1. Introduction  
   Recent studies have shown that breast cancer is the most common type of cancer among women [ 1 ],  
   accounting for about one third of newly diagnosed cancers in the US [2]. The mortality rate of breast  
   cancer is also high, accounting for 17% of deaths relating to cancer in general [3 ]. Accurate detections  
   and assessment of breast cancer in its early stages is crucial when it comes to reducing the mortality  
   rate. Mammography is until today the most useful tool for general population screening. However,  
   the accurate detection and diagnosis of a breast lesion solely based on mammography findings is  
   difficult and highly depends on the expertise of the radiologist, which leads to a high number of false  
   positives and additional examinations [4].  
   Computer-aided detection and diagnosis (CAD) systems are already being used to offer crucial  
   assistance in the decision-making process of radiologists. Such systems may significantly reduce  
   the amount of effort needed for the assessment of a lesion in clinical practice, while minimizing the  
   number of false positives that lead to unnecessary and discomforting biopsies. CAD systems regarding  
   mammography may address two different tasks: detection of suspicious lesions in a mammogram  
   (CADe) and diagnosis of detected lesions (CADx), i.e., classification as benign or malignant.  
   Deep learning is considered a significant breakthrough technology of recent years as it has  
   exhibited performance beyond the state-of-the-art in various machine learning tasks including object  
   detection and classification. Contrary to conventional machine learning methods, which require  
   a hand-crafted feature extraction stage, which is challenging as it relies on domain knowledge,  
   deep learning methods adaptively learn the appropriate feature extraction process from the input data with respect to the target output. This eliminates the tedious process of engineering and investigating the discrimination ability of the features while facilitating the reproducibility of the methodologies

Breast composition containing attenuating tissue is an essential element for evaluating mammogram reports to predict malignant and benign cases. Architectural distortion (AD) is the third most suspicious appearance on a mammogram representing abnormal regions that can be found visible on mammography projection [[3](https://word-edit.officeapps.live.com/we/wordeditorframe.aspx?new=1&ui=en-US&rs=en-IN&hid=B7dHp707ake0%2BXnPTgB1cw.0&wopisrc=https%3A%2F%2Fwopi.onedrive.com%2Fwopi%2Ffiles%2F7E059818549FE17B!336&wdnewandopenct=1669662615017&wdprevioussession=acca29c9-e144-49dc-bc6a-d85ca7429d05&wdorigin=OFFICECOM-WEB.START.NEW&wdo=2&wde=docx&sc=host%3D%26qt%3DDefault&mscc=1&wdp=0&uih=OneDrive&jsapi=1&jsapiver=v2&corrid=ff27fe7b-beb9-456f-a277-3e3a5aaf0e66&usid=ff27fe7b-beb9-456f-a277-3e3a5aaf0e66&newsession=1&sftc=1&wdredirectionreason=Unified_SingleFlush" \l "B3-biology-11-00015)]. The main parameters such as global asymmetry, focal asymmetry, and developing asymmetry of tissue can be calculated using machine and deep learning algorithms to track AD in mammograms. Asymmetries are the isodense tissues obscured by adjacent fibro glandular mass, representing true malignancy in mammograms. Architectural distortion tracking from mammograms is very difficult due to its subtle and varying asymmetry on breast mass and small size. Therefore, the manual interpretation of architectural distortion is a challenging task for radiologists to figure out abnormalities during the examination of mammograms. The leading types of cancer that can present architectural distortion on mammography are invasive lobular carcinoma (ILC) and invasive ductal carcinoma (IDC). The ILC and IDC on mammography having a star-shaped pattern are likely to be malignant, while the complex and radial sclerosing lesions architectural distortion having larger than 1 cm is probably benign [[4](https://word-edit.officeapps.live.com/we/wordeditorframe.aspx?new=1&ui=en-US&rs=en-IN&hid=B7dHp707ake0%2BXnPTgB1cw.0&wopisrc=https%3A%2F%2Fwopi.onedrive.com%2Fwopi%2Ffiles%2F7E059818549FE17B!336&wdnewandopenct=1669662615017&wdprevioussession=acca29c9-e144-49dc-bc6a-d85ca7429d05&wdorigin=OFFICECOM-WEB.START.NEW&wdo=2&wde=docx&sc=host%3D%26qt%3DDefault&mscc=1&wdp=0&uih=OneDrive&jsapi=1&jsapiver=v2&corrid=ff27fe7b-beb9-456f-a277-3e3a5aaf0e66&usid=ff27fe7b-beb9-456f-a277-3e3a5aaf0e66&newsession=1&sftc=1&wdredirectionreason=Unified_SingleFlush" \l "B4-biology-11-00015)].

Image data :

Several studies reported hand-crafted feature extraction techniques on mammogram images for AD ROI classification using machine learning and deep learning [[5](https://word-edit.officeapps.live.com/we/wordeditorframe.aspx?new=1&ui=en-US&rs=en-IN&hid=B7dHp707ake0%2BXnPTgB1cw.0&wopisrc=https%3A%2F%2Fwopi.onedrive.com%2Fwopi%2Ffiles%2F7E059818549FE17B!336&wdnewandopenct=1669662615017&wdprevioussession=acca29c9-e144-49dc-bc6a-d85ca7429d05&wdorigin=OFFICECOM-WEB.START.NEW&wdo=2&wde=docx&sc=host%3D%26qt%3DDefault&mscc=1&wdp=0&uih=OneDrive&jsapi=1&jsapiver=v2&corrid=ff27fe7b-beb9-456f-a277-3e3a5aaf0e66&usid=ff27fe7b-beb9-456f-a277-3e3a5aaf0e66&newsession=1&sftc=1&wdredirectionreason=Unified_SingleFlush" \l "B5-biology-11-00015)]. These methods successfully achieved remarkable accuracy in the diagnosis of breast cancer. However, many factors are involved in detecting architectural distortion, such as tinny size, subtle appearance inside mass, shape, noise, imaging artefact from digital mammograms. Due to a limited number of studies that reported AD ROI’s classification in the literature, this primarily discusses the most relevant studies in the first phase. The second phase discusses deep learning, machine learning, and mass segmentation, to determine the limitations of predicting breast cancer. There are many limitations in these studies for detecting architectural distortion ROIs and classification.

The AD cases are less in number- due to the early formation of tissue distortion and no discomfort with pain is not seen in the major of the cases of early developmet of AD. Due to this the availablity is less in the cases segmentation approach is considered for this as it involves with the intuitation of the negative images , sample of postive data are oversampled with the augementation for the better classification.

Bounding Box - Region of Mask :

The Mask played a cruial role in the detection of AD with the bounding region , During the Initial development the The mask is free hand annotation which resulted in the resultation in FP, FN higer in number due to incorrect region detection and gradular tissue visiblity in the type 3, type 4 tissue more. So in order to avoid the gradular tissue and also increase the region of the detection, the bounding box is considered for the better implementation. The unet model is choosen instead of the object detection due to the availablity of less image data .

Network Design :

Segmentation Model :

For the AD architeccutre is divided into two separate architectures which includes one for the region of the detection and folowed by roi classification. For the region detection an unet segmentation model is considered cause due to the simpler design in adding the negative images during the training process. The unet segmentation model is of x layers with downsampling and upsampling layer with a droput of x range,

The filter used in the model are x,x,x filter with the downsampling and using the same with x filter of upsampling the features. For the downsampling the maxpool is used with x kernal size, With all this together the architecture has x traniable parameters and x non trainable paramaters .

Architecture image

For each filter of the relu activation is used in the hidden layers for the last layer the activation is sigmoid. The segmatation model is of consider the image size of x is fed to the model, the output is a region of the AD segmentation. For the model the whole image is resized and fed to the model instead of patches which is addressed in the classification model.

The model is trained on the different vendors data and including the public data, which is fed to the model. The data is annotated with the radiologists for suspicious regions of AD of x cases , iimages are filtered and fed to the segmentation model

Classification Model

The segmentation model output roi of orgnial images is taken as a input and resized to x fed to the model. The model is developed on the basis of autoencoder architecture with the adding the previous weight sof the classification model.

Encoder and decoder model.

The autoencoder first trained for the achieving segmenation task with the same layer and weights are used for the encoder architecture and few layers added for the better feature detection, the model is trained on the few images of the with the segmentation layer weights in the encoder architecture for the grater accracy once, the encoder architecure is connected with the x Dense layer and Dropout for the classification task, the outpiut layer activation is softmax is used

The model is trained on the patches of the ROI and FP regions detected from the segmentation model, thus the model improved over the stage. In the later stage the encoder architecure with dense layers is taken and trained on the FP, TP regions for the better activation which resulted in the improvement of the accuracy

Architecture image

Training :

Both the models are trained with the data from the cases 300 cases and negative cases , equal distribution of cases are taken and trained on the -

Classification model -

The Bounding box patches are taken and model is trained

Hyper parameters :

Both models trained on the loss function of binary\_crossentropy with the learning rate of 1e-3 for the segmentation.

Each model is trained for 300 epochs and results of epoch weights iis tested on the few images for considered of the equalvent perfect pair in the detection and classification of AD

Workflow image

The auc of the model detection under the radiologists evaluation is as follows

Auc curve with metrics