Deep Learning Approaches for Segmentation and Classification of Breast Ultrasound Images

Megan Macrì[1], Pierangela Bruno[1,*] and Carmine Dodaro[1]

[1]Department of Mathematics and Computer Science, University of Calabria, Rende, Italy

Abstract

Deep Learning methods have become a powerful tool in medical imaging, with great potential to improve diagnostic accuracy and support early disease detection. This is especially critical for breast cancer, one of the most common cancers among women, where early detection of abnormal tissue is crucial to improve survival rates. In this paper, we explore the application of Deep Learning techniques to segment and classify breast masses as malignant or benign using ultrasound images, aiming to support breast cancer diagnosis. We propose a workflow that integrates two neural networks, i.e., a U-Net for image segmentation and an adapted version of SegNet for classification. Moreover, an ablation study was conducted for parameter tuning and determining their optimal configuration. Finally, our approach was tested on 780 ultrasound images and the results show promising improvements in diagnostic accuracy, demonstrating the potential of our approach to enhance early breast cancer detection.

Keywords

Deep Learning, Classification, Segmentation, Breast Cancer

1. Introduction

Artificial Intelligence (AI) includes a set of algorithms aimed at defining models able to mimic human behavior, such as Machine Learning (ML) and Deep Learning (DL). ML focuses on developing models that can learn from data to perform specific tasks without explicit programming. In particular, ML algorithms identify and extract patterns to make predictions based on the information they receive, representing a huge breakthrough in several fields such as finance, natural language processing, and medical diagnostics [1].

DL, which can be considered a more advanced area of ML, make use of deep neural networks that can automatically learn, represent data at multiple levels of abstraction and extract meaningful patterns directly from raw data, making them an effective tool in fields like image processing, speech recognition, and medical imaging [2].

One of the primary architectures in DL is the Convolutional Neural Networks (CNNs), specifically designed to analyze spatial relationships within images. This ability makes CNNs highly effective for tasks like classification, detection, and segmentation. In the context of medical imaging, CNNs have been widely used to enhance diagnostic accuracy and support clinical providers in the detection of diseases such as cancer [3].

Among the different types of cancer, breast cancer is the most prevalent one affecting women worldwide, making early diagnosis crucial for successful treatment outcomes and reducing mortality rates. According to the World Health Organization (WHO) report, breast cancer accounts for around 25% of all cancer cases in women globally [4]. Ultrasound imaging is a common non-invasive diagnostic tool, particularly useful to dense breast tissue identification where mammography may not be sufficient [5]. However, the interpretation of ultrasound images is highly dependent on the expertise of radiologists. Moreover, the differences between benign and malignant breast masses may be so subtle that the diagnosis becomes challenging. In this context, AI approaches can improve the detection and classification of breast masses, developing automated diagnostic workflow.

HC@AIxIA 2024: Third AIxIA Workshop on Artificial Intelligence For Healthcare, November 25–28, 2024, Bolzano, Italy

*Corresponding author.

macrimegan@gmail.com (M. Macrì); pierangela.bruno@Unical.it (P. Bruno); carmine.dodaro@unical.it/ (C. Dodaro)

In this work, we propose a hybrid framework that combines two well-known DL techniques, namely U-Net for segmentation and SegNet for classification. U-Net was designed for segmenting breast masses and SegNet to determine whether the segmented mass is benign or malignant. Our system enhances both segmentation precision and classification accuracy, providing a comprehensive tool for automated breast cancer diagnosis.

We tested our approach on ultrasound image dataset, collected at Baheya Hospital in Egypt [6], which includes both malignant and benign breast mass images. Given the limited size of the dataset, we employ techniques such as data augmentation and regularization strategies like dropout to mitigate issues related to overfitting and improve model generalization.

2. Related Work

In recent years, several studies employed DL techniques for both classification and segmentation tasks, addressing a broad range of challenges across various domains [7, 8, 9]. These methods have shown remarkable progress in image analysis, enabling more accurate classification and segmentation in medical imaging.

Specifically in breast cancer research, DL models have been widely used in supporting the diagnosis, classification, and detection of breast tumors. Chen et al. [10] proposed an adaptive attention U-Net (AAU-net) for breast lesion segmentation in ultrasound images. The approach is based on U-Net and incorporates a hybrid adaptive attention module (HAAM) to improve segmentation accuracy, particularly suitable for complex ultrasound images with similar intensity distributions and irregular tumor morphology. Several experiments were conducted on three public breast ultrasound datasets to evaluate the performance of the proposed AAU-net, achieving comparable performance over the state-of-the-art methods. Similarly, Zhang et al. [11] presented Swin-Net for performing breast tumor segmentation in ultrasound images. The method integrates Swin-Transformer and CNN techniques to enhance the accuracy of segmentation by leveraging global and multi-scale features. The authors conducted experiments on three datasets, achieving comparable performance w.r.t. literature methods. Zourhri et al. [12] introduced a DL approach for breast cancer diagnosis using ultrasound images, leveraging Transfer Learning with pre-trained models VGG16, VGG19, MobileNetV2, and ResNet50V2. Using a dataset of ultrasound images, the authors showed that VGG19 achieved the highest performance, highlighting the effectiveness of Transfer Learning in enhancing breast cancer classification in computer-aided diagnostic systems.

Segmentation followed by classification has also been used to ensure accurate diagnosis. Islam et al. [13] presented an Ensemble Deep Convolutional Neural Network (EDCNN) model for breast cancer detection and classification. The model was trained on ultrasound images with pre-processing techniques like resizing, normalization, and augmentation to improve performance. The study also integrated Grad-CAM for interpretability and used U-Net for image segmentation, enhancing the identification of breast cancer regions. Inan et al. [14] proposed an end-to-end automated pipeline for breast ultrasonography image classification, addressing the challenges of noisy images and the need for rapid, low-cost diagnosis. Their approach combines U-Net for segmentation, and VGG16 for tumor classification using transfer learning. The framework can help diagnose breast cancer by providing an accurate diagnostic tool. Our approach builds on these recent developments by integrating U-Net for precise segmentation and SegNet for robust classification. By applying data augmentation and pre-processing techniques, and combining two optimized networks, our proposed model aims to enhance both segmentation precision and classification accuracy, addressing common challenges like limited data and overfitting.

3. Proposed Approach

Our approach is composed of different steps, as shown in Figure 1. After pre-processing and reading the images (Step 1), we perform breast mass segmentation using the U-Net architecture (Step 2). When

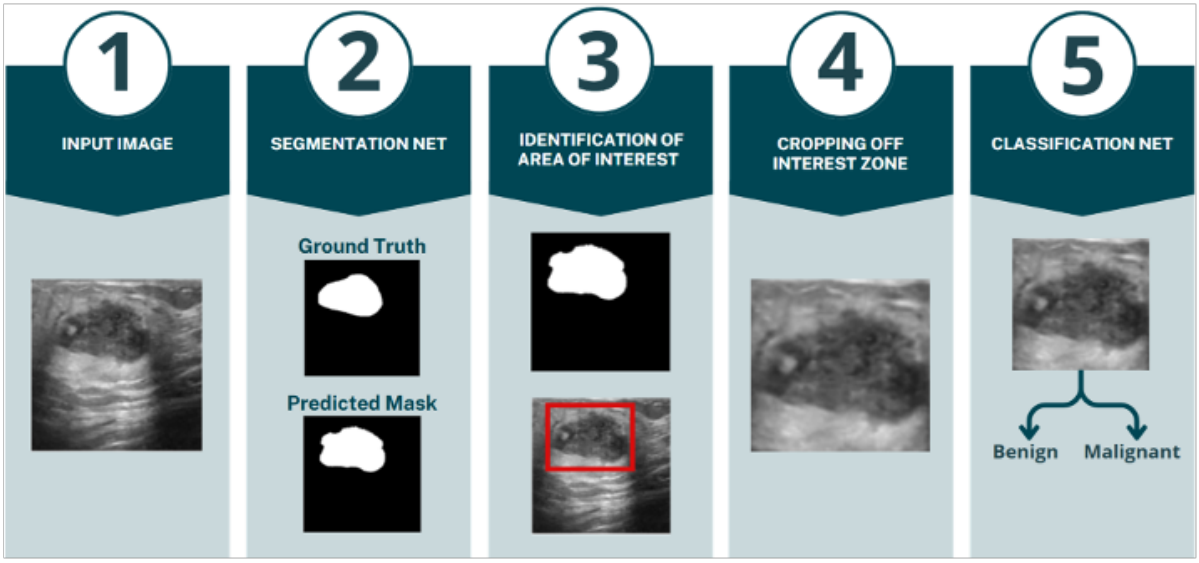

Figure 1: Workflow of the proposed approach

the masses are segmented, we isolate and extract the regions corresponding to the predicted masses (Steps 3-4). Finally, we classify these segmented regions using the SegNet architecture.

3.1. Segmentation

U-Net was chosen for the segmentation task due to its well-established success in medical imaging applications [15, 16]. The network consists of four blocks in both the contracting path and the expansive path. In the contracting path, each block has two convolutional layers with ReLU activation and "same" padding, and each convolutional layer is composed of kernels with a size of 3x3. At the end of each block, there is a Max Pooling layer with a size of 2x2. After the layers of the encoder, there are two convolutional layers with 3x3 kernels and another transposed convolutional layer with 2x2 kernels that leads to the layers of the decoder. The blocks in the expansive path have two convolutional layers with 3x3 kernels and a transposed convolutional layer with 2x2 kernels.

To improve generalization, we applied data augmentation techniques and incorporated a context extrapolation mechanism, ensuring accurate segmentation even at the image edges.

3.2. Classification

Once the tumor regions were segmented, the extracted regions were passed into a SegNet classifier [17]. SegNet uses a similar encoder-decoder structure but differs from U-Net in its approach to upsampling. Instead of using transpose convolutions, SegNet uses stored pooling indices from the encoder to guide the decoder in restoring spatial resolution, making it highly suitable for image classification tasks. This network also has four blocks in both the encoder and the decoder, and each block consists of two convolutional layers with ReLU activation and "same" padding, and at the end of each block, there is a Max Pooling layer and a Dropout with a rate of 0.5.

4. Experimental Analysis

4.1. Dataset description and processing

The dataset used in this study consists of 780 breast ultrasound images [6], collected at Baheya Hospital in Egypt. It includes ultrasound images from women aged between 25 and 75, with a total of about 600 patients. Specifically, 487 images contain benign tumors, the other 210 depict malignant nodules and the remaining are normal cases, meaning no lesion is present.

(a) Benign tumor (b) Malignant tumor (c) Normal case

Figure 2: An example of images for each class; from left to right, benign tumor, malignant tumor and normal case.

The images, originally stored in DICOM format, are converted to PNG format (500x500 pixels) for processing and are split into three categories: benign, malignant, or normal. An example of images for each category is shown in Figure 2. For each ultrasound image, a corresponding mask is provided, which serves as ground truth for the segmentation task. Furthermore, images are standardized and pre-processed to enhance contrast and clarity, ensuring consistency across the dataset. To increase the size of the dataset and, consequently, to mitigate the risk of overfitting, augmentation techniques were applied such as rotation, flipping, and scaling (see Figure 3).

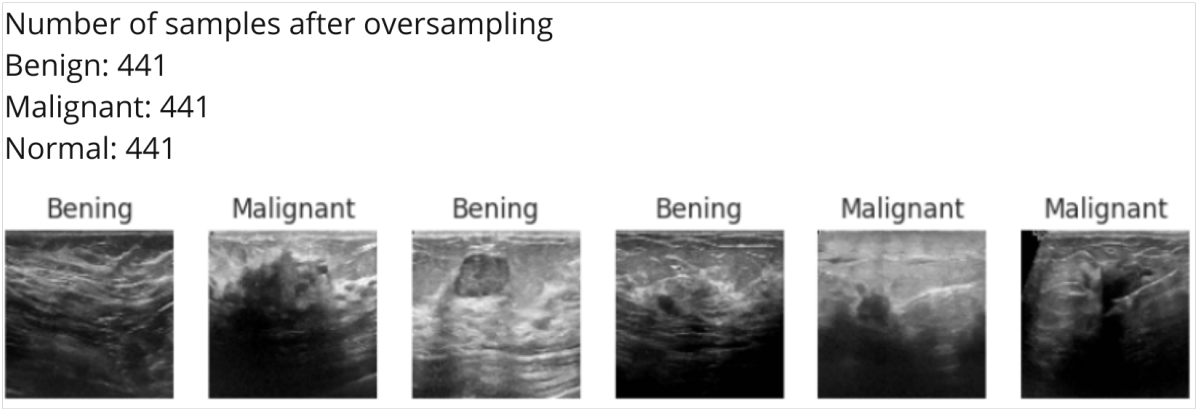

Figure 3: Number of ultrasound images after oversampling application

Specifically, in our experimental analysis, we focused only on *benign* and *malignant* classes, thus, reducing the number of images in 697. The images were pre-processed to reduce their dimensions to 256x256 resolution, converted to grayscale, and the corresponding mask were binarized.

4.2. Experimental setting

In the first experiment of our workflow (Step 2), we split the dataset into training, validation and testing set with a 15-5-80 ratio. For each experiment, we ensured that the dataset was split at the patient-level to prevent any potential information leakage. This means that all images from a particular patient were assigned to either the training, validation, or test set, ensuring that images from the same patient did not appear in both the training and test sets. We deliberately employed a smaller percentage of the dataset for training the segmentation model to ensure a fair evaluation process and reduce the number of images excluded from the subsequent classification step. Since the segmentation model is responsible for extracting regions of interest, it is essential to avoid using the same training data in the classification phase to ensure that the classification step is performed on truly unseen data, maintaining the integrity and robustness of the experimental process. However, this final dataset was still imbalanced and to reduce the risk of overfitting, we performed oversampling to equalize the number of samples in the majority class. Additionally, we employed data augmentation techniques on the training images, applying a 20° rotation, horizontal and vertical translations of 20%, zooming, and horizontal flipping to ensure that the model could generalize despite the reduced size of the initial dataset. The segmentation experiment was trained for 100 epochs with patience of 16 and a batch size of 8; also, we used the Adam optimizer, a learning rate of 0.0001, and binary cross-entropy loss function. To prevent the overfitting problem, we also used the early stopping strategy, finding the right number of epochs on which to set the patience value.

To perform image classification (Step 5), we rely on SegNet. The 80% of the remaining data was used to train the model and was split into training, validation, and testing set (60-10-30).

For classification, the network was trained for 50 epochs with patience 10 and a batch size of 16. We chose a learning rate of 0.00001 and binary cross-entropy as the loss function. Furthermore, dropout layers were used in the classification model to randomly deactivate neurons, and improve generalization ability.

Both models were implemented using TensorFlow and Keras.

Additionally, we performed an ablation study for parameter tuning to identify the most appropriate number of filters to use in each model. Specifically, filters varying from 32 to 128 channels, 16 to 64, and 64 to 512. The number of filters influences the model's ability to learn and represent features. We also experimented with various learning rate combinations.

4.3. Performance Metrics

We make use of accuracy and precision to evaluate the performance of our models. Specifically, accuracy refers to the proportion of correct predictions (both positive and negative) out of the total number of predictions. This is a key metric for assessing how well the model performs, giving an overview of its ability to classify samples correctly in the test set. Precision shows how often a model is correct when predicting the target class, it measures the ratio of true positives to the sum of true positives and false positives. In the medical context, this metric is important as high precision helps minimize the risk of false positives (i.e., misclassifying healthy or benign tissue as malignant) which can lead to unnecessary treatments and stress for patients.

5. Results and Discussion

The experimental results are summarized in Table 1 and Table 2, which detail the outcomes of the segmentation and classification tasks performed using U-Net and SegNet, respectively.

5.1. Segmentation Results

The U-Net architecture was applied to the segmentation of breast masses, with four experiments conducted using different learning rates and filter configurations. As shown in Table 1, the best

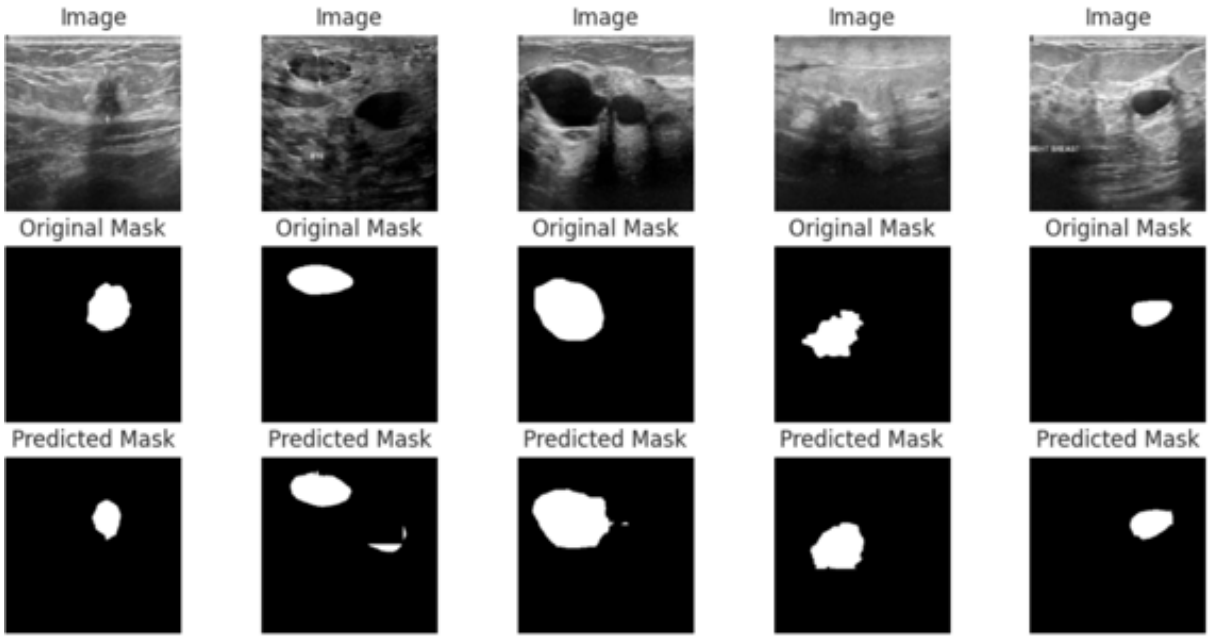

Figure 4: Example of the segmentation predictions along with raw images and ground truth masks achieved with configuration reported in Experiment 1.

performance was achieved in *Experiment 1*, where the learning rate was set to 0.0001 and filters were configured to range between 64 and 512 channels. A visual example of the results is shown in Figure 4.

This setup resulted in a test set accuracy of 93.23%. The smooth convergence of the training and validation accuracy curves in Figure 5 illustrates that the learning rate, in this case, was optimal, allowing the model to learn effectively without overfitting.

In *Experiment 4*, despite using similar filter configurations, the model exhibited signs of overfitting after epoch 60, as shown in Figure 6.

5.2. Classification Results

For the classification task, SegNet was employed to differentiate between benign and malignant breast masses. As shown in Table 2, *Experiment 2* obtained the best results, achieving 98% test accuracy with 99% precision for benign tumors and 97% precision for malignant tumors. The optimal performance was achieved with a learning rate of 0.00001 and filters ranging from 16 to 64 channels, demonstrating that a lower learning rate was suitable for this classification task. This setup allowed the model to achieve high precision without overfitting.

The performance of *Experiment 3* was also considerable, with a test set accuracy of 96%, although it showed slightly lower precision in identifying malignant tumors (93%).

5.3. Discussion

The results highlight the importance of selecting the appropriate learning rate and filter sizes for each task. For segmentation, larger filter sizes (64-512 channels) and a moderate learning rate (0.00001) provided the best performance. In contrast, for classification, smaller filter sizes (16-64 channels) and a lower learning rate (0.00001) were more effective, allowing the model to achieve high precision, particularly for malignant tumors.

The overfitting observed in *Experiment 4* during segmentation underscores the need for careful regularization; then, early stopping and data augmentation were applied to mitigate overfitting.

Overall, combining U-Net for segmentation with SegNet for classification demonstrated robust performance, with high accuracy in both tasks, making it a promising tool for automated breast cancer

diagnosis in ultrasound images. The results are highly encouraging, showing potential for clinical applications in early cancer detection.

Table 1
Results for U-Net segmentation experiments with varying learning rates and filter sizes.

Experiment	Learning Rate	Filters	Test Set Accuracy
1	0.0001	64-512	93.23%
2	0.0004	32-128	90.00%
3	0.00001	32-128	92.54%
4	0.00001	64-512	93.21%

Table 2
Results for SegNet classification experiments with varying learning rates and filter sizes.

Experiment	Learning Rate	Filters	Test Set Accuracy	Precision
1	0.0001	16-64	95%	88% Benign, 100% Malignant
2	0.00001	16-64	98%	99% Benign, 97% Malignant
3	0.00001	32-256	96%	100% Benign, 93% Malignant

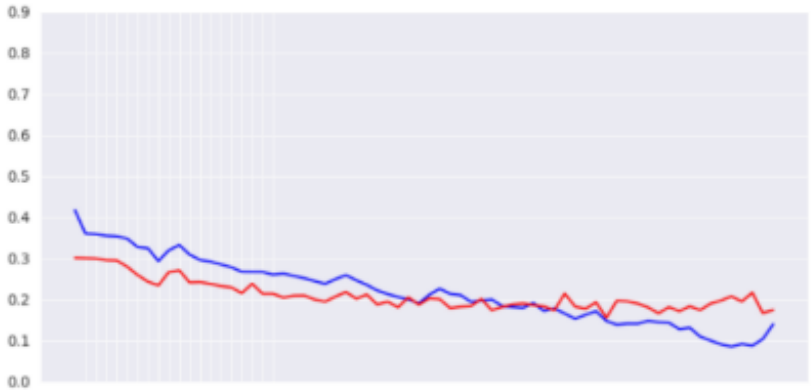

Figure 5: Plot of the accuracy curves for training and validation sets during segmentation (Experiment 1).

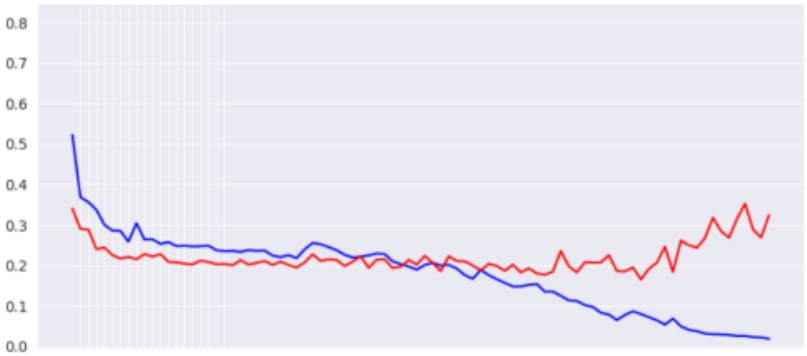

Figure 6: Overfitting observed during segmentation in Experiment 4 after epoch 60.

6. Conclusion

In this study, we proposed a hybrid DL framework that combines U-Net for breast mass segmentation and an adapted version of SegNet for the classification of benign and malignant tumors in ultrasound images. Our experiments demonstrated that the model achieved robust performance, with an accuracy of 93.23% in the segmentation task and a 98% accuracy in classification with high precision in identifying malignant tumors.

These results highlight the importance of carefully tuning hyperparameters, such as learning rate and filter sizes, to balance the complexity and performance of the model. Additionally, techniques like data augmentation and early stopping were crucial in preventing overfitting and improving generalization ability.

Our approach demonstrates a good capability in providing an automated and reliable tool for early breast cancer detection. Future work may focus on applying the model to larger and more diverse datasets, as well as exploring advanced methods such as transfer learning to further enhance its diagnostic capabilities.

Acknowledgments

Carmine Dodaro was supported by the European Union - NextGenerationEU and by Italian Ministry of Research (MUR) under PNRR project FAIR "Future AI Research", CUP H23C22000860006 and by the European Union - NextGenerationEU and by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.5, project "RAISE - Robotics and AI for Socio-economic Empowerment" (ECS00000035) under the project "Gestione e Ottimizzazione di Risorse Ospedaliere attraverso Analisi Dati, Logic Programming e Digital Twin (GOLD)", CUP H53C24000400006; and by GNCS-INdAM. Pierangela Bruno has been partially supported by PON "Ricerca e Innovazione" 2014-2020, CUP: H25F21001230004.

References

[1] M. I. Jordan, T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, Science 349 (2015) 255–260.

[2] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (2015) 436–444.

[3] R. Aggarwal, V. Sounderajah, G. Martin, D. S. Ting, A. Karthikesalingam, D. King, H. Ashrafian, A. Darzi, Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis, NPJ digital medicine 4 (2021) 65.

[4] L. Wilkinson, T. Gathani, Understanding breast cancer as a global health concern, The British journal of radiology 95 (2022) 20211033.

[5] R. Guo, G. Lu, B. Qin, B. Fei, Ultrasound imaging technologies for breast cancer detection and management: a review, Ultrasound in medicine & biology 44 (2018) 37–70.

[6] H. K. W. Al-Dhabyani, M. Gomaa, A. Fahmy, Dataset of breast ultrasound images, Elsevier 28 (2020).

[7] T. Roß, P. Bruno, A. Reinke, M. Wiesenfarth, L. Koeppel, P. M. Full, B. Pekdemir, P. Godau, D. Trofimova, F. Isensee, et al., Beyond rankings: learning (more) from algorithm validation, Medical image analysis 86 (2023) 102765.

[8] P. Bruno, M. F. Spadea, S. Scaramuzzino, S. De Rosa, C. Indolfi, G. Gargiulo, G. Giugliano, G. Esposito, F. Calimeri, P. Zaffino, Assessing vascular complexity of paod patients by deep learning-based segmentation and fractal dimension, Neural Computing and Applications 34 (2022) 22015–22022.

[9] H. Jiang, Z. Diao, T. Shi, Y. Zhou, F. Wang, W. Hu, X. Zhu, S. Luo, G. Tong, Y.-D. Yao, A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation, Computers in Biology and Medicine 157 (2023) 106726.

[10] G. Chen, L. Li, Y. Dai, J. Zhang, M. H. Yap, Aau-net: an adaptive attention u-net for breast lesions segmentation in ultrasound images, IEEE Transactions on Medical Imaging 42 (2022) 1289–1300.
[11] C. Zhu, X. Chai, Y. Xiao, X. Liu, R. Zhang, Z. Yang, Z. Wang, Swin-net: A swin-transformer-based network combing with multi-scale features for segmentation of breast tumor ultrasound images, Diagnostics 14 (2024) 269.
[12] M. Zourhri, S. Hamida, N. Akouz, B. Cherradi, H. Nhaila, M. El Khaili, Deep learning technique for classification of breast cancer using ultrasound images, in: 2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), IEEE, 2023, pp. 1–8.
[13] M. R. Islam, M. M. Rahman, M. S. Ali, A. A. N. Nafi, M. S. Alam, T. K. Godder, M. S. Miah, M. K. Islam, Enhancing breast cancer segmentation and classification: An ensemble deep convolutional neural network and u-net approach on ultrasound images, Machine Learning with Applications 16 (2024) 100555.
[14] M. S. K. Inan, F. I. Alam, R. Hasan, Deep integrated pipeline of segmentation guided classification of breast cancer from ultrasound images, Biomedical Signal Processing and Control 75 (2022) 103553.
[15] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer, 2015, pp. 234–241.
[16] N. Siddique, S. Paheding, C. P. Elkin, V. Devabhaktuni, U-net and its variants for medical image segmentation: A review of theory and applications, IEEE access 9 (2021) 82031–82057.
[17] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE transactions on pattern analysis and machine intelligence 39 (2017) 2481–2495.