

Computer Aided Detection Of Breast Cancer Using Bio Inspired Algorithm

Syed Rizwana, Peddi Kavya, Bolla Vinay Pooja, Bandi Poojitha, Shaik Khaja Mohiddin Basha, Dodda Venkatareddy

Department of Computer Science and Engineering, Narasaraopeta Engineering College, Narasaraopet, Palnadu, India.

Email: rizwana.nec@gmail.com

Abstract—Breast cancer still ranks among the most common causes of cancer-related deaths among women, hence the call for early diagnosis. Mammography is the most accepted screening test, but conventional computer-aided detection (CAD) has a high false positive rate (FPR) that gives rise to biopsy and false negatives (FN) where cancer is undetected. In solving these challenges, this paper provides a solution by employing the use of the Simple Genetic Algorithm (SGA), which is openly inspired by biological systems to enhance the performance of CAD systems for breast cancer detection. The SGA, which is based on the evolutionary process, can resolve problems in feature selection and classification of the mammogram by overcoming shortcomings of pattern recognition. By mimicking the genetic evolution process, ant colony optimization, and swarm intelligence, the SGA prevents noisy or variant images from anyhow decreasing the detection accuracy. Comprehensive tests on typical sets of mammograms confirm the effectiveness of the proposed approach regarding a twofold reduction of inappropriate positive and negative results. This enhanced accuracy of diagnoses can help radiologists to act early, combined with favorable outcomes for the patients, implying that early diagnosis may save lives.

Index Terms—Breast Cancer Detection, Genetic Algorithm (GA) Feature Selection Guided Filter Image Enhancement Convolutional Neural Networks (CNN) Neural Networks Image Preprocessing Medical Image Analysis

I. INTRODUCTION

Breast cancer is the leading cancer among women worldwide, of which it accounts for 25 % of all cancers and 685,000 fatalities yearly [1]. The time factor proves central to increasing the survival rate of performing minimally invasive procedures. Nevertheless, there are problems with conventional diagnostic methods even with the use of mammography, which is widely used in current practice. False positive results increase the number of biopsies that are done; false negatives delay treatments and shorten patients' lives. Further, current CAD systems are limited, especially in noisy image data, and do not have sufficient high cross-clinical variability generalization. In order to deal with these challenges, this work proposes a CAD framework based on the bio-inspired techniques, and more specifically, on the SGA. These algorithms are bio-inspired from the evolutionary biology in order to select features and classify mammography. Using guided image enhancement for preprocessing, LBP for feature extraction, and SGA for selecting features, the proposed approach provides a resolution to improve the reliability of diagnoses on a large scale. This work provides an organized methodology to approach the breast cancer detection problem and to eliminate false positives and

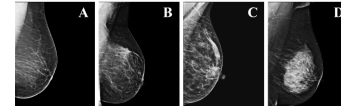


Fig. 1: Examples of the BI-RADS's four mammography density levels are as follows: From (A) a—virtually totally fatty; (B) b—fibroglandular density in scattered places; (C) c—heterogeneously dense; (D) d—very dense. [2]

negatives, increase classification performance, and overcome the drawbacks of current CAD systems. As a part of the continuous breakdown of the differences between research and practice concerning medical image analysis, our paper is primarily focused on the potential practical applications resulting from our findings.

Some well-known categories of bio-inspired algorithms encompass genetic algorithms (GA), particle swarm optimization (PSO), ant colony optimization (ACO), and artificial bee colony (ABC), which are derived from the mechanisms of living organisms or their parts. Therefore, in this work, the focus is on using bio-inspired algorithms in general and, more specifically, the Simple Genetic Algorithm (SGA) [3]. Genetic algorithms are overall a simulation of the natural procedure of evolution through selection, crossover, and mutations. Besides, they are very effective when it is required to search for the best solutions through large solution spaces. A simple genetic algorithm was preferred for this project because it is easy to implement, very stable, and provides good results for feature selection and classification. In this work, the SGA is effectively applied for the selection of the appropriate features from the images of mammograms for classification for enhancing the method of detection of breast cancer. These types of applications show that SGAs are highly proficient in increasing diagnostic reliability in medical image analysis in order to become the perfect tool for increasing cancer detection in mammograms. As noted in Araujo et al. (2017), the authors have also used genetic algorithms for classification of histology images for breast cancer. Second, the fitness function would be defined in the context of medical image analysis and classification applications only so that genetic feature selection [4] for TI would not be performed on the dataset while carrying out the proposed algorithm for designing the feature selection [5] procedure for mammogram images.

OBJECTIVES OF OUR WORK

- **Preprocessing: Guided Image Enhancement**
Enhance the mammogram images and, in the process, come up with better images that can aid in the detection of the diseases.
- **Feature Extraction: Local Binary Patterns (LBP)**
Propagation of the determined LBP to improve the features that are taken out of the mammography images to improve the likelihood of accurate categorization.
- **Feature Selection: Simple Genetic Algorithm (SGA)**
Using SGA for feature selection helps to improve the results but at the same time reduces the number of false positives and false negatives.

II. RELATED WORK

Del Ser, E. Osaba, D. Molina, X.-S. Yang, S. Salcedo-Sanz, D. Camacho, et al. [6] proposed “Bio-inspired computation: Where we stand and what’s next.” The bio-inspired computation has made research advances in the last few years, specifically in optimization, classification, and clustering. Thus, based on the bio-inspired algorithms, the primary advantages state that the method is very appropriate in approaching the solving of different complicated problems through flexibility and reliability in attaining the efficient solutions. Yousefi et al. [7] in their paper in 2021 employed a sparse deep convolutional autoencoder for early BC detection using dynamic thermography. This work is related to prior work on deep learning algorithms in medical images, autoencoders, reviews of dynamic thermography in detecting breast cancer, and sparse deep learning for feature extraction. Hirra et al. (2021) [8] apply a patch-based deep learning technique to distinguish between breast cancer types in histopathological images. Similar works encompass the works done on histopathology image analysis and cancer detection with deep learning algorithms by Cruz-Roa et al. (2017) and Cohen et al. on the use of a cheaper computer vision method for breast cancer diagnosis. Ronneberger et al. (2015) designed the U-Net architecture as part of their work. In the medical image segmentation domain, early cancer detection: A review on cost-effective methods by Zhao and Zhang (2019). Deep learning is applied in medical image diagnosis according to Rajpurkar et al. (2017), which is related to the diagnosis of breast cancer. Sinha and Singh (2020) discuss inexpensive imaging techniques, which are also cost-effective, much like the study by Sethy et al. [9]. Furthermore, Gong et al. (2018) offer a systematic review of computer vision for medical diagnosis, which is related to the paper under discussion. Nahid, Mikaelian & Kong (2018) [10] employ a restricted Boltzmann machine with a backpropagation algorithm to identify histopathological breast images.

III. PROPOSED METHODOLOGY

A. Preprocessing

The first phase of the methodology we proposed is to improve the quality of mammogram images using the guided image enhancement technique. It enhances signal contrast, the

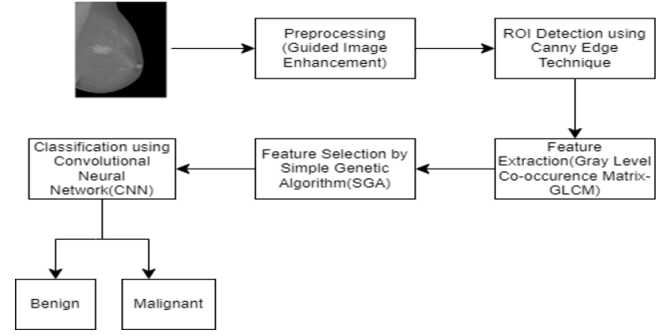


Fig. 2: Flowchart of Breast Cancer Detection

organizational borders of the images, and eliminates background noise, specifically in regions of interest (ROI), like possible tumors, which is shown in fig 3. This is important, for it enables subsequent steps in the CAD system to correctly identify abnormal regions. Further, we also employ the Contrast Limited Adaptive Histogram Equalization (CLAHE), where the contrast of regions in the image spatially varies, and it works well for spotting tiny details such as microcalcifications. This step is very vital in improving the quality of the image section as a way directed towards accommodating the various features extraction and processing. The preprocessing is to make the mammogram images clear, and the ROI detection is required to use the Canny Edge Detection to detect the likely suspicious parts. Subsequently, Gabor filtering is applied to capture edges, and next, feature extraction is done by using the Gray Level Co-occurrence Matrix (GLCM), which is used for capturing texture information. Finally, a simple genetic algorithm (SGA) is used to choose the highest discriminant features for a more enhanced classification.

B. ROI Edge Detection Using Canny Edge Technique

Canny edge detection is one of the essential methods used in the process of possibly separating the boundary of a tumor in the mammogram images. The next step in the considered approach is Canny edge detection, which helps to detect the contours of the tumors owing to intense changes in the image intensities. This step allows one to select the necessary areas that in all probability contain tumors for further work. These edges constitute the boundaries of the Regions of Interest (ROIs) that, once obtained, are subjected to feature extraction and classification. Thus, it can be concluded that, by providing correct definitions of these areas, Canny edge detection contributes significantly to improving the accuracy of stage detection in the overall deep learning stages.

C. Feature Extraction Using Gray Level Co-occurrence Matrix (GLCM)

In this step, the Gray Level Co-occurrence Matrix (GLCM) is used in the feature extraction of useful features within the regions of interest. GLCM studies the relationship of the intensity of two pixels and provides the number of occurrences when a given distance and direction are used. This technique

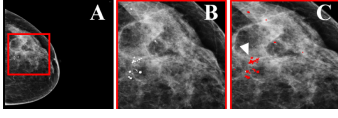


Fig. 3: A mammogram reveals a woman with moles that are both benign and worrisome. (B) The zoomed region that shows the MCs, denoted by the red square in (A). (C) The region in (B) where two extremely skilled radiologists have painstakingly highlighted the locations of the MCs (shown by the white arrow) [11]

allows for achieving quantitative measurements of secondary texture characteristics, including contrast, correlation, energy, and homogeneity, through which important data about the structure of the tissue is obtainable. In light of these advantages, we improve upon the model’s capability to distinguish between benign and malignant regions seen in mammogram images and further improve the detection rate for cancer.

D. Feature Selection with Simple Genetic Algorithm (SGA)

Here we apply the simple genetic algorithm (SGA) for the best feature selection. The SGA operates in the same way as the natural evolution when it chooses the most useful features in the dataset. In its successive generations, such an algorithm assesses existing options of features and refines them by excluding non-conveying features. This process not only helps in simpler manipulation of the dataset but also in enriching the feature space for the model by corresponding with the most informative and significant features to help achieve better accuracy in the presaging of benign and malignant regions in mammogram images.

ALGORITHM: SIMPLE GENETIC ALGORITHM (SGA)

Input: Dataset with extracted features.

Output: Optimal feature subset for classification.

- 1) Initialization: Create a set of potential feature subsets.
- 2) Evaluation: Each subset is evaluated using a classifier to determine its performance.
- 3) Selection: Choose the subsets with higher accuracy based on observations.
- 4) Crossover: Combine selected feature subsets to create new subsets.
- 5) Mutation: Randomly alter some feature subsets to introduce diversity.
- 6) Replacement: Generate the next generation by reproducing from the current population.
- 7) Termination: Stop the algorithm when a termination condition is met and select the best feature subset.

E. Classification Using Convolutional Neural Network (CNN):

To extend this process, we use a Convolutional Neural Network (CNN) [12] classification technique for mammograms. The CNN then uses the descriptive features existing in the Regions of Interest (ROIs) to obtain the right patterns and representations to enable classification between benign and malignant classes. By having numerous layers of convolution,

the proposed network recognizes multi-level features to extract the images’ patterns effectively.

TABLE I: Hyperparameter Settings for the Model

Hyperparameter	Value
Optimizer	Adam
Loss function	Binary Cross-Entropy
Batch Size	32
Epochs	11
Learning Rate	0.001
GA Feature Selection	622 features
GA Generations	11

Optimization and Training: This neural network was optimized with Adam Optimizer because it self-regulates the learning rate, thus making the process of training more efficient and faster. To quantify and calculate differences between the predicted and factual classification during the model training stage, a binary cross-entropy loss function was implemented. The overall process is explained in fig 2.

IV. EXPERIMENTAL SETUP

A. Dataset Description:

The dataset in use in this current research study is the DDSM, commonly known as the Digital Database for Screening Mammography.

TABLE II: Dataset Details for Breast Cancer Classification

Layer	Configuration
Dataset	DDSM (Digital Database for Screening Mammography)
Total Images	13,215
Training Split	80% (10,572 images)
Testing Split	20% (2,643 images)
Class labels	Benign, Malignant
Input Layer	Neurons=622 features (from GA)
Hidden Layer 1	Neurons=256, Activation=ReLU
Hidden Layer 2	Neurons=128, Activation=ReLU
Output Layer	Neurons=1, Activation=Sigmoid

B. Implementation

The implementation starts with the procurement of the DDSM dataset, which consists of 13,215 screen-fetched images categorized into benign and malignant. It is also divided into a training set and a test set with 80% and 20%, respectively. Some preprocessing was applied to the images, which were resizing and normalization, and then we applied Canny edge detection to isolate the region of interest (ROI) for extracting features. From each image, four features are obtained using GLCM, and from all the images, a total of 622 features are selected using SGA [13]. A convolutional neural network has been used on these features, and its efficient architecture includes an input layer; two dense layers of 256 neurons that adjust the ReLU function; a third layer with 128 neurons adjusted to the ReLU function for binary categorization; and lastly, an output layer with a sigmoid function. Recall, F1-score, accuracy, and recalculated precision.

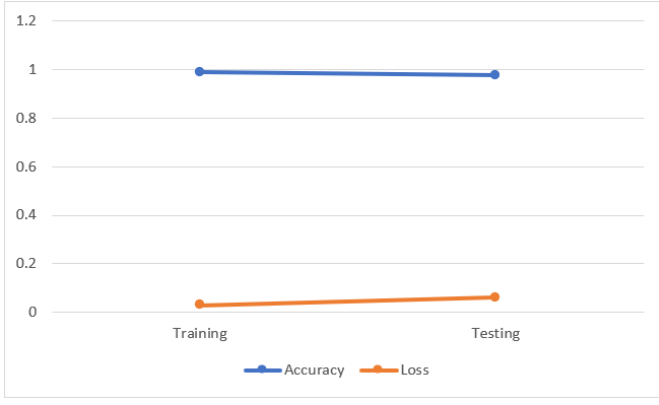


Fig. 4: Training and Testing Accuracy vs. Loss for Breast Cancer Classification Model

V. EVALUATION

We evaluated the process by using some formulas as mentioned

Loss Function (Binary Cross-Entropy): In classification tasks with binary outcomes (e.g., benign vs. malignant), the binary cross-entropy loss function is commonly used to measure the error.

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

Where:

N = No. of samples

y_i = Actual label (0 for Benign, 1 for Malignant)

p_i = Predicted probability for Malignant

AUC (Area Under the ROC Curve): In the context of a classification model, AUC holding different threshold values is applied for its evaluation. It determines the value of the curve of receiver operating characteristic (ROC).

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR})$$

TABLE III: Comparison of Feature Selection Methods in Breast Cancer Detection (2020-2023)

Reference	Year	Feature Selection Approach	Accuracy
Proposed Work	2024	Simple Genetic Algorithm (SGA)	98.88%
Bajer et al. [1]	2020	Genetic Algorithm (GA), Particle Swarm Optimization (PSO)	95%
Yousefi et al. [7]	2021	Dimensionality reduction using sparse encoding	93.5%
Hirra et al. [8]	2021	Deep learning-based feature extraction	96.2%
Sethy et al. [9]	2021	Feature extraction via CNN	94.3%
Alshayegi et al. [13]	2022	ANN-based feature extraction	96%

VI. RESULTS

Comparison of different feature selection algorithms helped in selecting features that could be suitable for mammographic image classification. Thus, among all algorithms used in

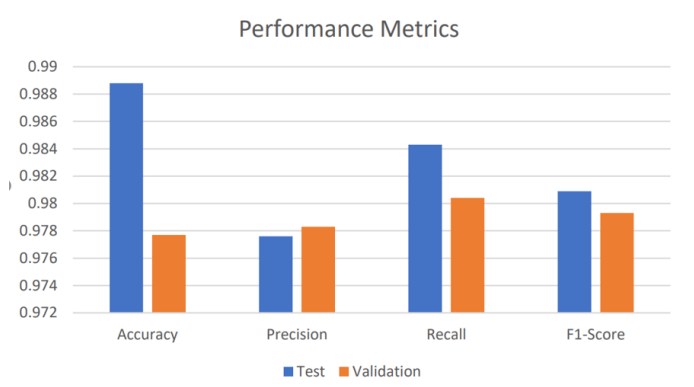


Fig. 5: Comparison of Performance Metrics for Test and Validation Data

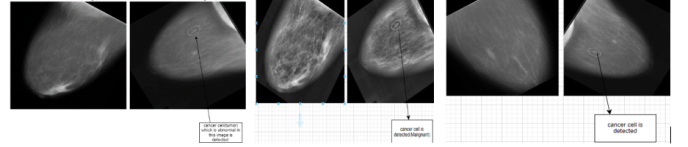


Fig. 6: Prediction of Breast Cancer Cells using Simple Genetic Algorithm

the experiment, the best result was obtained by the Genetic Algorithm (GA). That is why GA's ability to model adding evolutionary processes while selecting discriminative features produced significantly higher classification results. This was a much better improvement over Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) [14] [15]. As in the case with PSO, the algorithm offered well-rounded features with, however, less precise results than the ones found in GA. As for ACO, it was less helpful in increasing the classification accuracy, and this indicates some drawbacks of the method as the feature selection technique for the classification goals. Given the features chosen by GA, different classification models were compared, and CNN [16] provided the most accurate results. Through feature learning in mammographic images, complex patterns could be distinguished, and therefore, high classification was achieved by CNN.

Fig 5 shows the classification results based on the metrics identified as classification success rate, positive predictive value, and sensitivity with the harmonic mean of precision and recall using both the test and validation datasets. It can be seen that the model achieves good accuracy on both structured and unstructured sets, be it with a tad more recall for the test set data. F-measure and precision are quite similar in both the test and validation data, indicating acceptable performance of the method in identifying both benign and malignant cases. The evaluation assists in pointing out the feasibility of correctly diagnosing breast cancer using the chosen features and classification algorithms in the proposed model.

Figure 6 illustrates how the trained model and the Simple Genetic Algorithm (SGA) are used to classify breast cancer in its final form. At the left and right sides of the image, there are distinguishing characteristics between the benign and malignant cases, and a circle encloses the area where the

model picked out the cancerous region (tumor). Thus, using SGA for feature selection and classification, identified by the trained model, normal and cancerous tissues, thus achieving the correct identification of malignant areas in mammographic images.

VII. CONCLUSION

This paper proves that the incorporation of GA into CAD for breast cancer detection is possible, yielding good results. The use of GA has greatly improved the ability to identify regions of interest in mammography images, improving the likelihood that breast cancer will be discovered. Through GA, there has been a significant improvement in reducing false positives and thus the depression and anxiety of patients and reducing unnecessary imaging. The work stresses the benefits of the use of GA in mammographic image analysis, indicating the usefulness of this approach as compared to other CAD systems. The efficiency with which GA is capable of both searching and exploiting feature spaces has provided a boost toward increasing the diagnostic accuracy of CAD, which makes the tool vital for the continued development of CAD technologies. Moreover, applicability and flexibility for dealing with high amounts of data and different imaging environments prove the addition of GA to be valuable. This makes SGA appear as a potentially beneficial approach to improving breast cancer diagnosis accuracy as well as procedural effectiveness. Further studies must be directed at searching for other algorithms that are inspired by nature as well as improving the time consumption of the GA. Exploring other imaging modalities and other detection scenarios can improve the utilization of GA in medical sciences and extend the contributions to the development of CAD systems.

VIII. FUTURE WORK

Being under the class of heuristic algorithms, genetic algorithms (GA) hold tremendous prospects of expounding further within the different domains of study, especially in artificial intelligence and beyond. In the financial sector, it can be employed to enhance decision-making over investment, risk, and modeling. Trading algorithms, portfolio management, and fraud detection systems incorporated by the institutions can enhance decision-making as well as the overall accuracy of forecasts, thus minimizing operational risks. They also say that GA has good potential in the application of control, industrial automation, and robotics. GA is capable of influencing and optimizing sensor data analysis, control parameters, and system configuration to further improve the accuracy and flexibility of manufacturing and logistics, thereby improving the overall performance of a robotic system. In the sphere of environmental and resource management, GA can increase the efficiency of sensors, as well as data analysis of environmental and resource problems, and advance resource management methods. This can contribute towards improved utilization of natural resources, better control of pollution, and overall management and promotion of sustainable development. GA holds

potential for the telecommunication industry to achieve efficient networks, traffic management, and performance tuning. By further improving the configurations of the network as well as analyzing data, GA can improve the network throughput, reduce congestion, and therefore increase the reliability of the communications. In smart cities and infrastructure, Smart GA can contribute to traffic organization, energy provision, and other public services in general. The integration of such potential into the cities can make the cities better in terms of efficiency, sustainability, and responsiveness to the needs of the people.

REFERENCES

- [1] Bajer, D., Zorić, B., Dudjak, M., & Martinović, G. (2020). Benchmarking bio-inspired computation algorithms as wrappers for feature selection. *Acta electrotechnica et informatica*, 20(2), 35-43.
- [2] Spak, D. A., Plaxco, J. S., Santiago, L., Dryden, M. J., & Dogan, B. E. (2017). BI-RADS® fifth edition: A summary of changes. *Diagnostic and interventional imaging*, 98(3), 179-190.
- [3] Nakyeeyune, R., Ruan, X., Shen, Y., Shao, Y., Niu, C., Zang, Z., & Liu, F. (2021). Diagnostic Performance of SGA, PG-SGA and MUST for Malnutrition Assessment in Adult Cancer Patients: A Systematic Literature Review and Hierarchical Bayesian Meta-Analysis. *Nutrition and Cancer*, 74(3), 903-915. <https://doi.org/10.1080/01635581.2021.1942080>
- [4] Aalaei, S., Shahraki, H., Rowhanimanesh, A., & Eslami, S. (2016). Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets. *Iranian journal of basic medical sciences*, 19(5), 476.
- [5] Bolon-Canedo, V., & Remeseiro, B. (2020). Feature selection in image analysis: a survey. *Artificial Intelligence Review*, 53(4), 2905-2931.
- [6] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, 9(1), 12495.
- [7] Del Ser, J., Osaba, E., Molina, D., Yang, X.-S., Salcedo-Sanz, S., Camacho, D., ... & Herrera, F. (2019). Bio-inspired computation: Where we stand and what's next. *Swarm and Evolutionary Computation*, 48, 220-250. <https://doi.org/10.1016/j.swevo.2019.05.006>
- [8] Yousefi, B., Akbari, H., Hershman, M., Kawakita, S., Fernandes, H. C., Ibarra-Castanedo, C., Ahadian, S., & Maldague, X. P. V. (2021). Spaer: Sparse deep convolutional autoencoder model to extract low dimensional imaging biomarkers for early detection of breast cancer using dynamic thermography. *Applied Sciences*, 11(7), 3248. <https://doi.org/10.3390/app11073248>
- [9] Hirra, I., Ahmad, M., Hussain, A., Ashraf, M. U., Saeed, I. A., Qadri, S. F., Alghamdi, A. M., & Alfakeeh, A. S. (2021). Breast cancer classification from histopathological images using patch-based deep learning modeling. *IEEE Access*, 9, 24273-24287. <https://doi.org/10.1109/ACCESS.2021.3056842>
- [10] Sethy, P. K., Pandey, C., Khan, M. R., Behera, S. K., Vijaykumar, K., & Panigrahi, S. (2021). A cost-effective computer-vision based breast cancer diagnosis. *Journal of Intelligent & Fuzzy Systems*, 41(5), 5253-5263. <https://doi.org/10.3233/JIFS-189257>
- [11] Nahid, A.-A., Mikaelian, A., & Kong, Y. (2018). Histopathological breast-image classification with restricted Boltzmann machine along with backpropagation. *Biomedical Research*, 29(10).
- [12] Loizidou, K., Skouroumouni, G., Pitris, C., & Nikolaou, C. (2021). Digital subtraction of temporally sequential mammograms for improved detection and classification of microcalcifications. *European radiology experimental*, 5, 1-12.
- [13] Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., & others. (2017). Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE*, 12(6), e0177544. <https://doi.org/10.1371/journal.pone.0177544>
- [14] Gardezi, S. J. S., A. Elazab, B. Lei, and T. Wang, et al. 2019. Breast cancer detection and diagnosis using mammographic data: Systematic review. *Journal of medical Internet research* 21 (7):e14464. doi:<https://doi.org/10.2196/14464>.
- [15] Pritom, A. I., Munshi, M. A. R., Sabab, S. A., & Shihab, S. (2016). Predicting breast cancer recurrence using effective classification and feature selection technique. In 2016 19th International Conference on

Computer and Information Technology (ICCIT) (pp. 310–314). IEEE. <https://doi.org/10.1109/ICCITechn.2016.7860210>

- [14] Alshayegi, M. H., Ellethy, H., & Gupta, R. (2022). Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach. *Biomedical Signal Processing and Control*, 71, 1–10. <https://doi.org/10.1016/j.bspc.2021.103182>.
- [15] Tsehay, Y., et al. (2017). Biopsy-guided learning with deep convolutional neural networks for prostate cancer detection on multiparametric MRI. In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) (pp. 642–645). IEEE. <https://doi.org/10.1109/ISBI.2017.7950646>
- [16] Ting, F. F., Tan, Y. J., & Sim, K. S. (2019). Convolutional neural network improvement for breast cancer classification. *Expert Systems with Applications*, 120, 103-115.
- [17] S. Rafi and R. Das, "RNN Encoder And Decoder With Teacher Forcing Attention Mechanism for Abstractive Summarization," 2021 IEEE 18th India Council International Conference (INDICON), Guwahati, India, 2021, pp. 1-7, doi: 10.1109/INDICON52576.2021.9691681.