**An unstructured MRI compression technique using a convolutional approach**

Debkumar Chowdhury1\*, Tirtharaj Sinha1, Arnobrata Ghosh1, Anurag Unnikannan1, Susmit De1, Devesh Raj1and Siddhartha Bose1

1University of Engineering and Management, Kolkata, India

1[debkumar.cse@gmail.com](mailto:debkumar.cse@gmail.com),1[sinhatirtharaj@gmail.com](mailto:sinhatirtharaj@gmail.com) ,1[ghosharnobrata@gmail.com](mailto:ghosharnobrata@gmail.com),1[anuragunnikannan7@gmail.com](mailto:anuragunnikannan7@gmail.com),1[desusmit101@gmail.com](mailto:desusmit101@gmail.com), 1[kartik.sau@uem.edu.in](mailto:kartik.sau@uem.edu.in), 2[sanjuktamish@gmail.com](file:///C:\Users\TIRTHA\Downloads\sanjuktamish@gmail.com)

**Abstract.** ~~Currently, different methods are available for the purpose of breast cancer classification and detection. Most of these techniques are well appreciated by society and in response to the demand of society, almost every year the different techniques are introduced by different researchers. But it does not satisfy the demand of current requirements. Under such a situation, we are going to propose a new breast cancer classification and detection algorithm using a convolutional approach. This technique starts with the mammogram preprocessing step. It is followed by the convolutional model architecture design step. In the next step, the segregation of the dataset into the training and testing phase is performed. Then the convolutional model architecture is trained using the training dataset and pre-masked images. After that our proposed algorithm predicts, the breast cancer detection and classification result. We have found that our proposed algorithm can be used for breast tumor detection and classification from mammogram images with the average approximate accuracy of 98.5% and the average approximate F1 Score of 0.98. Novel preprocessing steps and modifications in the convolutional architecture make the proposed methodology unique. Due to high performance, novelty, ease of use, our proposed method is useful to develop any mobile or web application in the future.~~

**Keywords:** ~~Breast Cancer, Mammogram, Pre-masked image, Convolutional, Accuracy Score, F1 Score.~~

# Introduction

~~Breast Cancer is considered to be most common nowadays. According to the December 2020 statistics of the World Health Organization (WHO) among 24.5% of world's female cancer population is affected by this disease. It starts spreading from breast. At the early stage of this disease, the cancerous cells we may be found as a tumor or lump, which is clearly visible from the X-Ray images. Most of the breast tumors are found as non-cancerous in nature and often observed that they do not spread outside the breast, whereas malignant breast tumors are cancerous in nature and may spread outside the breast. Patients of this disease are often suffered from symptoms like breast or nipple pain, swelling of all or part of a breast, skin dimpling, nipple retraction, reddish nipple or breast skin, nipple discharge, swollen lymph nodes, etc. Among many types of this disease ductal carcinoma in situ (DCIS) and invasive carcinoma are considered to be the most common and phyllodes tumors and angiosarcoma are considered as rare.~~

~~Hence, it is considered as a serious issue of society and an automated computer-based method is required to detect and classify breast cancer with precision from mammogram images. This method helps the medical practitioner to start early treatment and also helps to reduce the fatality rate. The inconclusive, incomplete, and dissatisfactory results of the previously proposed techniques encourage us to develop a novel breast cancer classification and detection algorithm using a convolutional approach that works on unstructured data such as mammogram images. It displays the classified output using convolutional model architecture along with the satisfactory accuracy rate and F1 score.~~

~~Our paper is divided into various sections. In section 2, we explain the advantages and disadvantages of some pre-existing techniques, in section 3 we explain our proposed methodology which consists of the main architecture of the method and the algorithm, whereas the experiment result and analysis are explained in section 4.~~

1. **Literature Survey**

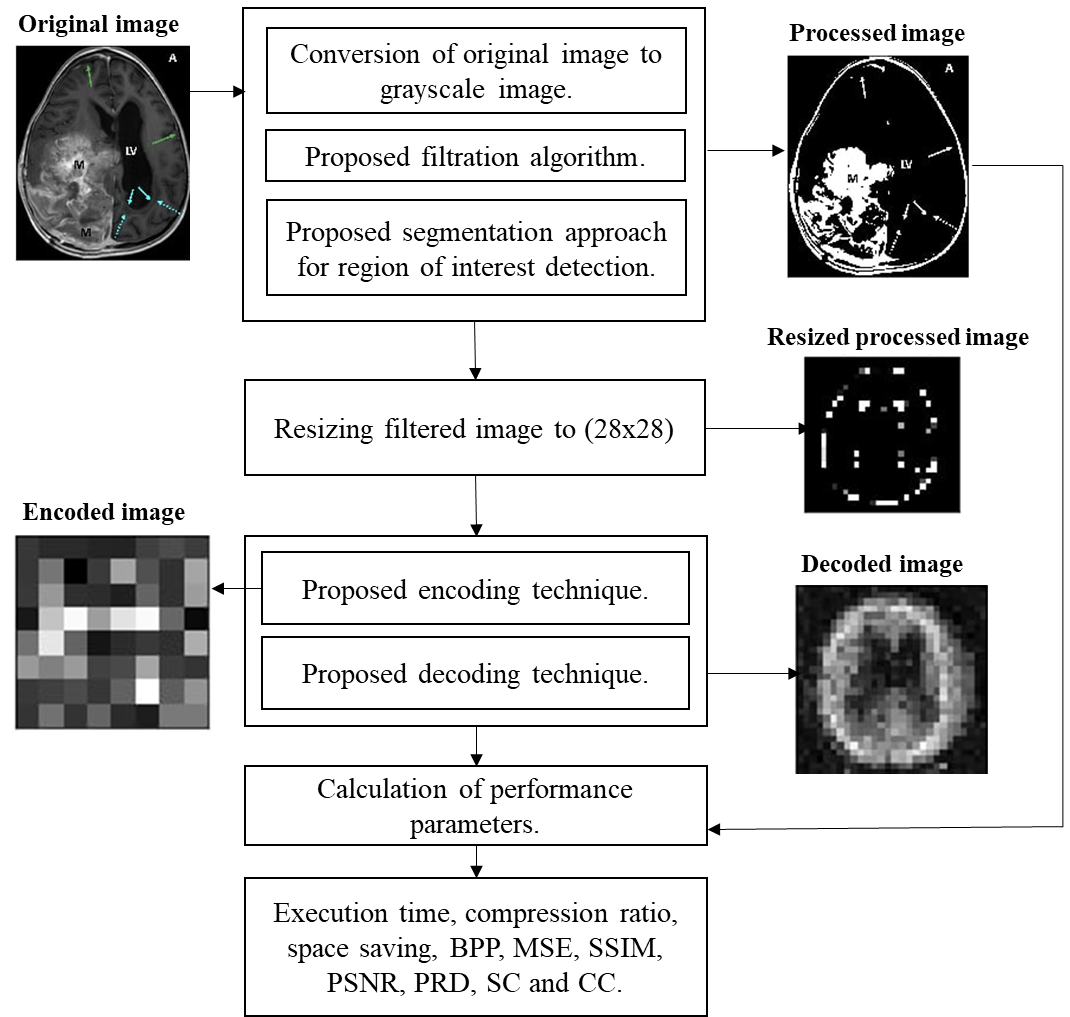
~~This section represents various pre-existing methods [7] [8] [9], along with their advantages and disadvantages as shown in the Survey Table (~~**~~Table. 1~~**~~). All these methods are well appreciated but in context with our problem they are producing inclusive, incomplete, dissatisfactory results. In-depth analysis of these methodologies has proven to be very competent to identify the downsides. Identification of these drawbacks helps us to update and modify our algorithm and code and to calculate the accuracy rate. The Survey Table is given below~~:

**~~Table. 1.~~** ~~Existing Methodology Analysis Table~~

|  |  |  |
| --- | --- | --- |
| **~~Methodology~~** | **~~Advantages~~** | **~~Disadvantages~~** |
| ~~Decision Tree [1]~~ | ~~Decision trees require less effort for data preparation during pre-processing.~~ | ~~A small change in the data can cause a large change in the structure. According to our problem, it causes instability.~~ |
| ~~K-Nearest Neighbor [2]~~ | ~~No assumption about data.~~ | ~~KNN needs a huge amount of memory.~~ |
| ~~Random Forest Classifier [3]~~ | ~~Random forest minimizes the overfitting issue and tries to increase the accuracy score.~~ | ~~This method needs high computational power and resources.~~ |
| ~~Support Vector Machine [4]~~ | ~~It is effective in high-dimensional spaces.~~ | ~~Due to a huge time consumption issue, the performance of this method is not satisfactory when we apply this strategy in a large dataset.~~ |
| ~~Gaussian Naïve Bayes [5]~~ | ~~Naive Bayes is more appropriate for categorical input variables than numerical variables. In the context of our problem, this property is suitable.~~ | ~~It presumes that every feature of the dataset is independent. This property limits the application of this technique in real-world cases.~~ |
| ~~Multi-Scale Fusion U-Net [6]~~ | ~~It solves the problem of multi-scale variation in breast lesions and boundary pixel blurring.~~ | ~~For the different modes of images, segmentation effects will be reduced.~~ |

1. **Proposed Methodology**

~~Our proposed methodology is focused on breast cancer classification and detection from an unstructured dataset (BUSI) [10] using breast mammogram classification and detection algorithm using a convolutional approach. We have proposed a block diagram to show the main concept of the methodology at a glance as shown in~~ **~~Fig. 1.~~**



**Fig. 1.** The built-in architecture of MRI compression technique using a convolutional approach

* 1. **Algorithm**

Our algorithm is divided into 16 steps. The algorithm takes a MRI, as an input from the dataset [10] and produces an encoded image and a decoded image. The algorithm is as follows:

**Algorithm:** An unstructured MRI compression technique using a convolutional approach

**Input:**

**Output:**  encoded image and decoded image

1. Read an image from the dataset.
2. Convert the image to a grayscale image,.
3. Preprocessed using the following formula and 5x5 kernel:

(1)

Where,

= Processed Image

= Standard Deviation

= th row in the grayscale image

= th column in the grayscale image

1. Perform segmentation on using the following equation:

If pixel then,

,

Otherwise,

, provided maximum pixel value = 255

Where,

= grayscale image

= a pixel of an image in position

= selected threshold value in our algorithm. It is tested after the trial-and-error method.

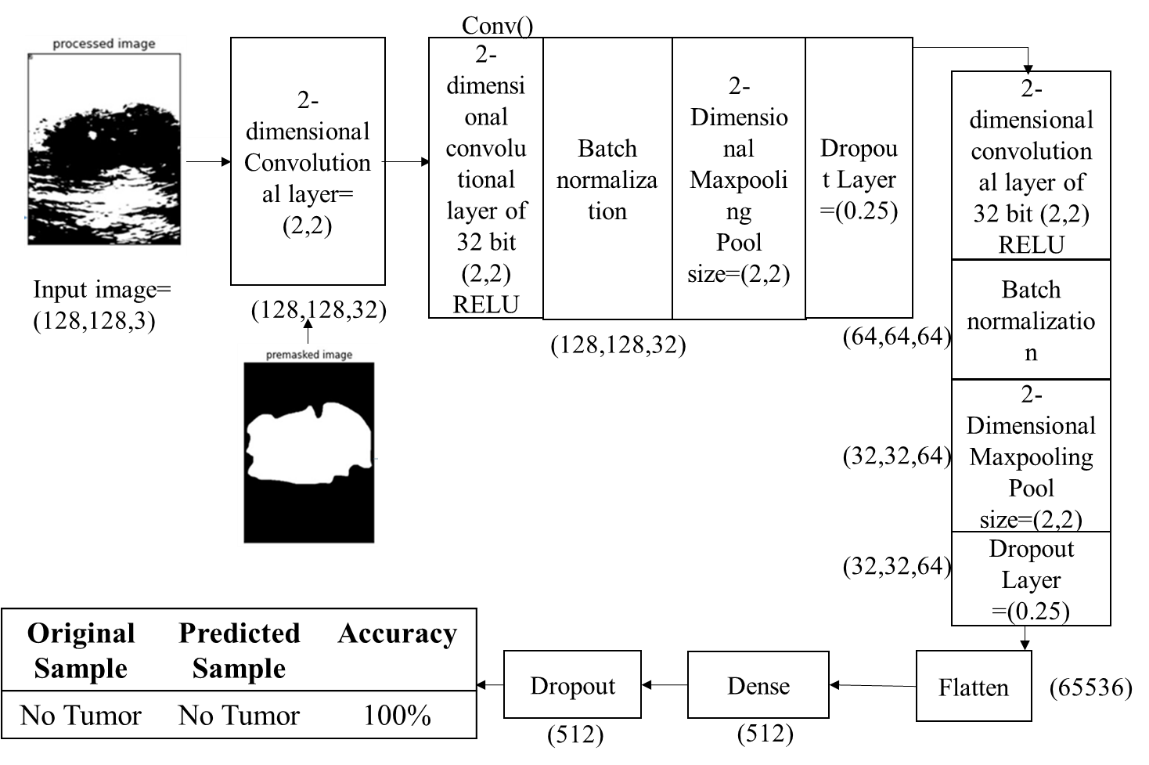
1 = Light

0 = Dark

1. Resize into a 28 X 28 image.
2. The proposed convolutional architecture each has an input x of size 28x28x3 is feed into it. In the convolutional architecture following layers are observed:
   1. Encoder Architecture
   2. One input layer which accepts input of size 784.
   3. One dense layer of output size 300 with an activation leakyReLU function.
   4. One dense layer of output size 64 with an activation leakyReLU function.
   5. Decoder Architecture
      1. One input layer which accepts input of size 64.
      2. One dense layer of output size 300 with an activation leakyReLU function.
      3. One dense layer of output size 784 with an activation leakyReLU function.
3. Segregate the whole dataset [10] into two sections. We select 80% MRI from the dataset [10] for training purposes and 20% MRI from the dataset [10] for testing purposes.
4. We feed the training and testing data into the designed convolutional model architecture to generate result.

**3.2 Model Architecture**

Training and testing phase can be explained with the help of a model architecture as shown in **Fig.2.** the architecture accepts preprocessed MRI image and displays encoded and decoded images along with performance parameters for the same.

~~~~

Premasked Image

Processed Image

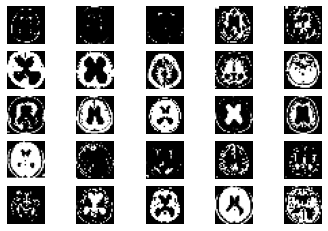
**Fig.2.** Training and testing phase using proposed model architecture

1. **Experimental results**

We consider the brain MRIs as a dataset, which consists of 132 images. The size, colour, and format of images in the dataset are similar in nature, whereas the resolutions of the images are different. The format of the images is ‘.jpg’ by nature.

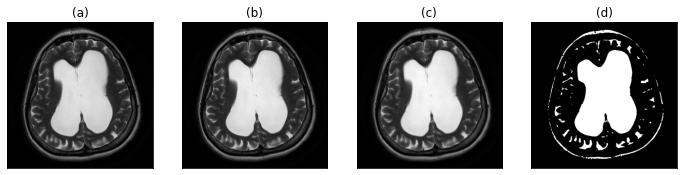
We have applied our algorithm in the python environment, version 3.8, with the hardware configuration of the Intel Core i3 5th Generation processor,4GB DDR3 primary memory (RAM), and an integrated graphics card. Anaconda as a distributor of Python version 3.8 is used. jupyter notebook version 6.3.0 as an open web interface is used as a programming platform for the implementation of our algorithm.

After reading the images from the local machine we have achieved the following results as shown in **Fig. 3.**



**Fig. 3.** MRI

After reading all the images from the dataset each image will pass through the various steps of the algorithm as shown in **Fig. 5.**

****

**Fig. 5.** (a) original image (b) gray-scale image (c) image after applying proposed filtration method (d) proposed segmentation approach for the region of interest detection

We are splitting our dataset [10] into training and testing sets. 80% of the MRIs are used for training and the rest is used for testing. After the execution of our proposed algorithm, we observed that the number of training samples are 104, number of testing samples are 26 whereas training shape values are (104, 28, 28). Our proposed convolutional layered architecture is trained. It is used to train on n number of samples (104 in our case). With the changes of several iterations or epochs, the total time consumed by the algorithm is calculated.

**Table 2:** An example of the training phase.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SNo.** | **Epoch No.** | **Time Taken (s)** | **Loss** | **Value Loss** |
| 1. | 1 | 1 | 0.2193 | 0.1604 |
| 2. | 2 | 0 | 0.1407 | 0.1219 |
| 3. | 3 | 0 | 0.1159 | 0.1086 |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
|  |  |  |  |  |
| 18. | 18 | 0 | 0.0517 | 0.0842 |
| 19. | 19 | 0 | 0.0494 | 0.0835 |
| 20. | 20 | 0 | 0.0473 | 0.0831 |

**Table 3:** Prediction result and accuracy calculation

|  |  |  |  |
| --- | --- | --- | --- |
| **~~SNo.~~** | **~~Original Sample~~** | **~~Predicted Sample~~** | **~~Accuracy (%)~~** |
| ~~1~~ | ~~No tumor~~ | ~~No tumor~~ | ~~100~~ |
| ~~2~~ | ~~Tumor~~ | ~~Tumor~~ | ~~100~~ |
| ~~3~~ | ~~Tumor~~ | ~~Tumor~~ | ~~100~~ |
| ~~4~~ | ~~No tumor~~ | ~~No tumor~~ | ~~100~~ |
| ~~.~~ | ~~.~~ | ~~.~~ | ~~.~~ |
| ~~.~~ | ~~.~~ | ~~.~~ | ~~.~~ |
| ~~.~~ | ~~.~~ | ~~.~~ | ~~.~~ |
| ~~66~~ | ~~No tumor~~ | ~~Tumor~~ | ~~0~~ |
| ~~67~~ | ~~No tumor~~ | ~~No tumor~~ | ~~100~~ |
| ~~68~~ | ~~Tumor~~ | ~~Tumor~~ | ~~100~~ |
| ~~69~~ | ~~No tumor~~ | ~~No tumor~~ | ~~100~~ |

~~In~~ **~~Table 3~~** ~~displays 69 randomly selected original samples from the dataset, corresponding predicted sample results and the accuracy score.~~~~In the end, we have found that our proposed algorithm can be used for breast tumor detection and classification from mammogram images with an average approximate accuracy of 98.5%. As the data is balanced, we are considering the results to be satisfied. Comparing it with other existing methodologies a satisfactory result is observed as shown in~~ **~~Table 4.~~** ~~We can conclude that our proposed methodology overpowers the existing classification algorithms [7] [8] [9] in terms of F1 and accuracy score.~~

**~~Table 4:~~** ~~Comparison chart~~

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **~~Serial No~~**~~.~~ | **~~Name of the classification algorithm~~** | **~~Accuracy Score~~** | **~~F1 Score~~** |
| ~~1.~~ | ~~Decision Tree [1]~~ | ~~97%~~ | ~~0.97~~ |
| ~~2.~~ | ~~K-Nearest Neighbor [2]~~ | ~~94.2%~~ | ~~0.94~~ |
| ~~3.~~ | ~~Gaussian Naïve Bayes [3]~~ | ~~97%~~ | ~~0.97~~ |
| ~~4.~~ | ~~Random Forest [4]~~ | ~~97%~~ | ~~0.95~~ |
| ~~5.~~ | ~~Support Vector Machine [5]~~ | ~~96%~~ | ~~0.96~~ |
| ~~6.~~ | ~~Multi-Scale Fusion U-Net [6]~~ | ~~96%~~ | ~~0.96~~ |
| **~~7.~~** | **~~Proposed Method~~** | **~~98.5%~~** | **~~0.98~~** |

1. **~~Conclusion~~**

~~In this paper, we have proposed a mammogram classification and detection algorithm using a convolutional approach. This algorithm is capable of preprocessing unstructured mammogram images from the BUSI dataset [10], collected from the web resource. It is responsible for the prediction of malignant and non-malignant mammogram images in terms of Yes (malignant sample) and No (non-malignant sample) values. It also generates an accuracy and F1 score through which we can compare our proposed method with existing methods [7] [8] [9]. The experimental result shows that after applying the proposed and existing methods [7] [8] [9] on the BUSI dataset [10], our technique is producing approximately 98.5% accuracy and 0.98 F1 scores on average. This result is considered to be satisfactory and based on this result we can say that the proposed algorithm overpowers the efficiency of the existing method as described in~~ **~~Table 4.~~** ~~Novel preprocessing steps and modifications in the convolutional architecture using multiple layers make the proposed methodology unique. Due to high performance, novelty, ease of use, our proposed method is useful to develop any mobile or web applications in the future. Our method can be tested on various mammogram datasets to identify the generic performance of the proposed method in the future. The performance of our method may be increased by making necessary modifications in algorithm and code.~~

# References

1. ~~L. Yi and W. Yi, "Decision Tree Model in the Diagnosis of Breast Cancer,"~~*~~2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC)~~*~~, 2017, pp. 176-179, doi: 10.1109/ICCTEC.2017.00046.~~
2. ~~Hadidi, moh'd & Alarabeyyat, Abdulsalam & Alhanahnah, Mohannad. (2016). Breast Cancer Detection Using K-Nearest Neighbor Machine Learning Algorithm. 35-39. 10.1109/DeSE.2016.8.~~
3. ~~B. Dai, R. -C. Chen, S. -Z. Zhu and W. -W. Zhang, "Using Random Forest Algorithm for Breast Cancer Diagnosis," 2018 International Symposium on Computer, Consumer and Control (IS3C), 2018, pp. 449-452, doi: 10.1109/IS3C.2018.00119.~~
4. ~~Shang Gao and Hongmei Li, "Breast cancer diagnosis based on support vector machine," 2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering, 2012, pp. 240-243, doi: 10.1109/URKE.2012.6319555.~~
5. ~~H. Kamel, D. Abdulah and J. M. Al-Tuwaijari, "Cancer Classification Using Gaussian Naive Bayes Algorithm," 2019 International Engineering Conference (IEC), 2019, pp. 165-170, doi: 10.1109/IEC47844.2019.8950650.~~
6. ~~J. Li, L. Cheng, T. Xia, H. Ni and J. Li, "Multi-Scale Fusion U-Net for the Segmentation of Breast Lesions," in IEEE Access, vol. 9, pp. 137125-137139, 2021, doi: 10.1109/ACCESS.2021.3117578.~~
7. ~~S. Ara, A. Das and A. Dey, "Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms,"~~*~~2021 International Conference on Artificial Intelligence (ICAI)~~*~~, 2021, pp. 97-101, doi: 10.1109/ICAI52203.2021.9445249.~~
8. ~~M. Amrane, S. Oukid, I. Gagaoua and T. Ensarİ, "Breast cancer classification using machine learning,"~~*~~2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)~~*~~, 2018, pp. 1-4, doi: 10.1109/EBBT.2018.8391453.~~
9. ~~B. Bılgıç, "Comparison of Breast Cancer and Skin Cancer Diagnoses Using Deep Learning Method,"~~*~~2021 29th Signal Processing and Communications Applications Conference (SIU)~~*~~, 2021, pp. 1-4, doi: 10.1109/SIU53274.2021.9477992.~~
10. ~~Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in Brief. 2020 Feb;28:104863. DOI: 10.1016/j.dib.2019.104863.~~