UGGNet: Bridging U-Net and VGG for Advanced Breast Cancer Diagnosis

Tran Cao Minh[1], Nguyen Kim Quoc[1], Phan Cong Vinh[1], Dang Nhu Phu[1], Vuong Xuan Chi[1], Ha Minh Tan[1,*]

[1]Faculty of Information Technology Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

Abstract

In the field of medical imaging, breast ultrasound has emerged as a crucial diagnostic tool for early detection of breast cancer. However, the accuracy of diagnosing the location of the affected area and the extent of the disease depends on the experience of the physician. In this paper, we propose a novel model called UGGNet, combining the power of the U-Net and VGG architectures to enhance the performance of breast ultrasound image analysis. The U-Net component of the model helps accurately segment the lesions, while the VGG component utilizes deep convolutional layers to extract features. The fusion of these two architectures in UGGNet aims to optimize both segmentation and feature representation, providing a comprehensive solution for accurate diagnosis in breast ultrasound images. Experimental results have demonstrated that the UGGNet model achieves a notable accuracy of 78.2% on the "Breast Ultrasound Images Dataset."

1. Introduction

Breast cancer is a type of cancer originating from the cells of the breast, often arising from the milk duct cells or surrounding cells. This is one of the most common types of cancer in women worldwide [11, 15]. To examine and detect the disease early, ultrasound imaging can be employed for diagnosis and monitoring of breast cancer. Ultrasound procedures can help determine the size of the tumor, and its characteristics, and identify suitable treatment options [20]. Based on the ultrasound images, the doctor remains the direct diagnostic authority regarding the patient's illness severity. There are two potential scenarios that a patient may face upon diagnosis: the ailment is either benign or malignant[8]. In recent years, deep learning models have demonstrated outstanding capabilities in predicting and classifying diseases, particularly when employing convolutional techniques [5, 18]. This is considered a crucial method in image information processing. The application of convolutions has expanded into various medical fields,

with the advantage of efficiently processing large image and data sets [10, 29]. In the domain of disease diagnosis, particularly in breast cancer, convolutional techniques have proven to be effective. In particular, research on cancer diagnosis utilizing medical imaging is a promising area, as it has been deployed to predict diseases based on the progression of patients over time, employing a self-attention-based model proposed by Aishik Konwer et al., incorporating a Temporal Convolutional Network (TCN) [16]. Manu Subramoniam et al. have suggested the application of the Resnet model to predict Alzheimer's disease from MRI images [26]. Additionally, Ahmet Solak et al. have proposed the use of a U-Net model for segmenting tumor masses within the adrenal gland [25].

In this paper, we propose the architecture of UGGNet and present experimental results on the BUSI dataset. Concurrently, this research discusses prospects and challenges in the future development direction of the topic, raising crucial issues that need to be addressed to enhance the efficiency and practical application of UGGNet in the field of medical diagnosis, specifically within the realms of computer vision and deep learning.

---

*Corresponding author. Email: hmtan@ntt.edu.vn

2. Literature Review

Since the advent of the Backpropagation algorithm by Professor Geoffrey Hinton in 1986 [23], research applying deep learning (DL) to various aspects of life and medicine has been burgeoning almost daily. Breast cancer, considered a typical disease affecting women, has become a crucial and promising research area. With the advancement of DL, machine learning and deep learning models have been developed to handle complex medical data and provide accurate predictions of patients' health conditions. Mesut TOĞAÇAR et al. [28] applied the Support Vector Machine (SVM) model to train on a dataset of 700 images, including benign and malignant variants of breast cancer images. The images were analyzed using a convolutional neural network (CNN) approach, and the input images were passed through the AlexNet model [17] for feature extraction. The extracted features were then combined with the SVM model for classification, achieving an impressive accuracy of 0.934. Ashutosh Kumar Dubey et al. [7] utilized the Breast Cancer Wisconsin (BCW) dataset, a structured dataset derived from digitized images of tumors. They described the characteristics of cell nuclei within the tumors and applied the K-means clustering algorithm, involving cluster initialization, distance measurement to the nearest clusters, and cluster optimization. The results yielded an average accuracy of 0.92 in experimental trials. Using the same BCW dataset, Omar Ibrahim Obaid et al. [21] experimented with SVM, K-nearest neighbors (KNN), and Decision tree models, achieving an accuracy of 0.981. Notably, the Quadratic SVM kernel $\{K(x, y) = (x \cdot y + c)^d\}$ demonstrated superior performance when dealing with complex relationships that cannot be well-classified by a simple linear boundary. Ensemble models have also emerged as a new approach. Mohamed Hosni et al. [13] combined three individual models, Decision tree + SVM + Artificial Neural Network (ANN), to create a comprehensive strength. Abien Fred M. Agarap [1] contributed a Multi-layer Perceptron (MLP) model, optimized the hyperparameters, and achieved an outstanding accuracy of 0.99, demonstrating the effectiveness of deep learning models.

In recent years, the "Dataset of breast ultrasound images - BUSI" [3] has emerged as a benchmark for researchers to experiment with their models and compare results. Michal Byra et al. [4] proposed a segmentation model based on U-Net [22] combined with Selective Kernel (SK), achieving a Dice score of 0.778. Jorge F. Lazo et al. [19] compared the effectiveness of different CNN architectures in classifying benign and malignant breast masses from ultrasound images. The two CNN architectures used were VGG-16 [24] and Inception-V3 [27]. Two training strategies were evaluated: using pre-trained models as feature extractors and fine-tuning pre-trained models. The dataset comprised 947 ultrasound images, including 587 images of benign masses and 360 images of malignant masses. Performance metrics used were accuracy and AUC (Area Under the ROC Curve). The results showed that fine-tuning VGG-16 achieved the best performance with an accuracy of 0.919 and AUC of 0.934. The comparison between the two training strategies indicated that fine-tuning the model generally outperformed using feature extraction.

Several data augmentation methods have also been implemented by Walid Al-Dhabyani et al. [2] to explore the use of deep learning in breast ultrasound image-based breast tumor classification. Specifically, the study focuses on two main aspects: identifying the significant problem of insufficient large datasets, which diminishes the performance of classification models. To address this challenge, the research proposes the use of data augmentation methods, including flipping, rotation, and adding noise, aiming to create a diverse and larger training dataset. In the realm of deep learning classification, the study concentrates on various architectures such as CNN, ResNet, and DenseNet to determine the benign or malignant nature of breast tumors. These architectures represent advancements in the field of deep learning and are effectively applied to medical image classification tasks. Additionally, the research investigates the performance of transfer learning models like VGG16 and MobileNet. Utilizing these models can leverage knowledge previously learned from large datasets, enhancing the classification ability of the model for breast ultrasound images. Among the studied architectures, DenseNet achieves the highest accuracy with an AUC of 0.976. In the study by Behnaz Gheflati et al. [9], the authors employed the Vision Transformer (ViT) architecture for breast ultrasound image classification. The results indicate that ViT achieves a classification accuracy of 0.79, with an AUC of 0.84, outperforming ResNet. UGGNet is constructed upon the architecture of U-Net, where the encoder is substituted with convolutional blocks from VGGNet. The study draws inspiration from the Capsule Network (CapsNet) model, integrating it with the U-Net architecture, resulting in a hybrid model termed CapUnet [12]. This integration enables UGGNet to acquire intricate features from breast images while simultaneously retaining the classification capabilities inherited from VGGNet.

3. Proposed method

The architecture UGGNet (Fig 1) consists of two main components: Segmentation with U-Net and
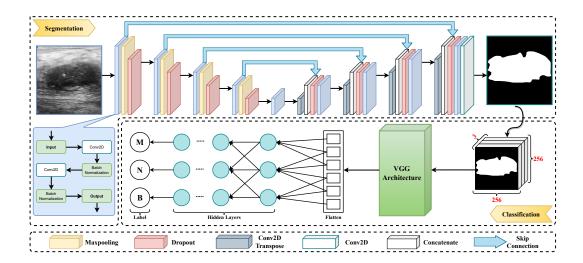
Figure 1. Proposed UGG Architecture

Classification with VGGNet. The encoder process includes the following steps: the input image, with dimensions $256 \times 256 \times 3$, undergoes a custom-defined Conv2D-Block (shown in blue), which consists of two Conv2D layers and two BatchNormalization layers alternately. Subsequently, a Max_Pooling layer (shown in yellow) with a kernel size of 2×2 follows the Conv2D-Block. At this point, the image is reduced in size by half compared to the image output from the Conv2D-Block, while the depth of the image is doubled. Assuming the previous image had dimensions $256 \times 256 \times 16$, the resulting image will have dimensions $128 \times 128 \times 32$. Next is a block (shown in red) with a 0.3 scaling factor, representing a Dropout layer. This process is repeated four times, starting from the input image $256 \times 256 \times 3$ and going through four layers, each defined as En_L_ (described in table 1). The layer specifications are as follows:

Table 1. Image shape during segmentation process

No.	Layer	Shape
1	Input	(256 x 256 x 3)
2	En_L_1	(128 x 128 x 16)
3	En_L_2	(64 x 64 x 32)
4	En_L_3	(32 x 32 x 64)
5	En_L_4	(16 x 16 x 128)
6	Last Image	(16 x 16 x 256)
7	De_L_1	(32 x 32 x 128)
8	De_L_2	(64 x 64 x 64)
9	De_L_3	(128 x 128 x 32)
10	De_L_4	(256 x 256 x 16)
11	Out Image	(256 x 256 x 1)

The "Last Image" is the final image after the Max_Pooling process in the 4th layer. At this point,

the image contracts to dimensions $16 \times 16 \times 256$, significantly increasing the depth to 256 to capture as many features as possible. Following this, the decoder process begins: the "Last Image" is decoded and its dimensions increased by a factor of 2 using Conv2DTranspose. Similar to the encoder, the decoder consists of four layers defined as "De_L_". The output image "Out Image" of the 4th decoder layer is an image with dimensions $256 \times 256 \times 1$. This image is referred to as the mask image in the segmentation process, signifying the completion of the segmentation process.

Following segmentation is the classification process. The image with dimensions $256 \times 256 \times 1$ cannot be directly input into the VGG architecture because the input requires dimensions of ($width$, $height$, 3). Therefore, the "Out Image" is passed through a Conv2D layer with a filter size of 3, which increases the depth of the image from $256 \times 256 \times 1$ to $256 \times 256 \times 3$. This image is then used as input for the VGG architecture. Subsequently, VGG produces output that stops at the layer Conv5-4 for VGG19 and Conv5-3 for VGG16. However, this output is not yet suitable for prediction.

At this point, a Flatten layer is needed to "flatten" the tensor into a 1D vector with a shape of (batch_size, 512). This 1D vector is then fed into fully connected layer, and its output passes through a softmax activation function with three elements corresponding to the three labels (normal - N), (benign - B) and (malignant - M).

4. Experiments

4.1. Dataset

The dataset used in this study is named "Dataset of Breast Ultrasound Images" [3], collected using the LOGIQ E9 ultrasound system [6]. It comprises 780

ultrasound images of female breasts, data was collected in 2018, focusing on the age group ranging from 25 to 75 years old. In total, this dataset comprises visual information of 600 female patients. The dataset is categorized into three labels: normal, benign, and malignant. Ultrasound images are associated with a corresponding mask image used for segmentation tasks. The average image size of the entire dataset is 500×500 pixels, stored in PNG format. 80% will be used to train the UGGNet model, the remaining 20% will be used for the final test set.
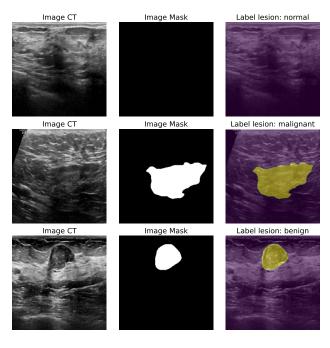

Figure 2. Ultrasound image with disease severity label

4.2. Metrics

Loss function. The categorical cross-entropy loss function, denoted as $H(y, \hat{y})$, is a method for measuring the difference between the actual distribution y (true labels) and the predicted distribution $\hat{y}$ (predicted probabilities) in a classification problem. The formula for the loss function is expressed as follows:

$$H(y, \hat{y}) = -\sum_i^N y_i \cdot \log(\hat{y}_i) \tag{1}$$

Here, y represents the actual distribution, where y_i is the probability of class i in the actual distribution. Conversely, $\hat{y}$ is the predicted distribution, and $\hat{y}_i$ is the predicted probability for class i. Through the formula, each element y_i combined with the natural logarithm of $\hat{y}_i$ is used to measure the discrepancy between the actual label and the corresponding prediction.

The detailed algorithm is presented in Algorithm 1. The goal is to minimize the value of the loss function,

---

Algorithm 1: Categorical Cross-Entropy Loss Calculation

Data: Actual distribution y, Predicted distribution $\hat{y}$
Result: Categorical Cross-Entropy Loss $H(y, \hat{y})$

1 **Initialization:** Set $H(y, \hat{y})$ to 0;
2 **foreach** *class i* **do**
3      $H(y, \hat{y}) \leftarrow H(y, \hat{y}) - y_i \cdot \log(\hat{y}_i)$;
4 **Output:** Categorical Cross-Entropy Loss $H(y, \hat{y})$;

---

which is synonymous with making the predicted distribution close to the actual distribution. This is an important tool in the process of training machine learning models, helping to shape and update the model's weights to optimize classification performance.

Evaluation Metrics. To assess the performance of a classification model, we use a Confusion Matrix. It operates on test data by categorizing predictions into four main types: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Table 2. Confusion Matrix

	True Class	
Predicted Class	True Positives	False Positives
	False Negatives	True Negatives

True Positive represents the number of cases where the model correctly predicts positive outcomes. True Negative is the number of cases where the model correctly predicts negative outcomes. False Positive is the number of cases where the model incorrectly predicts positive outcomes and False Negative is the number of cases where the model incorrectly predicts negative outcomes.

Based on these four values, we can calculate various important metrics to evaluate the performance of the model, such as Recall, Accuracy, and F1-score.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Accuracy is the percentage ratio of the number of correct predictions to the total number of data points. It provides an overall measure of the model's accuracy.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

Recall measures the model's ability to correctly identify cases that truly belong to a specific class. The Recall formula is the number of correct predictions divided by the total number of cases that belong to that class.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

F1-score is a combination of Precision, Recall. Precision is the model's ability to make accurate predictions for cases predicted to belong to a specific class.

4.3. Implementations

There are two main versions of UGGNet, namely UGG_19 and UGG_16. UGG_19 utilizes the U-Net architecture for feature extraction and employs VGG19 for classification, whereas UGG_16 also utilizes U-Net but performs classification using VGG16.

Training Parameters: During the training process of the machine learning model, the research utilized 80% of the data for training and 20% for model validation to ensure accuracy and performance. To initiate the training process, a learning rate (lr) of 0.0001 was employed, and a learning rate reduction was applied every 20 epochs if the validation accuracy (val_accuracy) did not improve. The learning rate decay factor (Factor_Decay_Lr) was set to 0.8.

For the training phase, a batch size (Batch_Size) of 64 was selected to optimize model weight updates. Additionally, to prevent overfitting, dropout with a rate of 0.3 was applied to avoid excessive learning of specific features in the training data. The training process iterated over 500 epochs; however, to avoid resource consumption without significant improvement, Early Stopping was employed. The model would stop training after 100 epochs if the validation accuracy did not increase. To evaluate the model's performance, 20% of the training data was used as a validation set (Validation_Split). This approach helped control the training process and assess the model more generally.

Finally, for model optimization, the research chose "adam" as the optimizer and "categorical_crossentropy" as the loss function to ensure effective learning and the best performance measurement on the test set in the field of computer vision and deep learning.

In the table 3, UGG_19 model, featuring a 7-layer architecture, appears significantly more complex compared to its predecessors. Designed with a total of 23,414,234 parameters, including 3,386,906 trainable parameters. With units per layer sequentially set at 1024, 512, 256, 128, 64, 32, and 16, the augmentation of both the number of layers and units may imply enhanced learning capability and improved representation of complex data.

4.4. Experimental Results

The experimental setup will encompass six cases, incorporating UGG_16 and UGG_19 architectures, each characterized by the number of layers and units in the fully connected layers, specifically 3, 5, and 7 layers. In the table 4 details of two models, UGG_19 and UGG_16, based on the number of layers. For UGG_19, training with 3, 5, and 7 layers concluded at epochs

229, 178, and 285, respectively. The corresponding accuracies were 0.7628, 0.7564, and 0.7821, recall rates were 0.7580, 0.7564, and 0.7821, and F1 scores were 0.7628, 0.7482, and 0.7754. For UGG_16, training with 3, 5, and 7 layers ended at epochs 292, 239, and 453. The accuracies were 0.7436, 0.7564, and 0.6923, recall rates were 0.7436, 0.7564, and 0.6923, and F1 scores were 0.7390, 0.7511, and 0.6991. Training durations varied, ranging from approximately 00:54:41 to 02:12:00.

Compared to 2020 studies, Lazo et al. [19] optimized Inception V3, achieving accuracies of 0.713 and 0.756. In 2021, Irfan et al. [14] proposed the di-Cnn Model, combining Densenet201 with a 24-Layer CNN, yielding an accuracy of 0.7961, detailed in the table 5

The provided table presents a comparative analysis of various deep learning models, with a specific emphasis on Di-CNN and UGG_19. Di-CNN, which employs DenseNet201 in conjunction with a 24-layer CNN, distinguishes itself with an impressive accuracy of 0.7961. However, a notable drawback lies in its extensive training duration, requiring 197 hours, 3 minutes, and 28 seconds. In contrast, UGG_19 adopts a different strategy, leveraging U-Net for feature extraction and utilizing VGG19 with an additional 7 layers for classification. Despite a shorter training period of 1 hour, 26 minutes, and 4 seconds, UGG_19 achieves a competitive accuracy of 0.7821. Remarkably, UGG_19 attains commendable results through the synergistic combination of U-Net and VGG19 with additional layers, underscoring the importance of thoughtful architectural selection. This highlights the effectiveness of UGG_19 as a model that strikes a balance between accuracy and training efficiency.

5. Conclusions

The UGGNet model is proposed for identifying features in medical images, leveraging a unique combination of the U-Net and VGGNet architectures to achieve high performance. There are two main versions of UGGNet: UGG_19 and UGG_16. UGG_19 employs the U-Net architecture for feature extraction and VGG19 for classification, whereas UGG_16 also utilizes U-Net but employs VGG16 for classification. Experimental results have demonstrated that UGG_19 achieves impressive performance, with the highest accuracy reaching 0.7821. Both Recall and F1 score are also impressive, at 0.7821 and 0.7754, respectively. The model was trained in an impressive time span of 1 hour, 26 minutes, and 4 seconds. The results of UGG_19 showcase an effective synergy between the detailed feature extraction capability of U-Net and the classification prowess of VGG19. This outcome underscores the strength of employing hybrid architectural approaches to create a practical and efficient model for identifying features in medical images.

Table 3. UGGNet model size

No.	Model	Num layer	Unit on layer	Total parameters	Training parameters
1	UGG_16	3	1024/512/256	18.061.498	3.343.866
2	UGG_16	5	1024/512/256/128/64	18.102.074	3.384.442
3	UGG_16	7	1024/512/256/128/64/32/16	18.104.538	3.386.906
4	UGG_19	3	1024/512/256	23.371.194	3.343.866
5	UGG_19	5	1024/512/256/128/64	23.411.770	3.384.442
6	UGG_19	7	1024/512/256/128/64/32/16	23.414.234	3.386.906

Table 4. Experimental results of UGGNet model

No.	Model	Num layer	Epoch Stop	Training time	Accuracy	Recall	F1
1	UGG_16	3	292/500	01:24:14	0.7436	0.7436	0.739
2	UGG_16	5	239/500	01:10:03	0.7564	0.7564	0.7511
3	UGG_16	7	453/500	02:12:00	0.6923	0.6923	0.6991
4	UGG_19	3	229/500	01:11:11	0.7628	0.758	0.7628
5	UGG_19	5	178/500	00:54:41	0.7564	0.7564	0.7482
6	UGG_19	7	285/500	01:26:04	0.7821	0.7821	0.7754

Table 5. Compare the results of UGGNet with previous research

No.	Model	Method	Training time	Accuracy
1	Inception V3 [19]	Feature extraction with CNN	-	0.7131
2	Inception V3 - Fine-tuning [19]	Feature extraction with CNN combined with hyperparameter optimization	-	0.7561
3	Di-CNN [14]	DenseNet201+ CNN 24 layer	197:03:28	0.7961
4	UGG_19	Feature extraction with U-Net and classification with VGG19 + 7 Layer	01:26:04	0.7821
5	UGG_16	Feature extraction with U-Net and classification with VGG16 + 5 Layer	01:10:03	0.7564

References

[1] Abien Fred M Agarap. On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In *Proceedings of the 2nd international conference on machine learning and soft computing*, pages 5–9, 2018.

[2] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Fahmy Aly. Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. *Int. J. Adv. Comput. Sci. Appl*, 10(5):1–11, 2019.

[3] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in brief*, 28:104863, 2020.

[4] Michal Byra, Piotr Jarosik, Aleksandra Szubert, Michael Galperin, Haydee Ojeda-Fournier, Linda Olson, Mary O'Boyle, Christopher Comstock, and Michael Andre. Breast mass segmentation in ultrasound with selective kernel u-net convolutional neural network. *Biomedical Signal Processing and Control*, 61:102027, 2020.

[5] Arkapravo Chattopadhyay and Mausumi Maitra. Mri-based brain tumour image detection using cnn based deep learning method. *Neuroscience informatics*, 2(4):100060, 2022.

[6] Erqiang Deng, Zhiguang Qin, Dajiang Chen, Zhen Qin, Yi Ding, Ji Geng, and Ning Zhang. Engan: Enhancement generative adversarial network in medical image segmentation. 2022.

[7] Ashutosh Kumar Dubey, Umesh Gupta, and Sonal Jain. Analysis of k-means clustering approach on the breast cancer wisconsin dataset. *International journal of computer assisted radiology and surgery*, 11:2033–2047, 2016.

[8] Oliver Faust, U Rajendra Acharya, Kristen M Meiburger, Filippo Molinari, Joel EW Koh, Chai Hong Yeong, Pailin Kongmebhol, and Kwan Hoong Ng. Comparative assessment of texture features for the identification of cancer in ultrasound images: a review. *Biocybernetics and Biomedical Engineering*, 38(2):275–296, 2018.

[9] Behnaz Gheflati and Hassan Rivaz. Vision transformers for classification of breast ultrasound images. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 480–483. IEEE, 2022.

[10] MJ Ghrabat, Zaid Alaa Hussien, Mustafa S Khalefa, Zaid Ameen Abduljabba, Vincent Omollo Nyangaresi, Mustafa A Al Sibahee, and Enas Wahab Abood. Fully automated model on breast cancer classification using deep learning classifiers. *Indonesian Journal of Electrical Engineering and Computer Science*, 28(1):183–191, 2022.

[11] Angela N Giaquinto, Hyuna Sung, Kimberly D Miller, Joan L Kramer, Lisa A Newman, Adair Minihan, Ahmedin Jemal, and Rebecca L Siegel. Breast cancer statistics, 2022. *CA: a cancer journal for clinicians*, 72(6):524–541, 2022.

[12] Yujuan Guo, Jingjuan Liao, and Guozhuang Shen. A deep learning model with capsules embedded for high-resolution image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:214–223, 2020.

[13] Mohamed Hosni, Ibtissam Abnane, Ali Idri, Juan M Carrillo de Gea, and José Luis Fernández Alemán. Reviewing ensemble classification methods in breast cancer. *Computer methods and programs in biomedicine*, 177:89–112, 2019.

[14] Rizwana Irfan, Abdulwahab Ali Almazroi, Hafiz Tayyab Rauf, Robertas Damaševičius, Emad Abouel Nasr, and Abdelatty E Abdelgawad. Dilated semantic segmentation for breast ultrasonic lesion detection using parallel feature fusion. *Diagnostics*, 11(7):1212, 2021.

[15] Kiran Jabeen, Muhammad Attique Khan, Majed Alhaisoni, Usman Tariq, Yu-Dong Zhang, Ameer Hamza, Artūras Mickus, and Robertas Damaševičius. Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion. *Sensors*, 22(3):807, 2022.

[16] Aishik Konwer, Xuan Xu, Joseph Bae, Chao Chen, and Prateek Prasanna. Temporal context matters: Enhancing single image prediction with disease progression representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18824–18835, 2022.

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[18] Neeraj Kumar, Ruchika Verma, Ashish Arora, Abhay Kumar, Sanchit Gupta, Amit Sethi, and Peter H Gann. Convolutional neural networks for prostate cancer recurrence prediction. In *Medical Imaging 2017: Digital Pathology*, volume 10140, pages 106–117. SPIE, 2017.

[19] Jorge F Lazo, Sara Moccia, Emanuele Frontoni, and Elena De Momi. Comparison of different cnns for breast tumor classification from ultrasound images. *arXiv preprint arXiv:2012.14517*, 2020.

[20] Vivian Man, Wing-Pan Luk, Ling-Hiu Fung, and Ava Kwong. The role of pre-operative axillary ultrasound in assessment of axillary tumor burden in breast cancer patients: a systematic review and meta-analysis. *Breast Cancer Research and Treatment*, 196(2):245–254, 2022.

[21] O Ibrahim Obaid, Mazin Abed Mohammed, Mohd Kanapi Abd Ghani, A Mostafa, Fahad Taha, et al. Evaluating the performance of machine learning techniques in the classification of wisconsin breast cancer. *International Journal of Engineering & Technology*, 7(4.36):160–166, 2018.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas 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*, pages 234–241. Springer, 2015.

[23] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[24] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[25] Ahmet Solak, Rahime Ceylan, Mustafa Alper Bozkurt, Hakan Cebeci, and Mustafa Koplay. Adrenal tumor segmentation on u-net: A study about effect of different parameters in deep learning. *Vietnam Journal of Computer Science*, pages 1–25, 2023.

[26] Manu Subramoniam, TR Aparna, PR Anurenjan, and KG Sreeni. Deep learning-based prediction of alzheimer's disease from magnetic resonance images. In *Intelligent vision in healthcare*, pages 145–151. Springer, 2022.

[27] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[28] Mesut TOĞAÇAR and Burhan ERGEN. Deep learning approach for classification of breast cancer. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, pages 1–5. IEEE, 2018.

[29] Saliha Zahoor, Umar Shoaib, and Ikram Ullah Lali. Breast cancer mammograms classification using deep neural network and entropy-controlled whale optimization algorithm. *Diagnostics*, 12(2):557, 2022.