Turkish Journal of Science & Technology
19(2), 457-473, 2024

Research Paper
https://doi.org/10.55525/tjst.1424371

The Design of Machine Learning-Based Computer-Aided System with LabVIEW For Abnormalities in Mammogram Images

Iman Hamadamın[1], Hasan Güler[2*]
[1,2] Electrical-Electronics Engineering Department, Firat university, Elazig/Turkey
[1]imanmuhammad1995@gmail.com, [2]*hasanguler@firat.edu.tr

(Geliş/Received: 23/01/2024; Kabul/Accepted: 10/05/2024)

Abstract: Mammogram is the best way of breast cancer detection nowadays, as breast cancer is the most common form of cancer in the female gender and this form of cancer usually causes death. Many scientists, doctors, and engineers are working together to deal with such serious issues in human life. This paper, it is aimed to develop a new computer-aided system with a graphical coded language to detect abnormalities in mammogram images by using machine learning technics such as ANN and SVM. The developed algorithm has a graphical user interface (GUI) and all results are shown in there. The algorithm was created using three different stages. These are image processing and mass segmentation, feature selection and extraction, and classification. To test the accuracy of the system as the sensitivity, specificity, and accuracy, mammogram images with forty benign and forty malignant masses were used. The obtained results for measuring the sensitivity, specificity, and accuracy are 95%, 97.5%, and 96.25% for ANN and 97.5%, 97.5%, and 97.5% for SVM, respectively. As can be said that the algorithm, user-friendly due to its user interface, can be preferred because it can detect many cancerous cells such as breast cancer with high accuracy.

Key words: Mammography, breast cancer detection, machine learning, graphical user interface.

Mamogram Görüntülerindeki Anormallikler İçin LabVIEW ile Makine Öğrenmesi Tabanlı Bilgisayar Destekli Sistem Tasarımı

Öz: Meme kanseri kadınlarda en sık görülen kanser türü olduğundan ve bu kanser türü genellikle ölüme neden olduğundan, günümüzde meme kanserini tespit etmenin en iyi yolu mamografidir. Birçok bilim insanı, doktor ve mühendis insan hayatındaki bu tür ciddi sorunlarla başa çıkmak için birlikte çalışmaktadır. Bu makalede, YSA ve DVM gibi makine öğrenmesi teknikleri kullanılarak mamogram görüntülerindeki anormallikleri tespit etmek için grafik kodlu bir dile sahip yeni bir bilgisayar destekli sistem geliştirilmesi amaçlanmıştır. Geliştirilen algoritma grafiksel bir kullanıcı arayüzüne (GUI) sahiptir ve tüm sonuçlar burada gösterilmektedir. Algoritma üç farklı aşama kullanılarak oluşturulmuştur. Bunlar görüntü işleme ve kütle segmentasyonu, özellik seçimi ve çıkarımı ve sınıflandırmadır. Sistemin doğruluğunu duyarlılık, özgüllük ve doğruluk olarak test etmek için kırk iyi huylu ve kırk kötü huylu kitle içeren mamogram görüntüleri kullanılmıştır. Duyarlılık, özgüllük ve doğruluk ölçümleri için elde edilen sonuçlar sırasıyla YSA için %95, %97,5 ve %96,25; DVM için %97,5, %97,5 ve %97,5'tir. Kullanıcı arayüzü sayesinde kullanıcı dostu olan algoritmanın, meme kanseri gibi birçok kanserli hücreyi yüksek doğrulukla tespit edebilmesi nedeniyle tercih edilebileceği söylenebilir.

Anahtar kelimeler: Mamografi, meme kanseri tespiti, makine öğrenmesi, grafiksel kullanıcı arayüzü.

1. Introduction

Recently, there is an increase in the rate of affected women with breast cancer. This type of cancer alone accounts for about 22% of female cancers and approximately 15% of mortality among women having cancer [1]. As a starting point toward a better understanding of breast cancer, it is important to know how cancer in general develops. Cancers occur when control of the division of normal cells is lost and they start to invade other healthy tissues which takes place when a single cell or a group of cells escapes from the usual control that regulates cellular growth when they start to multiply, spread and form a mass. When the mass is formed, it can be considered benign or malignant depending on its shape and behavior. When abnormal growth is restricted to a single and circumscribed mass of cells, it is known as benign. The term “cancer” is used to describe malignant masses which not only can invade surrounding tissues but also can spread or “metastasize” to distant areas of the body. When the breast masses reach a palpable size, this means that they are metastasized [2]. Properties such as margins and shapes help to define masses. For instance, masses with round and smooth margins indicate that they are benign while malignant masses have speculated, rough or blurry boundaries [3].

* Corresponding author: hasanguler@firat.edu.tr. ORCID Number of authors: [1] 0009-0001-2437-7262, [2] 0000-0002-9917-3619

The stages of breast cancer can be classified into three stages depending on the danger and the distance between cancer cells and the original tumor; local breast cancer, regional breast cancer, and distant breast cancer. In the first stage, breast cancer is still local. The cancer is still located inside the breast in lobules and ducts and it is not invading the neighborhood tissues. In other words, the normal tissues beyond the breast are not affected by this type of breast cancer [4]. The second stage of breast cancer is regional breast cancer. This stage occurs when the cancer cells start the invasion of neighbor tissues and try to reach the underarm lymph nodes. The lymph nodes are small organs that filter the body from foreign substances. The lymphatic system consists of lymph nodes and ducts that form a network. Its main work is to fight against the foreign substances in the body and filter them [5]. In the third stage, which is termed distant breast cancer, cancer cells are invasive and get into the lymph nodes. They also have a pathway into other parts of the body such as lungs, distant lymph nodes, skin, bones, liver, and brain [6].

The benign masses are considered to be in the first stage of breast cancer stages because they cannot metastasize and lack the invasive properties of cancer. Although benign masses can have side effects many kinds of this type of mass are not harmful to human health conditions and they are not life-threatening. The malignant masses are considered to be the second and third stages of breast cancer stages because they are not self-limited in growth. They have the capability of invading the neighboring tissues around the breast and spreading to distant regions of the body. Therefore, the term "cancer" is used for malignant masses which are usually more serious and more dangerous for human health condition.

The paper is organized into the following sections: The related work is determined in section 2, section 3 material and methods introduce the methods used in developed algorithms, in section 4, the results are given and the last section concludes the whole paper.

2. Related Works

Many scientists studying computer science, especially artificial intelligence, are working to detect cancerous cells in the early stages [7-19]. It is observed that mortality tends to decrease in line with the early detection studies that have increased in recent years [7]. Studies in which artificial intelligence methods such as machine learning are widely used determine with high accuracy in the diagnosis and prediction of various diseases such as breast cancer. The comparison of different approaches developed by many scientists in cancer detection can be seen in Table 1.

Table 1. Comparison of different approaches in cancer detection and classification.

Reference	Database	Segmentation	Classification	Results
[8]	DDSM	Texture based	Clustering	93%
[9]	MIAS	Watershed	SVM	98%
[10]	MIAS	Gabor Filter	k-means	99%
[11]	-	Entropy, mean, energy	ANN	90%
[12]	-	Statistical parameters	Triangulation	99.16%
[13]	WDBC	PCA	SVM, KNN	SVM-51.10% KNN-91.11%
[14]	WDBC	texture	SVM, NB	97.13%
[15]	WDBC	Accuracy, sensitivity	SVM	99.51%
[16]	DDSM cases	Texture descriptors	SVM	75%
[17]	WDBC	Sensitivity and specificity	MLP, NN, SVM	99.04%
[18]	MIAS	DCT and DWT	SVM	96.97%
[19]	WDBC	Wrapper method	SVM, KNN	SVM-97.18% KNN-95.12%

The authors of [8] proposed a machine-learning algorithm for extraction and clustering. While doing this, texture analysis was done for feature extractions. Another study [9] used SVM (Support Vector Machine) classifier for breast cancer detection. They reported 98% accuracy in their study. In [10], the authors presented a novel approach to breast cancer. They used threshold parameters to differentiate pixels of cancer regions. An automated technique using ANN (Artificial Neural Network) was proposed in [11]. Inputs of ANN were selected as entropy, mean, energy correlation, texture, and standard deviation, while ANN detects whether an image is cancerous or not.

The other study [12] used different algorithms in image segmentation, triangulation, binarizations, thinning, and Euclidean distance transformation in the detection of the cancer cell. Habib Dhahri et al [13] built an automated machine-learning workflow to optimize the list of data transformations. They proposed a genetic algorithm to optimize the data and control parameters.

A comparison for performance among various machine learning algorithms such as support vector machine (SVM), decision tree, k-nearest neighbors (k-NN), and naïve bayes was given in [14-19]. The authors declared that SVM gave the highest performance when considered with accuracy.

There are some classes of abnormalities found after the mammography images are taken. Expert radiologists can determine the type of abnormality just from its appearance and then they can determine whether the mass in the mammography image is benign or malignant mass. Breast cancer screening mammograms found these types of abnormalities: calcification (i.e., macro and micro-calcifications), spiculated masses, well-defined/circumscribed masses, architectural distortion, asymmetry breast tissues, and other miscellaneous findings [3]. In mammography images, the appearance of micro-calcifications looks like large white dots distributed randomly within the breast. It is found that half of the women over 50 and 10 women under that age have macro-calcifications in their breasts. Macro-calcifications are considered noncancerous and for that reason no need to do follow-up care [6]. Although Micro-calcifications are usually not an indicator or result of breast cancer, they can become dangerous when they are clustered in a group and appear in a certain pattern, when they are grouped, they are considered a starting indicator of breast cancer.

If calcifications are detected in the breast, doctors categorize them into three types to be treated. The first category is benign calcifications which are considered harmless and no need to do treatment for them. Another is probably benign calcifications. It is found that more than 98% of this type is noncancerous. Typically, they are monitored every six months for at least one year. If there is no change found after a year of follow-up, the doctor's recommendation is to have a routine mammogram once a year [6]. A spiculated mass is considered the most dangerous class of abnormality since it is one of the primary indicators of cancer [20]. This type of mass can be anywhere inside the human body but is often found in breasts or lungs. When these spiky masses are found anywhere inside the body even in the breasts doctor's recommendation is to give a biopsy to confirm whether they are malignant or benign. If they are malignant, the treatment can range from excision to radiation.

Well-defined/circumscribed masses are another class of abnormalities found in breast cancer screening and mammography. Another term that can be used for this type of mass is circumscribed carcinoma. This term refers to ductal carcinoma that appears as circumscribed on a mammogram. Although circumscribed carcinoma is less frequently seen than typical spiculated carcinoma, it has both types of severity of abnormality benign and malignant. Circumscribed carcinoma includes medullary, invasive ductal carcinoma, and other types [21]. Architectural distortion is the last class of abnormalities discussed in this chapter. It is considered the third most common class of abnormalities according to its appearance. It is found that 6% of abnormalities have this type. The incidence of architectural distortion is small compared to calcifications and visible mass. However, when it exists in the mammography image, it is difficult to be detected and diagnosed because of its variability in presentation [22]. A scar inside the breast formed from a previous surgery that is benign can be interpreted as architectural distortion. Although the reason for architectural distortion can be a result of benign disease, it is found that almost 80% of the detected masses are a result of invasive breast cancer [22]. The appearance of architectural distortion in mammography images seems like a disruption in the structure of the breast itself. The most interesting thing in this class of abnormality is that there is no mass to indicate and name it abnormal mass but the distortion appears as a stellate shape or with radiating speculation like the masses found in speculation cases.

Nowadays, many effective methods and devices are developed for detecting breast cancer. These methods are X-ray mammography, ultrasonography, trans-illumination, thermography, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) all of which are used for breast cancer diagnosis. It has been found that X-ray mammography is the best and the most effective method used for detecting breast cancer [23]. A mass is a lesion that occupies a space in the breast and it can be seen on at least two projections or viewpoints (Carnio-caudal CC and mediolateral-oblique MLO). The view of a mammogram of Carnio-caudal CC is taken from above while the view of a mammogram of mediolateral-oblique MLO is taken as an oblique or angled view [24].

3. Material and Methods

Materials are the database and the programming language used for designing the system, Mammographic Image Analysis Society (MIAS) database images are chosen to be used in the system, 62.5% of the data set (133 images) was selected for training, 37.5% of that (80 images) is chosen and used to test in the system [25]. The developed algorithm was created using the LabVIEW platform, which is a graphic code-based software. The reason why this platform is preferred is the ease of creating a user interface and the effort to create a new algorithm by creating completely mathematical expressions by ourselves instead of using the ready-made toolbox structure of machine learning algorithms, unlike the studies done so far.

In the image processing stage, image enhancement methods are implemented to process the image to make the result more suitable than the original image as shown in Fig.1. It brings out some features in the region of interest that are invisible or difficult to notice in the original image. Image enhancement techniques include histograms, image filtering, thresholding, morphological operations, removing undesired parts, and region segmentation.

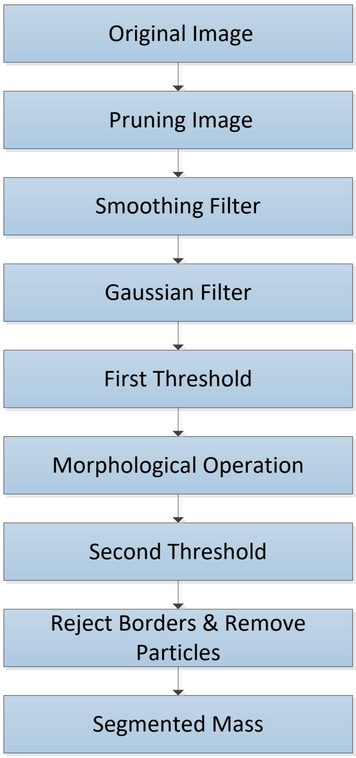

Figure 1. The steps of the Image Processing and Mass Segmentation Stage.

The histogram of a digital image in general is a discrete function that represents the number of pixels in each different gray level which is called the intensity level. It can be written as follows in Equation 1:

$$h(r_k) = n_k \tag{1}$$

Where r_k represents the k^{th} gray level and n_k is the number of pixels in an image having an intensity level r_k in equation 1. The range of intensity level for the 8-bit grayscale image will be [0, L-1]; i.e., the range will be [0 - 255] since$L = 2^k$. The histogram can be normalized and its range becomes [0 - 1] according to:

$$P(r_k) = n_k/n \tag{2}$$

Where n is the total number of pixels in the image in equation 2.

Kernel family filters contain four types of matrixes that can be used as different filters on the images which each type has some different sizes of the matrix such as (3*3, 5*5, and 7*7), the types are; Gradient, Laplacian, Smoothing, and Gaussian.

Thresholding is a simple method used for segmenting the breast in the mammography image. It is called global since it is based on the global information of the image like a histogram and a single threshold value is selected for the whole image. The global threshold value can be found easily because the intensity values of the abnormality regions are greater than the surrounding tissue [26]. It can be expressed as follows in Equation 3:

$$g(x,y) = \begin{cases} 1 & \text{if } f(x,y) \geq T \quad (\text{ROI} - \text{Breast}) \\ 0 & \text{Otherwise} \quad (\text{Background}) \end{cases} \tag{3}$$

Morphological operations are very useful in image processing for extracting, describing, and improving the shapes of regions of interest. They can be used in pre-processing and post-processing operations. Most of the time, morphological operations are used in binary images since they rely on the relative ordering of pixels and not on their numerical values [27]. Dilation and erosion are the two major morphological operations used in this work and most of the research done so far [28]. Opening and closing are two morphological operations resulting from the two basic morphological operations of dilation and erosion. They are a combination of dilation and erosion. The opening is an erosion operation followed by dilation while the closing is a dilation operation followed by erosion [29]. The expression of Dilation, Erosion, Opening and Closing respectively are as the following in Equation 4a-4d:

$$A \oplus B = \left\{ z \mid \left(\hat{B}\right)_z \cap A \neq \emptyset \right\} \tag{4a}$$

$$A \ominus B = \{ z \mid (B)_z \subseteq A \} \tag{4b}$$

$$A \circ B = (A \ominus B) \oplus B \tag{4c}$$

$$A \bullet B = (A \oplus B) \ominus B \tag{4d}$$

In LabVIEW there are two block diagrams for rejecting borders and removing particles in the image, rejecting border can be used for rejecting the pectoral muscle and the sticker in some images, else, the removing particles are for removing all small bright particles in the image except the suspicious mass, so in this way, the mass is segmented separated from all other parts of the image and background. Image processing and mass segmentation are done, and now some features can be selected and extracted from the mass, there are seven features selected and extracted from the suspicious mass that can be used as input for the classifier. The seven features are Contrast, Standard Deviation, Mean Intensity, Skewness, Entropy, Smoothness, and Uniformity. The expressions of the features from Standard Deviation to Uniformity respectively are as the following in equation 5a, 5b, 5c, 5d, 5e and 5f.

$$\sigma = \sqrt{\mu_2(z)} = \sqrt{\sigma^2} \tag{5a}$$

$$m = \sum_{i=0}^{L-1} z_i \, p(z_i) \tag{5b}$$

$$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 \, p(z_i) \tag{5c}$$

$$e = \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \tag{5d}$$

$$R = 1 - \frac{1}{1+\sigma^2} \tag{5e}$$

$$U = \sum_{i=0}^{L-1} P^2(z_i) \tag{5f}$$

All seven features are used as an input to be fed into the classifier to classify the mass; ANN and SVM which are ones of the machine learning methods and whose competencies in scientific studies are accepted were used in the classifier part of the system.

ANN is a classifier whose construction is composed of mathematical models similar to the nervous system [26]. ANN is a classifier used to determine whether the segmented suspected mass is benign or malignant according to the extracted features. Its construction is composed of mathematical models and algorithms similar to the nervous system of humans. The construction of ANN is three main layers; the Input layer, Hidden layers, and Output layer, as shown in Fig. 2.

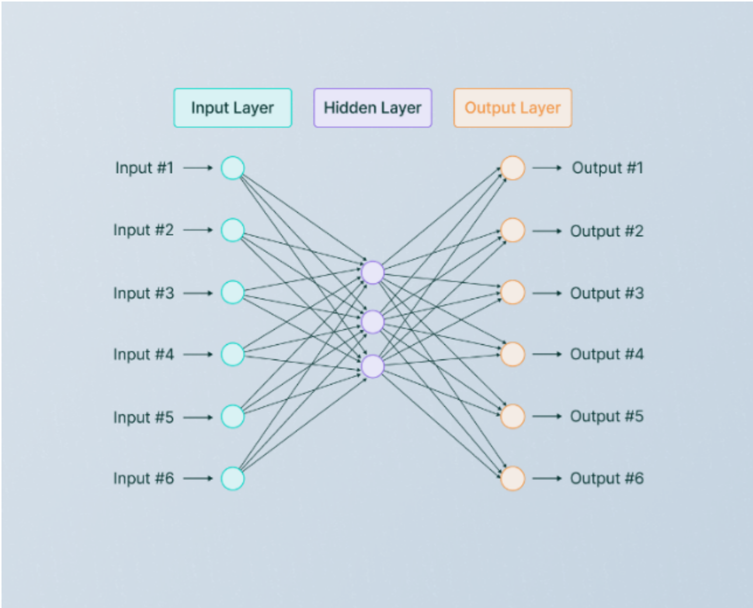

Figure 2. ANN simple Architecture.

One of the machine learning methods, SVM, also known as support vector networks, is a classification method with supervised learning algorithms that evaluate the data used for classification and regression analysis and recognize these patterns. Fig. 3 shows how SVM classifies benign and malignant masses in breast cancer [7].

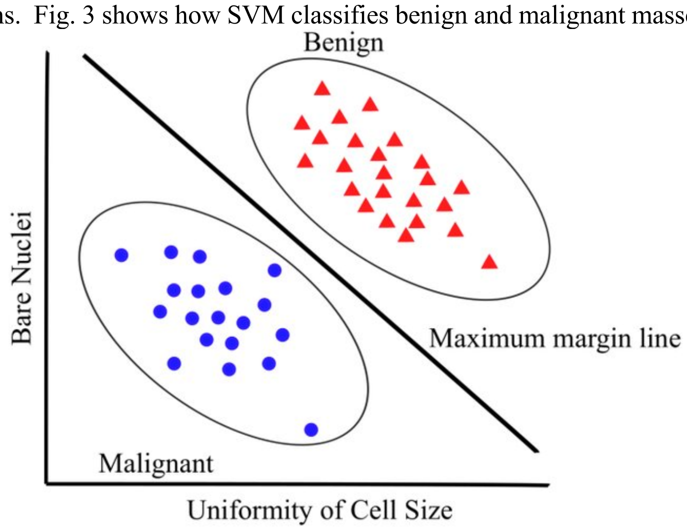

Figure 3. Demonstration of how SVM classifies benign and malignant masses in breast cancer.

The developed algorithm in the LabVIEW platform has some sub-VIs that are used to make the program work faster and easier. In the algorithm, there are five sub-VI GUIs. The GUI of the image processing and segmentation's sub-VI is given in Fig. 4.

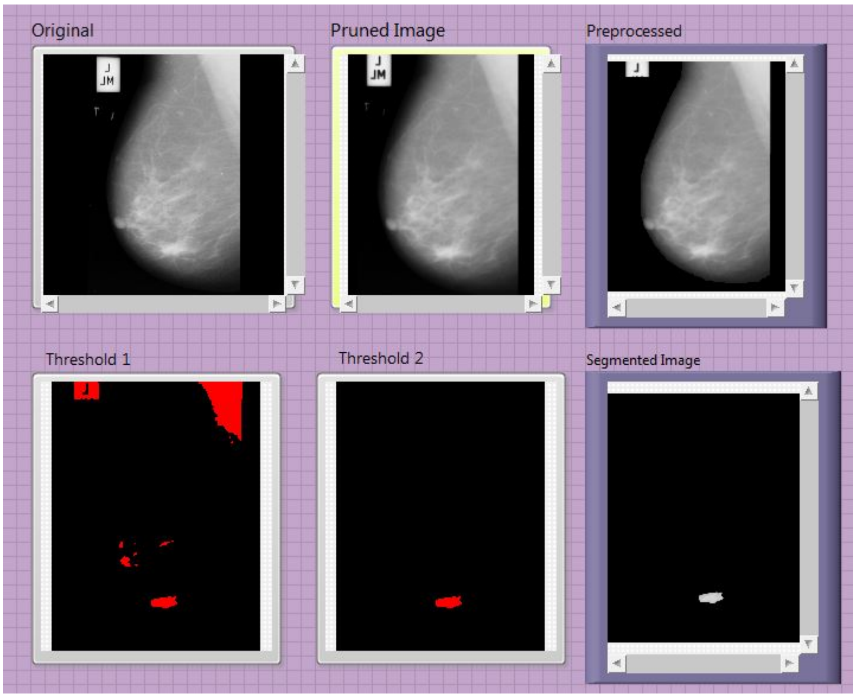

Figure 4. Image processing and Segmentation GUI.

After getting the segmented mass in the image which is the tumor of the breast, now it turns to finding its features of it for analysis and using them in the classification process. The features are Standard Deviation, Mean Intensity, Contrast, Skewness, Entropy, Smoothness, and Uniformity.

All the images in the database have to be trained to get their features, the result of every single image for all of the seven features will be like a table or like a matrix that can be safe as an excel file. For the classification stage, the ANN and SVM model can be used, and the options for input layers, hidden layers, and output layers can be chosen here for later to get the best result. The block diagrams and front panel of the last sub-VIs are shown in Fig 5-7.

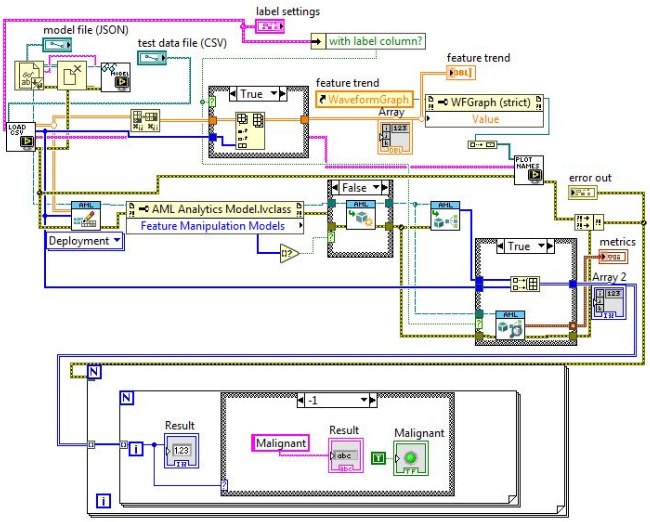

Figure 5. ANN Classification (block diagram).

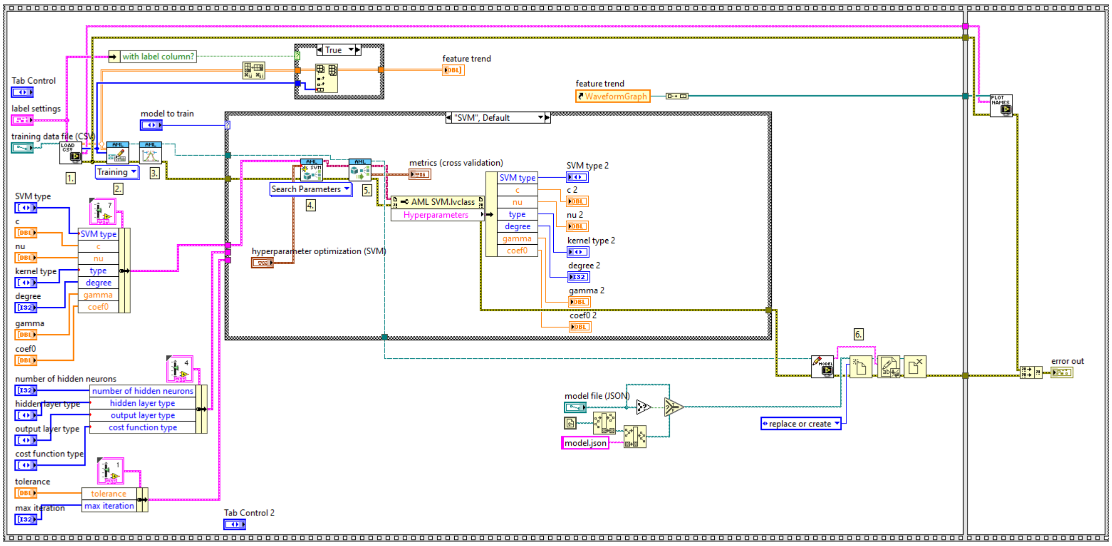

Figure 6. SVM classification (block diagram).

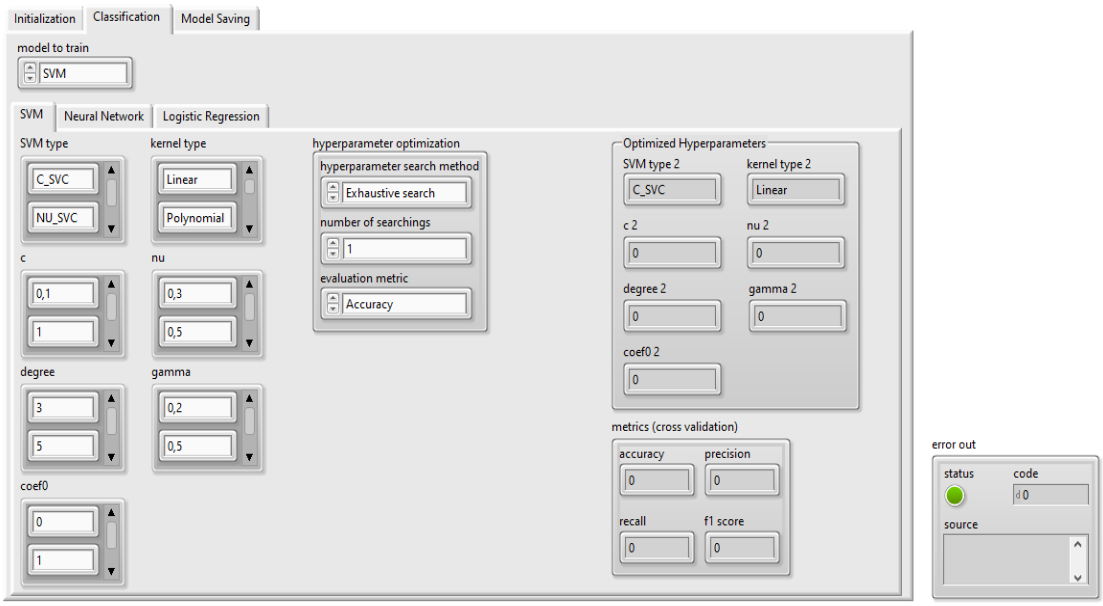

Figure 7. GUI screen for classification (front panel diagram).

With these created sub-VIs, the user will be able to determine whether the mass in the mammogram image is benign or malignant with 2 types of classification methods, which are ANN and SVM. The block diagrams given in Fig. 5 and 6 were created according to the mathematical forms of ANN and SVM algorithms.

4. Results

After applying all of the image processing algorithms one by one, a series of images are established which are shown in Fig. 8-10. The selected and extracted features from the segmented mass are shown in Table 2 and Table 3 with the result of classification with ANN and in Table 4 and Table 5 with the result of classification with SVM. The result of system classification is also compared with the provided results of the database reference, which is discussed more in the discussion section to show the efficiency of the system.

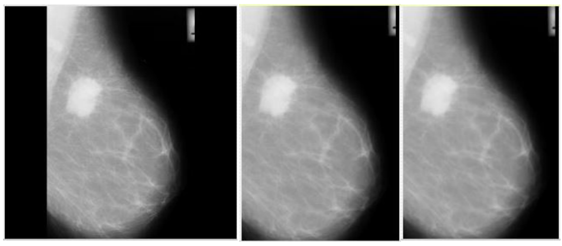

Figure 8. Original Image, Pruned Image, Applying smoothing filter on the Image.

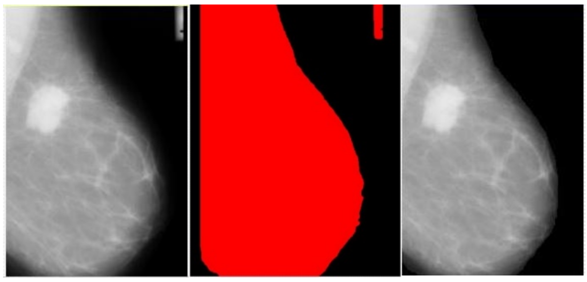

Figure 9. Applying Gaussian Filter on the Image, First Thresholding of the Image, Applying Morphological Operations on the Image.

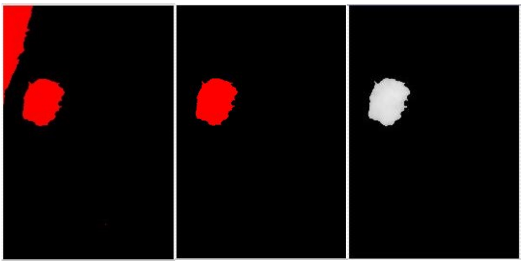

Figure 10. Second Thresholding of the Image, Removing all Undesired Parts, Segmented Mass.

The block diagram and front panel of the main program from which the 5 developed sub-VIs are run are shown in Fig. 11 and 12. In the block diagram, figure all three main stages are obvious to notice with the help of using sub-VI; Image processing and segmentation, Feature extraction, and Classification.

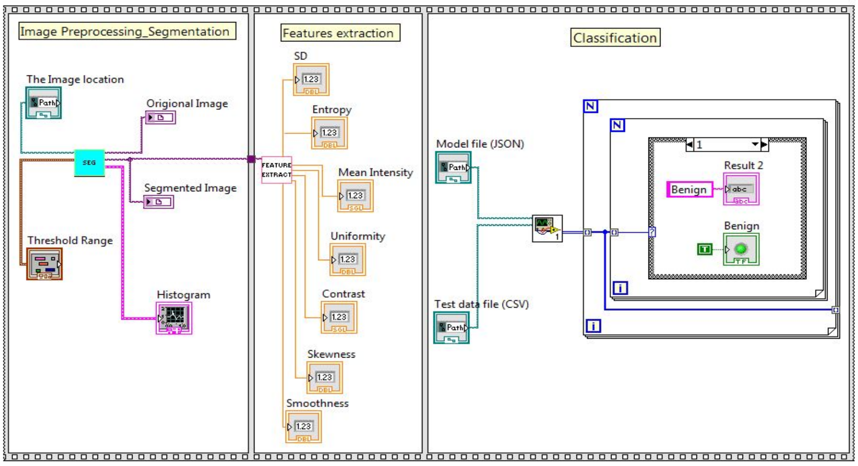

Figure 11. Main VI (block diagram).

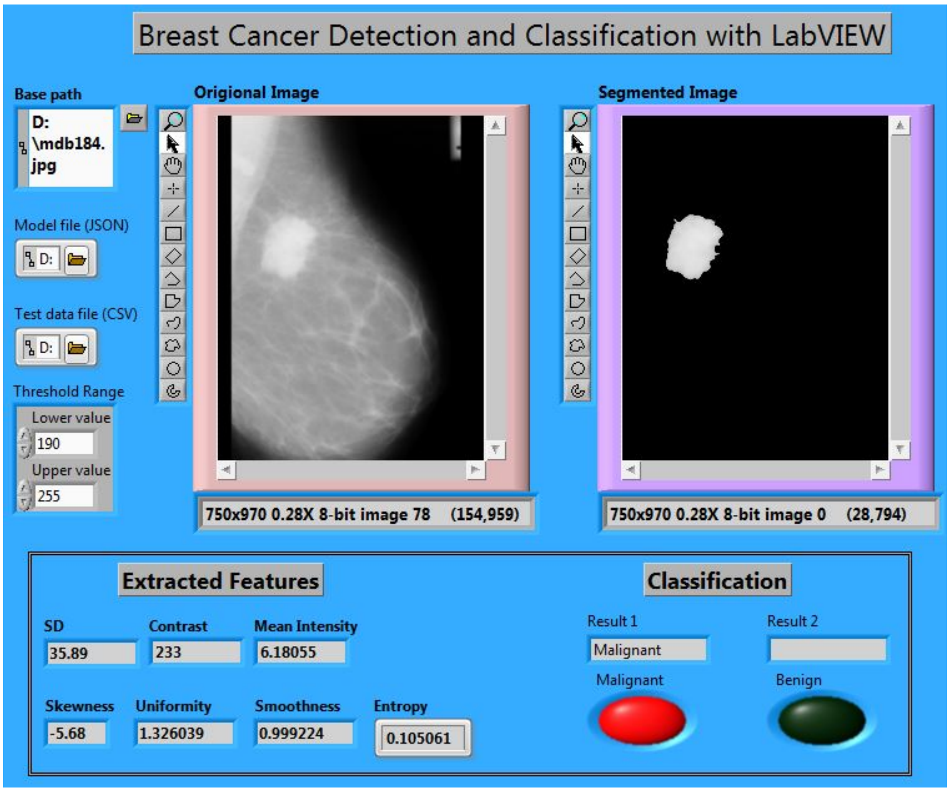

Figure 12. The developed main GUI.

From the database 62.5% of dataset images are trained into the system then the remained part of 37.5% which equals 80 images (40 benign and 40 malignant) are set to be test data. There is a matrix called the confusion matrix that shows the percentage of the correct result of any system, the matrix and the meaning of each element in the matrix is shown in Table 6 and in Equation 6.

From the database 62.5% of dataset images are trained into the system then the remained part of 37.5% which equals 80 images (40 benign and 40 malignant) are set to be test data. There is a matrix called the confusion matrix that shows the percentage of the correct result of any system, the matrix and the meaning of each element in the matrix is shown in Table 6 and in Equation 6.

$$\text{Confusion Matrix} = \begin{bmatrix} \text{TN} & \text{FN} \\ \text{FP} & \text{TP} \end{bmatrix} \tag{6}$$

Table 2. Extracted features from 40 benign images.

Image	Selected and Extracted Features							Classification with ANN	
Ref No.	Standard Deviation	Contrast	Mean Intensity	Skewness	Uniformi ty	Smoot hness	Entropy	Original	Estimat ed
mdb010	27.128	4.51	199	-6.003	0.999	0.756	0.095	B	B
mdb013	45.602	11.325	221	-3.826	1	2.208	0.184	B	B
mdb015	13.488	0.95	205	-14.309	0.995	0.183	0.021	B	B
mdb017	23.049	2.849	204	-8.091	0.998	0.539	0.054	B	B
mdb021	33.892	6.583	203	-5.047	0.999	1.192	0.123	B	B
mdb025	11.156	0.661	198	-17.056	0.992	0.125	0.014	B	B
mdb080	15.194	1.467	186	-10.5	0.996	0.233	0.038	B	B
mdb081	42.052	8.923	229	-4.539	0.999	1.848	0.15	B	B
mdb083	13.041	0.861	216	-15.303	0.994	0.171	0.019	B	B
mdb091	9.262	0.513	179	-18.336	0.988	0.086	0.013	B	B
mdb099	11.778	0.739	207	-16.169	0.993	0.139	0.017	B	B
mdb104	16.755	1.429	209	-11.76	0.996	0.283	0.028	B	B
mdb121	42.736	9.521	228	-4.309	0.999	1.917	0.16	B	B
mdb126	67.676	26.88	223	-2.157	1	5.303	0.388	B	B
mdb127	9.387	0.47	205	-20.301	0.989	0.088	0.011	B	B
mdb132	10.648	0.654	198	-16.518	0.991	0.114	0.018	B	B
mdb133	12.208	0.963	177	-12.954	0.993	0.15	0.026	B	B
mdb145	35.345	6.49	221	-5.323	0.999	1.291	0.115	B	B
mdb150	21.636	2.677	201	-8.047	0.998	0.475	0.058	B	B
mdb160	13.221	0.986	199	-13.638	0.994	0.176	0.024	B	B
mdb165	26.676	3.902	202	-6.739	0.999	0.727	0.073	B	B
mdb175	16.325	1.416	211	-11.636	0.996	0.269	0.03	B	B
mdb193	39.274	7.708	225	-4.936	0.999	1.602	0.123	B	B
mdb195	10.911	0.6	210	-18.427	0.992	0.119	0.013	B	B
mdb198	20.243	2.045	216	-9.908	0.998	0.414	0.039	B	B
mdb199	20.026	2.108	210	-9.522	0.998	0.405	0.043	B	B
mdb204	16.085	1.555	186	-10.473	0.996	0.261	0.037	B	B
mdb207	33.965	6.061	218	-5.479	0.999	1.19	0.107	B	B
mdb218	**34.461**	**6.102**	**223**	**-5.529**	**0.999**	**1.225**	**0.109**	**B**	**M**
mdb219	14.084	1.119	201	-12.807	0.995	0.2	0.027	B	B
mdb222	36.359	6.666	219	-5.318	0.999	1.366	0.107	B	B
mdb227	8.847	0.427	199	-21.162	0.987	0.078	0.01	B	B
mdb236	66.086	24.124	222	-2.404	1	4.949	0.325	B	B
mdb244	57.904	17.46	229	-3.047	1	3.658	0.254	B	B
mdb248	11.496	0.748	190	-15.602	0.992	0.133	0.018	B	B
mdb252	17.052	1.85	179	-9.308	0.997	0.294	0.045	B	B
mdb290	36.166	7.098	209	-4.931	0.999	1.358	0.126	B	B
mdb312	10.629	0.58	212	-18.689	0.991	0.113	0.013	B	B
mdb314	16.506	1.633	194	-10.263	0.996	0.275	0.039	B	B
mdb315	48.575	11.686	229	-3.942	1	2.496	0.176	B	B

Table 3. Extracted features from 40 malignant images.

Image	**Selected and Extracted Features**							**Classification with ANN**	
Ref No.	Standard Deviation	Contrast	Mean Intensity	Skewness	Uniformity	Smoothness	Entropy	Original	Estimated
mdb023	21.716	2.405	217	-9.07	0.998	0.477	0.046	M	M
mdb028	**18.364**	**1.696**	**214**	**-10.82**	**0.997**	**0.34**	**0.034**	**M**	**B**
mdb058	9.163	0.453	208	-20.71	0.988	0.084	0.011	M	M
mdb072	23.955	2.789	226	-8.558	0.998	0.582	0.052	M	M
mdb075	7.099	0.348	227	-21.42	0.981	0.051	0.012	M	M
mdb092	9.541	0.539	184	-17.99	0.989	0.091	0.014	M	M
mdb095	21.397	2.451	212	-8.811	0.998	0.464	0.052	M	M
mdb102	22.438	2.5	221	-8.982	0.998	0.51	0.048	M	M
mdb105	87.503	42.187	241	-1.605	1	9.437	0.515	M	M
mdb110	25.033	3.08	222	-8.075	0.998	0.636	0.056	M	M
mdb111	31.75	4.969	224	-6.283	0.999	1.033	0.084	M	M
mdb120	21.959	2.578	211	-8.555	0.998	0.489	0.052	M	M
mdb125	49.339	13.209	219	-3.512	1	2.609	0.212	M	M
mdb130	28.315	4.045	220	-6.944	0.999	0.818	0.071	M	M
mdb134	**13.344**	**0.953**	**202**	**-14.09**	**0.994**	**0.179**	**0.022**	**M**	**B**
mdb141	25.657	3.995	187	-6.384	0.998	0.674	0.082	M	M
mdb148	56.235	19.102	221	-2.649	1	3.527	0.327	M	M
mdb171	87.781	42.626	239	-1.587	1	9.523	0.512	M	M
mdb178	33.406	6.064	220	-5.408	0.999	1.153	0.117	M	M
mdb179	29.169	3.779	240	-7.736	0.999	0.865	0.05	M	M
mdb181	34.233	7.356	202	-4.516	0.999	1.226	0.152	M	M
mdb184	35.887	6.181	233	-5.681	0.999	1.326	0.105	M	M
mdb202	42.386	10.128	214	-4.008	0.999	1.899	0.18	M	M
mdb206	30.373	5.892	189	-5.026	0.999	0.957	0.129	M	M
mdb209	38.634	8.318	212	-4.476	0.999	1.562	0.155	M	M
mdb211	49.636	14.046	222	-3.293	1	2.661	0.247	M	M
mdb213	28.255	5.216	181	-5.283	0.999	0.826	0.113	M	M
mdb216	83.16	39.91	236	-1.624	1	8.508	0.511	M	M
mdb231	19.055	2.651	161	-7.218	0.997	0.37	0.07	M	M
mdb233	35.512	7.878	202	-4.339	0.999	1.323	0.165	M	M
mdb238	34.318	7.599	188	-4.364	0.999	1.235	0.151	M	M
mdb239	75.957	31.058	233	-2.056	1	6.734	0.379	M	M
mdb241	42.164	9.494	214	-4.247	0.999	1.868	0.152	M	M
mdb245	24.639	3.617	210	-6.81	0.998	0.62	0.079	M	M
mdb249	21.092	2.301	212	-9.17	0.998	0.45	0.046	M	M
mdb253	84.173	43.274	223	-1.448	1	8.958	0.51	M	M
mdb256	53.043	16.436	221	-2.955	1	3.084	0.283	M	M
mdb270	23.252	3.103	231	-7.478	0.998	0.55	0.065	M	M
mdb271	24.133	3.298	215	-7.337	0.998	0.593	0.074	M	M
mdb274	13.378	1.238	166	-10.87	0.994	0.181	0.035	M	M

Table 4. Extracted features from 40 benign images.

Image	Selected and Extracted Features							Classification with SVM	
Ref.No.	Standard Deviation	Contrast	Mean Intensity	Skewness	Uniformi ty	Smoothness	Entro py	Original	Estimate d
mdb010	28.352	4.245	211	-4.544	0.997	0.781	0.277	B	B
mdb013	49.446	11.416	219	-4.093	0.999	1.881	0.347	B	B
mdb015	9.541	0.539	184	-17.996	0.989	0.091	0.014	B	B
mdb017	21.397	2.451	212	-8.811	0.998	0.464	0.052	B	B
mdb021	23.955	2.789	226	-8.558	0.998	0.582	0.052	B	B
mdb025	9.541	0.539	184	-17.996	0.989	0.091	0.014	B	B
mdb080	12.525	4.219	219	-16.62	0.989	0.473	0.428	B	B
mdb081	29.316	7.521	218	-8.819	0.999	1.778	0.286	B	B
mdb083	61.764	22.34	229	-4.571	1	7.083	0.186	B	B
mdb091	28.817	4.78	245	-10.901	0.991	0.481	0.16	B	B
mdb099	19.448	8.641	228	-11.184	0.998	0.148	0.182	B	B
mdb104	21.081	3.639	217	-2.544	0.995	0.157	0.216	B	B
mdb121	31.445	16.94	229	-8.213	0.999	0.991	0.215	B	B
mdb126	31.102	18.744	214	-3.747	0.995	0.574	0.318	B	B
mdb127	11.271	7.846	199	-9.311	0.999	1.716	0.032	B	B
mdb132	19.746	11.042	219	-4.397	0.999	0.924	0.113	B	B
mdb133	14.215	9.406	217	-10.506	0.996	0.226	0.029	B	B
mdb145	**42.278**	**8.018**	**223**	**-6.436**	**0.998**	**2.101**	**0.147**	**B**	**M**
mdb150	19.181	0.614	202	-14.872	0.998	0.685	0.214	B	B
mdb160	12.345	8.051	218	-10.058	0.997	0.149	0.111	B	B
mdb165	19.824	7.203	213	-7.282	0.998	0.630	0.023	B	B
mdb175	11.113	1.335	196	-15.730	0.996	0.162	0.037	B	B
mdb193	24.605	4.064	210	-6.709	0.999	1.911	0.117	B	B
mdb195	12.515	5.029	219	-10.605	0.998	0.181	0.181	B	B
mdb198	14.134	6.411	213	-6.487	0.995	0.562	0.154	B	B
mdb199	17.536	10.022	201	-9.913	0.998	3.461	0.125	B	B
mdb204	36.515	5.216	181	-5.283	0.999	0.826	0.113	B	B
mdb207	83.16	39.91	236	-1.624	1	8.508	0.511	B	B
mdb218	19.055	2.651	161	-7.218	0.997	0.387	0.171	B	B
mdb219	35.512	5.178	212	-6.329	0.998	1.213	0.145	B	B
mdb222	34.318	7.599	188	-4.364	0.999	1.235	0.151	B	B
mdb227	43.547	30.858	231	-2.156	1	7.004	0.297	B	B
mdb236	34.614	19.412	218	-5.417	0.999	3.467	0.124	B	B
mdb244	12.319	13.110	212	-4.189	0.998	0.171	0.201	B	B
mdb248	12.982	8.051	219	-6.174	0.998	1.544	0.096	B	B
mdb252	78.713	34.714	223	-5.143	0.999	5.518	0.391	B	B
mdb290	56.043	12.136	214	-3.505	0.999	2.405	0.125	B	B
mdb312	20.512	13.033	229	-6.481	0.998	1.155	0.241	B	B
mdb314	9.312	9.418	210	-4.133	0.998	0.183	0.014	B	B
mdb315	18.748	11.481	185	-14.711	0.993	0.221	0.150	B	B

Table 5. Extracted features from 40 malignant images.

Image	**Selected and Extracted Features**							**Classification with SVM**	
Ref.No.	Standard Deviation	Contrast	Mean Intensity	Skewness	Uniformity	Smoothness	Entropy	Original	Estimated
mdb023	12.161	10.402	203	-5.283	0.999	2.241	0.152	M	M
mdb028	19.121	8.318	198	-7.624	0.996	0.896	0.121	M	M
mdb058	34.171	5.216	186	-1.218	0.997	1.011	0.177	M	M
mdb072	11.891	10.401	229	-4.339	0.998	1.805	0.366	M	M
mdb075	14.425	6.851	216	-4.364	0.999	0.412	0.118	M	M
mdb092	12.849	9.003	179	-2.056	0.998	0.458	0.365	M	M
mdb095	12.135	10.315	207	-11.121	0.998	0.099	0.145	M	M
mdb102	44.054	8.604	209	-7.419	1	0.891	0.312	M	M
mdb105	32.199	9.211	228	-4.471	0.999	0.955	0.114	M	M
mdb110	21.899	3.199	219	-14.416	0.997	0.129	0.123	M	M
mdb111	18.896	9.218	199	-8.181	0.999	1.611	0.163	M	M
mdb120	43.125	4.156	222	-5.108	0.998	1.363	0.178	M	M
mdb125	48.111	11.189	229	-5.214	0.999	5.414	0.499	M	M
mdb130	23.236	8.511	190	-1.287	0.997	2.487	0.189	M	M
mdb134	52.811	7.188	179	-9.629	0.998	2.011	0.196	M	M
mdb141	11.789	4.919	209	-6.344	0.999	0.395	0.163	M	M
mdb148	13.417	3.058	212	-5.256	1	4.489	0.275	M	M
mdb171	23.124	9.779	187	-8.017	0.999	1.955	0.201	M	M
mdb178	28.895	10.044	221	-8.849	0.998	1.718	0.255	M	M
mdb179	23.619	13.559	239	-3.316	0.999	1.611	0.102	M	M
mdb181	41.343	4.564	220	-9.156	0.999	2.428	0.132	M	M
mdb184	19.187	8.401	240	-8.851	1	3.216	0.458	M	M
mdb202	33.486	8.112	202	-7.818	0.999	4.191	0.211	M	M
mdb206	38.713	4.912	233	-6.201	0.998	1.045	0.254	M	M
mdb209	31.437	6.141	214	-2.145	0.999	2.612	0.176	M	M
mdb211	59.165	11.614	219	-11.003	1	3.866	0.321	M	M
mdb213	**18.957**	**9.916**	**212**	**-8.455**	**0.999**	**1.058**	**0.199**	**M**	**B**
mdb216	45.146	31.141	201	-5.244	0.997	6.004	0.301	M	M
mdb231	29.857	12.531	161	-5.811	0.997	1.778	0.457	M	M
mdb233	18.521	6.689	202	-4.352	0.999	3.653	0.257	M	M
mdb238	49.418	8.402	188	-4.619	0.999	2.514	0.326	M	M
mdb239	35.757	19.518	233	-3.151	1	4.437	0.233	M	M
mdb241	27.764	19.944	214	-2.141	0.999	3.618	0.304	M	M
mdb245	21.139	13.117	210	-1.991	0.998	1.121	0.039	M	M
mdb249	29.491	8.871	228	-6.571	0.998	1.415	0.099	M	M
mdb253	74.713	31.714	219	-10.844	1	4.257	0.465	M	M
mdb256	61.413	9.036	199	-7.544	0.998	5.104	0.196	M	M
mdb270	29.212	13.013	228	-4.718	0.999	1.956	0.132	M	M
mdb271	20.983	7.281	219	-6.157	0.998	4.913	0.156	M	M
mdb274	18.718	5.381	231	-9.711	0.995	3.801	0.099	M	M

Table 6. Measured and meaning of each part in the confusion matrix.

Measures	Meaning
True Negative (TN)	The mass is defined as benign by biopsy and is also classified as benign by the neural network.
False Negative (FN)	The mass is defined as malignant by a biopsy but it is classified as benign by the neural network.
False Positive (FP)	The mass is defined as benign by a biopsy but it is classified as malignant by the neural network.
True positive (TP)	The mass is defined as malignant by a biopsy and is also classified as malignant by the neural network.

After training and testing all 80 images, the results of the confusion matrix of ANN and SVM are as the following in Equation 7 and 8:

for ANN:

$$\text{Confusion Matrix} = \begin{bmatrix} 39 & 2 \\ 1 & 38 \end{bmatrix} \tag{7}$$

for SVM:

$$\text{Confusion Matrix} = \begin{bmatrix} 39 & 1 \\ 1 & 39 \end{bmatrix} \tag{8}$$

When examing Table 2 and Table 3, the missed images for ANN are the image with the reference number (mdb028 - mdb134) that are measured as a false negative, the misinterpreted image is the image with the reference number of (mdb218) that is measured as false positive. Also, while examing Table 4 and Table 5, the missed images for SVM are the image with the reference number (mdb213) that are measured as a false negative, the misinterpreted image is the image with the reference number of (mdb145) that is measured as false positive.

It is not good for the system to have a high value in FN and FP so to have an accurate diagnosis, the values of FN and FP should be small because a high value of FN means that malignant masses are missed in the handling process and high value of FP means that benign masses are misinterpreted as cancer. The obtained confusion matrix indicates some important performance measuring of the neural network classifier. These measures are sensitivity, specificity, and accuracy. They can be measured by the following expressions in equation 9a, 9b and 9c:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{9a}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{9b}$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{9c}$$

After measuring the sensitivity, specificity, and accuracy the results of ANN and SVM are given in Table 7.

Table 7. The sensitivity, specificity, and accuracy of ANN and SVM classification.

Classifier	The sensitivity (%)	The specificity (%)	The accuracy (%)
ANN	95	97.5	96.25
SVM	97.5	97.5	97.5

Also, it was observed that this study gave better results in terms of sensitivity, specificity, and accuracy than studies [9], [10] and [18] that studied breast cancer using the same dataset.

5. Conclusion

Breast cancer constitutes approximately 22% of cancer types seen in women. It is known that 15% of breast cancer cases result in death. Although many scientists carry out cancer-preventive studies, an acceptable treatment protocol has not been established in the current situation. For this reason, women over the age of 40 are required to have a mammogram every 6 months. The clinician, who examines these mammogram images, decides whether the mass in the images is benign or malignant. However, a mass that is overlooked by the clinician may cause irreversible results. For this reason, many scientists use these mammogram images with various artificial intelligence methods and perform computer-assisted mass detection studies to minimize the errors that may occur due to the clinician.

In this study, a graphical user interface was designed to analyze the mammogram images with machine learning using mathematical expressions instead of using a ready-made toolbox using a graphical code-based software platform and present the masses in the images to the clinician's approval. ANN and SVM algorithms were developed to detect abnormalities in mammograms. As seen in Table 7, the results for sensitivity, specificity, and accuracy, obtained in the ANN are 95%, 97.5% 96.25%, and the same values for SVM are 97.5%, 97.5% and 97.5%, respectively. SVM algorithms gave better accuracy than ANN and also the both developed algorithms gave better results than the studies which used the same dataset and algorithms.

In future studies, it is planned to demonstrate both machine learning and deep learning algorithms on the same GUI by using different datasets and more mammogram images.

Acknowledgements

A significant part of this paper includes the Master Thesis data of the first author. The authors confirm that the data supporting the findings of this study are available within the article.

References

[1] Heber D. Nutritional oncology Elsevier.2011; 393-404.
[2] Rangayyan RM, Neuman MR, Raton EB. Breast cancer and mammography. Biomedical Image Analysis 2005, 22-27.
[3] Vanel D. The American College of Radiology (ACR) breast imaging and reporting data system (BI-RADS™): a step towards a universal radiological language?. Eur J Radiol 2007; 61(2): 183.
[4] Smith RA, Saslow D, Sawyer KA, Burke W. Costanza ME, Evans III WP, & Sener S. American Cancer Society guidelines for breast cancer screening: update 2003, CA Cancer J Clin, 53(3), 141-169.
[5] Alteri R, Barnes, C, Burke A, Gansler T, Gapstur S, Gaudet M, Xu JQ. Breast cancer facts & figures 2013-2014. Atlanta: American Cancer Society.2013,1-38,
[6] Giuliano AE, Edge SB, Hortobagyi GN. Eighth edition of the AJCC cancer staging manual: breast cancer. Ann Surg Oncol, 2018; 25(7): 1783-1785.
[7] Divyavani M, Kalpana G. An analysis on SVM & ANN using breast cancer dataset. Aegaeum J, 8,2021, 369-379.
[8] Guzman- Cabrera R, Guzman-Sepulveda JR, Torres-Cisneros M, May- Arrioja D A, Ruiz-Pinales J, Ibarra-Manzano OG Avina Cervantes C, Gonzalez Parada A. Digital Image Processing Technique for Breast Cancer Detection, Int J Thermophys 2013, Springer Science Business Media New York 2012.
[9] Monica Jenefer B, Cyrilraj V. An efcient Image Processing methods for Mammogram Breast Cancer detection, JATIT, 2014, vol,69 No.1.
[10] Kumar AS, Bhupendra GA. Novel Approch for Breast Cancer detection and segmentation in a Mammogram. Procedia Comput Sci, 2015;54:676–82.
[11] Sonal N. Early detection of Breast Cancer using ANN, IJIRCCE, 2016;4, issue ,14008-14013.
[12] Angayarkanni N, Kumar D, Arunachalam G. The Application of Image Process- ing techniques for detection and classifcation of cancerous tissue in Digital Mammograms. JPSR. 2016;8(10):1179–83.
[13] Habib D, Eslam AIM, Awais M, Wail E, Mohammed FN. Automated Breast Cancer Diagnosis based on Machine Learning Algorithms, Journal of Health Engineering, volume 2019, Article ID 425341.
[14] Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst Appl 2009;36:3240–7.
[15] Muhic I. Fuzzy analysis of breast cancer disease using fuzzy c-means and pattern recognition, Southeast Europe J Soft Comput, 2013;2(1).
[16] Abien FMA. On breast cancer detection: an application of machine learning algorithms on the Wisconsin Diagnostic Dataset, ICMLSC 2018.
[17] Taha M. Classifcation of mammograms for breast cancer detection using fusion of discrete cosine transform and discrete wavelet transform features. Biomed Res. 2016;27(2):322–7.

[18] Ramik R. Breast Cancer Prediction using Machine Learning, JETIR, 2020;7(5)
[19] Onega T, Hubbard R, Hill D, Lee CI, Haas JS, Carlos HA, Tosteson AN. Geographic access to breast imaging for US women. Jour of the American Coll of Radiology, 11(9), 874-882,2014.
[20] Ball JE, Bruce LM. Digital mammogram spiculated mass detection and spicule segmentation using level sets. In 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,pp. 4979-4984,2007.
[21] Ramos RL, Armán FA, García MR, Fariñas IC, Perez EC. Well-circumscribed breast carcinoma. Keys to face the challenge of malignant tumors with a benign appearance. European Congress of Radiology-ECR 2015.
[22] Ayres FJ, & Rangayvan, R M. Characterization of architectural distortion in mammograms. IEEE Eng Med Biol Mag, 24(1), 59-67, 2005.
[23] Tang J. et. Al, Computer-aided detection and diagnosis of breast cancer with mammography, recent advances. IEEE Trans Inf Technol Biomed, 2009,13(2), 236-251.
[24] Smith A. Fundamentals of breast tomosynthesis. White Paper, Hologic Inc., WP-00007, 2008,8.
[25] Suckling J, Parker J, Dance D. Mammographic Image Analysis Society (MIAS) database v1.21. [Dataset]. Apollo - University of Cambridge Repository, 2015, https://www.repository.cam.ac.uk/handle/1810/250394
[26] Cheng HD, Shi XJ, Min R, Hu LM, Cai XP, Du HN. Approaches for automated detection and classification of masses in mammograms. Pattern Recognit, 2006, 39(4), 646-668.
[27] Mencattini A, Salmeri M, Lojacono R, Frigerio M, Caselli F. Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing. IEEE Trans Instrum Meas 2008; 57(7): 1422-1430.
[28] Tanyıldızı E, Orhan A. An introduction to variable and feature selection. Comput Appl Eeng Educ 2009; 17(2): 187-195.
[29] Mini MG, Thomas T. A neural network method for mammogram analysis based on statistical features. In TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region, 2003, Vol. 4, pp. 1489-1492