Early Prediction of Breast Cancer Risks Using Convolutional Neural Network

IJIMSR, Vol.1, No.1, (2023) 23.1.1.007

Dr. D. JAYA KUMARI*
Professor & HOD
Department of Computer Science & Engineering
Sri Vasavi Engineering College (Autonomous),
Pedatadepalli, Tadepalligudem, AP-534101.
Email: hod_cse@srivasaviengg.ac.in

Mr. K. PRAVEEN KUMAR
Assistant Professor
Department of Computer Science & Engineering
Sri Vasavi Engineering College (Autonomous)
Pedatadepalli, Tadepalligudem, AP-534101
Email: praveenkumar.cse@srivasaviengg.ac.in

***Corresponding author**

Received 28th August 2023; Accepted 8th September 2023
Keywords: Deep learning, Artificial intelligence, Breast cancer

Early Prediction of Breast Cancer Risks Using Convolutional Neural Network

Dr. D. JAYA KUMARI
Professor & HOD
Department of Computer Science & Engineering
Sri Vasavi Engineering College (Autonomous),
Pedatadepalli, Tadepalligudem, AP-534101.
Email: hod_cse@srivasaviengg.ac.in

Mr. K. PRAVEEN KUMAR
Assistant Professor
Department of Computer Science & Engineering
Sri Vasavi Engineering College (Autonomous)
Pedatadepalli, Tadepalligudem, AP-534101
Email: praveenkumar.cse@srivasaviengg.ac.in

ABSTRACT

Breast Cancer is one of the deadliest diseases in the world, among cancer categories, early detection is essential to better treatment outcomes and survival rates. Delay in diagnosis due to medical mistakes or incorrect interpretation of patient reports might have a negative effect on patient outcomes.Artificial intelligence (AI), a component of contemporary technology, has demonstrated encouraging potential for helping medical practitioners to increase the precision and effectiveness of cancer diagnosis. AI programs can be trained to examine several kinds of medical imaging data, including MRI and CT scans, and help spot anomalies or malignant cells. The study aims to establish a model for estimating the risk of breast cancer based on a combination of medical and behavior-related risk factors. The goal is to identify high-risk groups of patients who have undergone breast cancer surgery so that they can receive closer monitoring and more accurate screening.In this proposed work,we try to use CNN Model for classifying the risk of breast cancer.

Keywords: Deep learning, Artificial intelligence, Breast cancer

1. INTRODUCTION

One of the commonest and most serious forms of cancer afflicting women globally is breast cancer. Improving the prognosis and treatment results for those diagnosed with breast cancer requires early detection and prompt action. AI and other technologies, such as medical imaging, are crucial for the early detection and treatment of breast cancer. Here are some details about breast cancer specifically:

1.1 Breast cancer Prevalence: One of the commonest and most serious forms of cancer [1]-[3]afflicting women globally is breast cancer. Improving the prognosis and treatment results for those diagnosed with breast cancer requires early detection and prompt action. AI and other technologies, such as medical imaging, are crucial for the early detection and treatment of breast cancer. Here are some details about breast cancer specifically:

1.2 Early Detection: Breast cancer[4] is much more likely to be successfully treated and survive if it is discovered early. For the purpose of finding any abnormalities, routine screenings like mammograms, clinical breast exams, and self-exams are crucial.

1.3 Breast Cancer Screening and Diagnosis: Frequently involve the use of medical imaging techniques such as mammography, ultrasound, and MRI[5]. These imaging techniques support the detection of worrisome lesions or cancers in breast tissue.

1.4 Problems with Interpretation: Mammogram interpretation is particularly challenging and prone to human mistake. AI and computer-aided detection (CAD) systems can be used in this situation.

1.5 Artificial Intelligence for Breast Cancer Detection: AI[6] systems can examine mammograms and other imaging data to find patterns that may be symptomatic of malignant growths. These algorithms can help radiologists identify potential abnormalities, which will result in a more precise diagnosis.

1.6 Precision Medicine: AI[7] can also help with precision medicine by helping to customize treatment strategies based on unique patient traits like genetics and tumor activity. Precision medicine is the term for this method, which tries to improve the effectiveness of treatments.

1.7 Research and Development: Work on improving AI algorithms for breast cancer detection and diagnosis is ongoing. To increase their sensitivity and specificity, these algorithms undergo extensive training on big datasets.

1.8 Patient Empowerment: In addition to empowering medical professionals, AI may also empower patients by giving them access to informational materials, social networks, and self-care tools[8].

2. LITERATURE SURVEY

It would be necessary to summarize and analyze significant studies, journal articles, and other pieces of scientific literature that address various aspects of breast cancer as part of a literature review. An extensive literature review may encompass the following themes and findings, which are listed in general below:

2.1 Risk factors and Epidemiology: Breast cancer incidence and prevalence across various populations. Examination of risk elements like heredity, hormonal effects, way of life, and environmental exposures. Hereditary mutations like BRCA1 and BRCA2[9] and family history.

2.2 Early detection and Screening: Evaluation of several screening techniques, including as MRI, mammography, and upcoming technologies. Screening program effectiveness and restrictions in various age groups.Early detection is crucial for optimizing treatment results[10].

2.3 Imaging and Diagnosis: Tomosynthesis, mammography, ultrasound, MRI, and other advances in imaging techniques for the diagnosis of breast cancer.Systems for computer-aided detection (CAD) and their contribution to increasing diagnostic precision.

2.4 Subtypes of Molecular Pathology: Recognizing the various breast cancer subtypes (such as triple-negative, HER2-positive, and ER/PR-positive). Gene expression profiling and molecular markers for prognosis and treatment choices.

2.5 Treatment Methods: Overview of numerous therapeutic techniques, including surgery, chemo, radiation, hormone, and targeted therapy. Adjuvant and neoadjuvant procedures.Improvements in genetically-based personalized or precision medicine.

2.6 Life Quality and Survivability: Consequences of treatment throughout time on the health and quality of life of survivors.Techniques for dealing with difficulties that are physical, emotional, or psychological[11].

2.7 Innovations in science and medicine: Recent achievements in the study of breast cancer.Innovative targeted treatments and immunotherapy. Emerging ideas in tumor microenvironment and cancer stem cells.

2.8 Support and Information for Patients: The contribution of resources and patient support groups to improving patient wellbeing.Patient education is crucial for early diagnosis and treatment choices.

2.9 Health Disparities and Healthcare Policies: Examination of healthcare regulations and how they affect the treatment of breast cancer.Addressing the varying groups' health disparities in breast cancer outcomes.

2.10 Technology and AI in Breast Cancer: The detection, diagnosis, and treatment of breast cancer using artificial intelligence and machine learning.Big data and bioinformatics applications for breast cancer research.

3. PROPOSED DATASET

In this proposed work, we collected the dataset from: http://archive.ics.uci.edu/ dataset/17/breast+cancer+ wisconsin+diagnostic The dataset contains mainly 32 different attributes where each and every individual attribute has its own importance for identifying the breast cancer. The following are the attributes such as

Attribute Name	Role	Type	Demographic	Description	Units	Missing Values
ID	ID	Categorical				false
Diagnosis	Target	Categorical				false
radius1	Feature	Continuous				false
texture1	Feature	Continuous				false
perimeter1	Feature	Continuous				false
area1	Feature	Continuous				false
smoothness1	Feature	Continuous				false
compactness1	Feature	Continuous				false
concavity1	Feature	Continuous				false
concave_points1	Feature	Continuous				false
symmetry1	Feature	Continuous				false
fractal_dimension1	Feature	Continuous				false
radius2	Feature	Continuous				false
texture2	Feature	Continuous				false
perimeter2	Feature	Continuous				false
area2	Feature	Continuous				false
smoothness2	Feature	Continuous				false
compactness2	Feature	Continuous				false
concavity2	Feature	Continuous				false
concave_points2	Feature	Continuous				false
symmetry2	Feature	Continuous				false
fractal_dimension2	Feature	Continuous				false
radius3	Feature	Continuous				false
texture3	Feature	Continuous				false
perimeter3	Feature	Continuous				false
area3	Feature	Continuous				false
smoothness3	Feature	Continuous				false
compactness3	Feature	Continuous				false
concavity3	Feature	Continuous				false
concave_points3	Feature	Continuous				false
symmetry3	Feature	Continuous				false
fractal_dimension3	Feature	Continuous				false

From the above dataset we can see all the 32 attributes are having no missing values and hence this dataset is trained by our current model in order to check the performance of our proposed work.

4. PROPOSED CNN MODEL

4.1 Preparation of Data: Obtain a tagged dataset of medical images, usually those from histopathology or mammography. Make sure the dataset is balanced and inclusive of all types of breast cancer.

4.2 Data Augmentation: To improve the dataset's diversity and avoid overfitting, augment it by using transformations like rotations, flips, and zooms[13].

4.3 Data Preprocessing: Normalize the pixel values to a common scale and resize the photos to a constant size to preprocess the data[14].

4.4 Convolutional Layers: Create the CNN architecture using a number of convolutional layers. These layers continuously pick up details from the photos[15].

4.5 Insert Pooling Layers: (often max-pooling) to decrease the size of the scene and collect the most important data.

4.6 Convolutional layer: Output is flattened into a 1D vector using the flatten layer technique.

4.7 Dense (Fully Connected) Layers: To do classification based on the learned characteristics, add a number of fully connected layers.

4.8 Output Layer: A sufficient number of neurons, typically one for binary classification (cancer or non-cancer), should be present in the output layer.

4.9 Loss Function: Pick a suitable loss function for binary classification, such as binary cross-entropy, in the ninth step of the model training process.

4.10 Optimizer: Update the model's weights during training using an optimizer like Adam or RMSprop.

4.11 Training: Utilizing the annotated images for training, run the model on the training dataset. To prevent overfitting, keep an eye on the validation loss.

4.12 Testing and Evaluation

4.12.1 Metrics: Assess the performance of the model using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.

4.12.2 Testing: Evaluate the model's effectiveness using a different test dataset that it has not seen before.

4.13 Optimizing and modifying

4.13.1 Hyperparameter Tuning: To improve the performance of the model, experiment with various hyperparameters (such as learning rate and batch size).

4.13.2 Regularization: To avoid overfitting, use strategies like dropout and L2 regularization.

4.13.3 Visualizations: Use techniques like activation maps to depict the areas of the image that the model is concentrating on for its predictions when interpreting the results.

4.13.4 Misclassified Samples: Examine misclassified examples to find patterns or regions where the model may be having trouble.

Remember that developing an efficient CNN model for the diagnosis of breast cancer necessitates knowledge of both deep learning and medical imaging. The development of AI-based healthcare solutions must also take ethical issues, data protection, and the involvement of medical experts into account.

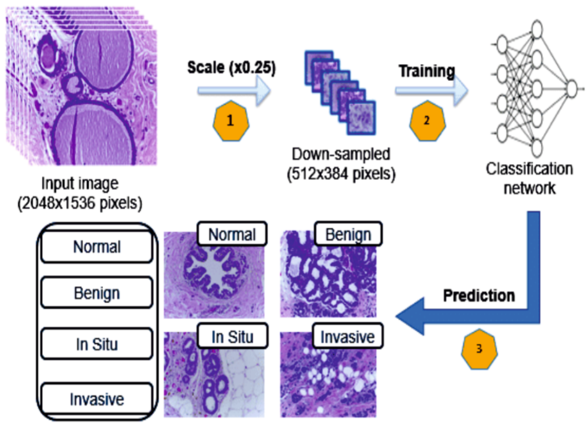


Figure 1. Represent the proposed architecture of CNN Model

From the above figure 1, we can clearly identify the proposed CNN model is constructed for prediction of breast cancer from input images.

Step 1: Obtain a labeled CT or MRI image dataset with instances of both malignant and non-cancerous tissue.Here from several fields we try to extract the image_id attribute as main attribute and then try to extract the image and its features before applying it for training the model.

	image_id
0	9075_idx5_x1801_y1201_class0.png
1	10282_idx5_x851_y1051_class0.png
2	9078_idx5_x3001_y1401_class0.png
3	9250_idx5_x1551_y51_class0.png
4	10278_idx5_x1451_y751_class1.png

Step 2: Create training, validation, and testing sets from the dataset. 70% for training, 15% for validation, and 15% for testing would be a typical split.

	image_id	target
0	9075_idx5_x1801_y1201_class0.png	0
1	10282_idx5_x851_y1051_class0.png	0
2	9078_idx5_x3001_y1401_class0.png	0
3	9250_idx5_x1551_y51_class0.png	0
4	10278_idx5_x1451_y751_class1.png	1
5	12895_idx5_x2251_y1751_class1.png	1
6	10288_idx5_x551_y1551_class0.png	0
7	12906_idx5_x801_y1601_class0.png	0
8	13401_idx5_x1601_y751_class0.png	0
9	12905_idx5_x1751_y1201_class0.png	0

Here we can see the sample images are assigned with labels 0 or 1.Which means the image which contain the cancer qualities are identified as 1 and those which are normal images are mapped with 0.

Step 3: Here we try to apply normalization by divide pixel values by the highest possible pixel value, such as 255, to normalize them to a range between 0 and 1.

Step 4: Build the Model Using a deep learning framework like TensorFlow or PyTorch, construct the CNN architecture.Multiple convolutional layers can be stacked to capture diverse aspects at varying levels of abstraction.Use activation functions to introduce non-linearity after each convolutional layer, such as ReLU (Rectified Linear Unit).

Add layers with a maximum amount of pooling to downscale the spatial dimensions and lighten the computational load.Convolutional layer output is flattened into a 1D vector using the flatten layer technique.Include one or more thick, fully linked layers to conduct categorization using the features that were previously learned.A single neuron with a sigmoid activation function for binary classification (cancer or non-cancer) should be present in the final dense layer, which is known as the output layer.

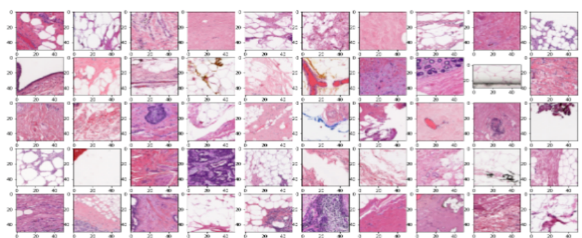


Figure 2. Represent the proposed architecture of CNN Model

From the above figure 2, Most of the mammograms are light pink, but there are some dark ones too. Cancerous patches appear more dense and violet than healthy ones. However, I believe that the model is able to recognize hidden patterns in these photos that allow us to determine the condition of each image. In fact, we were unable to distinguish between the two types with the naked eye.

Step 5: Apply Binary Objective Visualization

Using Binary objective visualization for each tissue slice, let's examine the pattern of patches and their distribution in each mammography in more detail.

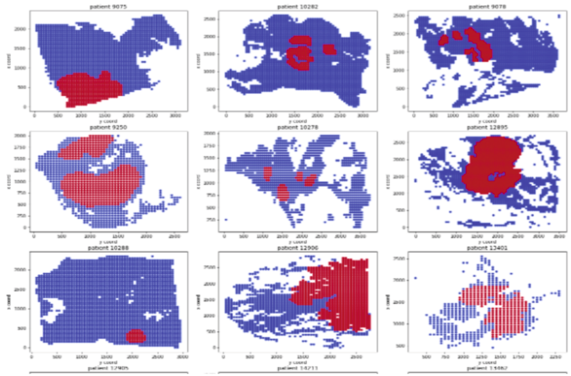


Figure 3. Represent the Binary Objective Visualization

After we apply the binary visualization,we can see the outliers of images containing the patches. There is a significant range in the number of cells present.We occasionally lack complete tissue data. It appears that tissue patches were lost or discarded while being prepared.

Step 6: Model Generation

```
my_model_im_norm =Sequential()
my_model_im_norm.add(Conv2D(filters=32,kernel_size=(4,4),inpu
t_shape=(50,50,3),activation='relu'))
my_model_im_norm.add(MaxPool2D(pool_size=(2,2)))


my_model_im_norm.add(Flatten())


my_model_im_norm.add(Dense(128,activation='relu'))
my_model_im_norm.add(Dense(2,activation='softmax'))


my_model_im_norm.compile(loss = 'categorical_crossentropy', o
ptimizer ='adam', metrics= ['accuracy'])
```

```
Epoch 5/60
24978/24978 [==============================] - 234s 9ms/st
ep - loss: 0.3759 - accuracy: 0.8404 - val_loss: 0.3487 -
val_accuracy: 0.8504
Epoch 6/60
24978/24978 [==============================] - 233s 9ms/st
ep - loss: 0.3650 - accuracy: 0.8442 - val_loss: 0.3627 -
val_accuracy: 0.8450
Epoch 7/60
24978/24978 [==============================] - 233s 9ms/st
ep - loss: 0.3591 - accuracy: 0.8480 - val_loss: 0.3655 -
val_accuracy: 0.8457
```

Finally we got an model accuracy of 84.57 % which is the current accuracy of CNN model for early identification of breast cancer.

5. EXPERIMENTAL FINDINGS

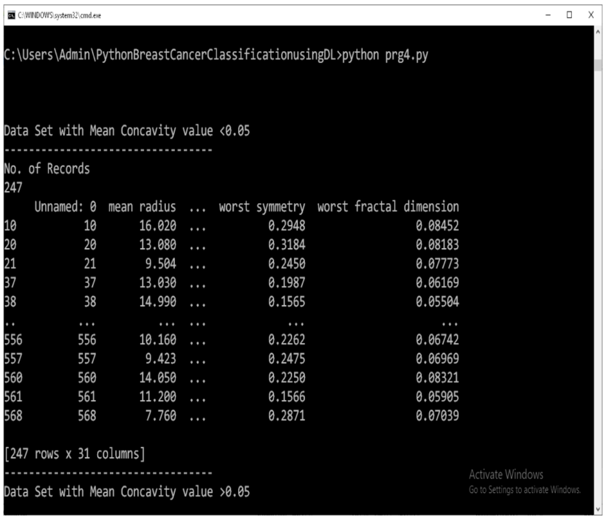
After a complete analysis of our proposed CNN model on the sample dataset we finally came to know the following findings.

The following are the findings made from the study:

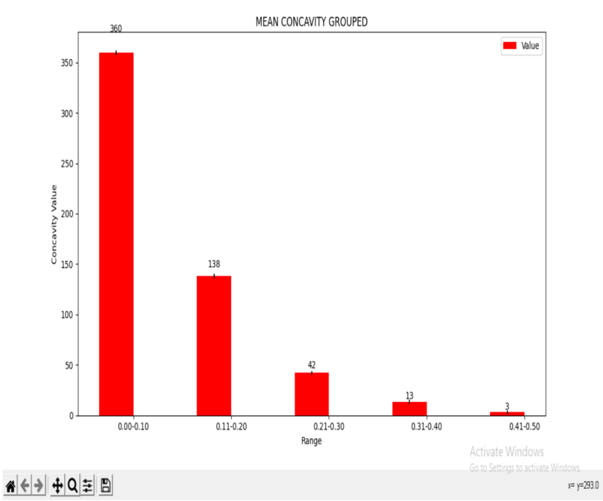
- It is found that MEAN CONCAVITY MAX VALUE is 0.4268.
- It is found that MEAN CONCAVITY RECORDS WITH VALUE <=0.05 is 247.
- It is found that MEAN CONCAVITY RECORDS WITH VALUE >0.05 is 322.
- It is found that MEAN CONCAVITY RECORDS ABOVE MID VALUE is 47.
- It is found that MEAN CONCAVITY RECORDS ARE FOUND IN 5 GROUPS FROM 0.0 TO 0.1, upto 0.41 to 0.50.
- It is found that MEAN SYMMETRY MAX VALUE is 0.304.
- It is found that MEAN SYMMETRY RECORDS BELOW MID VALUE (<0.152) is 71.
- It is found that MEAN SYMMETRY RECORDS ABOVE MID VALUE (>0.152) is 498.
- It is found that MEAN FRACTAL DIMENSION MAX VALUE IS 0.09744.
- It is found that MEAN FRACTAL DIMENSION RECORDS BELOW MID VALUE 0.
- It is found that MEAN FRACTAL DIMENSION RECORDS ABOVE MID VALUE (>0.04872) is 569.

- It is found that MEAN TEXTURE MAX VALUE IS 39.28.
- It is found that MEAN FRACTAL DIMENSION RECORDS BELOW MID VALUE (<19.64) is 325.
- It is found that MEAN FRACTAL DIMENSION RECORDS ABOVE MID VALUE (>19.64 is 244.
- CNN achieved maximum accuracy, i.e., 84.57 %.

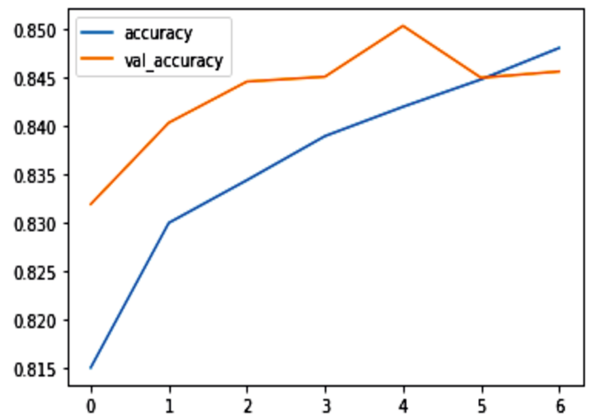
DATASET WITH MEAN CONCAVITY VALUE <=0.05



Mean Concavity Grouped



CNN Model Accuracy



- Model.evaluate_generator is out-of-date and will be eliminated in a subsequent release. Use 'Model.evaluate' instead, which is compatible with generators.
- Model.evaluate_generator is deprecated and
- val_loss: 0.36551591753959656
- val_acc: 0.8456743359565735,
- according to warnings.warn()

6. CONCLUSION

This project takes the breast cancer records in which data for various age group patients are taken with mean area, mean compactness, mean smoothness, mean concavity, mean symmetry, mean fractal dimension and other attributes. Here we try to take CNN model for identifying the breast cancer for a given patients based on dataset collected from Diagnostic Wisconsin Breast Cancer Database. The training data is taken 75% from the whole data set and model is predicted. Then the remaining 25% of the data is taken as test data and checked against the predicted model.By conducting various experiments on our proposed model,we finally came to an conclusion that our CNN model is giving an accuracy of 84.57 %, near to 85 % . We want to extend the same work on some other CNN models and increase the accuracy and reduce the time complexity.

7. REFERENCES

[1] P. K. Samanta, N. K. Rout and G. Panda, "A novel deep CNN model for improved breast cancer detection using ultrasound images," 2023 International Conference on Communication, Circuits, and Systems (IC3S), BHUBANESWAR, India, 2023, pp. 1-4, doi: 10.1109/IC3S57698.2023.10169383.

[2] B. Bılgıç, "Comparison of Breast Cancer and Skin Cancer Diagnoses Using Deep Learning Method," 2021 29th Signal Processing and Communications Applications Conference (SIU), Istanbul, Turkey, 2021, pp. 1-4, doi: 10.1109/ SIU53274.2021. 9477992.

[3] C. R. Prasad, B. Arun, S. Amulya, P. Abboju, S. Kollem and S. Yalabaka, "Breast Cancer Classification using CNN with Transfer Learning Models," 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1-5, doi: 10.1109/ICONAT57137.2023.10080148.

[4] M. N. Kumar, A. Jatti and C. K. Narayanappa, "Probable Region Identification and segmentation in Breast Cancer using the DL-CNN," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 1144-1149, doi: 10.1109/ICSSIT46314. 2019.8987818.

[5] M. Karatayev, S. Khalyk, S. Adai, M. -H. Lee and M. F. Demirci, "Breast cancer histopathology image classification using CNN," 2021 16th International Conference on Electronics Computer and Computation (ICECCO), Kaskelen, Kazakhstan, 2021, pp. 1-5, doi: 10.1109/ICECCO53203.2021.9663757.

[6] N. B. Nair, T. Singh, A. Thakur and P. Duraisamy, "Deployment of Breast Cancer Hybrid Net using Deep Learning," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-6, doi: 10.1109/ICCCNT54827.2022.9984513.

[7] S. Afaq and A. Jain, "MAMMO-Net: An Approach for Classification of Breast Cancer using CNN with Gabor Filter in Mammographic Images," 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2022, pp. 177-182, doi: 10.1109/CISES54857.2022.9844320.

[8] Anjana Lvanturi, Ankita Singh, B. Gunavitha, and K. S. Chethan. Soft Classification Methods for the Detection and Classification of Breast Cancer. The ICIEM 2020 conference is an international gathering on intelligent engineering and management. IEEE(2020).

[9] Jaafar Alghazo, Ghazanfar Latif, Muhammad O. Butt, and Faisal Yousif AI Anezi.Deep CNN is used for ultrasound image despeckling and breast cancer detection.International Computing and Communication Technologies Conference (RIVF) 2020.IEEE(2020).

[10] Nidhi Mathur, Aditya Prakash, Preeta Sharan, Sandin Kumar Roy, and Sumita Mishra. Thermal imaging for the detection of breast cancer. 2020 will see the seventh INDIACOM conference on computing for sustainable global development.IEEE2020.

[11] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert Syst. Appl., Vol. 36, No. 2, Mar. 2009, pp. 3240–3247.

[10] F. A. Cardillo, A. Starita, D. Caramella, and A. Cillotti, "A neural tool for breast cancer detection and classification in MRI," in Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 3, pages 2733–2736, October 2001.

[12] Breast cancer diagnosis in gadolinium-enhanced MR images by static region descriptors and neural networks was described by A. A. Tzacheva, K. Najarian, and J. P. Brockway in Journal of Magnetic Resonance Imaging, vol. 17, no. 3, 2003, pp. 337-342.

[13] "A tutorial survey of architectures, algorithms, and applications for deep learning," by L. Deng, APSIPA Transactions on Signal and Information Processing, vol. 3, 2014.

[10] J. Schmidhuber, "Deep learning in neural networks: an overview," in Neural Networks, vol. 61, no. 1, 2015, pp. 85–117.