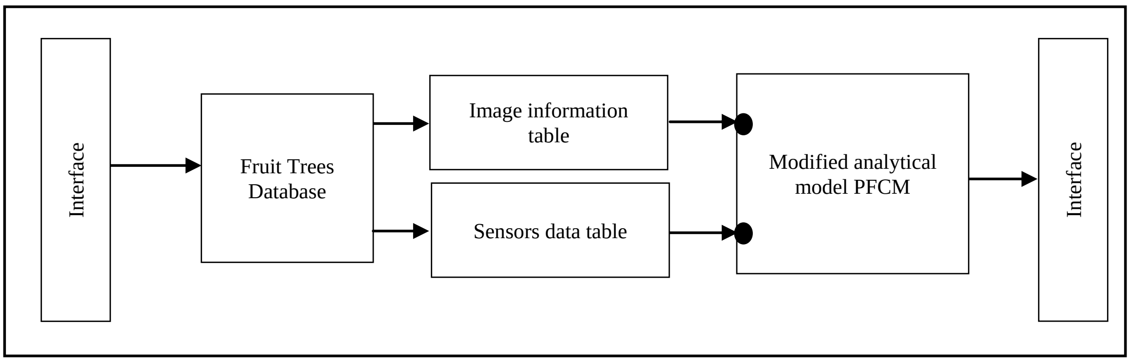


Figure 1. Structural model analysis of the development of diseases of fruit trees

## 4. RESULTS AND DISCUSSION

As a result of processing the experimental data of the drip irrigation system obtained during the performance of scientific research, a set of data was obtained processed using the clustering method. Three clusters were obtained corresponding to morning (M), afternoon (A) and evening (E) watering each one whith three factors: Pressure in the drip system (Z1), Drip line length (Z2 ), Water outlet diameter (Z3 ).

Table 1 shows that trees are watered in the morning, afternoon and evening. Figure 2 exhibit we use different volumes of water. The lack of watering, as well as its excess, leads to the emergence of a number of diseases, the presence of which inevitably leads to a change in the surface of the leaves.

Table 1. Data clustering depend on factors

| Factors | Clusters | | |
|---|---|---|---|
| | M | A | E |
| Z1 | 0.8 | 1.8 | 1.2 |
| Z2 | 120 | 150 | 180 |
| Z3 | 1.0 | 1.5 | 2.0 |

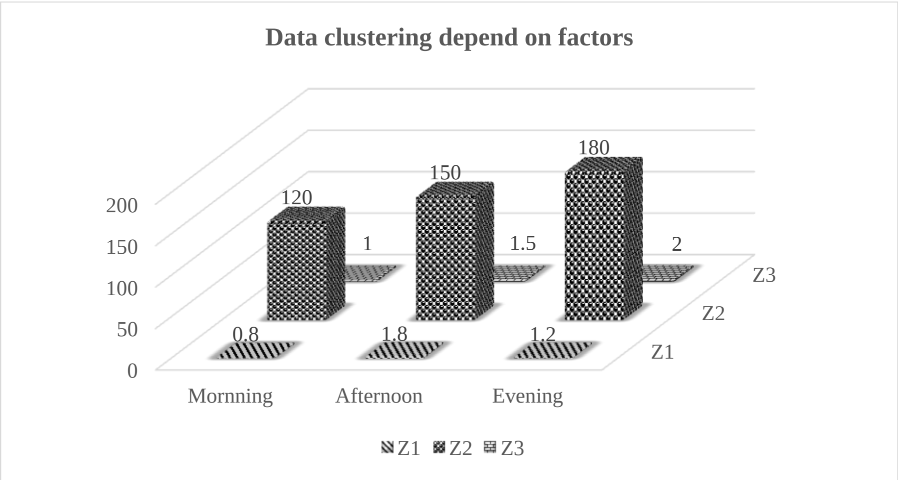


Figure 2. Clusters with different volumes of water

We used the similarity coefficient, percentage of false positives and percentage of false negatives for analysis of fragments of the results presented in Table 2 allows us to conclude that the introduction of the Chebyshev and Minkowski metrics is reasonable and provides better results.

Similarity coefficient shwo how similar clusters are to each other during clustering and when the value of this coefficient is large, it indicates the accuracy of the segmentation method. The value of a similarity relationship is measured by comparing the number of co-occurrence pixels with the number of individual pixels for each cluster in that relationship, where false positive (FP) shows how the pixel is incorrectly classified as not belonging to the cluster, but in fact belongs to cluster. A false negative (FN) means that a pixel is incorrectly classified as belonging to a cluster but actually does not.

Table 2. Segmentation accuracy for various clustering algorithms

| Clustering algorithm | Similarity | Percentage of false positives | Percentage of false negatives |
|---|---|---|---|
| FCM method | 85.23 | 21.25 | 7.50 |
| PFCM Method | 91.38 | 11.76 | 4.39 |
| PFCM method with Chebyshev metric | 93,8 | 11,60 | 3,20 |
| PFCM method with Minkowski metric | 93,5 | 11,80 | 3,35 |

## 5. CONCLUSION

In this paper, we have increased the percentage of correct diagnosis of fruit tree diseases, monitor their development as well as improving the efficiency of fruit tree diseases diagnosis by using a modified analytical model to process data and graphic information in the form of fruit tree leaf images allows us to conclude that the introduction of the Chebyshev and Minkowski metrics is reasonable and provides better results. An analysis of the results of paper showed that the proposed modified clustering method made it possible to reduce the percentage of false positives from 21.25 to 11.6, the percentage of false negatives from 7.5 to 3.2.

## ACKNOWLEDGEMENTS

Authors would thank university of kufa for support this research work.

## REFERENCES

[1] B. Yang, Y. Xu, "Applications of deep-learning approaches in horticultural research: a review," *Hortic Res*, vol. 8, no. 123, 2021. doi: 10.1038/s41438-021-00560-9.
[2] E. J. Pekkeriet, E. J. V. Henten, and J. B. Campen, "Contribution of innovative technologies to new developments in horticultur," *Acta Hortic*, vol. 1099, pp. 45-54, 2015. doi: 10.17660/ActaHortic.2015.1099.1.
[3] Z. Guan, Z. Jia, Z. Zhao, and Q. You, "Dynamics and distribution of soil salinity under long-term mulched drip irrigation in an arid area of Northwestern China," *Water*, vol. 11, no. 6, pp. 1225, 2019, doi: 10.3390/w11061225.
[4] S. M. Musinovich and K. I. Khaitmuratovich, "The Determination Of Soil Moisturization Dynamics Four Factor Experience," *The American Journal of Interdisciplinary Innovations and Research,* vol. 5, no. 676, pp. 1-8, 2021, doi: 10.37547/tajiir/Volume03Issue11-01.
[5] V. Phogat, M. A. Skewes, M. Mahadevan, and J.W. Cox, "Evaluation of soil plant system response to pulsed drip irrigation of an almond tree under sustained stress conditions," *Agricultural Water Management,* vol. 118, pp. 1-11, 2013, doi: 10.1016/j.agwat.2012.11.015.
[6] Y. Wang, D. Kou, M. Muneer, G. Fang, and D. Su, "The effects of irrigation regimes on soil moisture dynamics, yield and quality of lucerne under subsurface drip irrigation," *Applied Ecology And Environmental Research*, vol. 18, no. 3, pp. 4179-4194, 2019, doi: 10.15666/aeer/1803_41794194.
[7] P. Garcia-Caparros, J. Contreras, R. Baeza, M. Segura, and M. Lao, "Integral Management of Irrigation Water in Intensive Horticultural Systems of Almería," *Sustainability*, vol. 9, no. 12, pp. 2271, 2017, doi: 10.3390/su9122271.
[8] A. Cravero, A. Bustamante, M. Negrier, and P. Galeas, "Agricultural Big Data Architectures in the Context of Climate Change: A Systematic Literature Review," *Sustainability*, vol. 14, no. 13, pp. 7855, 2022, doi: 10.3390/su14137855.
[9] F. Giomi, *et al.*, "The Applicability of Big Data in Climate Change Research: The Importance of System of Systems Thinking," *Frontiers in Environmental Science,* vol. 9, pp. 619092, 2021, doi: 10.3389/fenvs.2021.619092.
[10] Q. Harpham, N. Tozer, P. Cleverley, D. Wyncoll, and D. Cresswell, "A Bayesian method for improving probabilistic wave forecasts by weighting ensemble members," *Environmental Modelling & Software,* vol. 84, pp. 494-504, 2016, doi: 10.1016/j.envsoft.2016.07.015.
[11] C. Zhang and Z. Liu, "Application of big data technology in agricultural Internet of Things," *International Journal of Distributed Sensor Networks*, vol. 15, no. 10, 2019, doi: 10.1177/1550147719881610.
[12] Y. Lvovich, Y. E. Lvovich, A. P. Preobrazhenskiy, Y. P. Preobrazhenskiy, and O. N. Choporov, "Optimization processes in the internet of things system at agricultural enterprises, " *IOP Conference Series: Earth and Environmental Science*, vol. 548, no. 2, pp. 022086, Aug. 2020, doi: 10.1088/1755-1315/548/2/022086.
[13] D. D. Koo, J. J. Lee, A. Sebastiani, and J. Kim, "An internet-of-things (IoT) system development and implementation for bathroom safety enhancement," *Procedia Engineering*, vol. 145, pp. 396–403, 2016, doi: 10.26102/2310-6018/2019.25.2.0.
[14] V. N. Kalaev, M. S. Nechaeva, A. Y. Vasilieva, I. Y. Vasiliev, A. P. Preobrazhensky, and I. Y. L'vovich, "Information training system for distance learning the methodology for carrying out a micronuclear test in a human buckle epitelium," *Modeling, Optimization and Information Technology*, vol. 7, no. 2, pp. 6–26, 2019, doi: 10.26102/2310-6018/2019.25.2.0.
[15] T. Talaviya, D. Shah, N. Patel, H. Yagnik, and M. Shah, "Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides," *Artificial Intelligence in Agriculture,* vol. 4, pp. 58-73, 2020, doi: 10.1016/j.aiia.2020.04.002.
[16] V. P. Kumar, K. C. Ramya, J. S. Abishek, T. S. Arundhathy, B. Bhavvya, and V. Gayathri, "Smart garden monitoring and control system with sensor technology," in *3rd International Conference on Signal Processing and Communication (ICPSC)*, Coimbatore, India, 2021, doi: 10.1109/ICSPC51351.2021.9451788.
[17] A. Jia, "Intelligent garden planning and design based on agricultural internet of things," *Complexity*, vol. 2021, pp. 1-10, 2021, doi: 10.1155/2021/9970160.
[18] T. H. Sardar and Z. Ansari, "Distributed big data clustering using MapReduce-based fuzzy C-Medoids," *J. Inst. Eng. India,* vol. 103, p. 73–82, 2004, doi: 10.1007/s40031-021.
[19] E. H. Ruspini, "Numerical methods for fuzzy clustering," *Information Sciences*, vol. 2, no. 3, pp. 319–350, Jul. 1970, doi: 10.1016/S0020-0255(70)80056-1.
[20] E. H. Ruspini, "A new approach to clustering," *Information and Control*, vol. 15, no. 1, pp. 22–32, Jul. 1969, doi: 10.1016/S0019-9958(69)90591-9.
[21] A. A. Al-Rammahi, F. A. Sari, and F. G. Al-Jelaihawi, "Covid-19 plasma monitoring based on clustering a large set of recovered patient data," in *2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, Vladivostok, Russia, 2020, doi: 10.1109/FarEastCon50210.2020.9271296.
[22] M.-Z. Chen, G.-L. Chen and S.-L. Chen, "Interval-valued examples learning based on fuzzy c-mean clustering," in *2006 International Conference on Machine Learning and Cybernetics*, Dalian, China, 2006, doi: 10.1109/ICMLC.2006.258596.
[23] Y. Liu, J. Chen, S. Wu, Z. Liu, and H. Chao, "Incremental fuzzyC medoids clustering of time series data using dynamic time warping distance," *PLoSONE*, vol. 13, no. 5, 2018, doi: 10.1371/journal.pone.0197499.
[24] M. S. Mahmud, J. Z. Huang, S. Salloum, T. Z. Emara, and K. Sadatdiynov, "A survey of data partitioning and sampling methods to support big data analysis," *Big Data Mining and Analytics*, vol. 3, no. 2, p. 85–101, 2020, doi: 10.26599/BDMA.2019.9020015.
[25] V. Rayala and S. R. Kalli, "Big data clustering using improvised Fuzzy C-Means clustering," *Revue d'Intelligence Artificielle*, vol. 34, no. 6, pp. 701-708, 2020, doi: 10.18280/ria.340604.

## BIOGRAPHIES OF AUTHORS


**Ali Abdulkarem Habib Alrammahi** obtained a master's degree in Information Technology from dr babasaheb ambedkar marathwada university in Aurangabad, India, in 2012. Currently, work at University of Kufa in the city of Najaf in Iraq. His research interests include data analysis, and image processing. He can be contacted at email: alia.alramahi@uokufa.edu.iq.

**Farah Abbas Obaid Sari** obtained a master's degree in Information Technology from dr babasaheb ambedkar marathwada university in Aurangabad, India, in 2012. Currently, work at University of Kufa in the city of Najaf in Iraq. Her research interests include image processing and artificial intelligence. She can be contacted at email: faraha.altaee@uokufa.edu.iq.

**Haidar Abdulwahab Habeeb Shamsuldeen** obtained a master's degree in Information Technology from Tambov state university in Tambov, Russia, in 2015. Currently, work at University of Kufa in Information Research and Training Center in training department in the the city of Najaf in Iraq. His research interests include data analysis, and datamining. He can be contacted at email: haidar.habeeb@uokufa.edu.iq.