U-Net: Breast Cancer Semantic Segmentation on Ultrasound Images using Grad-CAM

Abstract—**Breast cancer has remained as the most commonly diagnosed cancer worldwide, with the rate of infections among women continuously increasing with time. Hence, this issue requires appropriate tools for diagnosis to avoid casualties, misdiagnosis and much more. In regards to detection, classification and segmentation, significant progress has been made through breast cancer research facilities, and the accuracies received have been progressively improving through time, however interpretations and visualization methods can prove to be lackluster. Thus, Machine Learning (ML) classifiers are implemented to a higher degree to facilitate the early prediction process. In this paper, we will be dealing with the semantic segmentation aspect, and further refine the visualisation procedure in ML classifiers using Grad-CAM. To be more elaborate, we will implement U-Net Architecture and Adam Optimizer, which is a deep learning technique. Then we will demonstrate how Grad-CAM improves the accuracy by influencing the decision-making process of Convolutional Neural Networks (CNNs), validating model predictions and providing heat-maps to identify regions of interest, thus increasing the chances of a correct diagnosis. This model is not pre-trained and provides an output accuracy of 99.37%, which is significantly higher in comparison to other models.**

Index Terms—**Breast Cancer, Semantic Segmentation, U-Net Architecture, Grad-CAM, Adam Optimizer.**

I. Introduction

Breast cancer stands as a prominent cause of mortality among women on a global scale. According to the World Health Organization (WHO), breast cancer is the most commonly diagnosed disease on a global scale, with over 626,700 women succumbing to cancer-related ailments annually [22], as of 2018. Thus, it has been a very crucial topic for cancer research in multiple institutions throughout history. In terms of methodologies, there are treatments that can cure the cancer, at least to an extent, and precautionary processes that can increase the chances of detection at early stages or ideally, prevention from the cancer. So, how does Segmentation tie into this?

Segmentation is used in the precursor phases of detection for mammogram/ultrasound image analysis. There are many classifications for segmentation, from which we will implement deep learning segmentation. This will result in appropriate feature extraction, classification, visualization and relevant treatment. In the case of breast cancer, segmentation has the capability to detect and classify regions of interest from a mammogram or ultrasound image, using methods such as CNNs, U-Net etc. There is a vast array of visualization techniques and algorithms, but we will focus on Grad-CAM in particular.

Grad-CAM is used for creating a class-specific heat map based off of an image, i.e a mammogram, a trained CNN model and a class of interest. It is an observer-dependent measure; which means it visually assesses the correctness of a classification. For breast cancer segmentation, Grad-CAM will allow us to identify the difference and the severity of different tumors by processing the heat maps to classify and better understand the nature of the cancer.

Our main objective is to apply Deep-Learning Segmentation on ultrasound images and visualize our results using Grad-CAM, which is more class-discriminative and facilitates the identification process.

II. Related Work

A. Pre-Trained Convolutional Neural Networks

Masud et al. [1] used a data set of breast ultrasound pictures obtained from Rodrigues [2] which was publicly accessible. The data set had a total of 250 ultrasound scans, with 100 and 150 cases classified as benign and malignant respectively. The sizes of the photos are not uniform. The image resolutions range from at least 57 x 75 pixels to a topmost of 161 x 199 pixels using a grey colour scheme. The photographs are inputted into various models, where the dimensions are resized, depending on the specific pre-trained model. To preserve Gaussian-like distributions, data centralization is the primary focus, which also helps to speed up the convergence of the model. The data is collapsed into a unit of K equal sizes where each unit represents a validation data. The outcomes derived from the process of cross validation are afterwards aggregated to provide a singular estimation and to enhance the performance and promote generalisation of the models, the data set size is expanded using diverse data augmentation techniques. The researchers utilised numerous pre-trained predictive models based on CNNs to diagnose breast cancer from ultrasound pictures that includes AlexNet, DenseNet201, etc [1]. All pre-trained models go through fine-tuning on their top layers because their primary purpose is categorization. Furthermore, the final three layers of the pre-existing models are replaced in response to the current binary classification problem where it raised the weight and bias learning rate parameters to accelerate learning in the newly added layers.

B. Detection, Segmentation and Classification

Zahra Rezaei [5] investigated on image-based methods related to breast cancer where the reader is able to distinguish between many types of techniques, as well as their advantages and disadvantages, thanks to this paper's crucial significance in comprehending the processes of breast cancer detection, segmentation, and classification. She described the methodologies frequently utilized for tumor detection in detail, named the segmentation techniques that are currently being developed, and outlined the shortcomings of such various approaches. The most well-known imaging methods for detection include thermography, mammography, breast ultrasonography, MRI, computed tomography, and magnetic resonance imaging [5]. Although BU, mammography, and MRI are frequently used for detection, they have drawbacks, including reduced accuracy for late diagnoses, radiation exposure, and the inability to show calcification.

Since breast segmentation and post-detection is a complex process, there are certain algorithms which are followed to ensure the correct identification of benign and malignant tumors, i.e pre-processing to remove unwanted regions and extraction of the region of interest to facilitate localization. However, these come with drawbacks such as overlapping areas while thresholding, inability to cover complex shapes of mammography images, speckle noise and shadowing [5]. Feature extraction is carried out followed by feature selection to reduce the dimensionality of large images and to reduce the size of the features respectively. Although the problem of explicit segmentation is present, it can be avoided using texture representation. Tumor classification is done using classifiers developed utilizing machine learning methods. ML classifiers can classify the tumor into one of its three probable forms after features have been identified and chosen. However, when using SVM or hybrid classifiers, we can anticipate seeing a desirable accuracy. Nevertheless, there are some difficulties in the classification field, including the absence of standard image data set, scalability issues, the lack of vast amounts of data, low training samples, unbalanced class distributions and the presence of outliers, all of which may have an impact on the outcome [5].

C. Deep Transfer Learning

Chaudhary et al. [6] demonstrated an approach to identify breast cancer that can work from mammography images, outperforming traditional methods. Recent advancements in deep learning techniques have shown promise in improving breast cancer detection accuracy. This paper emphasizes the utilization of deep transfer learning, FastAI technique, and SqueezeNet architecture for detecting Invasive Ductal Carcinoma (IDC). Transfer learning approach has been shown to improve model performance by leveraging knowledge learned from one domain to another. In the context of breast cancer detection, it helps extract relevant features from medical images and aids in improving classification accuracy [6]. FastAI, a high-level deep learning library, simplifies the process of implementing complex neural network architectures and fine-tuning them for specific tasks. Its user-friendly interface and built-in functionalities allow researchers to efficiently develop and fine-tune models, making it suitable for medical image analysis like IDC detection. SqueezeNet is a compact convolutional neural network architecture designed for efficient image classification. Despite its small size, SqueezeNet maintains competitive accuracy by using fire modules that optimize the balance between model complexity and performance.

This architecture is particularly beneficial for resource-constrained environments and real-time applications. The proposed approach involves applying deep transfer learning using the FastAI library with the SqueezeNet architecture for IDC breast cancer detection. By transferring knowledge from a large medical image dataset to a smaller IDC data set, the model can learn discriminative features relevant to IDC classification. The SqueezeNet architecture's efficiency and effectiveness in feature extraction make it suitable for medical image analysis tasks. Several studies have explored the application of deep transfer learning, FastAI, and SqueezeNet for breast cancer detection. This can achieve up to 90 % success in identifying IDC cases [6]. Furthermore, by utilizing transfer learning and the SqueezeNet architecture in this IDC classification model, the sensitivity and specificity have increased substantially. The combination of deep transfer learning, the FastAI technique, and the SqueezeNet architecture present a promising avenue for enhancing IDC breast cancer detection accuracy. These approaches leverage transfer learning to extract meaningful features from medical images and efficiently fine-tune models for specific diagnostic tasks. [6] The reviewed studies showcase the potential of this approach in achieving higher accuracy and encouraging more efficient clinical decision-making in breast cancer detection.

D. Multiple Instance Learning

Abhijeet Patil [7] emphasized the need for improved breast cancer diagnostics due to rising incidences and its impact on female mortality. Traditional methods like analyzing histopathology slides are precise but time-consuming. The emergence of Computer-Aided Diagnostic (CAD) systems offers faster and consistent results. It focuses on the role of deep learning, especially CNNs like VGG16 and ResNet18, in enhancing diagnostics, introducing the Attention-driven Multiple Instance Learning (A-MIL) method for better image analysis. Within the sphere of medical imaging, deep learning had marked a paradigm shift [8]. Their competencies frequently rival, if not exceed, human expertise in areas like image categorization and object identification. A prominent concern is the perceived ambiguity of these models when applied to medical imaging. While existing methods such as guided backpropagation and Grad-CAM [11] aim to elucidate CNN functionalities, their precision on histopathology images can be inconsistent. A novel solution recently brought to the fore is A-MIL. Crafted to augment the identification of detrimental areas in histopathology images, the A-MIL mechanism assigns a single label to numerous image fragments. In doing so, the generated attention metrics signal the importance of each fragment, boosting detection accuracy. Preliminary tests on public data sets like BreakHIS and BACH have been encouraging. Impressively, the A-MIL approach paralleled, and in some aspects outperformed, esteemed models like VGG16 and ResNet18, especially concerning localization precision.

Traditional CAD solutions predominantly focused on extracting features connected to the appearance and texture of nuclei. Derived metrics from these elements were assessed through techniques as varied as SVM to Gaussian mixture models [12]–[14]. While these conventional methodologies held their ground for smaller data sets, they faltered with expansive ones. Conversely, CNN-centric deep learning models [7], [15] have consistently demonstrated superior adaptability across various data sets, eclipsing their traditional counterparts. While deep learning's role in transforming histopathology image analysis is undeniable, achieving precise localization remains elusive. The advent of the A-MIL technique promises to bridge this gap, heralding a new era of enhanced classification and visualization for breast cancer diagnostics.

III. Methodology

A. *Workflow*

The data set will be initially obtained via Kaggle [23], comprising a collection of images. Resizing is performed in order to optimize processing time and enhance efficiency. The process of image segmentation is accomplished by the utilization of a U-Net model, which is characterized by its U-shaped architectural design. The model undergoes testing using a set of example photos in order to obtain masks that accurately depict the segmented regions inside each image. Finally, Grad-CAM is employed for the purpose of weakly supervised localisation.

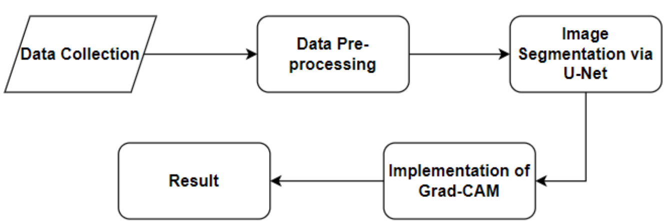

Fig. 1: Workflow of the Methodology

B. *Dataset*

The data set to be used in this iteration of a segmentation model, consists of breast ultrasound images among women within the ages of 25 and 75. The data was collected and compiled in 2018 by academics in Cairo University, Egypt [4]. With a sample population of 600 female patients, there are a total of 780 PNG images in the data set, with an average image size of 500 x 500 pixels. Ground truth images are presented with original images and all of them are categorised into normal, benign and malignant classes. In our case, these images will then need to undergo a pre-processing phase, after which the data set can be used for detection and segmentation of breast cancer.

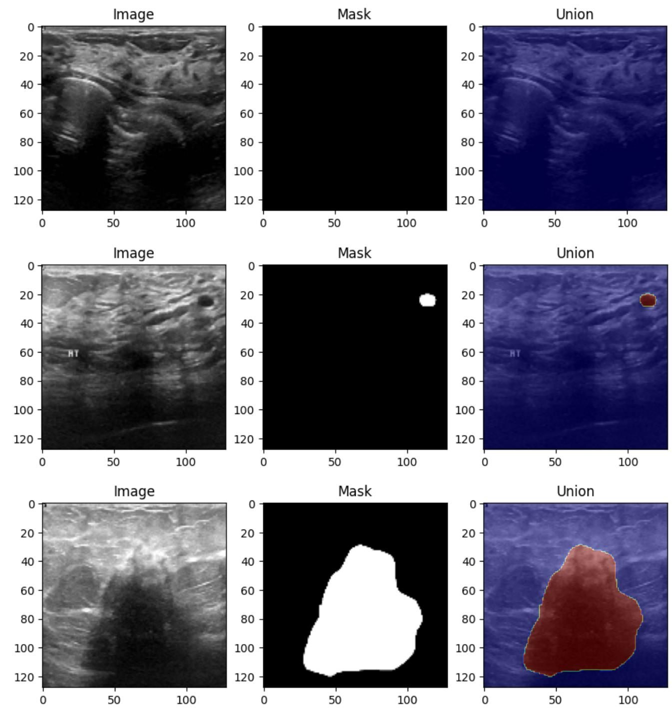

Fig. 2: Normal Class, Benign Class & Malignant Class (Top to Bottom)

C. *U-Net*

This architectural framework encompasses two distinct methodologies. The initial aspect to consider is the trajectory of compression employed by the image in order to capture a background, commonly referred to as the encoder. The encoder consists of a conventional stack of convolutional and maximum pooling layers [24]. The second approach involves a symmetrical expansion direction, commonly referred to as the decoder, which utilizes transposed convolutions to achieve accurate localization. In the process of up-sampling, U-Net employs a mechanism to enable the propagation of background information across a diverse set of feature channels into higher resolution layers. The expanded trajectory exhibits a degree of symmetry with the contracting counterpart, resulting in a u-shaped architectural configuration.

The network exclusively utilizes the valid portion of each convolution, without incorporating any fully connected layers. In order to effectively utilize the network for processing huge photos, it becomes imperative to employ a tiling technique. This strategic approach is essential as it circumvents the limitation imposed by the GPU memory, hence enabling the preservation of image resolution [24].

D. *Gradient-weighted Class Activation Mapping*

The production of a heat map for an input image is achieved through the utilization of the Gradient-weighted class activation map (Grad-CAM) technique. The technique approach employed by Ramprasaath et al. [25] successfully achieved localisation in a single iteration. Typically, the efficiency of

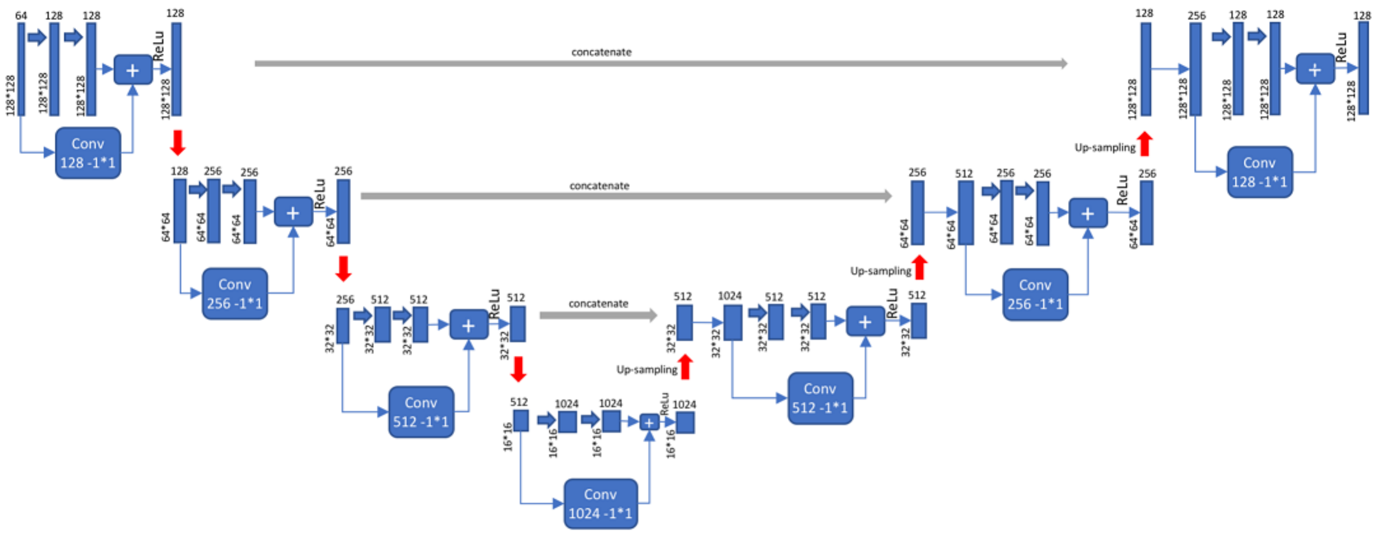

Fig. 3: Illustration of the U-Net Architecture

the process is enhanced by doing only one forward pass and a partial reverse pass per image, resulting in an improvement. Simonyan et al. [26] conducted a study where they employed visualizations of partial derivatives of predicted class scores with respect to pixel intensities. Additionally, they utilized techniques such as Guided Backpropagation [27] and Deconvolution [28] to introduce alterations to the "raw" gradients, leading to qualitative enhancements in the results. This Grad-CAM technique effectively emphasizes the significant regions within the image. Utilizing the gradients obtained from the final convolutional layer's goal allows for this. After removing the feature maps from the previous layer, each channel is given a weight based on the gradient of the class with respect to the channel. According to their relevance to the appropriate class, the observed phenomena shows the extent to which the input image elicits considerable activation in particular channels. There is no requirement to change the current architecture [3] or go through model retraining. This work used multiple class segmentation picture data set, where each image has a variety of items that need to be segmented. Within the classification system, there are three separate categories: normal, malignant, and benign. For each class, the Grad-CAM methodology is employed as a post-hoc approach to provide interpretability by visually representing the heat map of the last layer.

IV. Experimental Results and Analysis

A. Performance Analysis

Multiple performance evaluation criteria, such as accuracy, precision, and F1-score, are employed to assess and verify the effectiveness of the model. While accuracy is a widely used criterion in classification tasks, we conducted a comprehensive evaluation of our model by considering many metrics from different perspectives. The subsequent equations can be employed to articulate the many evaluation measures utilized in this investigation.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{4}$$

B. Comparative Analysis of Various Models

The provided table presents an evaluation of various models for a specific job, where each row corresponds to a model and the columns reflect important metrics used for assessment. It is evident that our model exhibits significantly greater percentages in Accuracy (99.37%), Recall (86.20%), Precision (81.30%), IoU (84.60%), and F1 Score (83.70%) when compared to the other models. Although there has been a slight improvement in the Accuracy and Recall scores, the Precision, Intersection over Union (IoU) and F1 scores have exhibited a substantial gain of at least 10% or more.

Model	Accuracy	Recall	Precision	IoU	F1 Score
MSGRAP [16]	97.79	80.41	74.59	62.26	76.58
PSPNet-18 [17]	97.74	80.88	70.58	60.58	75.20
ENCNet-18 [18]	97.60	79.90	68.59	57.70	72.66
FCN [19]	97.38	77.02	69.07	56.27	71.23
U-Net [20]	97.44	78.46	66.96	56.13	71.32
SegNet [21]	97.58	80.06	68.77	60.01	72.25
Our Model	99.37	86.20	81.30	84.60	83.70

Table I: Quantitative Evaluations of Various Models. The Results Obtained are in Percentage.

C. Training and Validation Results of Loss and Accuracy

The following graphs display the trajectory of loss and accuracy for both the training and validation data sets. A total of 100 epochs were conducted. It is observed that the validation loss initially dropped as the number of epochs increased. However, after reaching approximately the midpoint, the validation loss progressively began to climb. The trend in the train loss is a significant and rapid drop. This observed pattern arises as a consequence of overfitting problems.

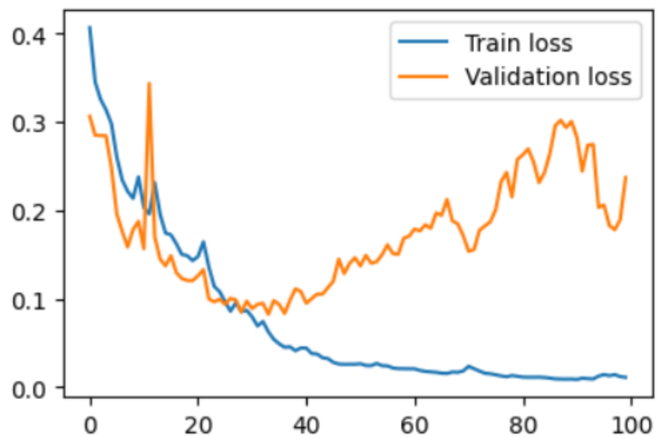

Fig. 4: Curve of Train and Validation of Loss

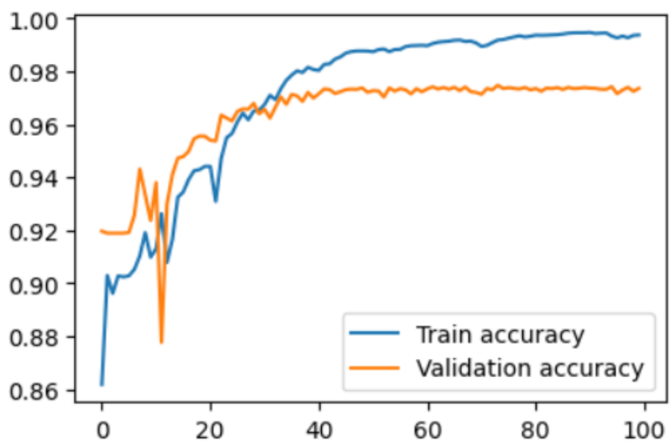

Fig. 5: Curve of Train and Validation of Accuracy

In the initial phase, the validation accuracy dipped for a short number of epochs but later spiked up. The accuracy curve for the training set exhibited a consistent rise during each period and rose over the validation accuracy curve.

D. Grad-CAM Results

The concept of ground truth enables the establishment of a correspondence between picture data and tangible characteristics. The acquisition of ground truth data facilitates the calibration of remote-sensing data, hence assisting in the interpretation and analysis of the sensed information. Predicted image pertains to the resultant feature maps that are generated by a neural network, representing the network's aim to capture significant features from the input image. The Grad-CAM technique leverages the gradients of the classification score in relation to the final convolutional feature map in order to ascertain the specific regions of an input image that exert the greatest influence on the classification score. The locations where the magnitude of this gradient is significant correspond precisely to the areas where the ultimate score is most influenced by the data.

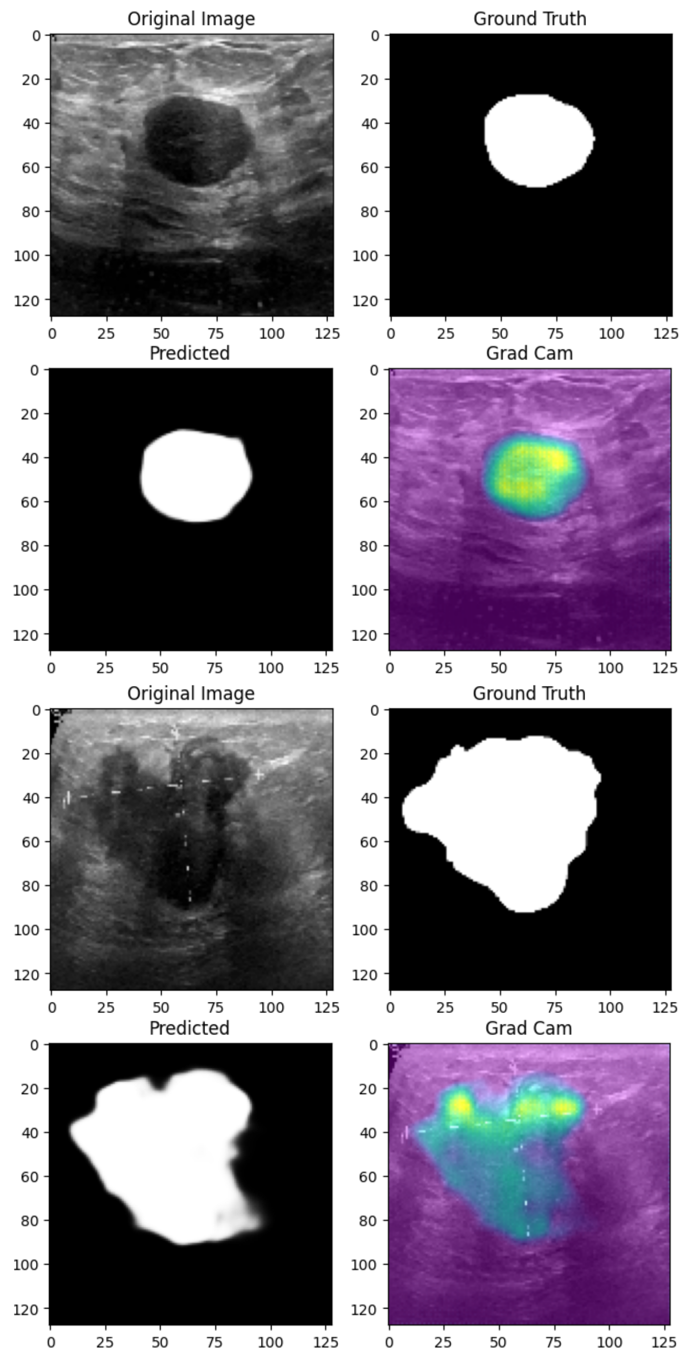

Fig. 6: Exhibition of Heatmaps Yielded using Grad-CAM

V. Future Plans and Challenges

The development and improvement of medical imaging segmentation techniques are crucial for the future of breast cancer detection and diagnosis. Without the assistance of medical professionals, future efforts will examine the possibilities of these techniques. As the existing machine learning techniques rely on expert-annotated images verified by pathology reports, this will provide a substantial barrier. Differentiating segmentation techniques based on the distinctive features, such as mass segmentation and microcalcification segmentation in mammography images, will be a future focus. Reducing incorrect diagnoses brought on by the excessive volume of mammography pictures and the scarcity of radiologists available to interpret them is a crucial concern as well. It is advised that radiologists and experts in artificial intelligence work together more closely to address this. In order to ensure patient safety and welfare, it is important to increase accuracy, cut back on unneeded treatments and expenses, and lower the chance of finding cancer in its late stages.

VI. Conclusion

This research presents a novel approach for segmenting breast cancer images using U-Net architecture, which is a fast scanning convolutional neural network, with the implementation of Grad-CAM. The suggested method is characterised by its robustness and scalability. The method under consideration is thoroughly assessed in terms of its performance and efficiency. The tool's time cost is sufficiently low, rendering it an attractive instrument for practical biological picture analysis.

References

[1] M. Masud et al., 'Pre-Trained Convolutional Neural Networks for Breast Cancer Detection Using Ultrasound Images', Volume 21, Issue 4, Article No.: 85, pp 1–17.

[2] P. S. Rodrigues (2017), "Breast Ultrasound Image", Mendeley Data, V1, doi: 10.17632/wmy84gzngw.1.

[3] H. G. Ramaswamy, "Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization". In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2020; pp. 983–991.

[4] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy (2018). Dataset of breast ultrasound images. Data in Brief. 2020 Feb;28:104863. DOI: 10.1016/j.dib.2019.104863

[5] Z. Rezaei, (2021). A review on image-based approaches for breast cancer detection, segmentation, and classification. Expert Systems With Applications, 182, 115204. https://doi.org/10.1016/j.eswa.2021.115204

[6] S. Chaudhury, K. Sau, M. A. Khan, & M. Shabaz, (2023). Deep transfer learning for IDC breast cancer detection using fast AI technique and Sqeezenet architecture, 20, 2023457. https://doi.org/10.3934/mbe.2023457 .

[7] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," in 2016 International Joint Conference on Neural Networks (IJCNN), Jul. 2016, pp. 2560–2567

[8] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," Insights into Imaging, vol. 9, no. 4, pp. 611–629, Aug. 2018.

[9] X. Zhang, J. Zou, K. He, and J. Sun, "Accelerating Very Deep Convolutional Networks for Classification and Detection," arXiv:1505.06798 [cs], May 2015, arXiv: 1505.06798.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," arXiv:1512.03385 [cs], Dec. 2015, arXiv: 1512.03385.

[11] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," arXiv:1610.02391 [cs], Oct. 2016, arXiv: 1610.02391.

[12] P. Filipczuk, T. Fevens, A. Krzyak, and R. Monczak, "Computer-Aided Breast Cancer Diagnosis Based on the Analysis of Cytological Images of Fine Needle Biopsies," IEEE Transactions on Medical Imaging, vol. 32, no. 12, pp. 2169–2178, Dec. 2013.

[13] Y. M. George, H. H. Zayed, M. I. Roushdy, and B. M. Elbagoury, "Remote Computer-Aided Breast Cancer Detection and Diagnosis System Based on Cytological Images," IEEE Systems Journal, vol. 8, no. 3, pp. 949–964, Sep. 2014.

[14] B. Zhang, "Breast cancer diagnosis from biopsy images by serial fusion of Random Subspace ensembles," in 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), vol. 1, Oct. 2011, pp. 180–186.

[15] A. Golatkar, D. Anand, and A. Sethi, "Classification of breast cancer histology using deep learning," in Image Analysis and Recognition. Cham: Springer International Publishing, 2018, pp. 837–844.

[16] H. Lee, J. Park, and J. Y. Hwang, "Channel Attention Module With Multiscale Grid Average Pooling for Breast Cancer Segmentation in an Ultrasound Image," IEEE Journals & Magazine, IEEE Xplore, https://doi.org/10.1109/TUFFC.2020.2972573.

[17] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.

[18] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, "Context encoding for semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7151–7160.

[19] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 12, pp. 2481–2495, 2017.

[22] World Health Organization. (n.d.). Breast cancer. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/breast-cancer.
[23] Kaggle: Your Machine Learning and Data Science Community. https://www.kaggle.com.
[24] M. Robin, Dr. J. John, and A. Ravikumar, "Breast Tumor Segmentation using U-NET," IEEE Conference Publication, IEEE Xplore. https://doi.org/10.1109/ICCMC51019.2021.9418447
[25] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," arXiv.org, Oct. 07, 2016. https://arxiv.org/abs/1610.02391v4
[26] K. Simonyan, A. Vedaldi, and A. Zisserman. "Deep inside convolutional networks: Visualising Image Classification Models and Saliency Maps," CoRR, abs/1312.6034, 2013. 3, 5
[27] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller. "Striving for Simplicity: The All Convolutional Net," CoRR, abs/1412.6806, 2014. 2, 3
[28] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In ECCV, 2014. 2, 3, 6