Bioengineering Research

A Proposed Algorithm for the Detection of Thyroid Cancer based on Image Processing

Mehrad Jaloli[1], Mohammad Fathi[1], Seyed Mohsen Mohammadi[2], and Reza Abbasi-Kesbi[3*]

[1]Department of Biomedical Engineering, Faculty of Medical Sciences and Technologies, Islamic Azad University Science and Research Branch, Tehran, Iran,

[2]Department of Mathematics, Faculty of Basic Sciences, Yadegar-e-Imam Khomeini (RAH) Shahre Rey Branch, Islamic Azad University, Tehran, Iran.

[3]MEMS & NEMS Laboratory, Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran

Correspondence to: Abbasi-Kesbi R. (E-mail: reza.abbasi@ut.ac.ir)

Abstract

Introduction: Today's world allows digital images to be downloaded and stored. Computer image processing systems have been developed to make these actions faster and more accurate, especially in medical engineering. Many diseases, including brain tumors, cancers are identified and estimated by benefiting of the image processing.

Objective: In this paper, a proposed algorithm for detection the thyroid cancer is presented.

Material and Methods: To this end, a few images of thyroid cancer are collected, and then these images are converted to jpeg format. After that, these images are called and converted into a matrix using MATLAB software. The obtained matrix is normalized and then, the noise of these images are removed. Next, the highlighted areas are defined by a threshold value. By summing up this highlighted value and negation of the original images, the cancerous area is determined.

Results: The results of the five images illustrate that the accuracy of the algorithm for the detection of thyroid cancer is 93.5% that show an improvement of 3-7 % than other works. Furthermore, sensitivity, specificity, and F-score, Mathew Correlation Coefficient (MCC) are 62%, 93.2%, 92.2%, and 86.4%, respectively, for the proposed method.

Conclusions: The method is a simple way and owns an acceptable accuracy that can be used for the detection of thyroid cancer in a portable computer. As the images show the cancerous parts have lighter pixels than non-cancerous parts that the feature is used for separating the cancerous part of the thyroid gland from non-cancerous parts.

Keywords: Thyroid cancer, image processing, rotational mean filter, static parameters

Received: 24 August 2019, Accepted: 22 September 2019

DOI: 10.22034/jbr.2019.199962.1013

1. Introduction

The thyroid gland is one of the vital glands of the body that indirectly affects all body organs such as the heart, kidneys, and so on [1]. The butterfly gland is located in the neck and in the front of the larynx [2]. The thyroid gland produces the thyroid hormone using iodine in foods and controls the rate of metabolism (the metabolism is the system by which the body uses its energy) [3]. The thyroid function is regulated by the pituitary gland located in the brain, which means that the thyroid hormone production is controlled by the pituitary gland [4].

Thyroid disorders can slow down the body metabolism. When the level of hormones is very low or very high, symptoms of this disorder are seen in the hypothyroidism or hyperthyroidism [5]. The most common thyroid disease comprises hypothyroidism, hyperthyroidism, thyroid enlargement (goiter), thyroid masses, thyroid cancer and thyroid inflammation (thyroiditis) [5]. Hypothyroidism or hyperthyroidism are the underlying causes of other illnesses in many cases. For example, high sugar and high fat can both be caused by hyperthyroidism, while hypothyroidism emerges with iodine deficiency in the body that leads to goiter disease [6]. Thyroid hypothyroidism is 3 times more common in Iranian newborns than in European countries [7]. Moreover, elderly people and women suffer from poor thyroid function than other people. Interpreting the results of laboratory tests for older people or pregnant women, or everyone who consumes other medicines or patients who are in special medical conditions is difficult [8-10].

Thyroid disease is one of the most commonly diagnosed diseases in the world today. Failure to check any of the thyroid disorders can lead to serious complications. Several studies have been done in this regard and have achieved a well-defined diagnostic accuracy [11-15]. In a study [14], using an algorithm that was called hierarchical method has been attempted to increase the accuracy of the diagnosis of thyroid disease. This algorithm is based on the behavior pattern that the specialist trained to diagnose thyroid disease and aimed to diagnose the type of thyroid disease (hypothyroidism, hyperthyroidism). Recently, data mining techniques have attracted many researchers to discover patterns for diagnosing diseases, in addition to industrial ones. Data mining systems can explore patterns in medical data and improve the decision-making process, thereby affecting the cost of treatment and improving the quality of health care [15-21].

Image processing is now increasingly referred to as digital image processing, a branch of computer knowledge that deals with digital signal processing, representing images taken with a digital camera or scanned by a scanner [22]. The processing images has two major branches improving images and vision of the machine. Improving the images involves techniques such as using a fader filter and increasing contrast to improve the visual quality of images and ensure that they are displayed correctly in the destination environment (such as a printer or computer display), while the machine vision is in ways that can perceive the content of the images and their means for using them in robotics and image axes [23]. In the specific mean of the image processing, it is any kind of signal processing that enters an image, such as a picture or a scene from a movie. The output of the image processor can be an image or a set of special characters or image variables. Most image processing techniques include dealing with the image as a two-dimensional signal and employing standard signal processing techniques on them [24]. Image processing often refers to digital image processing, but optical image processing and image analog also exist.

In this paper, a simple algorithm for the detection of thyroid cancer based on image processing is presented. At first, a few images of thyroid cancer are collect from [25] and then converted to jpeg format. After that, these images are called and converted them into a matrix. The obtained matrix is first normalized, and then, by defining the threshold values in original and negative images, the cancerous areas are separated than parts that are not cancerous. As the results prove, the cancerous parts have lighter pixels than non-cancerous parts that with having such character (brighter), the threshold values are selected. Moreover, the result illustrated having the accuracy of the proposed method is 93.5% that in comparison to [26-27] an improvement of 3-7% was made.

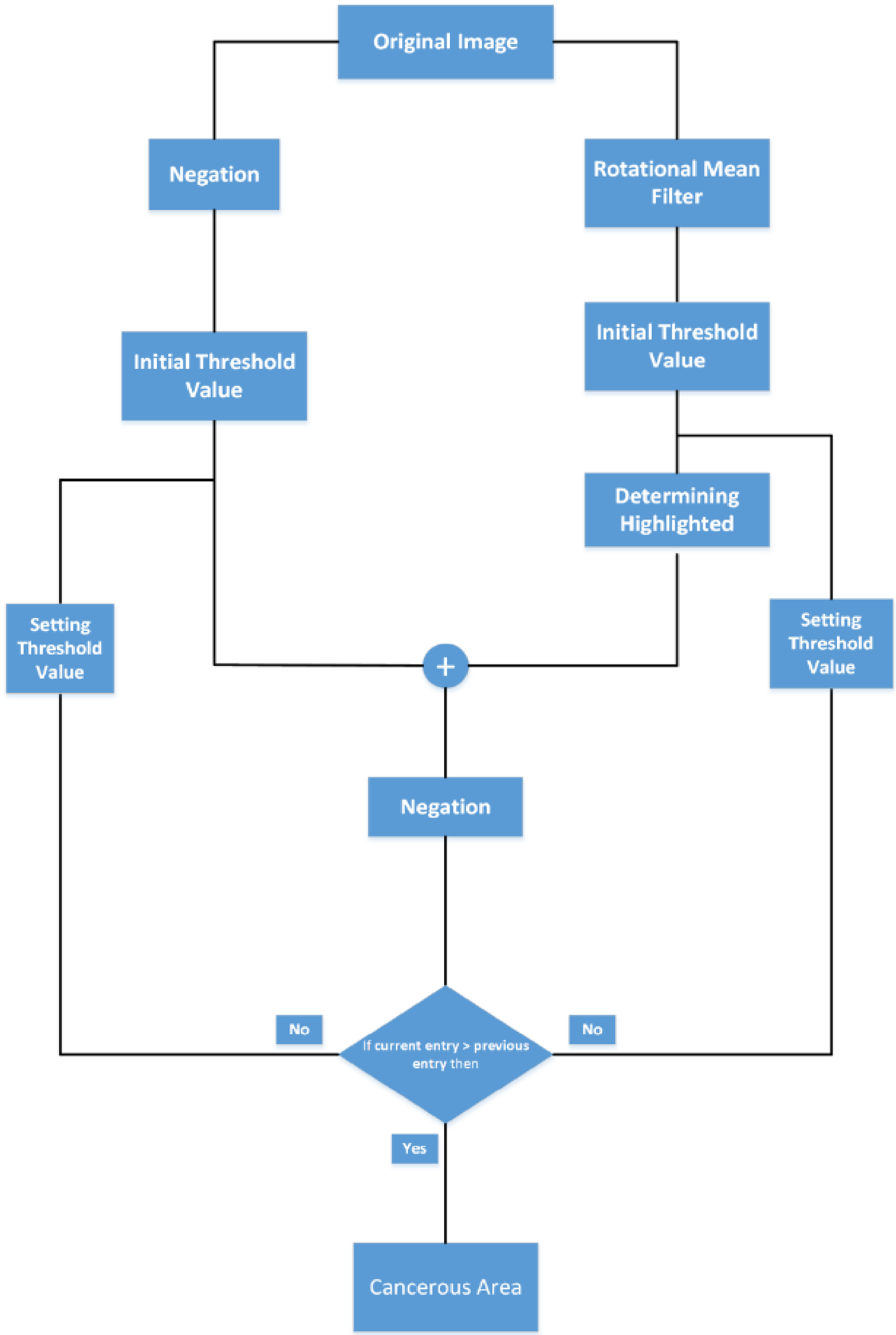

Figure 1: The proposed algorithm for cancer detection

2. Material and Method

2.1. The proposed Method

In order to detect thyroid cancer by image processing, first, several images of the thyroid cancer are collected with the help of special software. Five images are collected from [25] and then called in MATLAB software. After calling, it is necessary that these images reach to high-quality (pre-processed). To this end, the images are passed in a rotational mean filter. This rotational mean filter measures the average rotating neighboring points to reduce the noise effect. After that, the cancerous area is highlighted. In order to determine the highlighted area, the threshold value for this area is set. For this purpose, the area is started with an initial threshold value and the threshold change until maximum bright pixel is obtained (Segmentation).

As the obtained images show, the cancerous area appears white in color than in other areas. Thus, it should be better all the entry in the proposed algorithm are compared with the threshold value that is closer to the color of the light. In this paper, this value is set several times, and so it is experimental and its value is 190 (edge detection or edge detection). If the value of the entry is greater than this value, then it shows that the entry is related to the non-cancerous area, and the value to 255 is set in the entry. If the value of the entry of this value is less than the edge, then the value of the entry is converted to 0 in this new tab. According to this, a highlighted image have been made.

As it is mentioned earlier, the entries of the matrix contain values from 0 to 255. In order to the negation of the image, the original image is subtracted from the value of 255. Again, mandatory setting the threshold value in the negation image for having the highlighted image in negation state (segmentation). First, an initial threshold is chosen and then the value is repeated until having maximum light pixel in the stage. Now, this image, which is taken from the negation of the image, is summed with the image of the previous step. After negating this outcome, the cancerous area is extracted. Figure 1 shows the proposed algorithm to identify the cancerous area.

2.2. Determining important statically parameters

It should be noted that in order to obtain accuracy, a four-entry matrix after identifying the highlighted area is presented. As it is mentioned earlier, although the highlighted area is started with an initial value, it reaches an optimum threshold value after several repetitions (having maximum light pixels). In this situation, the four-entry matrix (confusion matrix) is defined that its entries differ for every image. In the mentioned matrix, the first entry (Tp) is the number of bright pixels, the second entry (Fp) is the pixels that are not bright but are inclined to be bright. In the third entry (Fn), the number of pixels that are inclined to dark, and in the fourth entry (Tn), the number of dark pixels is considered. Table 1 shows the confusion matrix and its entries for the method.

Table 1: Confusion matrix

	Estimated Value	
Real value	T_p	F_p
	F_n	T_n

The values of this matrix are for obtaining the following parameters: the accuracy, specificity, sensitivity, F-corse, and MCC can be defined as follows:

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (1)$$

$$Sensitivity = \frac{T_p}{T_p + T_n} \quad (2)$$

$$Specifcity = \frac{T_n}{F_p + T_n} \quad (3)$$

$$FScore = \frac{2T_p}{2T_p + F_p + F_n} \quad (4)$$

$$Mcc = \frac{T_p \times T_n - F_p \times F_n}{\sqrt{(T_p + F_p)(T_p + F_n)(T_n + F_p)(T_n + F_n)}} \quad (5)$$

Where T_P is true detection of bright pixels in images using test results, F_P is false detection of bright points in the image using the test results, T_N is true detection of dark spots in images using test results and FN is false detection of dark points in images using the test results.

Accuracy, in fact, indicates the degree of measurement to the actual value. Here, accuracy is the detection of more cancerous points or brighter pixels. Sensitivity means that our test how much have been able to select bright pixels. In other words, the sensitivity is the number of bright pixels is correctly selected by the test rather than the whole of the bright and dark pixels that are correctly selected. Specificity is the ability to test for the correct selection of images that do not have dark pixels.

In the statistical analysis of the binary classification, the F-score is a measure of the accuracy of the test. Both the accuracy p and the reminder r compute the test for calculating the score. In Fact, p is the number of correct positive results divided by the number of all classified positive results and r is the number of correct positive results divided by the number of all relevant instances (all examples should be positive). F-score is the mean harmony of accuracy, while the best F-score is at 1 (full accuracy) and the worst at 0.

The Mathew Correlation Coefficient (MCC) is used in machine learning as a two-class classification coefficient (two classes), introduced by W. Matthews in 1975. It considers positive and negative both true and false and is generally regarded as a balanced measure, which can be used even if the classes are of very different sizes. MCC is basically a correlation coefficient between the observed and proposed binary classifications and is expected to return values between -1 and +1. The +1 coefficient represents the complete prediction, 0 is not better than a random prediction, and -1 indicates a general disagreement between prediction and observation.

3. Results & Discussions

In order to detect thyroid cancer with the help of image processing, we need to collect the cancer images from [25] with the help of special software such as Snip tools. Then, we call the images in the m. file. As we know, photos are two-dimensional. For example, the photo shown in Figure 2, has 290 rows and 401 columns. Each of these entries of the matrix actually represents pixels of the image for us. It should be noted that these images are taken manually with the help of the Snip tool software, resulting in the difference of entry of the matrix. If we draw the original image, we will have the Figure2, in which the entry of this matrix will have different values from 0 to 255. The value of 255 means the color of black and the more distances than 0 means closer to the white color.

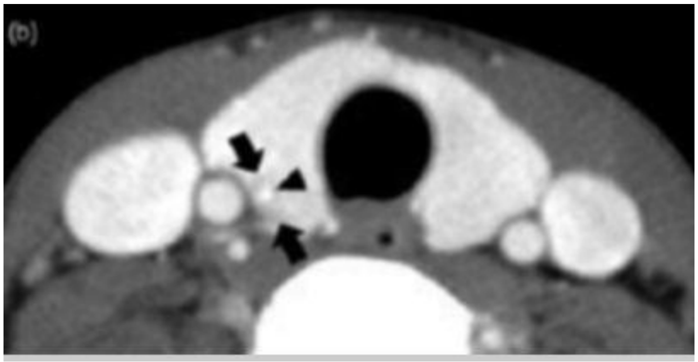

Figure 2: Image of thyroid cancer from [25].

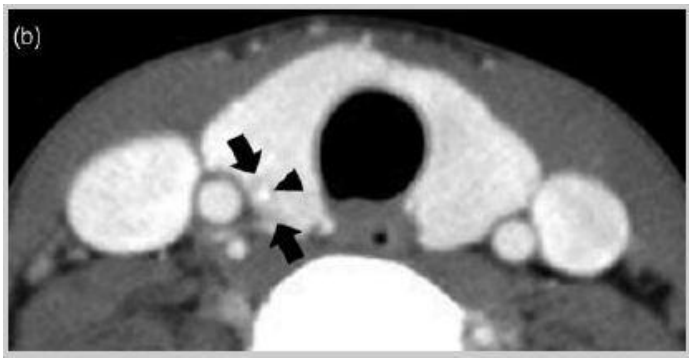

Figure 3: the filtered images using the rotational filter and setting image resolution.

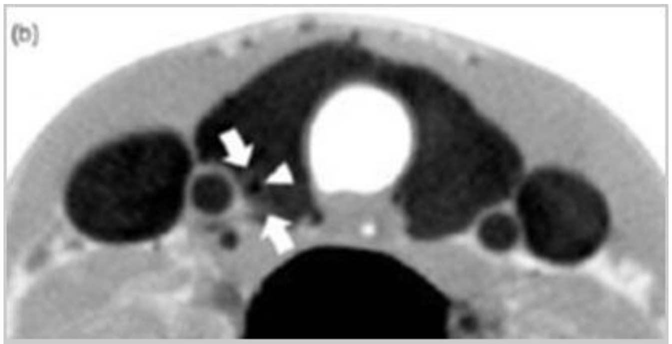

Figure 4: the negative image of the Figure 3

In order to create an image data set for the input of image processing, RGB images were used. The first step in the image processing method for detection of the thyroid cancer is to pre-process in which increases the quality of the input images for further analysis. Pre-processing is performed to eliminate any existing noise or interference. In other words, it is making images to get better results. In a similar way, the proposed method for detecting cancer relies on reducing the noise of the images at first, increasing the accuracy of the work and identifying the cancerous area with higher precision.

If the obtained images are colored, we turn them to the black and white for simplifying in the analysis. After that, we pass these images from a rotating filter to reduce their noise. The rotating filter is two orders, whose function is to improve image quality in order to increase accuracy. With the help of the filter, the image is filtered. The image after passing the filter with the order of 2 will be as Figure 3.

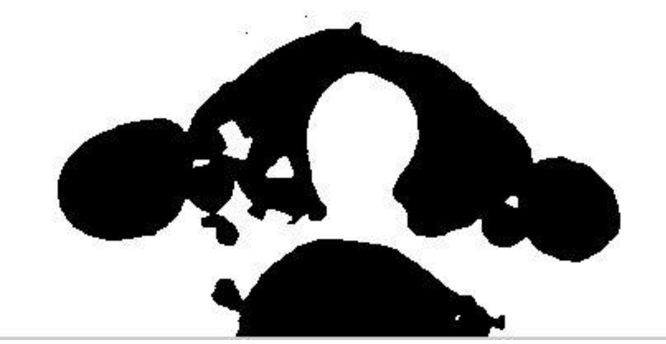

Figure 5: The highlighted area that is associated thyroid cancer

Figure 6*:* The negative image of Figure 5

Also, the quality of the image increases using the set of image resolution. In addition, the edges are sharpened, which will help to further enhance the image quality. Now, the existing noises are minimized and the image can be used for further analysis. If the image of Figure 3 is negated, Figure 4 will be obtained.

Then, the highlighted area is defined as that has mentioned in the material and method section. In other words, if one entry is more than 190, the entry is associated with black. If the entry value is lower than 90, the entry is associated with white. By considering Figure 4 and 3, two highlighted area is obtained. After acquiring the highlighted area from two figures, Figure 5 is extracted by summing up two highlighted area.

It should be noted that the other values from140 to 190 are for expanding the accuracy in the work for obtaining Fn in the mentioned confusion matrix while the values from 90 to 140 are associated with Fp. Also, the value of more than 190 is related to Tn and the value of lower than 90 is related to TP. The amount of white area is manually selected. By setting this value and checking it in a repeating loop, most white pixels are selected which indicate the accuracy of the proposed method. By negating the Figure 5, Figure 6 is obtained. Thus, the cancerous area was estimated.

Now, the number of bright pixels for determining precision was manually selected, that was described in the previous sections. In order to reach a high precision, the method was examined 100 times for finding and extracting maximum bright pixels based on threshold values. If the pixel is white, one unit is added to the entry of Tp. If the pixel is black, one value is added to the entry of Tn. If the pixel tends to white, a single unit is added in the entry of the Fn, and if the pixel tends to black, one unit is added in the entry of the Fp. Finally, the obtained values for this image will be as follows:

Tp=28868 ,   Fp= 3048 ,   Fn=3078 ,   Tn=81006

The obtained values with the help of the equations 1-5 at the end of material and method will be as follows:

Ac=%95 Se=81 % Sp=96 % Fc=90% Mcc=87%

If the proposed method is repeated for other images from [25], the obtained results are as follow. The final results are considered here because of the reputation procedures and preventing redundancy.

The result from all images will be as Table 2. As Table 2 shows, the average accuracy, sensitivity, specificity, Fc, and MCC are, respectively, 93.5%, 62%, 93.2%, 92.2%, and 86.4%. Therefore, the algorithm can be an appropriate method in the field. In comparison to other works in [26-27], the proposed algorithm has higher accuracy in image processing for detecting thyroid cancer and improved 3-7% than them.

Table 2: Static values obtained for 5 images from [25] using the proposed algorithm

No.	Ac (%)	Se (%)	Sp (%)	FC (%)	**MCC (%)**
1	95	81	96	90	87
2	92	59	86	93	86
3	93	56	91	94	87
4	91	57	95	92	83
5	95	57	98	92	89
Mean	**93.5**	**62**	**93.2**	**92.2**	**86.4**

4. Conclusions

In this paper, a proposed method to determine thyroid cancer was presented based on image processing. In this method, first the noise of the images are deleted and then the highlighted area is set based on the threshold values in both original and negation original images. By summing up the highlighted area that extracted from these two images, the cancerous areas were determined. This work was done 100 times for every image with purpose reaching maximum accuracy. The mean accuracy was obtained 93.5 % for five images while mean sensitivity, specificity, F-score, and MCC were obtained 62%, 93.2%, 92.2%, and 86.4%, separately. The results gives an improvement 3-7% than other works.

Conflict of interest

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

Acknowledgements

Not applicable.

References:

[1] Brander, Antti, Pertti Viikinkoski, Juha Nickels, and Leena Kivisaari. "Thyroid gland: US screening in a random adult population." Radiology 181, no. 3 (1991): 683-687.

[2] Carrasco, Nancy. "Iodide transport in the thyroid gland." Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes 1154, no. 1 (1993): 65-82.

[3] Mandel, Susan J., and Louis Mandel. "Radioactive iodine and the salivary glands." Thyroid 13, no. 3 (2003): 265-271.

[4] Huszenicza, G. Y., M. Kulcsar, and P. Rudas. "Clinical endocrinology of thyroid gland function in ruminants." VETERINARNI MEDICINA-PRAHA- 47, no. 7 (2002): 199-210.

[5] Singer, Peter A., David S. Cooper, Elliot G. Levy, Paul W. Ladenson, Lewis E. Braverman, Gilbert Daniels, Francis S. Greenspan, I. Ross McDougall, and Thomas F. Nikolai. "Treatment guidelines for patients with hyperthyroidism and hypothyroidism." Jama 273, no. 10 (1995): 808-812.

[6] Levey, Gerald S., and Irwin Klein. "Catecholamine-thyroid hormone interactions and the cardiovascular manifestations of hyperthyroidism." The American journal of medicine 88, no. 6 (1990): 642-646.

[7] Veisani, Yousef, Kourosh Sayehmiri, Shahab Rezaeian, and Ali Delpisheh. "Congenital hypothyroidism screening program in iran; a systematic review and metaanalysis." Iranian journal of pediatrics 24, no. 6 (2014): 665.

[8] Abbasi-Kesbi, Reza, Atefeh Valipour, and Khadije Imani. "Cardiorespiratory system monitoring using a developed acoustic sensor." Healthcare technology letters 5, no. 1 (2018): 7-12.

[9] Valipour, Atefeh, and Reza Abbasi-Kesbi. "A heartbeat and respiration rate sensor based on phonocardiogram for healthcare applications." In 2017 Iranian Conference on Electrical Engineering (ICEE), pp. 45-48. IEEE, 2017.

[10] Abbasi-Kesbi, Reza, Alireza Nikfarjam, and Hamidreza Memarzadeh-Tehran. "A patient-centric sensory system for in-home rehabilitation." IEEE Sensors Journal 17, no. 2 (2016): 524-533.

[11] Zarkesh, Maryam, Mahdi Taghaddosi, Fereidoun Azizi, Azita Zadeh-Vakili, and Mahdi Hedayati. "Importance of epigenetic changes in the thyroid cancer incidence and their therapeutic applications." Pathobiology Research 17, no. 3 (2014): 1-24.

[12] Shukla, Shruti, Ashish Lakhmani, and Ambuj Kumar Agarwal. "Approaches of artificial intelligence in biomedical image processing: A leading tool between computer vision & biological vision." In 2016 International Conference on Advances in Computing, Communication, & Automation (ICACCA)(Spring), pp. 1-6. IEEE, 2016.

[13] Mukherjee, Manali, and Mausumi Maitra. "Application of Biomedical Image Processing in Blood Cell Counting using Hough Transform." In Biometrics: Concepts, Methodologies, Tools, and Applications, pp. 1175-1194. IGI Global, 2017.

[14] Salem, Christian, Danielle Azar, and Sima Tokajian. "An Image Processing and Genetic Algorithm-based Approach for the Detection of Melanoma in Patients." Methods of information in medicine 57, no. 01/02 (2018): 74-80.

[15] Chen, Hui-Ling, Bo Yang, Gang Wang, Jie Liu, Yi-Dong Chen, and Da-You Liu. "A three-stage expert system based on support vector machines for thyroid disease diagnosis." Journal of medical systems 36, no. 3 (2012): 1953-1963.

[16] Hsu, Wynne, Mong Li Lee, and Ji Zhang. "Image mining: Trends and developments." Journal of intelligent information systems 19, no. 1 (2002): 7-23.

[17] Abbasi-Kesbi, Reza, Hamidreza Memarzadeh-Tehran, and M. Jamal Deen. "Technique to estimate human reaction time based on visual perception." Healthcare technology letters 4, no. 2 (2017): 73-77.

[18] Alireza Nikfarjam, Abbasi-Kesbi, Reza, and Ardalan Akhavan Hezaveh. "Developed wearable miniature sensor to diagnose initial perturbations of cardiorespiratory system." Healthcare technology letters 5, no. 6 (2018): 231-235.

[19] Abbasi-Kesbi, Reza, and Alireza Nikfarjam. "A miniature sensor system for precise hand position monitoring." IEEE Sensors Journal 18, no. 6 (2018): 2577-2584.

[20] Nikfarjam. A., and Abbasi-Kesbi, R. "Denoising MEMS accelerometer sensors based on L2-norm total variation algorithm." Electronics Letters 53, no. 5 (2017): 322-324.

[21] Parsa, Y., Mirmalek, S.A., Kani, F.E., Aidun, A., Salimi-Tabatabaee, S.A., Yadollah-Damavandi, S., Jangholi, E., Parsa, T. and Shahverdi, E., 2016. A review of the clinical implications of breast cancer biology. Electronic physician, 8(5), p.2416.

[22] Deserno, T.M. ed., 2011. Biomedical image processing. Springer Science & Business Media.

[23] Webb, A. and Kagadis, G.C., 2003. Introduction to biomedical imaging. Medical Physics, 30(8), pp.2267-2267.

[24] Unser, M. and Aldroubi, A., 1996. A review of wavelets in biomedical applications. Proceedings of the IEEE, 84(4), pp.626-638.

[25] Kim, Dong Wook, Soo Jin Jung, and Hye Jin Baek. "Computed tomography features of benign and malignant solid thyroid nodules." Acta Radiologica 56, no. 10 (2015): 1196-1202.

[26] Li, X., Zhang, S., Zhang, Q., Wei, X., Pan, Y., Zhao, J., Xin, X., Qin, C., Wang, X., Li, J. and Yang, F., 2019. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. The Lancet Oncology, 20(2), pp.193-201.

[27] Keramidas, E.G., Maroulis, D. and Iakovidis, D.K., 2012. TND: a thyroid nodule detection system for analysis of ultrasound images and videos. Journal of medical systems, 36(3), pp.1271-1281.