



Investigation of the eczema and skin cancer disease diagnosis by using image processing techniques

Yusuf Süer Erdem ¹, Özhan Özkan ²

(Alınış / Received: 23.05.2022, Kabul / Accepted: 30.06.2022, Online Yayınlanma / Published Online: 30.06.2022)

Keywords

Eczema
Matlab
Image processing
Skin Diseases
Skin Cancer

Abstract: It is seen that many diseases, especially dermatological diseases, arise due to bad weather conditions such as high temperature, dust, smoke, and sun in the environment. The most common diseases are eczema caused by malnutrition, soil, bacteria, bad food, and other factors, and risky moles, which are usually caused by excessive sunlight or during childbirth. Due to all these environmental, physiological, and chemical factors, it is important to quickly detect all existing skin diseases, especially eczema and risky moles, and it has become inevitable to establish a less costly diagnostic system with the help of doctors to prevent the aggravation of the diseases. If eczema and risky skin problems progress, they will be difficult to treat and take a long time. For this reason, the research aims to take an image from the infection site and then process this image in many ways in a MATLAB environment to obtain an output that can help doctors in their work. Decision trees supported the currently used image processing and classification method, and the results of both methods are also compared with this method. According to the results obtained, the accuracy, sensitivity, and selectivity ratios of decision trees are obtained compared to image processing. The software used gives a warning when the image processing and decision tree methods give conflicting results. If this occurs, it is necessary to stick to the doctor's data. The system in this study aims to improve human life and make it safe by recognizing eczema and risky moles. In this study, samples were selected from various layers of the body. In addition, a different interpretation can be made in the normal situation. When this approach technique is applied, more appropriate results have emerged in the process of detecting eczema and risky moles. In addition, normal skin is also involved in the process. Being able to define the normal state has been a contribution to science and it is aimed in this study to facilitate the work of medical personnel.

¹ Author 1 Yusuf Süer Erdem (Department of Electrical and Electronics Engineering, Sakarya University, 54050 Sakarya, Turkey)

² Author 2 Özhan Özkan (Department of Electrical and Electronics Engineering, Sakarya University, 54050 Sakarya, Turkey)

1. Introduction

When skin diseases are examined, it is seen that there are many causes. Examples of physical skin diseases are wounds in the human body, burns, and cracks caused by cold. In addition, there are also skin diseases caused by parasites. In addition, another type of skin disease is seen as allergic skin disease [1].

Dermatological diseases are frequently seen in the community as diseases with low mortality. These diseases are thought to be important diseases due to their high morbidity, but it is seen that the number of patients with skin diseases who apply for medical help is low. As the years progressed, it has been determined that there has been an increase in the applications of patients with dermatological complaints to health institutions, of which society has become conscious. Since dermatological diseases are long-term diseases and these diseases are costly diseases, their importance is increasing [2].

When the studies on the subject are examined, it has been determined that the use of artificial intelligence and data mining applications is used effectively in the field of health. At this point, in many areas; It is seen that all activity takes place in many fields such as medical disease diagnosis, follow-up, cost estimation, imaging analysis, resource planning, emergency management, processing of unstructured data [3, 4, 5].

Artificial intelligence models, which are used as a result of bringing patient data in high dimensions to the functional point, play an active role in increasing data reliability and quality [6, 7]. Regarding the use of artificial intelligence in the field of health, legal and ethical processes are used effectively regarding the accuracy of clinical data, data management, and data protection [8]. The processes regarding this issue differ from country to country [9].

The use of artificial intelligence is especially used in the detection of diabetes. Different applications are used concerning artificial intelligence in the detection of diabetes. In the study of Mujumdar and Vaidedi in 2019, large data sources in the health management system are used to increase the accuracy rate in the detection of diabetes [10].

A decision tree is defined as a model used in data mining to show classifiers and regressions. The decision tree consists of some nodes and branches. In the classification carried out in the decision tree, leaves and classes are specified. It has been determined that the decision is made according to one or more attributes in another node. A decision tree is used extensively in data mining because it is simple and understandable. To interpret the output in the decision tree, the entire content is explained one by one, independently of an expert. A graphical method is applied in the decision tree. Since this method is used in the decision tree, it is seen that it is simpler than other classification methods. Due to a large number of nodes in the decision tree, it can be difficult to display the graph [11].

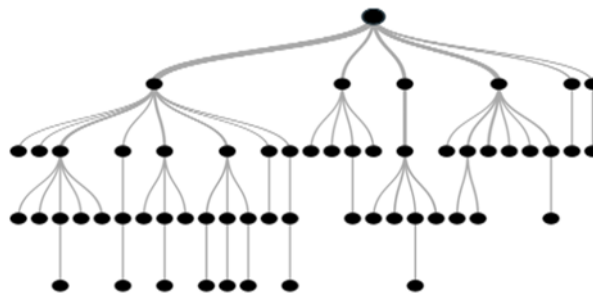


Figure 1. Example of Decision Tree

When the decision tree model was examined, it was determined that the model consisted of several parts. These parts:

- Creating node definitions to assign each record of the data set in the decision tree to a leaf node,
- Rules for assigning each record of the dataset to a leaf node,
- Revealing the posterior probabilities of the nodes of each leaf,
- A target level is assigned to each leaf node in the decision tree.

In this part, it was determined that node definitions were developed using the training data set. And it is specified in terms of input ranges. Using the training dataset, posterior probabilities are calculated for each node. Finally,

the target level found is assigned to each node and then the process is carried out using the training datasets within the training phase [12].

Many methods are used in the image analysis process. These methods are morphologically processing the image, extracting the image features, realizing the separation of the image into layers, identifying the patterns, and edge detection methods.

When morphological image processing is examined, it has been determined that it is a type of image processing in which the structure of objects in an image is changed. It is seen that three basic morphological processes, expansion, erosion, and skeletonization method are used in this species.

In enlargement, it is ensured that an object is reduced uniformly by erosion as a result of uniform enlargement within the spatial scale. In the skeleton method, especially an object is represented in the form of sticks [13].

It has been determined that there are activities to extract information such as high-level color and shape from images to obtain the characteristics of an image. For this reason, when the applied processes are examined, shadow detection, edge detection, contrast detection, motion detection, slope detection, and determination of the sequence of movements are made to examine the image properties caused by the density difference. In the operations applied in this framework, high- and low-level feature extraction operations are carried out. The basis of these processes is based on the principle of shadow, edge, contrast determination, detection of motion, and slope and determination of motion sequences depending on the image properties caused by the density difference [14].

It has been determined that especially machine learning increases the accuracy rate in the determination of diseases. When the studies on this subject are examined, a study has been made on the determination of skin diseases by using supervised machine learning algorithms because of the use of the data set on the patients, especially in the Indian sample [15].

When the health management processes in which machine learning methods are used are examined, a study related to the management of diabetes has been made. As a result of this examination, the prediction and determination of the disease, the complications, the genetic history of the patients, management, and treatment processes were examined. In this study, it is seen that 85% of classification and supervised algorithms are used in prediction regarding the machine learning process. In this study, the most preferred algorithm is the Support Vector Machine.

The reasons for the widespread use of machine learning techniques in the field of health sciences are listed as follows. These are:

- Estimation,
- Diagnosis and determination of post-disease complications,
- It is determined as providing quality service to patients by saving time and workload [16].

It is seen that a patient-oriented and artificial intelligence-based patient follow-up system is recommended for providing comfort for patients with diabetes and improving treatment processes within the health management processes [17].

In another study investigating the opportunities of machine learning in healthcare, a model classified according to the risk groups of patients using hospital data was proposed, and it was determined that there was an improvement in patient follow-up and related actions [18].

In another study, it is seen that an artificial intelligence model was created that identifies high-cost and risky patients to determine the necessary treatments and steps in the treatment of diabetes [19].

In the study of Mercaldo et al. in 2017, it was determined that the Hoeffding Tree algorithm was used to diagnose and classify diabetes [20].

Despite the availability of advancing technology, precisely diagnosing the symptoms of any skin disease is still a major challenge for physicians. Many people suffer from serious skin conditions that require them to go to hospitals and undergo expensive medical examinations that take days. In this study, a method that solves skin diseases based on image processing and data science is used. Codes prepared on MATLAB help the above problem to a certain extent, and can also identify two different skin diseases, eczema, and skin cancer. In addition to previous studies, it can also provide an interpretation for normal skin. The study detects a range of symptoms in

a matter of seconds and acts as a switch. It also makes the diagnosis more intuitive and realistic. This study aims to classify different diseases and normal conditions based on images given as input. The study uses the MATLAB software program. Images, Dermnet and DermWeb etc. It is collected from various publicly available databases such as First, sample images of the two skin diseases and the normal condition need to be processed. Second, the diseased parts of the images are identified, and the corresponding mathematical operations are performed. Based on these, the characteristics of the two types of skin diseases and normal conditions are extracted, and the associated feature tissue parameters of these diseases and pixels of the lesion areas are collected through image segmentation. Finally, eczema and skin cancer symptoms are defined using the multiple support vector machine (MultiSVM) method to increase the identification accuracy and provide an interface to the physicians, and the normal state is not neglected either. It was created with 25 of 72 images in the Multi SVM image folder. However, the system gives consistent results when any new image is introduced that does not form the database. Clinical data of images were also collected from public databases such as Dermnet and DermWeb. Adding the decision tree system to the study helps the other party in the accuracy analysis part. Decision trees use the clinical data of the images to indicate that if the image is sick, it is sick or if it is normal, it is normal. The results of image processing and decision trees are compared with each other and in case of conflict, the system warns the system administrator.

1.1.Description of Diases

Eczema is in the form of red, inflamed, and itchy patches that appear during the heat increase in the skin. Skin cancer, which occurs in the form of malignant moles, can suddenly appear on normal skin without any warning, or it can develop on a pre-existing mole. Source of vitally serious skin cancer; cells called melanocytes. Melanocytes produce the pigment melanin, which gives our skin its color and enables us to tan. Skin cancer occurs because of the abnormal and uncontrolled proliferation of melanocytes.

1.1.1.Eczema

The best definition for eczema is an inflammatory skin reaction to various exogenous and/or endogenous stimuli [21].

The main manifestations and symptoms of eczema in infants are itching, rash and extensors, and in older children, skin deformities may occur. The occurrence of symptoms depends on the type of chronic recurrent eczema and whether there is another atopic disease in the family, and dryness is another manifestation of this disease. Facial pallor, various allergic shines, dermatographism, conjunctivitis, cataract, and skin prick test positivity can be seen in these people.

Clinical findings consist of a severe itchy rash that presents as red dots, scaling, and papulovesiculation. This is a generalization, but eczema progresses step by step, depending on many of the previously mentioned characteristics such as the patient's disease history, age, immunity, and skin type.

1.1.2.Skin Cancer

It is necessary to know the location and appearance of moles in the body well to perceive the changes in the following processes. It is possible to confuse malignant moles with other atopic moles. Sometimes skin cancer may be mistaken for a normal mole, but there is a possibility that that mole has been skin cancer since its inception [22].

Other warning signs; peeling, leakage, bleeding or swelling, or blistering on the mole. In addition, there may be changes in feeling related to me such as itching, tenderness, and pain. However, the relevant physician (usually a dermatologist) should advise on the necessity of removing an abnormal mole or following it closely.

It is recommended to consult a doctor if the following features are observed in a mole. These features are:

- If one half of the mole is not similar to the other half (in color and/or shape),
- If the borders of the mole are irregular (indented),
- If the color of the mole is not homogeneous (two or more of the colors such as brown, black, red, gray, white is combined or have a mottled appearance),
- If the mole is larger than 6 mm in diameter (roughly larger than the diameter of an eraser pencil).

2.Methods

2.1.Methodology

The method to be used for this study is the iterative model, which is a part of the software development lifecycle. The iterative model is a method of breaking down the software development lifecycle of large applications into smaller segments [23].

In addition, the iterative method starts with a small application of the software requirements branch and includes stages that develop until the entire system is complete. In addition, the iterative approach gives the system the ability to make some changes and add new practical capabilities. The main idea of this approach is to develop the system over iterative loops in smaller chunks over a period of time.

“The key to successful use of an iterative software development lifecycle is rigorous validation of requirements and validation and testing of each version of the software with those requirements in each cycle of the model.” [24].

As the main features of the iterative method; In the model, it can be shown that results are obtained periodically and early, can measure progress, develop in parallel, and do not have to change requirements or scope. On the negative side, the methodology has a few disadvantages such as having too many resources and their management. Figure 2 below shows an example of an iterative pattern diagram.

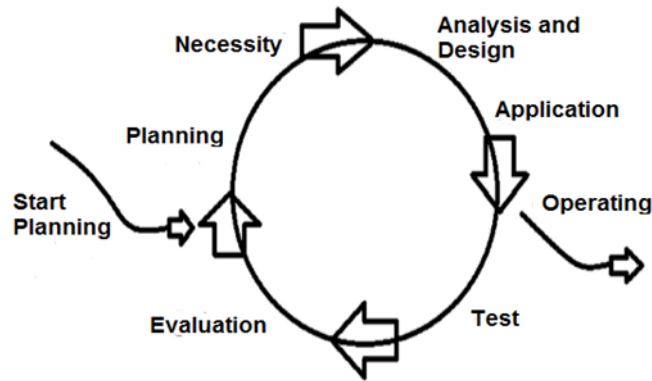


Figure 2. Iterative model diagram

The main stages of the iterative model in this study were determined as follows. These are:

- The stage of collecting and analyzing the requirements.
- Design phase: This phase analyzes the system requirements collected in the first phase and processes the system design by designing the block diagram, adjusting the MATLAB software to be used in this study to identify diseases.
- Implementation phase: In this phase, the work is implemented according to the current requirements and information, and while the system is being implemented, the work is divided into small programs called units to facilitate the implementation in MATLAB.
- Integration and testing phase.
- System deployment: After all functional and non-functional tests of the system are completed, the product is run with real data in the working environment.
- Care.

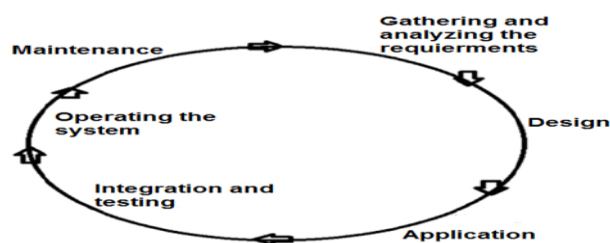


Figure 3. Iterative model stages of this study

Progress with iterative means that bug fixes are done continuously, so the result is more likely to be delivered on time and in higher-order as a result.

A function as an example: $f(x)=x^3-x-1$
 If we first write the equation $x=\phi(x)$
 $a < b$ and $f(a) \cdot f(b) < 0$ the points a and b found.
 If $f(a)$ is closer than 0'af(b)
 then $x_0=a$, otherwise $x_0=b$.
 $x_1=\phi(x_0)$
 $x_2=\phi(x_1)$
 $x_3=\phi(x_2)$
 ...
 is repeated until $|f(x_i)-f(x_{i-1})| \approx 0$

In the current algorithm, the newly added image is included in the function created by MultiSVM and the aim is to obtain results with minimum error with the iterative model. The system repeats itself every time a new image is uploaded to the system [25].

A comparative research method has been applied to this study, which helps the original image acquisition, feature extraction, and classification process based on MultiSVM. Previous studies discuss the use of colors and their effect in recognizing the surface of the human skin using the RGB YCrCb-HSV technique, distinguishing changes in the ratio of colors present on all four sides of the skin.

As the work progresses, there are clear differences in the feature extraction method part. Previous studies aimed to identify the skin area directly, which means that the primary goal was to know the location and surface of the skin. In recent studies, it is possible to follow the method of finding colors and extracting energy values from the image to detect diseases. The system in this study, which we have done, has the capacity to do this. What will be achieved in this work is to consider the extraction of multiple features, to improve the performance of the study, and to increase the capacity of the detection system to extract color features. Unlike other studies, the skin with eczema and skin cancer is also determined as normal skin. Unlike other studies, MultiSVM was used. In addition, unlike other works, MultiSVM was created with only lines of code without using MATLAB GUI. This system aims to increase the success of image classification every time the user uploads a new image.

2.2.Design and Analysis

Using the flowchart in Figure 4, the full functioning of the system can be seen.

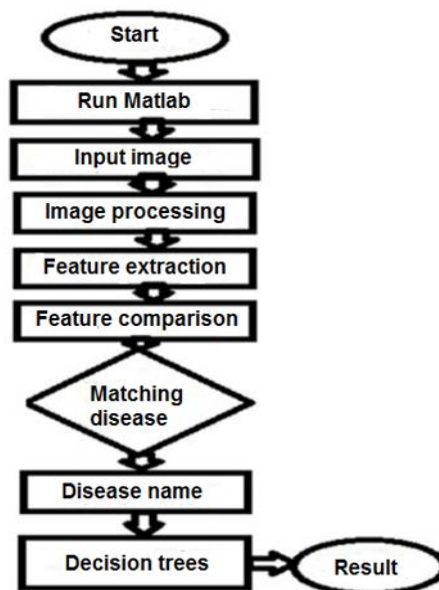


Figure 4. System flow chart

2.3. Operation and algorithm

The algorithm runs the system regularly with piecemeal progression logic. First, it is necessary to define RGB and grayscale images.

A digital color image is a type of image in which each pixel contains color information. Visually acceptable. For results, it is important to declare three color channels for each pixel, which in some color spaces are also interpreted as coordinates. In the computer environment, RGB color space is widely used. In a color image, each pixel has three values (or channels) that measure the chromaticity and intensity of the light. Luminance information in the spectral band is considered as real information stored in digital image data [26].

In grayscale, the value of each pixel is a single sample. A grayscale image is different from a single-bit two-tone black and white image. The grayscale digital image only gives density information. Also known as black and white, this image type consists of shades of gray ranging from black to white, with white at the strongest and black at the weakest. Grayscale images have many shades of gray, usually, grayscale images are produced by measuring the intensity of light in each pixel of a single band of the electromagnetic spectrum and exhibit monochromatic behavior only when a certain frequency is captured. Grayscale images can also be synthesized from a full-color image [26].

Converting RGB to grayscale; The main reason for this process is to process the brightness information of the image or to make an operation with the color information, for example, it can be examined whether an object is the same color in two different images or edge detection can be performed. To do this, it is necessary to separate the luminance information from the color information; this is mostly done using RGB to grayscale. Comparing colors in the grayscale color space is much easier than in RGB. For this reason, programmers prefer grayscale over RGB color space [27].

Feature extraction; means extracting features from images of processed lesion regions. The complexity of the classification problem in the processed images can be reduced by this method. The problem area can be distinguished from the background through certain features such as geometry and color. Feature extraction is very important to improve the accuracy of results in biomedical image processing, it is a necessary step to improve the quality of the processed image [28].

Image segmentation: It is defined as the process of dividing the image into several parts by converting the representative image into an easy form to make it more meaningful and easier to analyze [29].

The segmentation steps using K-means are as follows:

Step (1): Image reading.

Step (2): Converting the image from RGB color space to $L^*a^*b^*$ color space: The image used contains different colors. To distinguish these colors, it is necessary to convert the image from color space to $L^*a^*b^*$ space. All the color information of the space consisting of the L luminosity layer and $a^* b^*$ chromaticity layer are in a^* and b^* layers.

Step (3): Classification of colors in $a^* b^*$ space using the K-means clustering system. Clustering is a method of dividing the data into a set of clusters, this is done when the data is used without defined groups, the purpose of this algorithm is to find the groups in the data used, then the image will be divided into 3 clusters using the euclidean distance metric.

Step (4): An array is created to store the results of the clustering, the cluster containing the ROI to be converted to grayscale is selected, and then the GLCM is generated.

ROI: This name means "region of interest" containing a subset of samples selected from a dataset defined for a specific purpose [30]. The concept of ROI is widely used in many application areas. A region of interest (ROI) is a piece of the image on which filtering or other processing can be performed. More than one ROI can be identified in a single image [31].

The Gray Level Co-occurrence Matrix (GLCM) method is a method by which second-order statistical texture properties can be extracted. A GLCM is a matrix in which the number of gray levels in the image is equal to the number of rows and columns. A matrix element of two pixels with relative frequency $(a, b | \Delta c, \Delta d)$ is separated from these pixels by a pixel distance $(\Delta c, \Delta d)$, this matrix belonging to a certain neighborhood is formed at the density "a" and the other has density "b" [32].

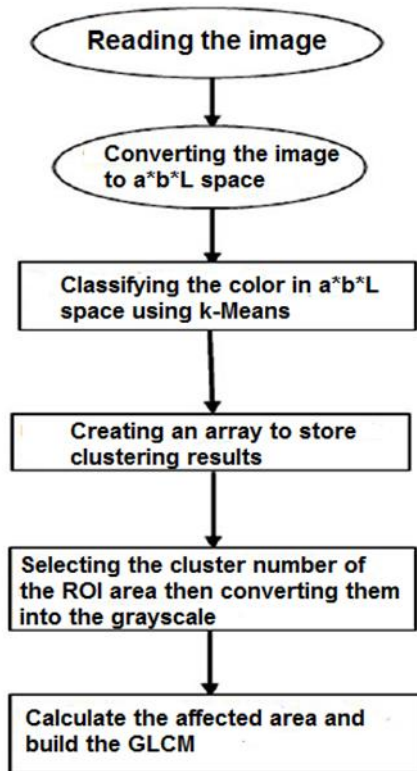


Figure 5. Block diagram of image segmentation

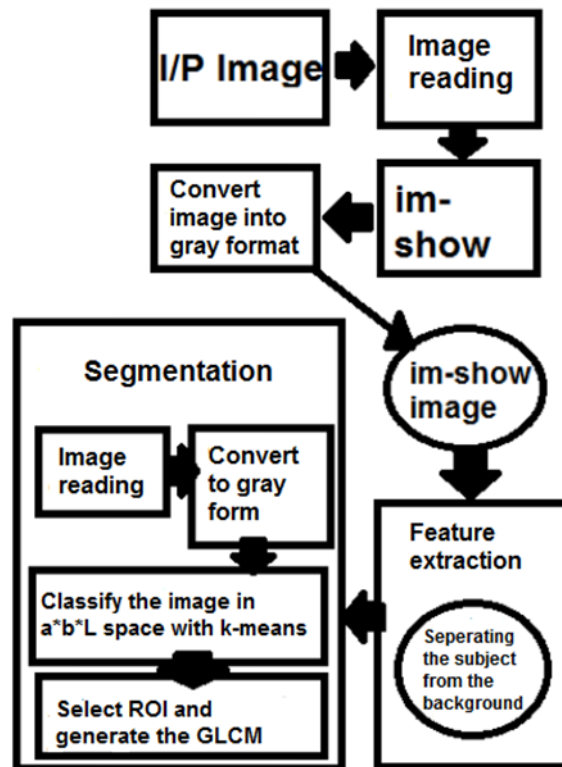


Figure 6. General block diagram of the proposed technique

2.4. MultiSVM and feature extraction

The mathematical background of the multiclass SVM (MCSVM) problem, typically with noise: $(x_1, y_1), \dots, (x_N, y_N)$, where $x_i: i = n$ length vectors from 1 to N and $y_i \in \{1, \dots$ represents the class of $M\}$ instances. The classical approach strategy for solving MCSVM classification problems is to think of the problem as a kind of combination of binary

classification problems. In the “One Support Vector Machine All” method, M classifiers are created, one support machine one for each class. The Mth classifier creates a hyperplane between the m class and the remaining M – 1 class. A new test sample is assigned to the class at which the distance from the positive boundary is maximum. By minimizing the optimization problem, it can be generalized to the following [33]:

$$\begin{aligned} \Phi(w, \varepsilon) &= \frac{1}{2} \sum_{m=1}^M (w_m^T \cdot w_m) + C \sum_{i=1}^N \sum_{m \neq y_i} \varepsilon_i^m, \\ \text{s.t. } (w_{yi}^T \cdot X_i) + b_{yi} &\geq (w_m^T \cdot X_i) + b_m + 2 \cdot \varepsilon_i^m, \\ \varepsilon_i^m &\geq 0, \text{ for } i = 1, \dots, N: \quad m \in \{1, \dots, M\} \setminus \{y_i\}. \end{aligned}$$

The decision boundary can be written as:
 $f(x) = \arg \max_m [(w_m^T \cdot x) + b_m]$ for $m = 1, \dots, M$;

The optimization problem will be solved by finding Lagrange's saddle point. The final classifier is as follows:

$$F(x) = \sum a_i x_i^T + b_m$$

There are only N coefficients in the above equation (regardless of the number of classes, M), and the editing directly tries to reduce the nonzero number of α . In this study, the MultiSVM Struct structure is obtained by processing the 2 disease states and normal states with the obtained features without extracting the features. The extracted features show different properties from each other for the disease states and the normal state found in the study. Therefore, this is very important for classification. It is necessary to examine these attributes and how they are extracted.

When examining the properties of the gray level formation matrix, we obtain the values of contrast, correlation, energy, and homogeneity. These values show different properties for each image [29].

Contrast: Returns a measure of the intensity contrast between a pixel and its neighbor over the entire image.

$$\sum_{i,j} |i - j|^2 \cdot p(i, j)$$

Correlation: Returns a measure of how related a pixel is to its neighbor over the entire image.

$$\sum_{i,j} \frac{(i - ui)(j - uj)p(i, j)}{\sigma_i \sigma_j}$$

Energy: Returns the sum of the squared elements in the gray level formation matrix;

$$\sum_{i,j} p(i, j)^2$$

Homogeneity: Returns a value that measures the closeness of the distribution of elements in GLCM to the GLCM diagonal.

$$\sum_{i,j} \frac{p(i, j)}{1 + |i - j|}$$

Apart from these four features, there are different features obtained by using the RGB matrix form of the picture. For instance:

Kurtosis: A measure of how outlier-prone a distribution is.

$$k = \frac{E(x-u)^4}{a^4}$$

Here μ is the mean of x, σ is the standard deviation of x, and E(t) represents the expected value of quantity t.

Skewness: A measure of the asymmetry of the data around the sample mean.

$$s = \frac{E(x-u)^3}{a^3}$$

Here μ is the mean of x, σ is the standard deviation of x, and E(t) represents the expected value of quantity t.

Variance: Returns the variance of the A elements along with the first array size whose size is not equal to 1. For a random variable vector, A consisting of N scalar observations, the variance is defined as:

$$v = \frac{1}{N-1} \sum_{i=1}^N |A_i - \mu|^2$$

where μ is the mean of A,

$$\mu = \frac{1}{N} \sum_{i=1}^N A_i$$

Some definitions of variance use a normalization factor N instead of N-1, which can be specified by setting w to 1. Smoothness: The sum of all elements of the RGB matrix is set to a value and the formula $1 - (1/(1+a))$ is used. For all kinds of diseases and normal conditions, the operation where the distribution of color values in the RGB image is different will produce different results.

RMS (Root-mean level): Root-square-mean level of a vector x;

$$X_{rms} = \sqrt{\frac{1}{N} \sum_{n=1}^N |x_n|^2}$$

with aggregation carried out along the specified size.

Entropy: A statistical measure of randomness that can be used to characterize the texture of the input image.

Entropy is defined as $-\sum(p \cdot \log_2(p))$; p contains the normalized histogram numbers returned from imhist.

Standard deviation: Returns the standard deviation of A elements along with the first array size whose size is not equal to 1.

$$s = \frac{1}{N-1} \sum_{i=1}^N |A_i - \mu|^2$$

where μ is the mean of A,

$$\mu = \frac{1}{N} \sum_{i=1}^N A_i$$

The standard deviation is the square root of the variance. Some definitions of standard deviation use the normalization factor N instead of N-1, which you can specify by setting w to 1.

Mean: For a random variable vector A consisting of N scalar observations, the mean is defined as:

$$\mu = \frac{1}{N} \sum_{i=1}^N A_i$$

IDM: Gets the measure of the distribution of GLCM elements to the GLCM diagonal.

2.5.Evaluation of the methods used

Image Processing and Decision Trees; The mean value of each feature in all images was calculated to examine the extracted features and to determine which disease they correspond to. In Table 1 below, the average value was found by moving through only 3 pictures of eczema. This shows the reader how MultiSVM is running in the background.

Table 1. Feature Inference Results and Averages of Three Skin Infected With Eczama

	A	B	C	D	E
1	Parameters				Average
2	Contrast	0,67704	0,43435	0,77974	0,63037
3	Correlation	0,87741	0,88214	0,9015	0,88701
4	Energy	0,57749	0,63881	0,2442	0,56794
5	Average	32	23	65	40
6	Standard Deviation	66	53	73	64
7	Entropy	3	2	5	3,33333

8	RMS	7	6	11	8
9	Homogeneity	0,95079	0,96476	0,90031	0,93
10	Variance	3,81E+18	2,20E+18	4,71E+18	3,58E+18
11	Smoothness	1	1	1	1
12	Kurtosis	2	2	0,60633	1,33333
13	Distortion	6	7	2	5
14	IDM	255	255	255	255

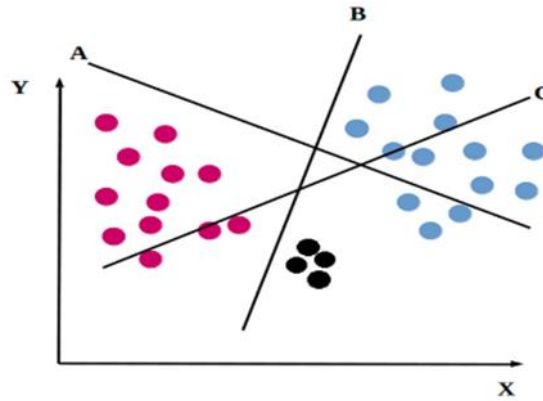


Figure 7. MultiSVM's determination of planes and boundaries [32]

Figure 7 represents the operating system of MultiSVM. In the real system, there are data taken from 25 images. Three planes have been created: Eczema, Skin Cancer, and Normal condition. Since the skin condition with the highest number of 25 images that make up the Multi SVM is skin cancer, the blue dots symbolize the skin cancer, the pink dots symbolize the normal state, and finally, the black dots symbolize the eczema disease. For each point, the health status of the skin is written in the relevant matrix. MultiSVM classification (fitcecoc) is performed.

The parameters appearing in TABLE 1 are extracted again for each added image, whether or not it is out of 25 photographs in the database, and its location in the MultiSVM plane is determined by the 'predict ()' estimation method. As a result, when a new image is given to the system, the system will determine how close it is to these three regions and accordingly classify whether it is one of the diseases or normal.

There are 51 stored images for skin cancer, 11 for eczema, and 10 for normal skin. It has the advantage of having many images. The values obtained from the images of these two diseases and normal skin in the database were compared with the data in the 'Diseasesetyeni. mat' file.

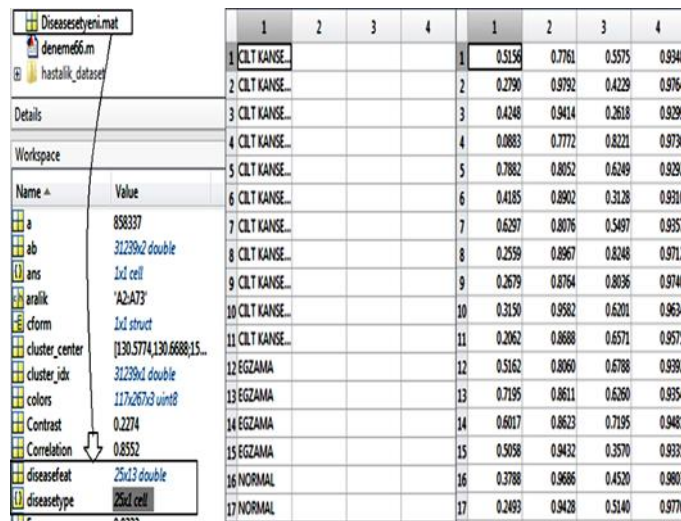


Figure 8. The matrices and their contents in Diseasesetyeni.mat

Footnote: Skin cancer= Cilt kanseri, Eczama= Egzama

'Diseasesetyeni. mat' and its matrix and their values were created with 25 of the 72 images already in the system. However, the system gives consistent results when any new image that does not create the database is introduced. As mentioned before, the predict () function is used during the simulation.

Studies are carried out to find the feature with similarities, to know the obtained similarities, and to reveal the percentages. There is a limited amount of data on 'diseasesetyeni. mat'. However, this data set was sufficient to reveal the purpose and results of the current study. In the future, the study can be made more detailed by increasing the data set. The data of all images are not saved in 'Diseasesetyeni. mat'. In the study, it has been tried to prove that it gives correct results for the newly added images. For each image entered, the values are always extracted again and compared with 'Diseasesetyeni. mat'.

The values of skin cancer and eczema disease and normal skin are found in the matrices in the 'Diseasesetyeni. mat' file. The database contains all the necessary information that helps the user in the search process. The process consists of checking the values from 1 to 25 for each image and using the MultiSVM planes created from them to classify the image with the appropriate group and achieve the required similarity ratio.

The situations processed in the background of the code are:

Image is taken from the outside and displayed inside the system. The features of the image are extracted and saved in the workspace. The 'disease feat' and 'disease type' matrices in the 'Diseasesetyeni. mat' file is subjected to the MultiSVM training function and the 'svmStructDisease' structure is created. It is shown in Figure 8. Finally, the predict the disease function is performed, a match is made between the new image and the database, and the results of the predictions of the diseases or the normal state are given.

As a result, the system will notify the other party of the disease type or normal state. If the result of the image processing differs from the result of the decision trees, this will also be reported as an upstream message.

Table 2. Decision Tree Method

KAY	T3	TST	TSTT	TSH	MADTSH	Görüntü İşleme	Karar ağac
1	113	9,9	3,1	2	5,90	cilt kanseri	eczama
2	105	7,3	1,5	1,5	0,10	cilt kanseri	cilt kanseri
3	114	9,9	2,4	1,5	5,70	cilt kanseri	cilt kanseri
4	106	4,2	1,2	1,6	1,40	cilt kanseri	normal
5	112	8,1	1,9	3,7	2,00	cilt kanseri	cilt kanseri
6	111	8,4	1,5	0,8	1,20	cilt kanseri	cilt kanseri
7	115	7,1	1,3	1,3	2,00	cilt kanseri	cilt kanseri
8	109	10,4	1,9	0,4	0,10	cilt kanseri	cilt kanseri
9	100	6,1	2,4	1,8	3,80	cilt kanseri	cilt kanseri
10	105	11,1	2	1	1,00	cilt kanseri	cilt kanseri
11	116	11,1	2	1,2	2,30	cilt kanseri	cilt kanseri
12	117	12,2	1,9	1,2	3,90	cilt kanseri	cilt kanseri
13	106	9,4	1,5	0,8	0,50	cilt kanseri	cilt kanseri
14	121	10,1	2,4	0,8	3,00	cilt kanseri	cilt kanseri
15	121	13,5	1,5	1,6	0,50	cilt kanseri	cilt kanseri
16	109	10	1,3	1,8	4,30	cilt kanseri	cilt kanseri
17	118	8,1	1,9	1,5	13,70	cilt kanseri	normal
18	115	8,1	1,7	0,6	2,20	cilt kanseri	cilt kanseri

Footnote: Skin cancer= Cilt kanseri

Eczama= Eczama

Image processing= Görüntü işleme

Decision tree= Karar ağacı

Record= Kayıt

Decision tree	Image processing			
	Normal	Cilt kanseri + Egzama	Sum	
Normal	DT1	YT1	DT1+YT1	
Cilt kanseri + Egzama	YT2	DT2	YT2+DT2	
Sum	DT1+YT2	YT1+DT2	DT1+YT1+YT2+DT2	
Accuracy=	(DT1+DT2)/	(DT1+YT1+YT2+DT2)		
Sensitivity=	(DT1)/	(DT1+YT2)		
Selectivity=	(DT2)/	(YT1+DT2)		
	ROC1(Cilt K.)	ROC2(Egzama)	ROC3(Normal)	Sum
ACCURACY:	0,916666667	0,958333333	0,958333333	0,94
SENSIVITY:	0,94	0,833333333	0,9	0,89
SELECTIVITY:	0,863636364	0,983333333	0,967741935	0,94
DT1:	47	10	9	
DT2:	19	59	60	
YT1:	3	1	2	
YT2:	3	2	1	
	72	72	72	

Figure 9. Calculation of Sensitivity, Accuracy and Selectivity

Footnote: Skin cancer= Cilt kanseri, Eczama= Egzama

Table 2 and Figure 9 show the use of the decision tree method. DT1 is the number of images in which image processing for normal skin images yields parallel and accurate results against decision trees. DT2 is the number of images where image processing for skin cancer and eczema skin images yields parallel and accurate results across decision trees. YT1 is the custom of images where image processing for skin cancer and eczema skin images yields inconsistent and incorrect results against decision trees. YT2 is the custom of images where image processing for normal skin images gives opposite and wrong results against decision trees. Relevant clinical data used for decision trees Sakarya Uni. Taken from the lecture notes of the Data Mining course. In addition, the code lines in the system are the version of the Plant Disease Classification study in MathWorks File Exchange, developed with the decision tree method. In the excel data in Table 2, only some of the clinical data of 72 images in the picture folder are shown as representative. The health status of the skins was started to be classified by placing the limit of 5.2000 on the TST data found in clinical data. Decision trees progress step by step. Obtaining entropy information at each stage shows the user how well the separation progresses, and the closeness of entropy to 0 shows the opposite side that the process steps are approaching the end. Of the values included in the classification, T3 is the result of the blood test used when measuring the level of thyroid hormone, TST is the thyroid gland serum hormone test, TSH is the thyroid-stimulating hormone test, TSTT is the tubing scopic threshold test, MADTSH is the absolute difference from the thyroid-stimulating hormone test based on.

Entropy: It is mostly defined as randomness and disorder in a system and is used in many fields from statistics to technology. R is seen as a resource. It is assumed that this resource can generate n messages as {m1, m2, m3,..., mn}. All messages are generated independently of each other and the probability of generating mj messages is pj.

The entropy H(R) of the R source that produces messages with probability distribution P= {p1, p2, p3,..., pn} is as follows.

$$H(R) = -\sum_{i=1}^n p_i * \log_2(p_i) \text{ or } H(R) = -\sum_{i=1}^n p_i * \log\left(\frac{1}{p_i}\right)$$

Finally, when the classification with decision trees is completed, the sensitivity, accuracy, and selectivity analysis of the system are performed according to the image processing by using the above formulas. It describes the results of the comparison of simulative data and numerical data.

3.Results

The operation of any system depends on the probability of occurrence of a particular event, this study includes a wide range of possibilities, it is necessary to compare images of a particular disease and define the category to which it belongs. Reference is made to studies that can compare a database of pathological images with images from outside the database and work with several models. While working on the identification and classification of diseases, it was observed that insufficient information was collected for the database, and the success of the process was low when only the image processing technique was used in the process. The database is very important and dermatology medicine needs to focus on improving the content of the database so that it is easier to identify any disease with images stored in its system. Moreover, adding the decision tree system to the study helps the user in the accuracy analysis part.

This problem was solved by extensive research on various websites associated with the idea of the study, and a sufficient number of images were found in terms of the number and quality of images. A series of images have been collected to be used directly in the database and by comparing them according to the inputs to be presented, the desired results can be obtained. In addition, all data and images were collected from public databases.

Skin diseases seriously affect people's lives and health. This study proposes an effective approach to identifying skin diseases. To increase the diagnostic accuracy of many types of dermatological diseases, it is necessary to replace traditional methods and develop automated methods. In this study, two types of skin diseases were defined, namely Eczema and Skin Cancer diseases, and it can also determine normal skin. The decision tree method adds strength to this study and distinguishes it from other studies. The program will detect and identify dermatological diseases. It was designed through the MATLAB program. This study will serve a large segment of society and will solve the problems in identifying and distinguishing the types of skin diseases in humans due to the similar effects and symptoms that humans are exposed to and will certainly contribute to the early diagnosis and treatment of skin diseases.

The detection of skin diseases has gone through several testing stages to make more accurate image processing. Simulations were made on two different diseases and normal conditions using the MATLAB program. This study is very useful in increasing the understanding of programming by using MATLAB software. After relying on the similarity rate for eczema and skin cancer diseases, the system was tested with new images. As expected, the result of the system will be an assistant for all users, including doctors and nurses. It is used with MATLAB software technology. The system can be used by any person, be it a professional, a programmer, or a regular person.

In the first part, the aim is to match the images to one of the selected diseases and to reach the basic principles for estimating the similarity ratio when an image is uploaded to the system, and a training model that can always be improved has been successfully obtained. The number of items in the database has been increased and since this study will help many people in the hospital, the system has been designed with an easy-to-use interface to make it much easier to use, and this section has been completed very successfully. Apart from the professionally working GUI interface of MATLAB, uniquely written codes were used, and this helped a lot in recognizing the backside of the process. Energy feature extraction and extraction of other features were used on images of different diseases taken from the user and not included in the database system, but which will be compared with this particular database.

The use of skin diseases detection methods will positively affect the field of medicine in terms of early detection of the disease. Since the patients will not have to stay in the hospital for a long time to perform the examination procedures, the number of patients will be improved. It also saves time for nurses and doctors, making the process go faster, and more time is allocated for other diseases that require more than one procedure to complete the examination. In all its aspects, the importance of developing a skin disease detection system is how to extract features from the images, including the acquisition phase of the images, and increase the accuracy of the study, which will make a significant difference to community users. It certainly has benefits and reduces diseases and it is recommended that government or private organizations working in the medical field support and invest in this method for the following reasons. These reasons:

1. Increasing number of skin diseases
2. Ability to use multiple features for higher resolution such as contrast,
3. Ability to extract entropy, correlation, and homogeneity features,
4. Purpose of using a camera that takes pictures of the patient directly in the future
5. Application of different images of a different disease,
6. It is listed as the purpose of having the function of determining the amount and type of drug-assigned to the patient in the future.

Acknowledgement

In this study, gratefully thanking to other studies with similar ideas that supported and cited as references for this article. And would also like to thank to the JIET community for supporting the publication of this work.

Author Contribution

In the study carried out, Yusuf Süer Erdem contributed to the creation of the model, literature review, data collection and processing, while Özhan Özkan contributed to the formation of the idea, the evaluation of the results and the spelling of the article.

Conflict of Interest

The authors confirm that there is no known conflict of interest or common interest with any institution/organization or person. Ethics committee approval is not required for the prepared article. There is no conflict of interest with any person/institution in the prepared article.

References

- [1] Kaplan, M. Implementation of an Auxiliary System for the Diagnosis of Skin Diseases Using Artificial Neural Networks, TC Firat University, Institute of Science, Master Thesis, 2016.
- [2] Season, V. (2010). Epidemiology of dermatological diseases, disease burden, and place in primary care. *Turkey Clinics J. Fam Md-Special Topics*, 2010; 1(2): 15-20.
- [3] Alonso, DH., Wernick, MN., Yang, Y., Germano, G., Berman, DS., Slmoka, P. Prediction of cardiac death after adenosine myocardial perfusion SPECT based on machine learning. *J Nucl Cardiol*. <https://doi.org/10.1007/s12350-017-0924-x>, 2018.
- [4] Narula, S., Shameer, K., Salem Omar, AM., Dudley, JT., Sengupta, PP. Reply Deep learning with unsupervised features in echocardiographic imaging. *J Am Coll Cardiol*; 2017, 69: 2101–2.
- [5] Esteva, A., Kupre, B., Novoa, RA., et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017, 542:115–8.
- [6] Cichosz, SL., Johansen, MD., Hejlesen, O. Toward big data analytics: a review of predictive models in the management of es and its complications. *J es Sci Technol*, 2015, 10(1):27-34.
- [7] Tran, BX., Latkin, CA., Giang, VT., et al., The Current Research Landscape of the Application of Artificial Intelligence in Managing Cerebrovascular and Heart Diseases: A Bibliometric and Content Analysis. *Int. J. Environ. Res. Public Health*, 2019, 16:2699.
- [8] Char, DS., Shah, NH., Magnus, D. Implementing Machine Learning in Health Care, Addressing Ethical Challenges. *N. Engl. J. Med.*, 2018, 378: 981–983.
- [9] Celebi, V., Inal, A. Problem of Ethics in the Context of Artificial Intelligence. *The Journal of International Social Research*, 2019, 12, 66.
- [10] Mujumdar, A., Vaidehi, V. Diabetes Prediction Using Machine Learning Algorithms. *Procedia Computer Science*, 2019, 165: 292–299.
- [11] Farid, D., Sadeghi, H., Hajjgol, E. ve Parirooy, N. Classification of Bank Customers by Data Mining: a Case Study of Mellat Bank branches in Shiraz, *International Journal of Management Accounting and Economics*, 2016, 3: 534-543.
- [12] Walsh, S. Applying Data Mining Techniques Using SAS® Enterprise Miner- Course Notes, SAS Institute Inc., North Carolina, 2005.
- [13] Pratt, W. K., Digital Image Processing. USA: John Wiley & Sons, 2007.
- [14] Nixon, M. S., Aguado, A. S., Feature Extraction, and Image Processing. Newnes, UK, 2002.
- [15] Kaur, H., Kumari, V. Predictive modeling, and analytics for diabetes using a machine learning approach. *Applied Computing and Informatics* <https://doi.org/10.1016/j.aci.2018.12.004>, 2018.
- [16] Kavakiotis, I. et al. Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology*, 2017, 15: 104–116.

- [17] Araújo F.H.D. et al. Using machine learning to support healthcare professionals in making preauthorization decisions. *International Journal of Medical Informatics*, 2016, 94:1–7.
- [18] Parikh, R.B., Kakad, M., Bates, DW. Integrating predictive analytics into high-value care: the dawn of precision delivery. *JAMA*, 2016, 315: 651-652.
- [19] Bates, DW., Saria, S., Ohno-Machado, L., Shah, A., Escobar, G. Big data in healthcare: using analytics to identify and manage high-risk and high-cost patients. *Health Aff*, 2014, 33: 1123-1131.
- [20] Mercaldo, F., Nardone, V., Santone, A. Diabetes Mellitus Affected Patients Classification and Diagnosis through Machine Learning Techniques. *Procedia Computer Science*, 2017, 112: 2519-228.
- [21] Al-Khafaji, Muallah, S. K., Ibraheem, M. R. Detection of Eczema DISEASE by using Image Processing. *The Eurasia Proceedings of Science, Engineering & Mathematics (EPSTEM)*, 2018, Volume 2: 2602–3199.
- [22] Acıbadem Web and Medical Editorial Board. Skin (Skin) Cancer. Acıbadem Healthcare Group. 2019, from <https://www.acibadem.com.tr/ilgi-alani/cilt-deri-kanseri/#signs>
- [23] Al Shahibi, I. S. S., Koottala, S., Detection of Skin Diseases Using Matlab. Journal of Student Research Fourth Middle East College Student Research Conference, Muscat, Sultanate of Oman, 2020.
- [24] Mathworks Help Center, Retrieved December 6, 2021, from <https://www.mathworks.com/help/stats/fitcecoc.html>
- [25] <https://atozmath.com/example/CONM/Bisection.aspx?he=e&q=it>
- [26] Houcque D., Introduction To MATLAB For Engineering Students, <https://www.mccormick.northwestern.edu/documents/students/undergraduate/introduction-to-matlab.pdf>, 2005.
- [27] Kumar, B., Rai, S.P., Saravana Kumar, U., Verma, S.K., Garg, Pankaj K., Vijaya Kumar, S.V., Jaiswal, R., Purendera, B.K., Kumar, S.R. and Pande, N.G. Isotopic characteristics of Indian Precipitation. Published online in *Water Resources Research*, 2010, Vol. 46, DOI: 10.1029/2009WRSR008532, 2010.
- [28] Yurtay, N., Adak, M. F., Dural, D., Serttaş. S. A study on use of decision tree method in the diagnosis of thyroid disease. International Science and Technology Conference, Retrieved 28 November 2021. Published, 2012.
- [29] "Dermatitis" defined, Suzanne Smith, Susan Nedorost, 2012.
- [30] MathWorks Help Center, Retrieved December 6, 2021, from <https://www.mathworks.com/help/stats/fitcecoc.html>
- [31] Image Texture Feature Extraction Using GLCM Approach, P. Mohanaiah, P. Sathyanarayana, L. GuruKumar, 2013.
- [32] Data Scientist Website, Retrieved December 12, 2021, from <https://veribilimcisi.com/2017/07/19/destek-vektor-makineleri-support-vector-machine/>
- [33] <https://www.researchgate.net/publication/221608588>
- [34] Yurtay, N. Data Mining Applications Lecture Notes, Week 4, Sakarya University, Retrieved 06 December 2021.