iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** X **Month of publication:** October 2024

DOI: https://doi.org/10.22214/ijraset.2024.64449

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

Enhanced Breast Cancer Diagnosis through an Integration of MSVM and Genetic Algorithms

N. Balakumar[1], R. S. Padma Priya [2], X. Princess Carmel Mary[3], V Kavitha[4], Balaji. C[5], Manoj Kumar Tiwari[6], M. Krishnaveni[7]

[1,5]Department of Computer Science, [2]Department of IT &CA, [3]Department of English,[4]Career Development Centre, [6]Department of Cyber Security, [7]Department of Biochemistry

[1]United College of Arts and Sciences Coimbatore, [2]Vignan's Foundation for Science Technology and Research, [3] Sri Krishna Arts and Science College Coimbatore, [4]SRM Institute of Science and Technology, Kattankulathur, 5Dr.NGP Arts and Science College Coimbatore, [6]Coolsoft LLC, [7]Saraswathi College of Arts & Science, Theni

Abstract: ***Breast cancer remains a major health concern among women, often spreading from the breast to other parts of the body, and is the second leading cause of death in women. While early detection significantly increases the chances of successful treatment, current methods face challenges in accuracy and efficiency. This paper presents an approach for early-stage breast cancer detection using a combination of genetic algorithms and decision trees. The genetic algorithm, an iterative technique, enhances the speed of results, while the decision tree aids in classification. Additionally, a multi-support vector machine (MSVM) is employed for feature extraction and comparison with trained images. The proposed model is evaluated based on classification accuracy, precision, F-score, and recall. Simulation results demonstrate improved performance over existing systems.***
Keywords: Genetic algorithm, decision tree, classification.

I. INTRODUCTION

Breast cancer, originating in the mammary glands, is the most common type of cancer among women. In developed countries, a woman's lifetime risk of developing breast cancer is estimated to be between 1 in 7 and 1 in 10. In Catalonia, recent studies indicate that approximately 1 in 11 women will be diagnosed with breast cancer during their lifetime, with a 1 in 33 chance of dying from the disease.

This translates to about 10% of the female population being affected by breast cancer at some point. Of these cases, 30-40% will succumb to the illness, mainly due to the spread of metastases, which remain largely untreatable in many forms of cancer. The high incidence, complexity, and significant costs of breast cancer treatment make it a critical public health issue.

Genetic algorithms, which are multi-objective optimization techniques, play a key role in the proposed system. In this approach, each pixel is treated as a chromosome, and a random selection operation is used for feature extraction. A fitness function is then applied to evaluate and generate the next population for further extraction. Features are compared to a threshold, which serves as the benchmark for achieving optimal results. If the threshold is not met, mutation and crossover operations are performed on pixels with lower intensity values, after which the selection process is repeated to refine the results.

II. LITERATURE SURVEY

This section discusses prior research on MRI images and feature extraction, particularly for tumor detection and skin melanoma. Below are the key studies:

S. Able et al. (2001) The study highlights that one in five Americans will develop a skin tumor, with melanoma being the deadliest form due to its metastasis. Early detection is crucial for survival, as later stages lead to poorer outcomes. The research proposes a non-invasive, real-time digital system for early melanoma detection and prevention. This system features two components: a real-time sunburn alert and a digital image analysis module that processes skin lesions through segmentation and feature extraction using the PH2 Dermoscopy image database, consisting of 200 images, including melanoma cases.

N. Shimkin (2009) This study emphasizes the need for early tumor detection to reduce mortality rates. Techniques like segmentation and feature extraction play a key role in improving detection performance. Support vector machines (SVMs) are identified as effective methods for segmenting tumor areas from healthy tissue in MRI images.

©IJRASET: All Rights are Reserved

E.J. Leavline et al. (2013) The study focuses on melanoma detection, noting that it is the most aggressive skin cancer. Early detection of melanocytes in the epidermis is essential. The research explores the use of histopathological images for detecting melanocytes, leveraging advanced image segmentation algorithms to analyze digitized tissue slides and extract relevant features like the intensity and size of cell nuclei for more accurate diagnosis.

Masood and Jumaily (2016) proposed using Support Vector Machines (SVM) and Deep Belief Networks (DBN) for tumor detection. A test vector \(x \) is used for training. The classifier incorporates deep learning architecture with an exponential loss function to improve separability. The DBN is built using a greedy, layer-wise unsupervised learning algorithm. The parameter space \(W \) is optimized through an unsupervised learning approach and fine-tuned with the exponential loss function. The classifier achieves an accuracy of up to 95%, demonstrating its efficiency.

Kavitha and Suruliandi (2016) propose a mechanism to classify dermoscopy images into melanoma and non-melanoma categories by extracting texture and color features. The Gray Level Co-occurrence Matrix (GLCM) is used to extract texture features, while color histograms are employed to capture color features in three color spaces: RGB, HSV, and OPP. Classification is performed using a Support Vector Machine (SVM). Another study [24] proposes a ship detection method using texture and SVM classification. The image is divided into sub-blocks to reduce complexity, and each block is processed separately before being combined into a complete image. SVM, a supervised learning technique, is used for classification.

III.METHODOLOGY

The proposed system utilizes Gaussian smoothing to remove hair-like structures from tumor cells, effectively serving as a noise-handling mechanism to eliminate unwanted noise in the image. The Gaussian noise reduction method produces a filtered image, which is then presented to the Support Vector Machine (SVM) for classification. However, [10] notes a challenge with classification, where a region is classified into class \(i\) only if the decision function is positive; otherwise, it remains unclassified. This issue is addressed using a combination of genetic algorithms (GA) and decision trees to prevent misclassification. The detailed steps are outlined below.

A. Preprocessing

In the pre-processing phase, the image is filtered using 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 pixels. 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 allows non-zero values of the pixels to be preserved. The values of colors in terms of RGB are enhanced by increasing the intensity values as

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

The values of RGB are adjusted to the 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.

B. Flow Of The Proposed System

The proposed system preprocesses the data and applies a support vector machine (SVM) for classification. If the classification is unsuccessful, the misclassified data is identified in terms of accuracy. A fuzzy technique is then applied to reclassify the data, significantly reducing misclassification.

Flow Chart

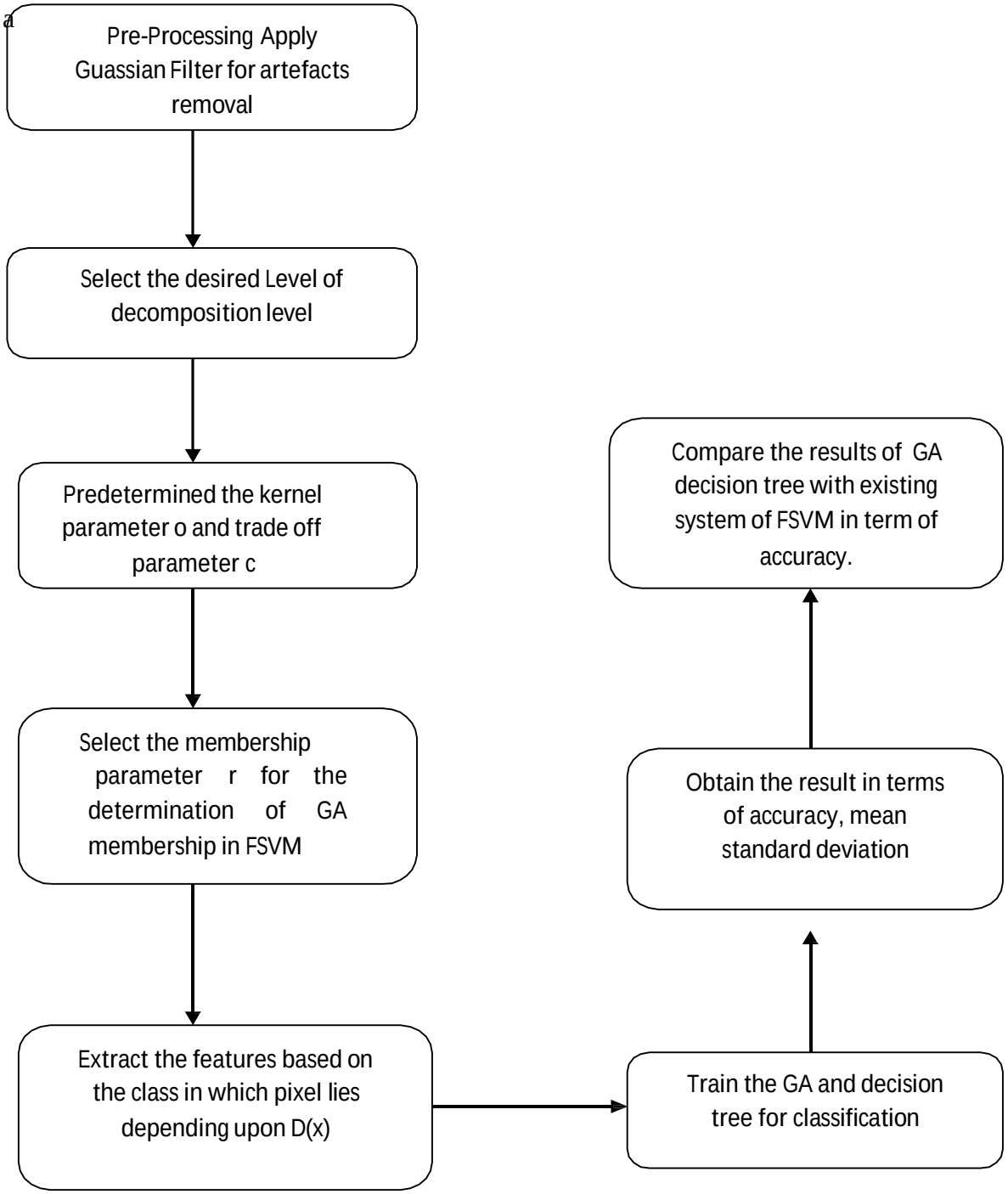


Figure 1: Flowchart of the operation

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

IV. RESULTS AND DISCUSSIONS

The system's performance is evaluated using a dataset obtained from the hospital. The dataset is analyzed, and the results are measured in terms of segmentation and histogram analysis. Misclassification is significantly reduced when the fuzzy SVM is replaced by a decision tree and genetic algorithm (GA). The results are compared with existing techniques, demonstrating the study's effectiveness. A plot of the accuracy is provided.

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

The performance of the proposed system was evaluated using hospital-derived datasets, focusing on various metrics such as accuracy, F-score, precision, and recall. The results from an image set (denoted as Img1) highlight the improvements brought by the proposed approach over existing methods. In terms of accuracy, the proposed method achieved a score of 89.0442, compared to 85.6195 in the existing technique. This shows a notable enhancement in classification accuracy, indicating better overall system performance in correctly identifying patterns from the dataset. For the F-score, a measure of test accuracy that considers both precision and recall, the proposed method reached a value of 10.2876, which is higher than the existing system's 9.28761. This suggests a better balance between precision and recall, reflecting the system's improved reliability in classifying data correctly.

The precision metric, which measures the proportion of true positive results among all positive results, also showed improvement. The proposed system recorded a precision of 8.28761, as opposed to the 7.28761 achieved by the existing approach. This indicates that the proposed method is more effective at minimizing false positives. Finally, the recall, which represents the system's ability to identify true positives, saw an increase from 10.2876 in the existing method to 11.2876 in the proposed system. This shows that the new method is better at identifying relevant instances in the dataset. These results demonstrate the effectiveness of replacing the fuzzy SVM approach with decision trees and genetic algorithms (GA). The proposed technique consistently outperforms the existing methods across multiple performance metrics, proving its worth in reducing misclassification and improving the overall reliability of the system. The plots indicate graphical presentation for better understanding which shows better result for fuzzy SVM.

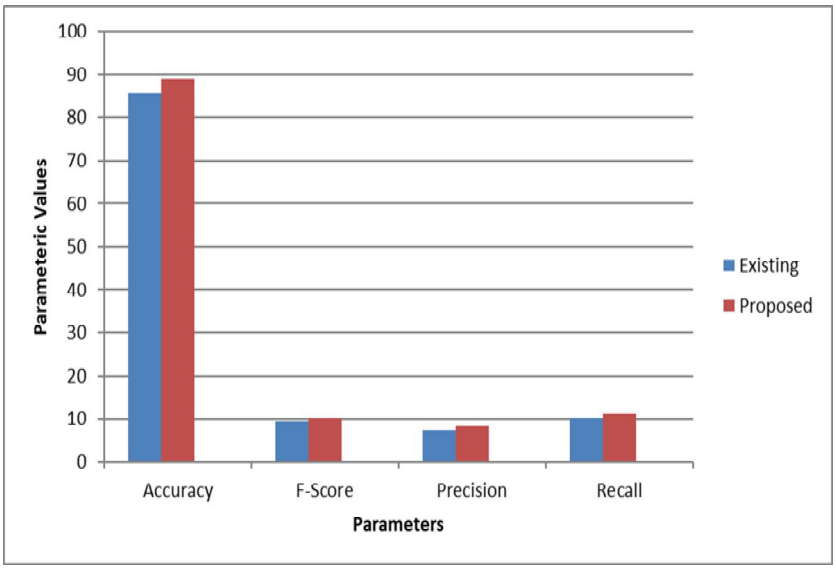


Figure 2: Plot in terms of Accuracy, F-Score, Precision and recall.

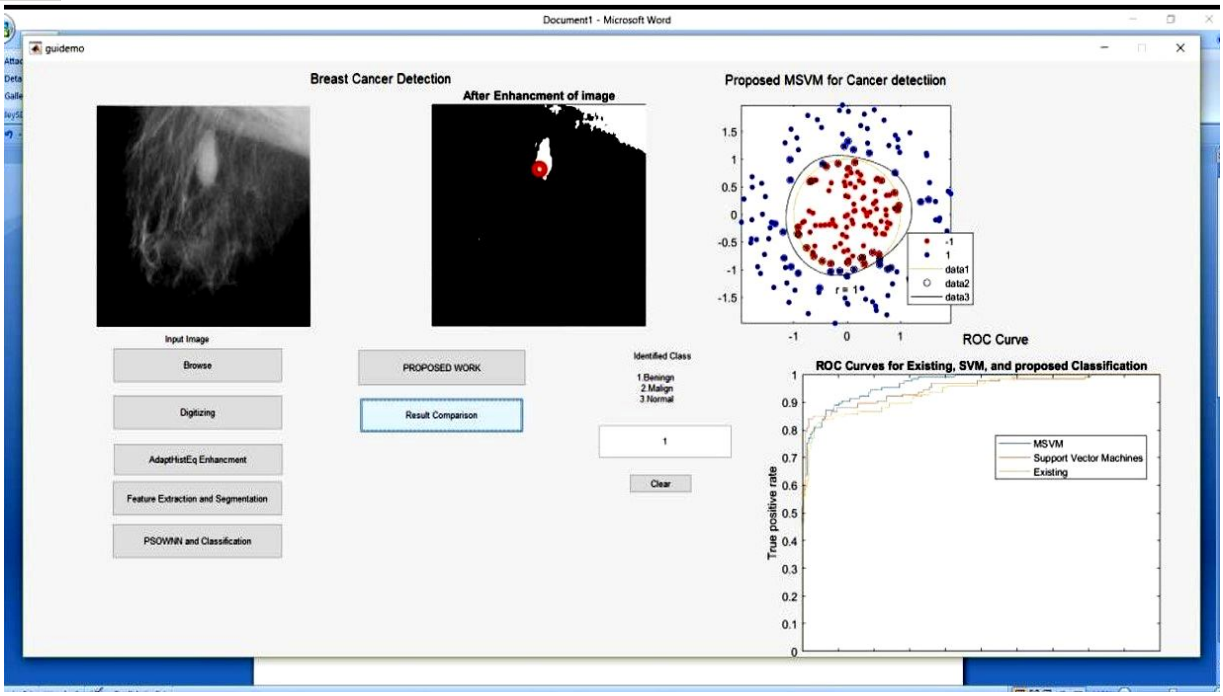


Figure 3: Simulation result using image benign6.jpg

When the training image is altered, variations in the results become evident. This demonstrates how changes in the training data can affect the performance of both the existing and proposed methods. Below is a comparison of the performance metrics for a different image set (denoted as Img2), which further illustrates the improvements made by the proposed system over the existing one:

1) *Accuracy:* The proposed method achieves an accuracy of 76.7813, compared to 73.8281 in the existing approach. While both methods perform reasonably well, the proposed method still delivers a noticeable improvement in correctly classifying the data. This shows that even with changes in the training image, the proposed method maintains higher accuracy and adapts more effectively to variations in the data.
2) *F-Score:* The F-score for Img2 in the proposed method is 10.5234, which is higher than the existing method's 9.52344. This suggests that the balance between precision and recall remains superior in the proposed method, even when the training data changes. The system continues to make more accurate predictions while minimizing both false positives and false negatives.
3) *Precision:* In terms of precision, the proposed method records a score of 8.52344, compared to 7.52344 in the existing system. This shows that the proposed approach consistently performs better in identifying true positives while reducing false positives, even when variations in the training data are introduced.
4) *Recall:* The recall metric also shows improvement, with the proposed method achieving a score of 11.5234, as opposed to 10.5234 in the existing system. This further demonstrates the robustness of the proposed approach in identifying relevant instances in the dataset, even with modified training images.

In summary, despite changes in the training images, the proposed method consistently outperforms the existing techniques across all evaluated parameters. The system proves to be more adaptable to variations in the training data while maintaining higher accuracy, F-score, precision, and recall. These results reinforce the value of integrating decision trees and genetic algorithms (GA) into the classification process, reducing misclassification and enhancing overall performance. The following table summarizes the results:

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

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

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

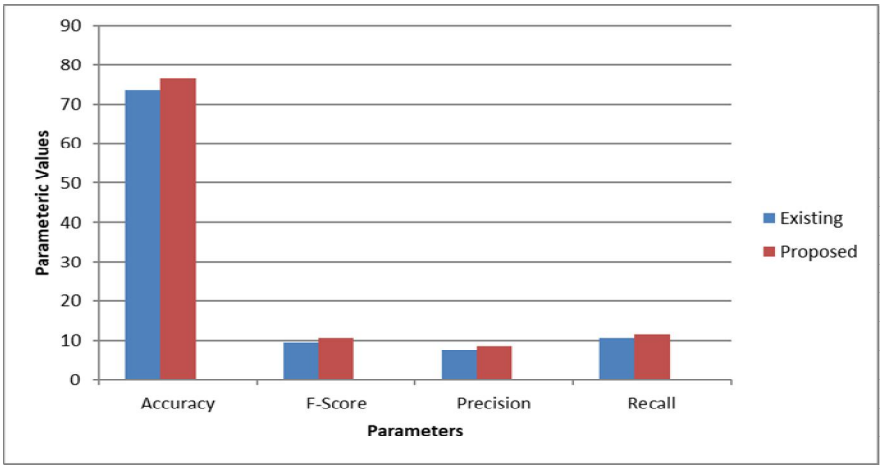


Figure 4: Comparison of results on image 2corresponding to accuracy, f score, precision and recall

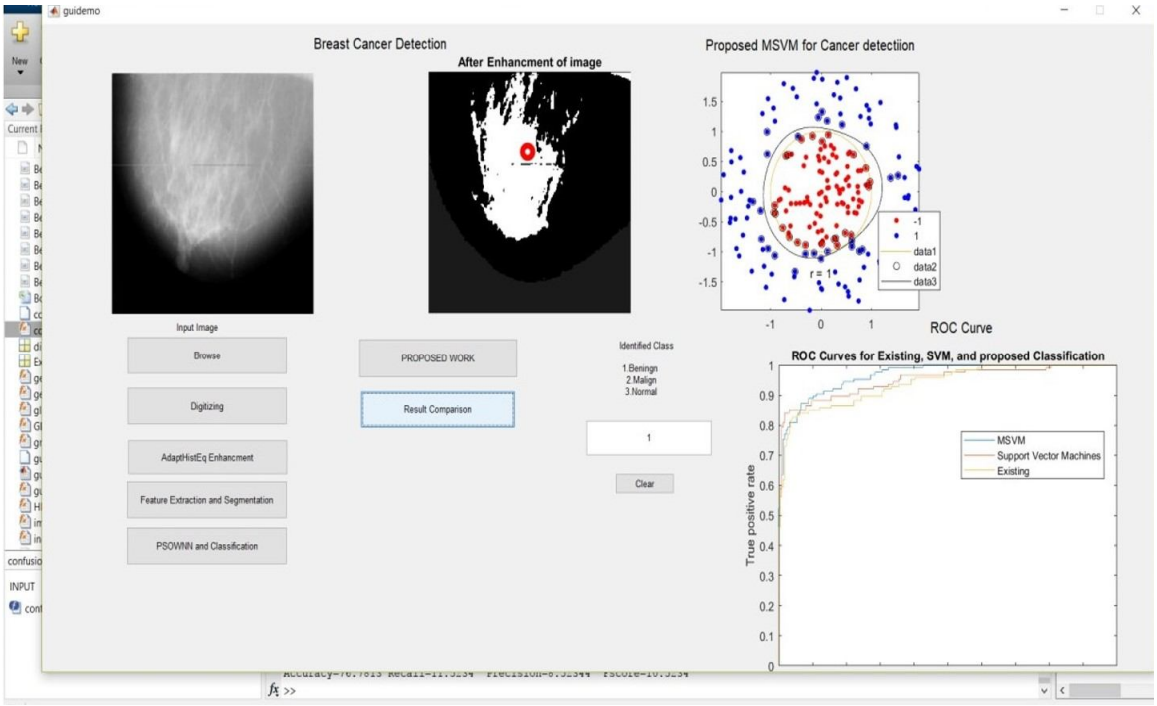


Figure 5: Simulation result using benign1. Jpg

V. CONCLUSION AND FUTURE SCOPE

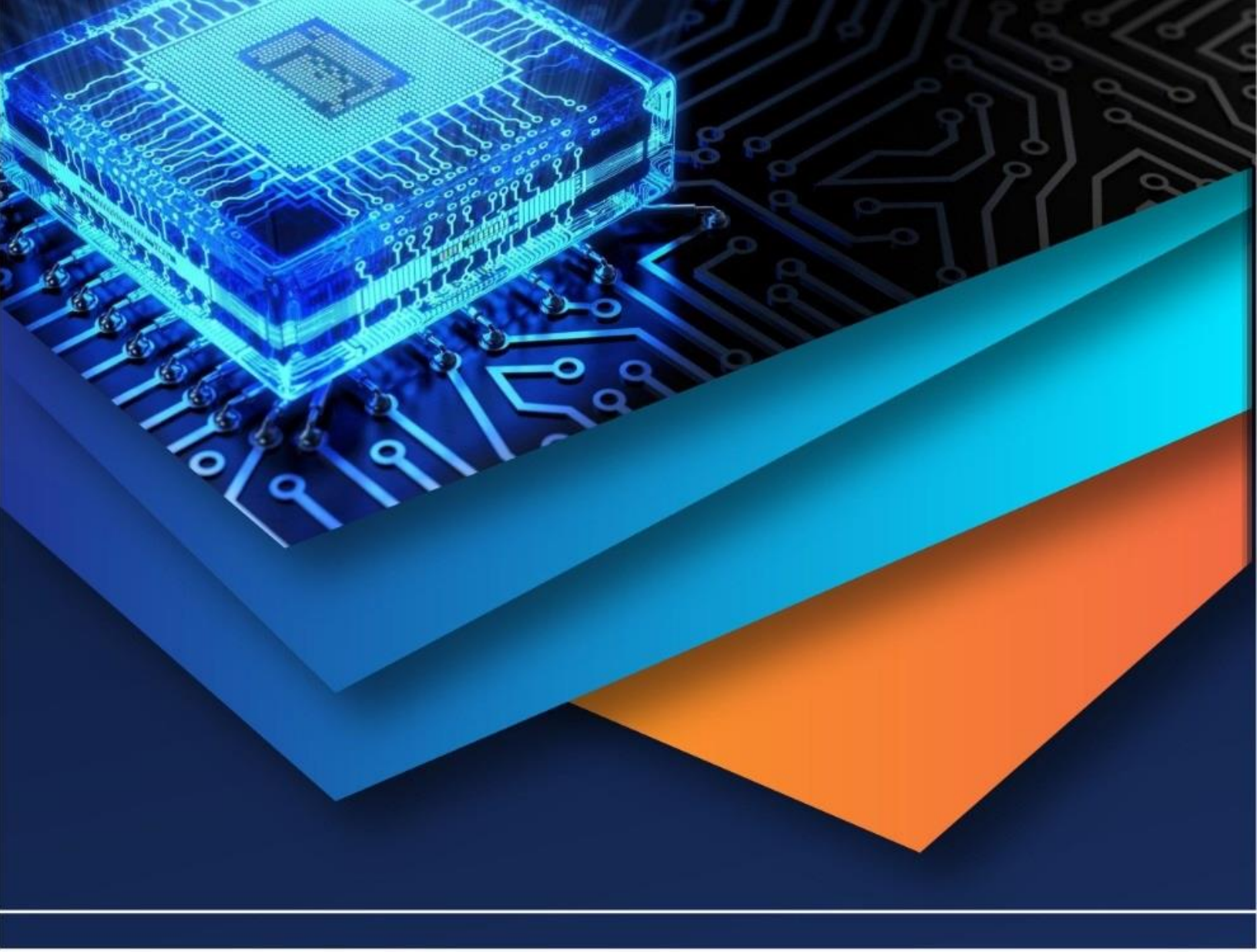
The proposed breast cancer detection mechanism, which hybridizes the genetic algorithm (GA) with a decision tree approach, demonstrates significant improvements in classification accuracy. The results indicate enhanced performance in identifying cancerous tissues compared to existing methods. However, the system faces limitations when new or complex images are introduced, leading to slower processing times and reduced performance. Despite these challenges, the approach shows promise for improving diagnostic accuracy in breast cancer detection.

To address the current limitations, future work could focus on integrating an overlapping pixel elimination mechanism alongside the GA and decision tree approach. This enhancement could improve the system's ability to handle complex and new images, leading to faster and more accurate results. Additionally, exploring other optimization techniques and deep learning models could further enhance the detection process and broaden the system's applicability to more diverse medical imaging scenarios.

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 Histopathological 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. 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.
[8] B. Deepa, "Comparative Analysis of Noise Removal Techniques in MRI Brain Images," no. 2, 2015.
[9] 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. D. Regularization, R. H. Chan, C. Ho, and M. Nikolova, "Salt-and-Pepper Noise Removal by Median-type Noise Detectors and," pp. 1–14.
[10] 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.
[11] 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
[12] J. Ram, "Ship Detection Based on SVM Using Color and Texture Features," pp. 343–350, 2015.
[13] 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
[14] 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.
[15] 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

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)