ISSN NO: 2770-4491
Date of Publication:26-03-2023

Skin Cancer Detection Using K-Means Clustering-Based Color Segmentation

Wameedh Raad Fathel[1]
wamed81@gmail.com
[1]Ninawa Health Directorate, High Institute of Health in Mosul, Mosul, Iraq
Ahmed Saeed Ibrahim Al-Obaidi[2]
ahmedsaeed@ntu.edu.iq
[2]Technical Engineering College, Northren Technical University (NTU), Mosul, Iraq
Maysaloon Abed Qasim[2]
maysloon.alhashim@ntu.edu.iq
[2]Technical Engineering College, Northren Technical University (NTU), Mosul, Iraq
Marwa Mawfaq Mohamedsheet Al-Hatab[2]
marwa.alhatab@ntu.edu.iq.
[2]Technical Engineering College, Northren Technical University (NTU), Mosul, Iraq

Abstract

Approximately 75% of all cancers are found on the skin. Due to its high mortality rate, Skin cancer (SC) must be treated immediately after detection. SC, in fact, results from abnormalities in the skin's surface. In spite of the fact that most people who have SC make a full recovery, this disease remains a major source of anxiety for the general public. Most SCs develop only locally and invade surrounding tissues, but melanoma, the rarest form of SC, may move throughout the body via the bloodstream and lymphatic system. in this study k mean algorithm and $L'a'b'$color space has been used to detect melanoma skin cancer. The data set used has been obtained from Kaggle skin cancer collection challenge. The input image has been changed to $L'a'b'$color space, k mean clustering algorithm was used to cluster the image into three clusters and return an index corresponding to each cluster then the a'b' layers have been used as image clusters and l layer has been used to detect the exact part of the image.

Key words: skin

1. Introduction

Skin cancer (SC) is a type of cancer that start in the skin. It is caused by the growth of cells that don't work right and can spread to other parts of the body. Non-melanoma (NMSC) and melanoma are the two main types of SC [1].

Sunlight and artificial sources of ultraviolet radiation like tanning beds are the primary culprits in the development of SC. More than 1.5 million cases of SC were detected worldwide in 2020, and more than 120 000 fatalities were attributed to the disease. Children and teenagers who spend too much time in the sun are at increased risk of developing SC as adults [2].

More than million new cases of NMSC excluding Basal Cell Carcinoma (BCCs) and 65,000 deaths were reported worldwide in 2018, making it the fifth most prevalent form of cancer overall [3]. In comparison, approximately 300,000 new cases of melanoma and 60,000 deaths were reported in 2018. Extreme proliferation of melanocytes (The basal layer of the epidermis has cells that make pigment) is a hallmark of Pigmented Skin Lesions (PSLs). Some PSLs are harmless, whereas others are cancerous [4]. BCC and Squamous Cell Carcinoma (SCC) are the most frequent NMSC.

Because of its poor spreading potential, BCC is associated with low mortality (and is therefore typically excluded from general cancer statistics) [5]. Nodal metastases are uncommon in people without SCC, but individuals with this disease are at an increased risk of getting them [6]. Malignant melanoma, conversely, is the worst kind of SC because it is so difficult to diagnose before spread throughout the body. A dermatologist can make a preliminary diagnosis of SC by visually inspecting the PSL for the ABCDE (Asymmetry of the mole, Border irregularity, Color uniformity, Diameter and Evolving size, shape or color)

ISSN NO: 2770-4491
Date of Publication:26-03-2023

guideline [7]. According to this evaluation, if the dermatologist has reason to believe the lesion is cancerous, a biopsy will be conducted. Once a sample has been collected, it is analyzed pathologically to determine a final diagnosis.

To help dermatologists in their daily work, numerous technologies have been developed that use dermoscopic images and algorithms to implement the ABCD rule (without considering changing characteristics) for PSL evaluation and classification [8][9]. However, present methods are not precise enough, leading to a number of mistakes in diagnosis. Due to the lack of certainty in existing diagnoses, therefore more efficient methods of SC diagnosis is requested to reduce the number of unnecessary surgical procedures.

In the last few years, hyperspectral imaging has become a non-invasive, non-ionizing, and label-free imaging modality in the medical field (HSI). This imaging method can combine digital imaging with spectroscopy techniques to get more information about a scene's spectral properties, both in the visible range of the electromagnetic spectrum and beyond [10]. In a hyperspectral (HS) image, each pixel shows the so-called spectral signature of the material or substance at its corresponding spatial coordinates. It has been shown that spectral signature analysis can be used to get quantitative information about how a tissue works [11]. A lot of research has been done on the basics of this technology and the tools used to collect this kind of data for in-vivo medical applications [12]. But, the review by Johansen et al. [13] shows that there isn't much research on using HSI to find SC when the person alive.

For this paper using segmentation of image means breaking it up into distinct regions that share similar characteristics.

Image segmentation is a crucial, technique in digital image processing, and the accuracy of segmentation directly influences the effectiveness of a tasks. Considering its complexity present segmentation algorithm has achieved particular to varied degrees, but the study on this facet confronts numerous problems. Since the clustering analysis algorithm categorizes data sets into subsets according to predetermined criteria, it finds extensive use in the field of image segmentation.

In this study the luminance component in color space is fixed to filter the influence of background. A new way to figure out the value of K in the K-means clustering algorithm was also suggested. The image segmentation method we proposed in our paper has shown good results in the field of SC images.

2. Literature review

In the most recent few years, image segmentation approaches based on clustering have been presented. The k-means clustering technique is a popular option among the many different clustering algorithms available.

In 2017, Aljawawdeh, Imraiziq and Aljawawdeh. investigated the detection of melanoma in skin images and suggested a model for melanoma segmentation and classification. Utilizing the genetic algorithm and particle swarm optimization, the proposed method enhanced the segmentation produced by the k-means clustering algorithm. Features for classification were retrieved from segmented images, while neural networks were employed for categorization. The upgraded k-means algorithm produced the highest accuracy of classification [14].

In 2019, Roberta, *et al.* Suggest the use of themes to examine the diagnosis of skin lesions on the computer. Four color spaces are used, including CIE $L'a'b'$, Red Green and Blue (RGB),Hue, Saturation and Value (HSV) and CIE L*u*v*. They were used to extract the color and properties of the texture. Six distinct classifiers were used with the aim of evaluating this approach, including k-nearest neighbors (kNN), Bayes Networks, C4.5 decision tree, Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Optimal Power Flow (OPF) [15].

In 2020, Ahmad, B., *et al.* Proposed a model for better dermatological categorization using deep CNN with Triple loss function. To fix the issues with facial skin images, ResNet152 and InceptionResNetV2 layers were added. In the first of several stages, using a training set, draw out the image's salient elements. representations in the Euclidean metric. A subsequent step is to categorize skin disorders. Acne, Blackheads, Dark Circles, and Spots were considered as four types of skin problems [16].

ISSN NO: 2770-4491
Date of Publication:26-03-2023

3. Methodology

The aim of this study is to segment color in an image using $L'a'b'$ color space and k-means clustering. Figure (1) shows the block diagram of the proposed system, the entire processes are as follow:

1-read the input image

2-convert image from RGB to $L'a'b'$color space

3-performe k mean clustering using k=3 to cluster the image into three clusters using Euclidian distance. three clusters exist in a'b' layer

4- use l layer to separate the region of interest.

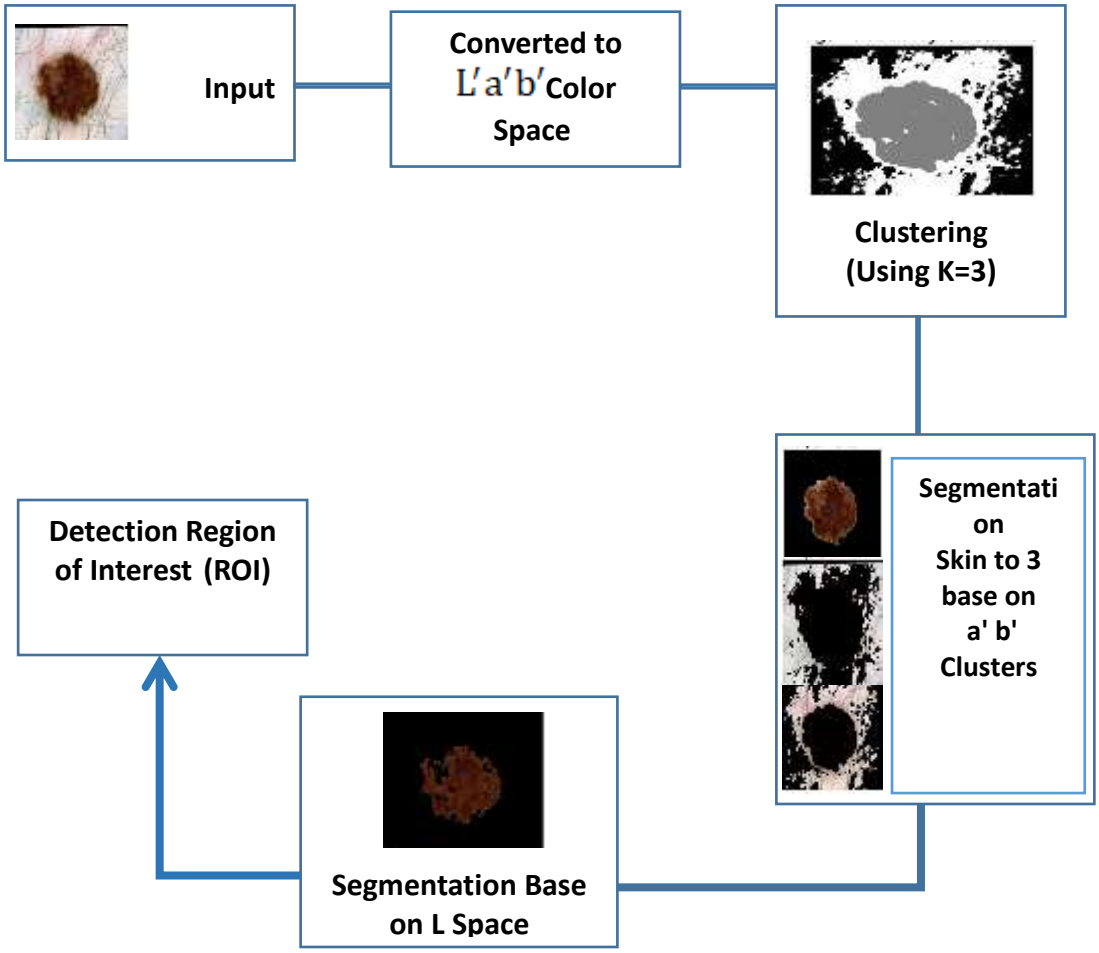


Figure 1: the block diagram of the proposed system

2.1 $L'a'b'$Color Space Conversion

The broad gamut of $L'a'b'$ allows it to display not just the whole gamut of colors that RGB and CMYK can produce, but also those which can't even come close to [6]. The $L'a'b'$model is a way to describe how the human eye interprets colors. Also, the RGB color model has too much transition color between blue and green, but the $L'a'b'$color model adjusts for this by reducing the amount of red, green, and blue in the spectrum. However, the range of colors from green to red is missing yellow. Because we need to preserve as much as possible of the original color space when working with skin images, $L'a'b'$ has been chosen.

After extensive testing, we discovered that diverse SC images will invariably result in an uneven background due to variances in conditions such as the light and the color of the skin itself which may have a significant impact on the segmentation findings. In lab color photographs, luminance or lightness (L) of white-black and chromatic color components is used to determine how to color each pixel (a) red-green and (b) yellow-blue [17].

ISSN NO: 2770-4491
Date of Publication:26-03-2023

First and foremost, we must realize the conversion of the image's $L'a'b'$color space and RGB color space. Because RGB cannot be transformed directly into $L'a'b'$, it must first be turned into MNO and then into $L'a'b'$, represented as RGB to MNO then to $L'a'b'$ [18][19].

As a result, our conversion is divided into two stages:

First stage is the conversion from RGB to MNO. The RGB format has three channels of pixels, with a range (0 to 255) for each channel.

Equations 1-3 give the conversion formulas:

$$R = gamma\left(\frac{r}{255}\right) \quad (1)$$

$$G = gamma\left(\frac{g}{255}\right) \quad (2)$$

$$B = gamma\left(\frac{b}{255}\right) \quad (3)$$

$$\text{Gamma (z)} = \begin{cases} \left(\frac{z+0.055}{1.055}\right)^{2.4} & \text{if } z > 0.04045 \\ \frac{z}{12.92} & \text{otherwise} \end{cases} \quad (4)$$

Equation 4 can be used to find gamma function, which is not fixed and it is used for nonlinear tone editing to increase contrast in the image.

Second stage is the conversion from ZUV to $L'a'b'$. Equations 5-7 represent this conversion [20].

$$L' = 116t\left(\frac{U}{U_n}\right) - 16 \quad (5)$$

$$a' = 500\left[\left(\frac{Z}{Z_n}\right) - t\left(\frac{U}{U_n}\right)\right] \quad (6)$$

$$b' = 200\left[\left(\frac{U}{U_n}\right) - t\left(\frac{V}{V_n}\right)\right] \quad (7)$$

Where *t* represents the calibration function similar to gamma.

The most important calculation to indicate the cluster's compactness is the average intra-cluster distance, which can be determined by using equation 8.

$$AveaIntra = \frac{1}{N}\sum_{i=1}^{k}\sum_{x \in C_i} \|X - Z_i\| \quad (8)$$

So when two results are being compared, the one with the lower value for this measure is the better choice. Where k represents the number of region in the segmented image and *N* represents the number of pixels in the i region [21].

Based on the development of a between cluster scatter matrix S_b and a within cluster scatter matrix S_w, we may obtain the scatter matrix j(x). The scatter matrix S_b between clusters can be calculated by equation 9

$$S_b = \sum_{i=1}^{K} \frac{N_i}{N} (z_i - z)(z_i - z)^T \quad (9)$$

ISSN NO: 2770-4491
Date of Publication:26-03-2023

Where N_i is the number of patterns in cluster. C_i, N_i the total number of patterns, z_i is the cluster center of cluster C_i and z is the mean of all patterns.

The cluster scatter matrix S_w can be calculated using equation 10.

$$S_w = \frac{1}{N}\sum_{i=1}^{K}\sum_{\mathrm{x}\in C_i}(\mathrm{x}-\mathrm{z_i})(\mathrm{x}-\mathrm{z_i})^T \quad (10)$$

Where X represents the patterns [21]. the basic criteria based on the scatter matrix can be described by equation 11.

$$J(x) = \mathrm{tr}\,(S_b + S_w) \quad (11)$$

Where tr represents the summation of diagonal elements of the matrix.

2.2 Image segmentation

The procedure of isolating the sick area from normal skin based on the uniformity of the pixels is known as segmentation. Color and texture characteristics can influence pixel homogeneity. We divided the skin lesion sections using color attributes of the affected area.

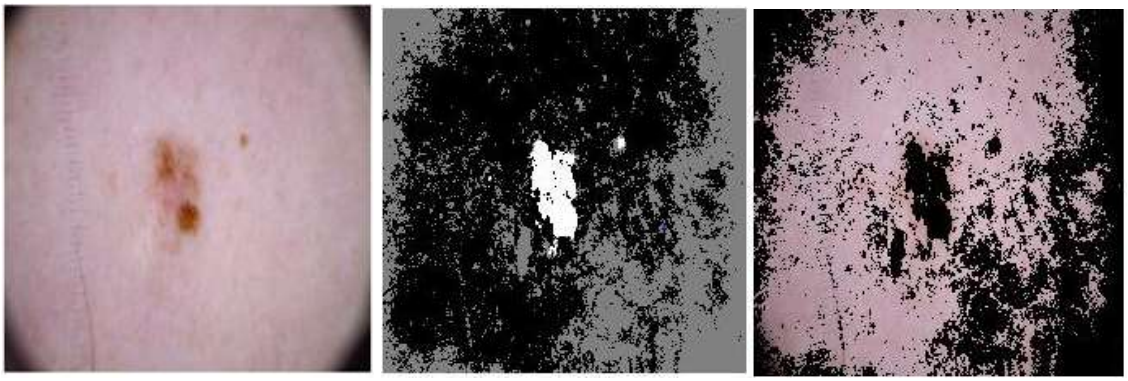
4. Results and discussion

In this paper SC detection has been implemented using color based segmentation with k mean clustering. The dataset used have been collected from the Kaggle skin cancer collection challenge [22]. first of all, the input image has been converted from RGB color space to $\mathrm{L'a'b'}$ color space.

In $\mathrm{L'a'b'}$color space image luminosity and color are separated which make regions segmentation by color easier and independent of lightness. Moreover, this kind of color space is suitable with human visual perception of the regions of image which is distinct with white.

The $\mathrm{L'a'b'}$color space consists of three layers, luminosity layer L', the chromaticity layer a' which indicates where a color falls along the red-green axis, and the chromaticity layer b' which indicates where a color falls along the blue-yellow axis. The k mean cluster has been applied with (k=3) to separate the image pixels into three clusters, every pixel in the image has been labeled with its cluster index.

Segmentation has been done to separate the areas in image using a'b' layer resulting three images (cluster 1 ,cluster 2 , cluster 3) , figure (2) shows sample of results , (a) is the input image , (b) refers to image labeled by cluster index ,(c) shows segmentation with a'b' layers and (d) shows segmentation with l layer which is the actual affected area. This is one of the technologies used in healthcare and the medical field [23-27].



(a) (b) (c) cluster1

ISSN NO: 2770-4491
Date of Publication:26-03-2023

(c) cluster2 (c) cluster3 (d)

Figure 2: Cancer detection (a) input image (b) image labeled by cluster index (c) segmentation with a'b' values (cluster 1, cluster 2 and cluster 3) (d) segmentation with l value

5. Conclusion segment color in an image using $L'a'b'$ color space and k-means clustering

In this paper, SC detection using segmentation with k mean clustering was presented, image segmentation using $L'a'b'$ color space and using k mean clustering with k=3 has been used to detect the exact sick area .

The obtained results proved the effectiveness of this method

References

1. J. Kato, K. Horimoto, S. Sato, T. Minowa, and H. Uhara, "Dermoscopy of Melanoma and Non-melanoma Skin Cancers," *Frontiers in Medicine*, vol. 6, no. August, pp. 1–7, 2019, doi: 10.3389/fmed.2019.00180.
2. "Ultraviolet radiation." https://www.who.int/news-room/fact-sheets/detail/ultraviolet-radiation (accessed Dec. 13, 2022).
3. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018, doi: 10.3322/caac.21492.
4. A. Di Stefani, I. Zalaudek, G. Argenziano, S. Chimenti, and H. P. Soyer, "Feasibility of a two-step teledermatologic approach for the management of patients with multiple pigmented skin lesions," *Dermatologic Surgery*, vol. 33, no. 6, pp. 686–692, 2007, doi: 10.1111/j.1524-4725.2007.33144.x.
5. I. Nindl, M. Gottschling, and E. Stockfleth, "Human papillomaviruses and non-melanoma skin cancer: Basic virology and clinical manifestations," *Disease Markers*, vol. 23, no. 4, pp. 247–259, 2007, doi: 10.1155/2007/942650.
6. V. Madan, J. T. Lear, and R. M. Szeimies, "Non-melanoma skin cancer," *The Lancet*, vol. 375, no. 9715, pp. 673–685, 2010, doi: 10.1016/S0140-6736(09)61196-X.
7. H. Tsao *et al.*, "Early detection of melanoma: Reviewing the ABCDEs American Academy of Dermatology Ad Hoc Task Force for the ABCDEs of Melanoma," *Journal of the American Academy of Dermatology*, vol. 72, no. 4, pp. 717–723, 2015, doi: 10.1016/j.jaad.2015.01.025.
8. S. Jain, V. Jagtap, and N. Pise, "Computer aided melanoma skin cancer detection using image processing," *Procedia Computer Science*, vol. 48, no. C, pp. 735–740, 2015, doi: 10.1016/j.procs.2015.04.209.
9. R. Kasmi and K. Mokrani, "Classification of malignant melanoma and benign skin lesions: Implementation of automatic ABCD rule," *IET Image Processing*, vol. 10, no. 6, pp. 448–455, 2016, doi: 10.1049/iet-ipr.2015.0385.
10. T. Nagaoka, A. Nakamura, Y. Kiyohara, and T. Sota, "Melanoma screening system using hyperspectral imager attached to imaging fiberscope," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 30, pp. 3728–3731, 2012, doi: 10.1109/EMBC.2012.6346777.
11. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, vol. 19,

ISSN NO: 2770-4491
Date of Publication:26-03-2023

no. 1, p. 010901, 2014, doi: 10.1117/1.jbo.19.1.010901.
12. M. Halicek, H. Fabelo, S. Ortega, G. M. Callico, and B. Fei, "In-vivo and ex-vivo tissue analysis through hyperspectral imaging techniques: Revealing the invisible features of cancer," *Cancers*, vol. 11, no. 6, pp. 1–30, 2019, doi: 10.3390/cancers11060756.
13. . H. Johansen *et al.*, "Recent advances in hyperspectral imaging for melanoma detection," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 12, no. 1, pp. 1–37, 2020, doi: 10.1002/wics.1465.
14. A. Aljawawdeh, E. Imraiziq, and A. Aljawawdeh, "Enhanced K-mean Using Evolutionary Algorithms for Melanoma Detection and Segmentation in Skin Images," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 8, no. 12, pp. 477–483, 2017.
15. O. B, Roberta, P. Aledir S, and T. João Manuel R.S, "Computational diagnosis of skin lesions from dermoscopic images using combined features," *Neural Computing and Applications*, vol. 10, pp. 6091–6111, 2019.
16. B. Ahmad, M. Usama, C. M. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, "Discriminative Feature Learning for Skin Disease Classification Using Deep Convolutional Neural Network," *IEEE Access*, vol. 8. pp. 39025–39033, 2020, doi: 10.1109/ACCESS.2020.2975198.
17. A. S. R. Sinaga, "Color-based Segmentation of Batik Using the L*a*b Color Space," *SinkrOn*, vol. 3, no. 2, p. 175, 2019, doi: 10.33395/sinkron.v3i2.10102.
18. L. V. Ri and U. R. S. Xvlqj, "<Lhog (Vwlpdwlrq Ri &Klool &Urs Xvlqj ,Pdjh 3Urfhvvlqj 7Hfkqltxhv," no. February, pp. 200–204, 2019.
19. N. K. EL Abbadi and E. Saleem, "Automatic gray images colorization based on lab color space," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 3, pp. 1501–1509, 2020, doi: 10.11591/ijeecs.v18.i3.pp1501-1509.
20. S. Saifullah, R. Dreżewski, A. Khaliduzzaman, L. Karlo Tolentino, and R. Ilyos, "K-Means Segmentation Based-on Lab Color Space for Embryo Detection in Incubated Egg," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 8, no. 2, pp. 175–185, 2022, doi: 10.26555/jiteki.v8i2.23724.
21. H. Turi and B. Comp.(Hons), "Clustering-Based Colour Image Segmentation," *School of Computer Sciience and Software Engineering Monash University,Australia*, vol. 4, no. 1, pp. 88–100, 2001.
22. K. SCOTT MADER, "Skin Cancer MNIST: HAM10000" kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000, Accessed 12 November 2020.
23. Marwa MohamedSheet AL-Hatab, Raid Rafi Omar Al-Nima, Ilaria Marcantoni, Camillo Porcaro, and Laura Burattini, "Classifying Various Brain Activities by Exploiting Deep Learning Techniques and Genetic Algorithm Fusion Method", *TEST Engineering & Management*, Vol. 83, pp. 3035-3052, 2020.
24. Marwa MohamedSheet AL-Hatab, Raid Rafi Omar Al-Nima, Ilaria Marcantoni, Camillo Porcaro, and Laura Burattini, "Comparison Study Between Three Axis Views of Vision, Motor and Pre-Frontal Brain Activities", *Journal of Critical Reviews*, Vol. 7, Issue 5, pp. 2598-2607, 2020. doi:10.31838/jcr.07.05.430
25. Marwa Mawfaq Mohamedsheet Al-Hatab, Raid Rafi Omar Al-Nima, Maysaloon Abed Qasim,"Classifying healthy and infected Covid-19 cases by employing CT scan image", *Bulletin of Electrical Engineering and Informatics*, vol.11,Issue.6,pp. 3279-3287,2022.
26. Marwa Mawfaq Mohamedsheet Al-Hatab, Mohammed Abid Alhashim, Maan Ahamed Fadhil, AL-Jadir Ali R Hasan, Tahani Ghanim Al-Sultan," Innovative Non-Invasive Blood Sugar Level Monitoring for Diabetes Using UWB Sensor", *Journal of Optoelectronics Laser*, vol.41,Issue.4, pp. 422-437,2022.
27. Marwa Mawfaq Mohamedsheet Al-Hatab, and Mohammad Zaid Shuaib AlNima. "Hematological Classification of White Blood Cells by Exploiting Digital Microscopic Images." *Eurasian Research Bulletin* , vol.18 , 2023.