DETECTION OF BREAST CANCER IN MAMMOGRAPHY IMAGES USING THE NOVEL TECHNIQUE OF GENETIC ALGORITHM AND DECISION TREE

Daljit Kaur
M.Tech Student
Computer Science Engineering
KCIET, Nawanshehar, Punjab, India.

Dr. Vijay Kumar Joshi
Principal
Computer Science Engineering
KCIET,Nawanshehar,Punjab,India

Abstract:Now day's breast cancer is prevailing in women. It can be spread to other areas as it started from the breast and it is second largest disease that causes death of women. It is curable if it is detected at the early stages but lack of proper methodology is still a challenge. In this paper we proposed an approach to detect breast cancer at early stages that is based on genetic algorithm and decision tree. The genetic approach is an iterative technique that gives results faster and decision tree is utilized for classification. The multi support vector machine is also utilized in order to extract the feature and match it against the trained image. The parameters which are utilized in proposed work includes classification accuracy, precision, F-score, recall. The simulation results show better performance as compare to existing system.

***Key words*: Genetic Algorithm, Decision Tree, Classification**

1. INTRODUCTION

Cancer starting in the mammary organ is the most well-known kind of cancer in ladies. The lifetime danger of breast cancer for a lady in created nations has been computed at around 1 of every 7 to 1 out of 10. With regards to Catalonia, the most recent investigations report a collected lifetime danger of creating breast cancer of 1 out of 11 with a 1 out of 33 likelihood of death because of breast cancer. This implies around 10% of the female populace will be determined to have breast cancer sooner or later of their life.

Out of these patients, around 30-40% will inevitably pass on of this ailment, essentially because of the refinement of metastases, a hopeless condition in many sorts of cancer. This high occurrence, the many-sided quality and the overall expenses of the treatment for this infection make breast cancer the important medical issues in our general public.

Genetic algorithm is multiheuristic algorithm having multiple objectives associated with it. Genetic algorithm is associated with different phases. The pixels correspond to chromosomes. To perform feature extraction, selection operation takes place. The proposed system uses random selection operation. The fitness function evaluation is used to generate next population for feature extraction. The extracted features are compared against the threshold value. The threshold value is assumed to be base value above which optimality is achieved.

$$Threshold = Optimal_{base-value}$$

The fitness function evaluation takes place in order to obtain optimal results for classification.

$$F(t) = Pixel_i(Features_j > Threshold)$$

The mutation and crossover is performed only if threshold value is invalidated. Mutation and crossover is accomplished by identifying pixels having intensity values lower than desired levels. After which selection operation is performed again.

2. LITERATURE SURVEYS

This section of the chapter provides details of earlier research carried out on the study of MRI images with feature extraction. There is enormous information available on the experimental work related to MRI images, which is discussed as under:

S.Able et al. [2001] One in five Americans will create skin tumor in their lifetime, and by and large, one American bites the dust from skin growth consistently. Melanoma spreads through metastasis, and can frequently be lethal. Factual proof has uncovered that the lion's share of passings coming about because of skin growth are accordingly of melanoma. Facilitate examinations have demonstrated that the survival rates in patients rely on upon the phase of the growth at the time it is first analyzed; early discovery and mediation of melanoma shows higher odds of cure. Clinical analysis and visualization of melanoma are trying since the procedures are inclined to misdiagnosis and mistakes because of specialists' subjectivity.

Harmful melanomas are deviated, have unpredictable outskirts, scored edges, and shading varieties, so breaking down the shape, shading, and surface of the skin sore is vital for melanoma early discovery and aversion. This paper proposes the two noteworthy segments of a noninvasive ongoing computerized skin sore investigation framework for melanoma early recognition and aversion. The primary segment is a continuous alarm to enable clients to avert skin consume brought on by daylight; a novel condition to figure the ideal opportunity for skin to consume is in this way presented.

The second part is a computerized picture examination module which contains picture securing, hair discovery and rejection, injury division, highlight extraction, and order. The proposed framework utilizes PH2 Dermoscopy picture database from Pedro Hispano Hospital for advancement and testing purposes. The picture database contains an aggregate of 200 dermoscopy pictures of sores, including considerate, atypical, and melanoma cases. An examination of the execution of all capabilities is exhibited in this paper with a specific end goal to figure out what highlight sets give the best order comes about.

N.Shimkin[2009] Effective techniques are required to detect tumor at early stage. Technology helps to detect cancer as early as possible so that cure is within the reach of humans. Segmentation and feature extraction is critical towards this aspect. This section discusses mechanisms that enhance performance of Tumors detection process.

Tumor detection is compulsory at early stage to avoid deadly effects within human body. Death rate is enhanced considerably if detection is at 4^{th} stage. Recovery rate is greatly enhanced if it is detected at 2^{nd} or early stage of lesion. Support vector machine is one of the effective images processing segmentation mechanism used to detect distinguished part from the original part.

E.J. Leavline et al. [2013] Skin growth is the most successive and dangerous kind of disease. Melanoma is the most forceful sort among skin malignancies and in the event that they are recognized at an early stage, they can be totally cured. In melanoma analysis, the location of the melanocytes in the epidermis territory is an essential stride. For the discovery of melanocytes, utilization of histopathological pictures can be utilized. With the current appearance of entire slide advanced scanners, tissue

histopathology slides can now be digitized and put away in computerized picture shape. The digitized pictures are broke down with cutting edge picture division calculations and components, for example, power and size of the cell cores is utilized to filter out the hopeful cores locales.

A.MasoodAndA.Jumaily [2016]proposed SVM and Deep belief network for detection of Tumour. A test vector x is considered for training purpose. Final classification through classification model is given through the following function

$$F_{final} = sign(w.(f_i(x)_{k*i})$$

The classifier includes deep learning architecture and exponential loss function used to enhance seperability. Deep belief network is constructed using greedy layer wise unsupervised learning algorithm and parameter space of W is constructed by the use of unsupervised learning approach along with exponential loss function for fine tuning the classifier. Accuracy of the classifier is up to 95% hence is efficient.

J.C. KavithaAndA.Suruliandi [2016] propose mechanism to classify demography image into melanoma and non melanoma images. Texture and colour features are extracted for analysing the same. GLCM is used to extract the texture features of an image. Colour histograms are effective mechanism proposed to extract the colour features in three colour spaces with primary colour collaboration including RGB, HSV and OPP. Classification is generated by the use of SVM(Support Vector Machine).[24]proposes ship detection by the use of texture and SVM classification. Image is characterized into sub block to reduce the complexity of the image. Each block is processed separately and then collaborated together to form complete image. Supervised learning technique SVM is used for classification. From the mathematical point of view, feature extraction mechanism utilized following equation

$$forlabelpair(i,j)x = 1 - - - l$$

$$\min(y) = \frac{1}{2}w^t w + C\sum \delta$$

W is known as weight factor, δ known as misclassification, C is known as regularization parameter.

3. PROPOSED METHODOLOGY

Proposed system uses Gaussian smoothening to remove the hairy part out of the cells in body that have tumor. In other words noise handling mechanism is used to handle any noise present within the image. Gaussian noise handling mechanism provide filtered image which is presented to SVM. [10]describes problem with classification is that region is classified in class i only if ith decision function is positive. The value if not positive then it is not classified. This problem of unclassified region is resolved by the use of genetic algorithm and decision tree mechanism. Hence GA and decision treeis proposed to overcome this problem of misclassification. The detailed steps are described as unders.

4. PREPROCESSING

In the pre-processing phase image is filtered using the Gaussian noise removal to remove any artifacts present within the image.the transformation associated with the Gaussian filter is applied to every pixel present within the image. the transformation equation used is given as

$$G(x) = (1/\sqrt{2\pi\sigma^2})^{e^{-\frac{x^2}{2\sigma^2}}}$$

This equation is implemented in one dimension. In two dimensions the equation is altered as

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-x^2+y^2/2\sigma^2}$$

Where X is the distance of pixel from horizontal axis and y is the distance of pixel from vertical axis. Σ is the slandered deviation.

The image is resized to make critical portion of the image clearer. The clipped image is enhanced using histogram equalization mechanism. Histogram equivalence technique on colored image is applied. The equation used is as follows

$$Hq(v) = round(\frac{cdf(v) - cdf_{min}}{(MxN) - 1} x(L-1)$$

Cdfmin is the minimum non zero value in cumulative frequency. MxN indicates number of pixel. To scale pixels in the original image having pixel L-1 and the equation is modified as

$$Hq(v) = round(\frac{cdf(v) - cdf_{min}}{(MxN) - 1} x(L-2)$$

Scaling of pixels allow non zero values of the pixels to be preserved. The values of colours in terms of RGB is enhanced by increasing the intensity values as

RGB(R,G,B)=RGB(255,254,100)

The values of RGB are adjusted to desired levels to increase the intensity levels. This mechanism is followed to increase contrast levels. After the intensity is enhanced pre-processed image is presented to the next phase.

5. FLOW OF PROPOSED SYSTEM

The proposed system performs pre-processing and the support vector machine is applied to classify the data. The classification if not successful then misclassification is obtained in terms of accuracy and fuzzy technique is implemented in order to classify the data. The misclassification hence is considerably reduced.

Flow Chart:-

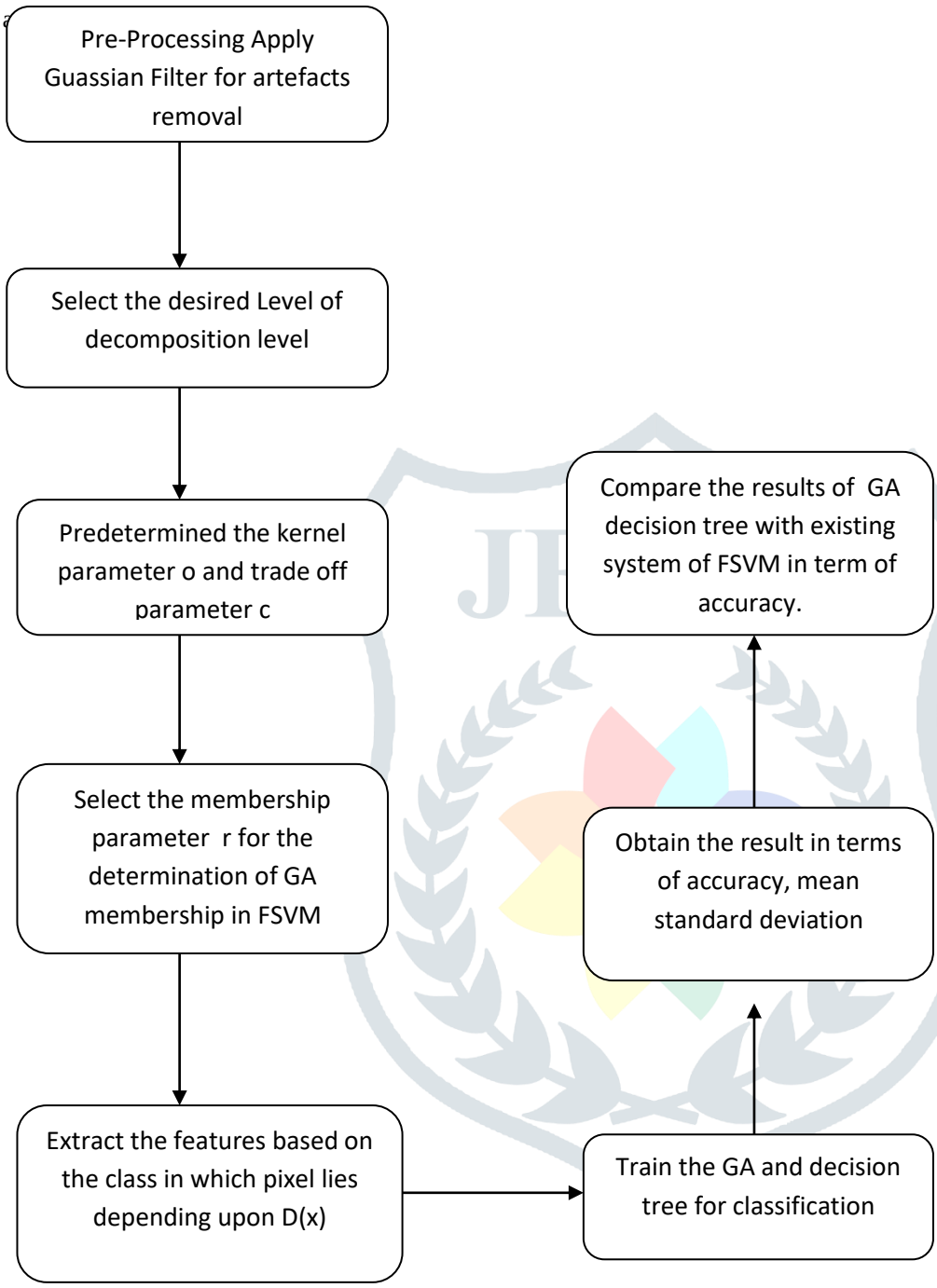

Figure 1: Flow chart showing the flow of operation

6. Results and discussion

The performance is evaluated by suing the dataset derived from the hospital. The listed dataset used and results are obtained in terms of segmentation and histogram as:

Misclassification is reduced considerably when fuzzy SVM is replaced with decision tree and GA. The results are compared against the existing techniques prove the worth of the study. Plot in terms of accuracy is given as

Table 1: Results in terms of various features using image benign6.jpg

Image Set	Attribute	Existing	Proposed
Img1	Accuracy	85.6195	89.0442
Img1	F-Score	9.28761	10.2876
Img1	Precision	7.28761	8.28761
Img1	Recall	10.2876	11.2876

Plots in terms of various attributes on image set img1 are shown as follows:

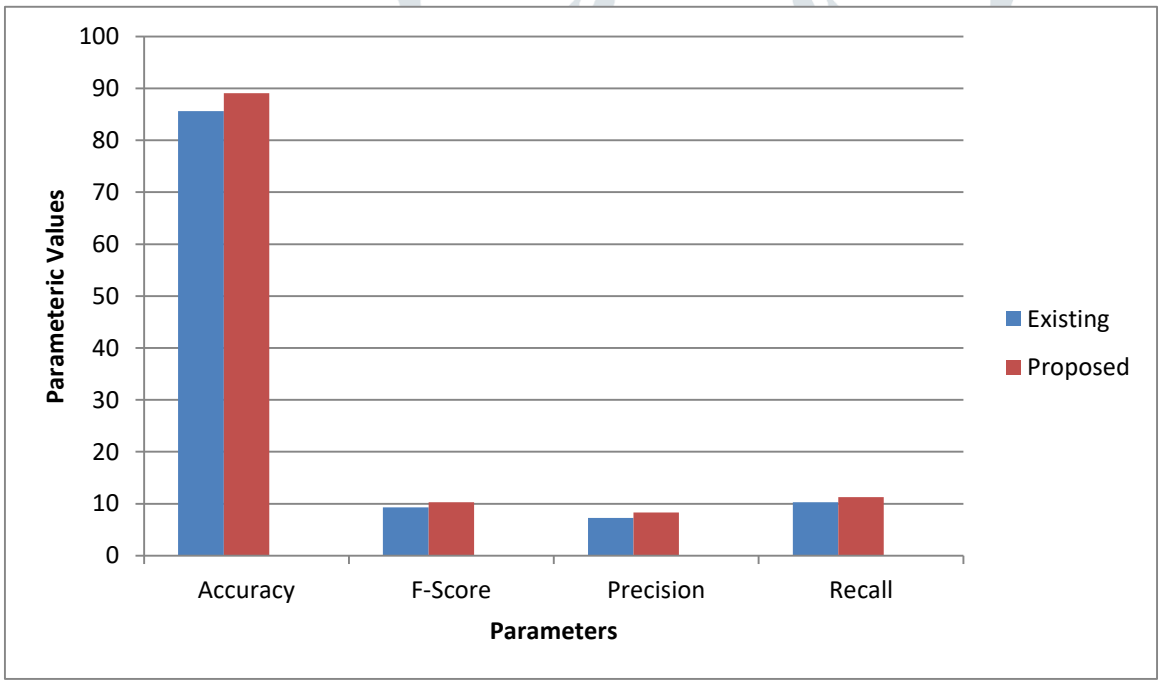

Figure 2:- Plot in terms of Accuracy, F-Score, Precision and recall.

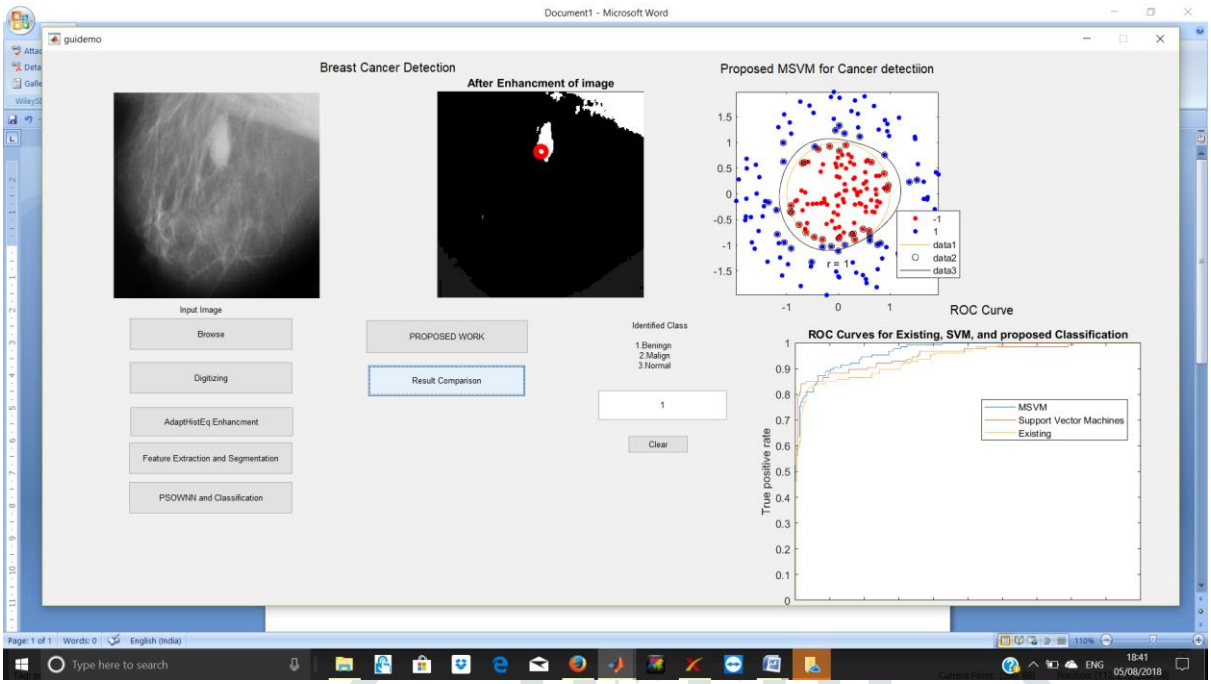

Figure 3: Simulation result using image benign6.jpg

As the training image is altered the results shows variation. The results in tabular form is given as under

Table 2:Results in terms of various features using image benign1.jpg

Image Set	Parameters	Existing	Proposed
Img2	Accuracy	73.8281	76.7813
Img2	F-Score	9.52344	10.5234
Img2	Precision	7.52344	8.52344
Img2	Recall	10.5234	11.5234

The plots indicate graphical presentation for better understanding which shows better result for fuzzy SVM.

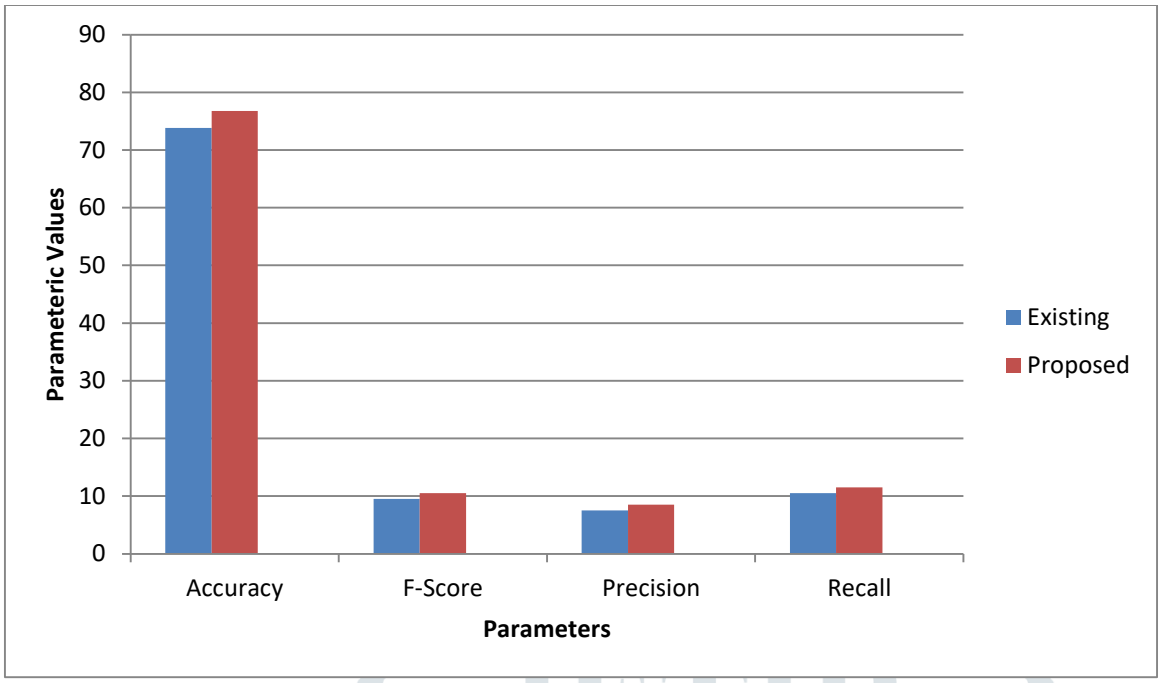

Figure 4: Comparison of results on image 2corresponding to accuracy, f score, precision and recall

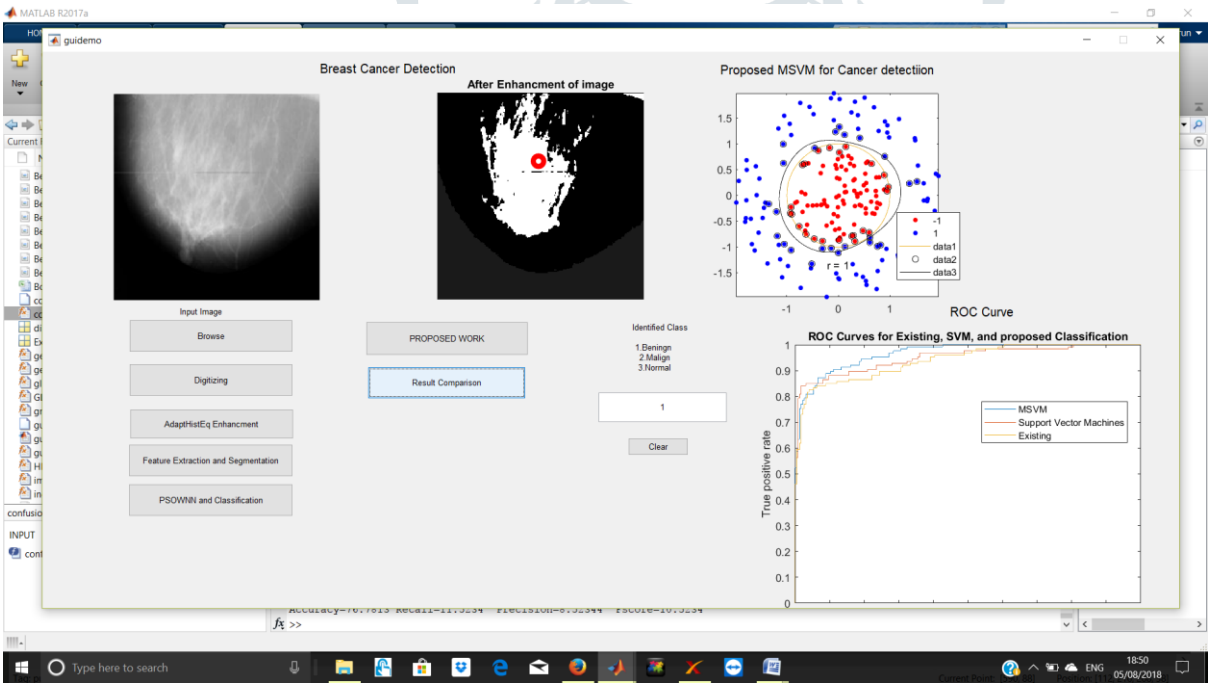

Figure 5: Simulation result using benign1. Jpg

7. Conclusion and Future Scope

Breast cancer detection mechanism is implemented through the proposed literature. The mechanism used includes hybridization of Genetic algorithm with decision tree approach. The result indicates betterment in terms of classification accuracy.

- Classification accuracy is improved through the proposed hybridization
- Feature extraction through genetic approach fetches features with precision from the dataset images
- Supervised learning ensures accurate classification but new image requires training process to be performed again.

- Precision indicates specific area of corruption accurately identified through this system.
- Comparison with existing literature shows worth of the study.

In current thesis classification accuracy is improved. The improvement however suffers setback in case new image is introduced. The rate at which result is obtained in case of complex image is slow. In the future overlapping pixel elimination mechanism can be used along with GA and Decision Tree to improve performance further.

8. REFERENCES

[1] A. K. Gupta, "Speckle Noise Reduction Using Logarithmic Threshold Contourlet," pp. 291–295, 2013.

[2] A. Sri Krishna, G. Srinivasa Rao, and M. Sravya, "Contrast Enhancement Techniques Using Histogram Equalization Methods on Color Images With Poor Lightning," *Int. J. Comput. Sci. Eng. Appl.*, vol. 3, no. 4, pp. 15–24, 2013

[3] A. Masood and A. A.- Jumaily, "Semi - advised Learning Model for Skin Cancer Diagnosis based on Histopathalogical Images," pp. 631–634, 20

[4] A. Taeb, S. Gigoyan, and S. Safavi-Naeini, "Millimetre-wave waveguide reflectometers for early detection of skin cancer," *IET Microwaves, Antennas Propag.*, vol. 7, no. April, pp. 1182–1186, 2013.

[5] A. Shenbagarajan, V. Ramalingam, C. Balasubramanian, and S. Palanivel, "Tumor Diagnosis in MRI Brain Image using ACM Segmentation and ANN-LM Classification Techniques," *Indian J. Sci. Technol.*, vol. 9, no. 1, Feb. 2016.

[6] B. Deepa, "Comparative Analysis of Noise Removal Techniques in MRI Brain Images," no. 2, 2015.

[7] C. Tippanna Madiwalar, S. K. Babu, R. K. B, and V. K. R, "Compression Based Face Recognition Using Dwt and Svm," *An Int. J.*, vol. 7, no. 3, pp. 444–449, 2016[8] N. V. S. Malothu Nagu[1], "Image De-Noising By Using Median Filter and Weiner Filtering," *Int. J. Innov. Res. Comput. Commun. Eng.*, pp. 5641–5649, 2014.

[8] D. Regularization, R. H. Chan, C. Ho, and M. Nikolova, "Salt-and-Pepper Noise Removal by Median-type Noise Detectors and," pp. 1–14.

[9] D. J. Sawicki and W. Miziolek, "Human colour skin detection in CMYK colour space," *IET Image Process.*, vol. 9, no. 9, pp. 751–757, 2015.

[10] E. J. Leavline, D. A. Antony, and G. Singh, "Salt and Pepper Noise Detection and Removal in Gray Scale Images : An Experimental Analysis," vol. 6, no. 5, pp. 343–352, 2013

[11] J. Ram, "Ship Detection Based on SVM Using Color and Texture Features," pp. 343–350, 2015.

[12] J. C. Kavitha and A. Suruliandi, "Texture and color feature extraction for classification of melanoma using SVM," *2016 Int. Conf. Comput. Technol. Intell. Data Eng. ICCTIDE 2016*, 2016

[13] K. Gu, G. Zhai, S. Wang, M. Liu, J. Zhoi, and W. Lin, "A general histogram modification framework for efficient contrast enhancement," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 2816–2819.

[14] M. A. Farooq, M. A. M. Azhar, and R. H. Raza, "Automatic Lesion Detection System (ALDS) for Skin Cancer Classification Using SVM and Neural Classifiers," *2016 IEEE 16th Int. Conf. Bioinforma. Bioeng.*, pp. 301–308, 2016.

[15] N. Shimkin, "The Wiener Filter," *Estim. Identif. Dyn. Syst.*, no. 48825, pp. 1–15, 2009.

[16] N. Shimkin, "The Wiener Filter," *Estim. Identif. Dyn. Syst.*, no. 48825, pp. 1–15, 2009.

[17] O. Abuzaghleh, B. D. Barkana, and M. Faezipour, "Noninvasive real-time automated skin lesion analysis system for melanoma early detection and prevention," *IEEE J. Transl. Eng. Heal. Med.*, vol. 3, no. March, 2015.

[18] P. Punhani, "Noise Removal in MR Images using Non Linear Filters," 2015.

[19] P. G. Jose, S. Chatterjee, M. Patodia, S. Kabra, and A. Nath, "Hash and Salt based Steganographic Approach with Modified LSB Encoding," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 4, no. 6, pp. 2257–2263, 2016.

[20] R. Balakrishnan, "Brain tumor diagnosis from MRI feature analysis - A comparative study," in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2015, pp. 1–4.

[21] S. Khosravi, "A Hybrid Neural Network Using ICA and CGA for Skin Detection in RGB Images," 2016.

[22] S. Caorsi and C. Lenzi, "Skin artifact removal technique for breast cancer radar detection," *Radio Sci.*, vol. 51, no. 6, pp. 767–778, 2016.

[23] S. Bianco, F. Gasparini, and R. Schettini, "Adaptive skin classification using face and body detection," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4756–4765, 2015.

[24] T. Xu, Y. Wang, and Z. Zhang, "Pixel-wise skin colour detection based on flexible neural tree," *Image Process. IET*, vol. 7, no. April, pp. 751–761, 2013.

[25] T. I. and S. Abe, "Fuzzy Support Vector Machines for Pattern Classification," no. 1, 2001

[26] T. Nadu and T. Nadu, "Detect Breast Cancer using Fuzzy C means Techniques in Wisconsin Prognostic Breast Cancer (WPBC) Data Sets," vol. 2, no. 5, pp. 614–617, 2013.

[27] V. G. Moruzzi, "An Inpainting Technique Based on Regularization to Remove Bleed-Through from Ancient Documents I . Gerace , C . Palomba Dipartimento di Matematica e Informatica Universit ` a degli Studi di Perugia A . Tonazzini Istituto di Scienza e Tecnologie dell ' Inf

[28] Y. Ma, D. Lin, B. Zhang, Q. Liu, and J. Gu, "A Novel Algorithm of Image Gaussian Noise Filtering based on PCNN Time Matrix," in *2007 IEEE International Conference on Signal Processing and Communications*, 2007, pp. 1499–1502.