Electronic Letters on Computer Vision and Image Analysis 23(2):85-104, 2024

A Study on CNN-Based and Handcrafted Extraction Methods with Machine Learning for Automated Classification of Breast Tumors from Ultrasound Images

Mohamed Benaouali*, Mohamed Bentoumi*, Mansour Abed*,
Malika Mimi* and Abdelmalik Taleb Ahmed+

* *Electrical Engineering Department, Signals and Systems Laboratory, University Abdel Hamid Ibn Badis of Mostaganem, Algeria*
+ *IEMN DOAE UMR CNRS 8520, Université Polytechnique des Hauts de France, Valenciennes, France*

Received 16th of April, 2024; accepted 5th of November 2024

Abstract

In this paper, we present an efficient procedure for automatically classifying ultrasound images of benign and malignant breast tumors. We evaluated our approach using four openly available datasets and investigated two categories of feature extraction methods: handcrafted methods (Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG)) and methods based on convolutional neural network (CNN) models. For classification, we explored three classifiers: linear support vector machines (SVM), k-nearest neighbors (KNN), and artificial neural networks (ANN). Two experiments were conducted: the first aimed to design a classifier for each individual dataset, whereas the second aimed to develop a unified classifier for the ensemble datasets. The obtained results demonstrate that the ANN classifier associated to the early stopping (ES) criterion, is very effective in both experiments, outperforming KNN and SVM with 100% accuracy. Additionally, using CNN models as feature extraction methods proved effective. Among these CNNs: ResNet50, InceptionV3 and DenseNet201 achieve 100% accuracy in the first experiment, while DenseNet201 allows achieving 100% accuracy in the second experiment. Comparative analysis with existing research demonstrates the competitiveness or superiority of the proposed procedure.

Key Words: breast tumor, ultrasound images, feature extraction, handcrafted methods, pretrained CNN model, classification.

1 Introduction

Breast cancer has become one of the most common cancers worldwide, accounting for approximately 12% of all new cancer cases with approximately 2 million new cases and 685,000 deaths worldwide each year [1]. Early detection of this type of cancer is an important factor in reducing mortality in women by 40% [2]. An

Correspondence to: mohamed.benaouali@univ-mosta.dz

Recommended for acceptance by Angel D. Sappa
https://doi.org/10.5565/rev/elcvia.1887
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

early diagnosis allows starting the adequate tumor treatment in time to considerably increase the chances of healing.

Medical imaging is the main tool for the detection and diagnosis of breast cancer. There are different medical imaging modalities to note: Mammography, magnetic resonance imaging (MRI), echography (ultrasound image) and tomography (CT: computed tomography). Mammography and ultrasound are the two most used methods for the early detection of breast cancer in women [5] [6]. Mammographic imaging has limitations that depend on the ages of the patients, and the density of the breasts in addition to the modality being radioactive. On the other hand, ultrasound is a non-radioactive and non-invasive modality for the early detection of breast cancer in women under 40 years of age. This technique is particularly well-suited for dense breasts compared to other modalities [3] [6]. Ultrasound imaging (US), or echographic imaging, is the best alternative to mammographic imaging which is a painful examination for a woman. Previous work has shown the effectiveness of ultrasound images for the detection and classification of breast tumors [7] [8]. Nevertheless, processing a large number of medical images directly can be a tedious task. Additionally, the quality of ultrasound imaging depends on how the radiologist manipulates the hardware device, which represents a significant limitation.

The approaches for automatic detection and classification of breast cancer from US images come from the field of image processing and classification. Currently, two classes of approaches belong to this field can be distinguished: Classical approaches based on machine learning (ML) techniques and approaches based entirely on deep learning (DL).

These so-called classical approaches are mainly divided into three steps: image preprocessing, feature extraction and classification. The preprocessing consists of the detection and localization of the region of interest (ROI), in order to reduce the size of the image to be analyzed and to be able to apply adjustments either to the contrast or the brightness of the image. This is done for quality enhancement purposes. Feature extraction is a transformation operation that transforms an image into a vector of features representative of the information contained in the ROI. In general, two categories of feature extraction methods can be distinguished: “Handcrafted” methods and methods based on deep learning (DL). Handcrafted-based methods can be divided into two groups: Texture-based methods and geometry-based methods. The former methods extract texture descriptors from the ROI, while the latter describe the shapes present in the ROI [4] [5]. Deep learning (DL)-based feature extraction approaches use models built from convolutional neural networks (CNNs) to enable feature extraction through learning. The use of CNN models pre-trained by Transfer Learning for the extraction of features from US images is an interesting alternative method for the classification of breast tumors.

The second class of approaches based entirely on DL model have been very popular in recent years. In particular ”end-to-end” CNNs, offer powerful capabilities for image processing and classification, allowing advanced automation and automatic feature extraction from raw data. This has contributed to their growing popularity in the field of medical image analysis. On the other hand, the implementation of approaches based on DL model requires large training datasets, often called ”big data”. This is due to the nature of deep neural networks, which are complex models with many parameters to estimate. These models require a large number of training examples to generalize effectively and avoid overfitting. This is why we opted in this work for the use of approaches based on ML techniques belonging to the first class. In addition, these approaches may be more suitable when the available data are small or of limited quality.

Even there are many research works dealing with the problem of detection and classification of breast cancer tumors using different types of imaging modalities such as mammography, ultrasound breast imaging, however, has attracted less interest due to the quality of US images and the lack of public reference datasets. In this paper, we present a study on the application of ultrasound image processing procedures using four open source datasets of US breast images. The main objective is to classify breast cancer tumors using the classical approaches. The classification under consideration allows to distinguish between malignant and benign tumors. In this work, our major contributions are summarized as follows:

- We explore and use two categories of feature extraction methods: Handcrafted methods based on texture feature extraction and Deep Learning methods based on pre-trained CNN models.

- We use pre-trained CNN models as feature extractors. We adapt these models by adding a Global Average Pooling (GAP) layer to simultaneously perform flattening and feature vector reduction. For the handcrafted methods, the values of the internal parameters were selected heuristically to obtain a reduced vector for the classification phase.

- Six pre-trained CNN models are adapted and used as feature extractors, along with two hand-crafted methods, namely LBP and HOG. These eight extractors are used in combination with three different classifiers: SVM, KNN and ANN, resulting in a total of 24 combinations for the evaluation process of the adopted procedure.

- We evaluate and compare the performance of different extraction methods using three types of classifiers to determine the most effective approaches for breast lesion classification.

- We use four open-source ultrasound image datasets to test and evaluate the different procedures. This allows us to assess the generalizability of these methods and the quality of the datasets used.

The remainder of this paper is organized as follows. In section 2, we present the state of the art of work carried out in the field of detection and classification of breast tumors from ultrasound images. Section 3 is devoted to the presentation of the procedure considered in this research by describing in detail each step of the proposed procedure. The experimental results and discussions are presented in section 4. Finally, Section 5 concludes the paper.

2 Related Work

In recent years, several research teams have been confronted with the problem of detecting and classifying breast masses from ultrasound images. It is difficult to establish an exhaustive list of the methods developed in the field. Much of the work in the literature has focused on solving detection and classification separately [19] [24]. As already mentioned, from feature extraction point of view, the methods developed in this context can be classified into two categories, namely Handcrafted methods and CNN based methods.

In the context of the first category, the authors in [10] proposed the extraction of the shape and texture features of the ROI after a segmentation step of this ROI from ultrasound images. The global feature vector was obtained by concatenating the vector of twenty texture features obtained by the GLC Matrix method (GLCM) with a vector of seven shape features. Three types of classifiers were used and compared, namely: The decision tree (DT), the k-nearest neighbors (KNN) and the ensemble classifier for the classification of benign and malignant tumors. An accuracy of 96.6% was obtained by the ensemble classifier applied on the BUSI (Bresat ultrasound image) dataset [20]. In [11], the authors presented a procedure for classifying breast tumors from ultrasound images based on different methods of extracting features from the ROI. Feature extraction is preceded by a preprocessing step which consists of using the ground truth mask as a priori information in order to obtain three images from the original image. The extracted features were classified using six different classifiers. The procedure has been applied and tested on the BUSI dataset and has obtained an accuracy of 97.4%.

The authors in [12] applied different methods of extracting texture features and morphology features on regions of interest in US images of female breast. Several combinations and several classifiers were tested in order to obtain better performance. The texture features are combined and the dimension of the obtained feature vector is reduced using the principal component analysis (PCA) method. Then, a support vector machine (SVM) classifier was applied to the vector for the classification of breast tumors. The morphological feature vector is classified by the Naive Bayes classifier. Finally, the outputs of the two classifiers are merged to give a single decision. This yielded an accuracy of 91.11%. In view of the classification of breast tumors from US images, the authors in [13] presented a procedure to combine the features extracted by the hadcrafted methods (texture and morphological) with those obtained using a pre-trained CNN model (DL-based features). Texture features and DL-based features were combined to feed an SVM classifier. The morphological characteristics

were classified by a Naive Bayes (NB) classifier. The outputs of the two classifiers (SVM and NB) were weighted and merged to give a single decision which improves the overall classification performance. The 5-fold cross-validation provided an accuracy of 89.17%. In [14], the authors proposed a feature extraction based on a CNN model, more precisely, the DarkNet-53 model pre-trained after an augmentation of ultrasound image data from the BUSI dataset. The best features were selected by two methods and merged to be classified by an SVM classifier which gave an accuracy of 99.1%. In [15], the authors proposed an approach derived from deep learning based on the transfer learning (TL) technique for the classification of breast masses from US images. They introduced a technique for learning a pre-trained CNN model by adding intermediate layers referred to as: deep representation scaling (DRS). Only these last layers were updated during learning. Then, the CNN model is trained on a base of US BUSI images for the classification of breast masses. The proposed method of additional layers (DRS), combined with fine-tuning techniques, yielded a classification accuracy of more than 92%. In [16], the authors generated composite images obtained by decomposition of the original US images and their masks. The composite images were then merged through the RGB channels of the input image of a CNN model for feature vector extraction. Several CNN models have been tested and evaluated for this purpose. A neural classifier was used to classify the feature vectors yielding a better classification rate of 95.48%.

The authors in [17] propose an increase in the number of ultrasound images in a public datasets using the Generative Adversarial Network (GAN) model. The authors used and compared two DL approaches: The complete learning of a CNN model and the use of a CNN model by transfer learning for the classification of US images of breast cancer. The authors concluded that employing the NASNet model and tranfer learning together with GAN-based image augmentation provided an accuracy of 99%. In the same framework, the researchers in [18] used a CNN model, based on the Inception V3 architecture. In order to improve the classification performance, the image learning dataset was augmented using synthetic images generated by a semi-supervised GAN model. The proposed GAN architecture generated high-quality synthetic US images that were verified by two experienced radiologists. This enabled the authors to obtain breast mass classification accuracy reaching 90.41% based on the BUSI dataset. Wilding et al. [24] developed a breast tumor classification and segmentation procedure from a set of US images. Before that, they have found that 3.92% of the images in the used datasets were incorrectly labeled. An expert radiologist evaluated all the dataset images used to correct these labeling anomalies. The procedure consisted of two stages of successive classifications to discriminate the three classes of US images: normal, benign and malignant. The last step was a segmentation algorithm of any masses present in an image. The classification procedure allowed to discriminate healthy images from abnormal images with an accuracy of 96%, and it also allowed to distinguish benign from malignant images with an accuracy of 85%. Labcharoenwongs et al. [25] conducted a study to automatically detect and classify breast tumors using deep learning model. The data was provided by the Department of Radiology at Thammasat University and the Queen Sirikit Breast Cancer Center in Thailand. Different methods of data augmentation have been used. Tumor detection, localization and classification were performed using the DL-based YOLO7 architecture. An estimate of tumor volume was made using a simple method. Promising results were obtained with lesion classification accuracy of 95.07%, sensitivity of 94.97% and specificity of 95.24%. The authors in [26] developed a new end-to-end CNN architecture based on the InceptionV3 model to classify benign and malignant tumors from US breast images. They used a dataset designed from five datasets (three public and two private datasets) to train and evaluate the end-to-end CNN model. An accuracy of 81% was obtained.

The authors in [27] developed a benchmark for deep learning-based BUS (Breast ultrasound) image classification to compare it with current approaches and gain insights into techniques that improve classification generalization. They used a dataset composed of five public datasets (3,641 BUS images). A novel multi-task learning approach (segmentation and classification) has been proposed that incorporated a small tumor sensitive network as a core network. The latter consists of a main task (tumor classification) and a secondary task (tumor segmentation). The proposed approach achieved a sensitivity of 90.4% and a specificity of 89.8%, in the classification task.

It is evident that there is not really a comprehensive comparative study to effectively evaluate feature extraction approaches for the classification of benign and malignant tumors from breast ultrasound images. This

shortcoming is mainly due to a lack of large datasets of breast ultrasound images, as well as differences in the quality of images from different sources and some anomalies in the labeling of these images. Consequently, our motivation is to propose a design and evaluation approach for classical approaches which makes it possible to overcome these shortcomings and to be able to carry out a valid comparative study to evaluate the performance of the different approaches for classifying breast tumors from US images. In our work, we explore two categories of feature extraction methods for breast tumors classification: Handcracted methods and methods based on pre-trained CNN models. We evaluate and compare these methods using three types of classifiers namely, SVM, KNN and ANN. Additionally, we use four open source datasets of US breast images to test and assess our approaches. In the next section, we detail each contribution related to these aspects.

3 Materials and Methods

In this section, we present the approach adopted for the classification of malignant and benign breast tumors in women from US images. As shown in Figure 1, the procedure is divided into three phases schematized by three blocks: a preprocessing block, a feature extraction block and a classification block. In this paper, we focused on the last two blocks that contain the main contributions of our present work.

The three phases are detailed in the following paragraphs after a brief description of the considered ultrasound image datasets.

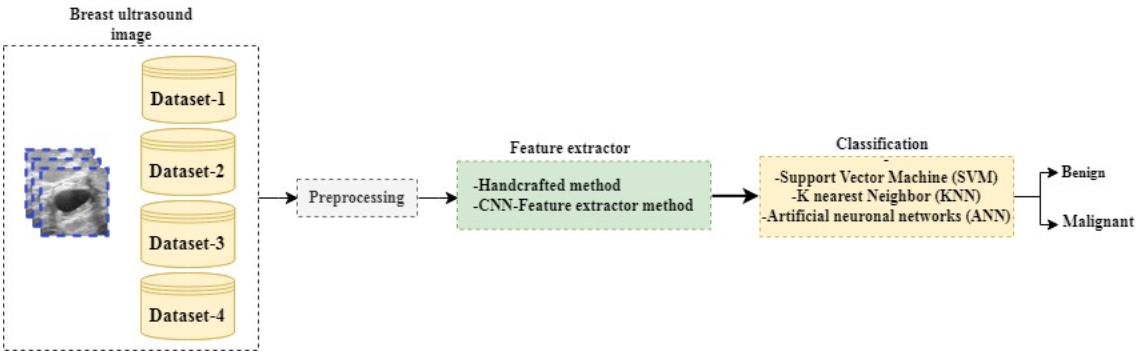

Figure 1: The overall framework of the proposed procedure.

3.1 Data description

To evaluate and validate the methods employed in this work, four distinct public datasets are used, referred to as: dataset-1 (BUSI: Breast Ultrasound Images) [20], dataset-2 (Dataset B: Breast Ultrasound Lesions) [21], dataset-3 [44] and dataset-4 [45]. Figure 2 presents example images from each dataset.

- **Dataset-1**: The BUSI dataset [20] comprises breast ultrasound images of women aged between 25 and 75. Collected in 2018, it includes a total of 600 patients. This database contains 780 images, with an average size of 500×500 pixels, saved in PNG format. These images are categorized into three classes: 133 normal images, 445 benign images, and 210 malignant images.
- **Dataset-2**: The images in this dataset were gathered from the UDIAT diagnostic center of the Corporation Parc Tauli in Sabadell, Spain, in 2012 [21]. This dataset consists of 163 breast ultrasound images, including 53 images of malignant tumors and 110 images of benign tumors. The average size of the images in this dataset is 760×570 pixels.
- **Dataset-3**: This dataset was obtained and provided by the Department of Radiology at Thammasat University and the Queen Sirikit Breast Cancer Center of Thailand [44]. Dataset-3 comprises 263 breast images, with 120 benign and 143 malignant cases. The image sizes vary from 1246×864 pixels to 386×456 pixels.

- **Dataset-4**: Dr. Geertsma, a skilled radiologist from Gelderse Vallei Hospital in the Netherlands, collaborated with Hitachi Medical Systems Europe to collect the HMSS dataset [45]. This dataset contains 1992 breast ultrasound images, with 845 images depicting benign tumors and 1147 images depicting malignant tumors. The image sizes range from 600×450 pixels to 300×225 pixels.

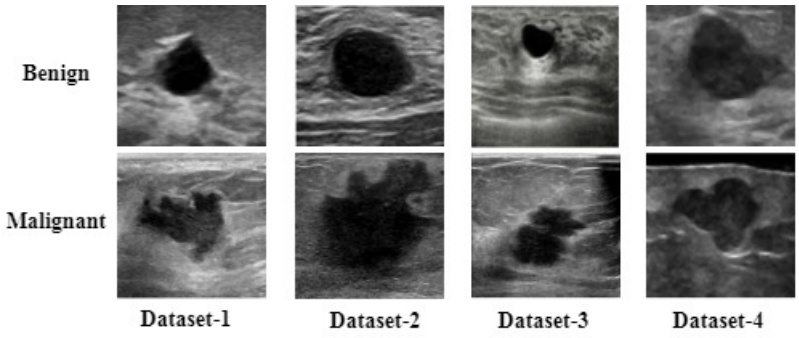

Figure 2: Examples of ultrasound images illustrating benign and malignant tumors in each dataset.

3.2 Preprocessing

Ultrasound imaging is a non-invasive technique that allows for real-time examination of various internal organs and tissues of the human body. However, this imaging modality has a significant drawback: the quality of the images is degraded due to the presence of speckle noise. This noise arises from the interference of acoustic waves scattered by the different tissues within the examined medium. Speckle noise creates a grainy appearance in the ultrasound images, which can mask the details of tissue structures and complicate image interpretation. To address this issue, we employed the anisotropic diffusion filter [28], an effective method for reducing speckle noise in ultrasound images while maintaining a good balance between noise smoothing and contour preservation.

Although filtering methods have been extensively utilized in the preprocessing of breast ultrasound images [29] [30], contrast enhancement methods have been relatively underused. In this study, we applied the histogram equalization technique to improve the contrast of breast ultrasound images. This technique is particularly effective for enhancing image quality when details are hidden in overly dark or bright areas. Histogram equalization operates by redistributing the intensities (grayscale levels) of the image pixels to achieve a histogram that is as uniform as possible.

Figure 3 shows an example of the result of applying this preprocessing step, which consists of two successive operations: denoising by filtering and contrast enhancement by histogram equalization.

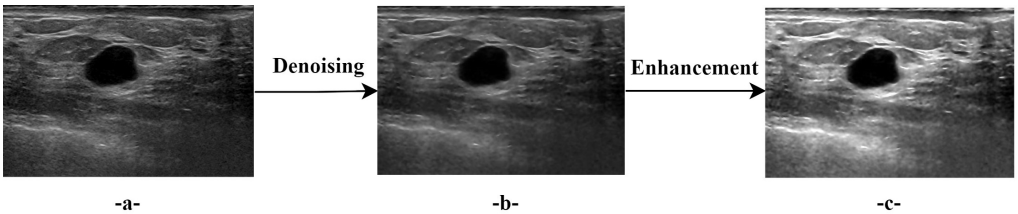

Figure 3: An illustrative example of the performed image preprocessing : (a) the original image, (b) the denoised image, and (c) the result after enhancement.

After filtering the original image and enhancing its contrast, the region of interest (ROI) is extracted using a mask derived from the reference images from datasets-1 and datasets-2. However, the images from datasets-3

and datasets-4 have already been cropped to the ROIs. Finally, all images are resized to a uniform size of 128×128 pixels for the rest of the procedure. By thoroughly browsing the literature dedicated to deep feature extraction and classification of breast ultrasound images [46], it is observed that three types of input images have been employed: full images, manually extracted ROI images and segmented tumor images. In our study, we specifically used manually extracted ROI images.

3.3 Features extraction

Feature extraction plays a crucial role in the classification of breast tumors, as it has a significant impact on the results of this classification [31]. Features are the specific signatures of a given image that allow it to be described in a distinct way. Ultrasound images of benign and malignant breast tumors typically exhibit noticeable differences in both visible features and hidden texture features. These features depict specific signatures that describe an image and enable us to differentiate between benign and malignant tumors.

For this, we used two different methods for feature extraction: Handcrafted methods based on texture feature extraction, such as Local Binary Patterns (LBP) [38] and Oriented Gradient Histogram (HOG) [39], and Deep Learning methods based on pre-trained CNN models namely: ResNet50 [32], Xception [33], DenseNet201 [34], EfficientNetB7 [35], Inception V3 [36], and VGG16 [37]. A brief overview of these methods is provided in the following paragraphs.

3.3.1 Handcrafted methods

Benign and malignant tumors present in ultrasound images can be distinguished based on their morphology. Benign tumors typically exhibit a regular, round, or oval shape with a smooth contour, while malignant tumors display an irregular shape, an ill-defined contour, and lobules [5] [7]. However, relying solely on morphological characteristics is insufficient for achieving accurate results with classifiers. In this study, we chose to employ two texture feature extraction methods: the Local Binary Patterns (LBP) method [38] and the Histogram of Oriented Gradients (HOG) method [39].

- **The LBP descriptor**: The fundamental concept of LBP is to summarize the textural information around each pixel in an image by comparing this pixel with its immediate neighbors. An LBP is calculated for each pixel in the image according to the following operator:

$$\begin{aligned} LBP\left(x,y\right) &= \sum_{p=1}^{8} 2^{p-1} s\left(i_p - i_c\right) \\ s\left(z\right) &= \begin{cases} 1 & : \ z \geq 0 \\ 0 & : \ z < 0 \end{cases} \end{aligned} \tag{1}$$

  where i_c is the gray level value of the central pixel (x, y), p is the number of the adjacent pixel, i_p is the gray level value of the adjacent pixel, and $s\left(z\right)$ is the symbolic function. To characterize an image, a histogram of LBP values is constructed. This histogram represents the distribution of binary patterns in the analyzed image and serves as a feature vector for classification.

- **The HOG descriptor**: To extract HOG descriptors, an image is first divided into small regions called cells. For each cell, the brightness gradients are calculated using differential operators such as the Sobel operator. Next, the orientations of the local gradients are quantified by grouping the orientations into several intervals, forming orientation histograms. To account for variations in scale and position of objects, neighboring cells are grouped into blocks. The orientation histograms of the cells in each block are concatenated to form the final HOG descriptor vector. Generally, normalization is also performed to enhance the robustness of the descriptor against variations in lighting and contrast.

3.3.2 Deep learning methods

The CNN architecture is the basis of many deep learning models applied to image processing. As depicted in Figure 4, the CNN model architecture is organized as a sequence of blocks, with the initial blocks primarily comprising two types of layers: convolutional and pooling layers. The final blocks consist of fully connected (FC) neural networks functioning as a traditional classifier. Convolutional layers perform feature extraction by scanning the input image and applying convolutional filtering operations as described in Eq. 2. The filtering is performed using multiple convolution filters grouped together in a bank of filters [9], thereby enabling the generation of a feature map.

$$S_l\,(m,n) = (I\ *\ h_l\,)(m,n) = \sum_{i}^{M}\sum_{j}^{N}\sum_{k}^{C} I(i,j,k)\ \cdot\ h_l\,(m-i,n-j,k) \tag{2}$$

where S_l represents the l^{th} feature map corresponding to the l^{th} filter h_l, M and N are, respectively, the height and width of the image I, C represents the number of channels (for an RGB image, $C = 3$), and (m,n) are the coordinates of one pixel. The number of feature maps $\{S_l\}$ is equal to the number of filters (h_l) in the corresponding layer. These feature maps are then subjected to a nonlinear function such as the Rectified Linear Unit (ReLU). The pooling layer allows for reducing the feature map size (downsampling) while preserving their essential features. The output of one layer is the input of the next layer. Subsequently, the final block of the CNN network consists of fully connected (FC) layers, which act as a classical neural network for executing classification tasks. The first FC layer receives the flattened outputs of the last convolutional layer as input. CNN models are designed for end-to-end image classifcation and trained on large datasets (big data).

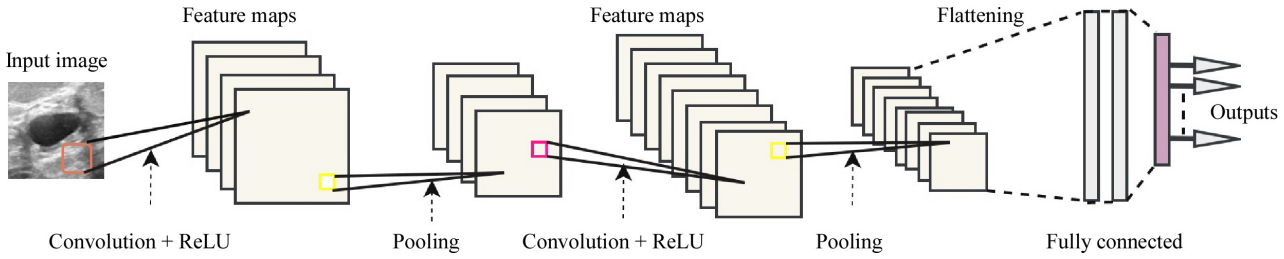

Figure 4: Example of architecture of a CNN network.

- **Adaptation of CNN models for feature extraction**:

Pre-trained CNN models are designed for end-to-end image classification. In order to use them as feature extractors, an adaptation is required. For this purpose, we removed all fully connected layers as well as the last pooling layer from all used pre-trained CNN models. The other layers were kept for feature extraction. Consequently, the last layer of the remaining block consists of p feature maps ($\{S_l\}$). We added a global pooling layer to flatten and reduce the dimension of the output vector. There are several ways to perform this operation, among the most used that we cite: the global average pooling (GAP) and the global maximum pooling (GMP). Previous works [4] [22] [23] showed that using GAP provides better performance. For this reason, we propose to use the GAP described by:

$$x_l = \frac{\sum_m^M \sum_n^N S_l\,(m,n)}{M \times N} \quad ;\ \ l \in [1,p] \tag{3}$$

where M and N are the size of the last p feature maps $\{S_l\}$, and and (m,n) are the coordinates of one pixel. As a result, each CNN-FE (CNN-Feature Extractor) model produces a feature vector (X) of dimension p which represents the features of the input image. And thus, all the CNN models used in this study (see Table 1) were adapted to perform the extraction task.

Table 1: Characteristics of CNN architectures used in this work.

CNN-FE	Architecture	Image input size	# of parameters (in Millions) before adaptation	# of parameters (in Millions) after adaptation
ResNet 50	50 layers	$(224 \times 224) \times 3$	25.5	23.5
Xception	71 layers	$(299 \times 299) \times 3$	22.9	20.8
DenseNet201	201 layers	$(224 \times 224) \times 3$	20.2	18.3
EfficientNetB7	37 layers	$(600 \times 600) \times 3$	66.7	64.09
Inception V3	48 layers	$(229 \times 229) \times 3$	23.8	21.7
VGG16	16 layers	$(224 \times 224) \times 3$	138.3	14.7

3.4 Classification

In the field of tumor diagnosis from medical images, classification plays a crucial role in predicting the nature of tumors based on a set of features extracted from the image. As previously mentioned, the feature extraction phase involves extracting useful characteristics from the BUS image to provide a 1D feature vector (X), enabling class discrimination, which, in our case, differentiates between benign and malignant tumors.

In a classification problem, the objective is to assign an observation (feature vector (X)) to a predefined class or category. There are two contexts for implementing classification methods: supervised methods and unsupervised methods. Supervised methods are employed when the class of the data (observations) is known, while unsupervised methods are applied when the class is unknown. In our work, we have labeled datasets, which leads us to use supervised methods. Consider the case with two classes (Ω_1: benign, Ω_2: malignant) and a training set $D = \{(X_1, y_1), \ldots, (X_N, y_N)\}$, where $X_i \in R^p$ and $y_i \in \{-1, +1\}$, with:

$$\begin{array}{l} y_i = +1 \quad \text{if} \quad X_i \in \Omega_1 \\ y_i = -1 \quad \text{if} \quad X_i \in \Omega_2 \end{array} \tag{4}$$

The purpose is to construct a decision function (f) (Eq. 5) from the N examples (set D) that assigns any new data (X_n) to one of the two predefined classes:

$$\hat{y}_n = f(X_n) \tag{5}$$

The classification methods can be divided into two primary categories. In the first category, the shape of the separation surface (f) between classes is well defined. The corresponding methods aim to estimate the parameters of this separating surface and optimize the decision-making process. There are two subcategories: linear classifiers and nonlinear classifiers. Conversely, the methods of the second category do not consider the shape of the separation surface but evaluate distances relative to the predefined classes.

In this study, three different supervised classifiers were utilized: Support Vector Machines (SVM) [40], KNearest Neighbors (KNN) [41] and Artificial Neural Networks (ANN) [47]. These classifiers were selected due to their proven effectiveness in medical image classification, providing diverse solutions to address the specific needs of our study. The ANN are inherently nonlinear, which is why we employed them as nonlinear classifiers. The SVM method, chosen in its linear form for its simplicity and efficiency, falls under the subcategory of linear classifiers. The KNN classifier is an emblematic example of the second category of nonparametric classifiers, where the classification of an example (feature vector) is determined by the proximity of its K nearest neighbors in the feature space.

4 Experimentation

In this work, we implement an automatic classification procedure of benign and malignant tumors based on breast ultrasound images from still images. We will present the methodology used for classifying ultrasound

images of benign and malignant breast tumors in the following section. The adopted procedure's implementation methodology is a significant contribution to this work, which we highlight. Afterwards, we will present and discuss the performance results obtained from the various methods used.

4.1 Methodology

First of all, a data preparation step was carried out. For this purpose, we performed the extraction of the region of interest (ROI) corresponding to a tumor, whether benign or malignant. All images from the four datasets were scaled to a size of 128×128 pixels. To address the imbalance issue between the two benign and malignant classes for each dataset, we applied data augmentation to the images of the minority class (see Table 2). Data augmentation in this case was achieved by a rotation operation with four different degrees of rotation.

Table 2: Summary of datasets before and after data augmentation

Datasets	Before data-augmentation	After data-augmentation	Total BUS Images
Dataset-1	Benign: 135, Malignant: 141	Benign: 141, Malignant: 141	282
Dataset-2	Benign: 92, Malignant: 45	Benign: 92, Malignant: 92	184
Dataset-3	Benign: 56, Malignant: 90	Benign: 90, Malignant: 90	180
Dataset-4	Benign: 427, Malignant: 803	Benign: 427, Malignant: 427	854

In the following, we aim to recall the implemented procedure for breast tumor classification from US images. It is worth noting that the procedure consists of three steps (see Figure 1): preprocessing, feature extraction, and classification.

4.1.1 Image preprocessing

This preliminary preprocessing step aims to optimize the contrast of the images by removing noise and improving them compared to their original versions. To do this, all images undergo a denoising process through the application of an anisotropic diffusion filter specially designed to reduce speckle noise. Subsequently, these denoised images are subjected to a histogram equalization technique to further enhance the contrast.

4.1.2 Feature extraction

When the images are processed in the previous phase, they are transformed into feature vectors using two different types of methods: handcrafted methods, including LBP and HOG extractors methods, as well as pre-trained CNN models: ResNet50, Xception, DenseNet201, EfficientNetB7, InceptionV3 and VGG16. For each method, a feature vector with fixed dimension is obtained as output (see Figure 5). Feature vector reduction is seamlessly integrated into the feature extraction process. In handcrafted methods, internal parameter values are heuristically chosen to achieve a reduced vector for classification.

For pre-trained CNN models, a Global Average Pooling (GAP) layer is incorporated to handle both flattening and feature vector reduction. As previously stated, all CNN models utilized in this work were adapted for the extraction task. Figure 5 illustrates the inclusion of the GAP layer employed for feature extraction. For each CNN-FE model, a feature vector (x) of dimension p is computed for each image. There is no necessity to retrain or fine-tune the model for this feature extraction process.

Note that CNN-FE models should be fed with color images (RGB images: 3 matrices), while handcrafted methods operate on grayscale images (1 matrix). The four datasets include grayscale images. Thus, we overcome this problem by duplicating each grayscale image to obtain three matrices. These will feed the pre-trained CNN-FE models in order to extract the features.

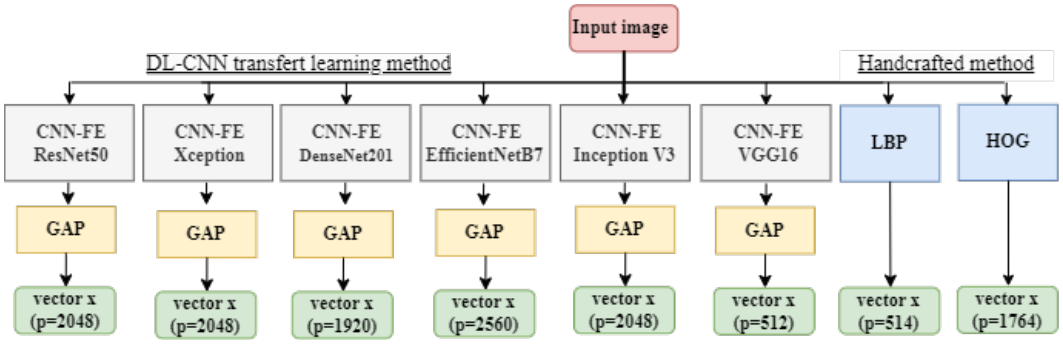

Figure 5: Dimensions of the feature vector extracted by each feature extractor

4.1.3 Classification

In this phase, three distinct types of classifiers were designed: linear SVM, KNN and ANN. For the SVM classifier, a linear kernel was used to design a linear SVM classifier. For the KNN classifier, the number of neighbors was set to 5 (K =5), the choice of 5 neighbors is a compromise commonly used in KNN applications, offering a reasonable balance between flexibility and generalization.

Regarding the ANN, we used a multi-layer perceptron with a single hidden layer composed of 20 neurons, each associated with the 'Relu' activation function, and an output layer composed of a single neuron with 'sigmoid' activation function. The latter choice is motivated by the nature of our binary classification problem (two classes: "malignant" and "benign" (positive class and negative class)). In the training phase, we used the 'binary crossentropy' loss function to measure the disparities between the ANN classifieur predictions and the actual class labels. To avoid overfitting, a certain amount of misclassified training samples can be accepted. For the ANN classifier training, the use of the early stopping (ES) [4] criterion can be shown as a very interesting alternative to avoid overfitting and improve the generalization error.

Each of these three classifiers requires training and testing phase in order to evaluate its performance. Two experiments were set up to evaluate the performance of the classifiers on four datasets :

- The first experiment consists of designing a classifier for each dataset. This implies that the training and testing phases would exclusively utilize a single dataset at any given time. This experiment allows us to assess and compare how different feature extraction methods perform when paired with three classifiers (SVM, KNN, and ANN).

- The second experiment aims to create a classifier by merging the datasets. This involves the exhaustive use of available data to train a single classification model. By merging these datasets, the objective is to fully exploit the diversity of information contained in each in order to improve the performance of the resulting classifier.

4.1.4 Performance Metric

In this work, in order to assess the proposed procedure, the classification performance is measured quantitatively by the use of three metrics: accuracy, sensitivity, specificity [4][42][43]. These metrics are introduced as following:

- **The accuracy** expresses the overall rate of good classification which is formally defined as the proportion of images classified correctly in the total number of images:

$$\underset{on\ test\ set}{Accuracy} = \frac{number\ of\ images\ correctly\ classified}{number\ of\ all\ images} \tag{6}$$

- **The specificity** is a metric that expresses the rate of negative examples correctly predicted by the classifier:

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

where **TN** is true negative and **FP** is false positive.

- **The sensitivity** is a metric that expresses the rate of positive examples correctly predicted by the classifier:

$$Sensitivity\ (recall) = \frac{TP}{TP + FN} \tag{8}$$

where **TP** is true positive and **FN** is false negative.

In these works, those metrics are estimated by the k-fold cross-validation method. The entire dataset is divided into k subsets. We perform the learning on $k-1$ set and the test (accuracy estimation) on the remaining set. The operation is repeated k times and each time the test set is changed. The rate of good classification (accuracy) is the average of the k rates obtained.We applied 10-fold cross-validation to test our procedure [4].

4.2 Results and discussions

This section presents the outcomes of the adopted procedure shown in Figure 1 to classify breast tumors in women using ultrasound images. Four different datasets were used. As already explained in Section 4.1, the proposed procedure includes three stages: preprocessing, feature extraction and classification into two classes, benign and malignant. Six pre-trained CNN models were adapted and used as feature extractors, along with handcrafted LBP and HOG methods. These eight extractors were used in combination with three different classifiers: SVM, KNN and ANN, producing 24 distinct combinations to evaluate the adopted procedure.

Two experiments were conducted to evaluate the effectiveness of the proposed methodology in classifying breast tumors. Various combinations of feature extractors and classifiers were employed on four US image datasets. The following subsections present the results obtained from each experiment.

4.2.1 Experiment 1

As shown in Table 3 and Figure 6, the results of the first experiment clearly indicate that the performance of the procedure integrating the ANN classifier is significantly higher than that obtained with the SVM and KNN classifiers, whatever the feature extraction method used. This finding demonstrates that the use of artificial neural networks is a robust and effective approach which benefits from regularization of learning via the Early Stopping criterion, and shows good performance. This also indicates that the classification of US breast tumor images constitutes a non-linearly separable problem.

It should be noted that the use of the ANN classifier in combination with the CNN-FE extractors: ResNet50, DensNet201 and Inception-V3, led to an accuracy of 100% for dataset-1 and dataset-3. The CNN-FE extractors: Xception and EfficientNetB7, combined with an ANN classifier, yielded very satisfactory classification rates for dataset-1, with accuracies of 99.55% and 99.31% respectively however, the CNN-FE VGG16 extractor combined with an ANN classifier achieved accuracies of only 97.24%, 97.89%, and 97.22% for dataset-1, dataset-2, and dataset-3 respectively. It's worth mentioning that the handcrafted methods LBP and HOG, when combined with the ANN classifier, also achieved high accuracy. The combination of the HOG extractor with the ANN classifier yielded classification rates of 99.47% and 99.44% for dataset-2 and dataset-3 respectively. Using the LBP extractor combined with the ANN classifier, the accuracy reached 98.97% for dataset-1 and 98.89% for dataset-3. Significantly, we see in Figure 6 that the most remarkable classification performance was achieved with dataset-1, which highlights the relevance of the data contained in this dataset for this specific task. In the case of dataset-2 and dataset-3, we also obtained good results for certain cases of combination between extraction methods with the ANN classifier. However, a contrasting observation emerges when analyzing

Table 3: Accuracy results of the classification procedure for each dataset (Experiment 1)

	Classifiers											
FE method	SVM	KNN	ANN	SVM	KNN	ANN	SVM	KNN	ANN	SVM	KNN	ANN
Resnet50	98.94	96.82	**100**	91.40	88.65	98.95	90.56	88.33	97.22	86.06	84.54	93.47
Xception	98.21	93.95	99.55	91.32	83.16	98.89	85.56	82.22	97.78	81.14	74.01	91.25
DensNet201	98.62	98.58	99.31	93.48	92.37	98.42	88.33	89.44	**100**	85.83	87.23	91.94
EfficientNetB7	98.61	97.18	99.31	93.51	94.04	96.84	85.00	86.11	93.89	78.33	77.51	93.47
Inception V3	97.52	96.45	98.62	92.87	88.48	99.47	86.67	85.00	**100**	80.67	80.68	84.09
VGG16	96.83	94.70	97.24	85.26	86.40	97.89	88.89	86.11	97.22	77.87	79.16	85.97
LBP	93.60	88.66	98.97	84.82	79.94	91.37	77.22	81.11	98.89	63.34	66.52	71.66
HOG	96.08	88.66	98.28	84.18	81.52	99.47	86.11	61.11	99.44	78.33	77.86	84.26
Datasets	Datasets-1			Datasets-2			Datasets-3			Datasets-4		

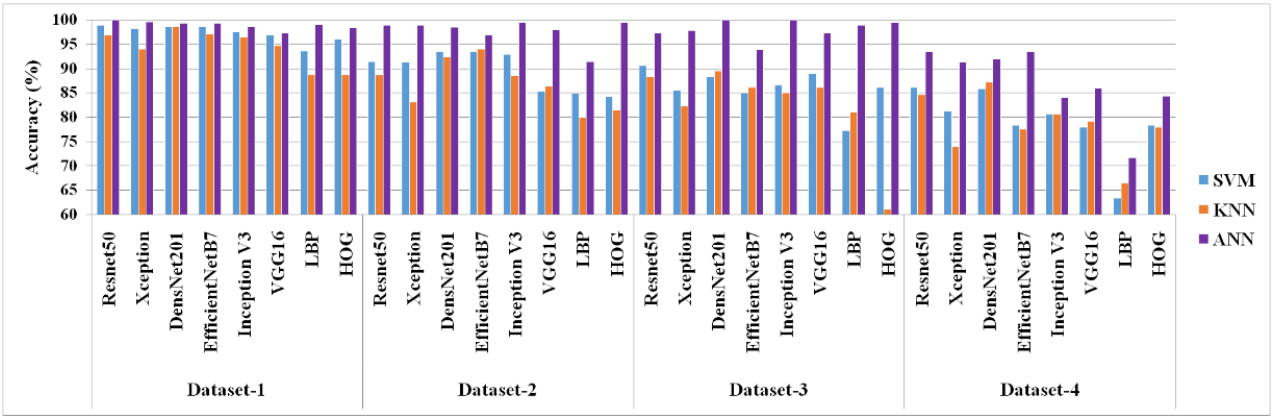

Figure 6: Accuracy of each classifier evaluated for each dataset (Experiment 1).

the results of dataset-4, where less satisfactory performances were obtained in the two experiments carried out. This variation between datasets highlights the critical importance of data quality and representativeness to ensure robustness of feature extraction and classification methods. Thus, it is essential that an expert radiologist conducts a thorough analysis of the dataset-4 to detect and eliminate any outliers, with the aim of improving the quality of this dataset. Consequently, we opted for the exclusion of dataset-4 in the remainder of this work.

4.2.2 Experiment 2

In this second experiment, dataset-4 was excluded while dataset-1, dataset-2 and dataset-3 were kept and merged to form a single set consisting of 646 images (323 benign images and 323 malignant images). The main objective is to fully exploit the wealth of information they contain, to improve the overall performance of the procedure adopted in our study. The results of this second experiment are presented in Table 4 and Figure 7 demonstrate that the combination of the DenseNet201 CNN-FE with the ANN classifier constitutes the best combination, giving 100% accuracy.

It is also evident that the ANN classifier showed the highest performance compared to the SVM and KNN classifiers (see Figure 7), regardless of its combination with the feature extractors. The best performance obtained with the LBP extractor is produced when combined with the ANN classifier, giving an accuracy of 86.58%, specificity of 96.23% and sensitivity of 75.66%. This case presents an imbalance between specificity, which measures the procedure's ability to correctly classify negative cases, and sensitivity, which evaluates the procedure's ability to correctly classify positive cases. Yet, combining the HOG extractor with the ANN classifier led to significantly higher accuracy, reaching 98.46%, with a specificity of 98.18% and a sensitivity

Table 4: Performances of the classification procedure for the three merged datasets (Experiment 2)

Feature Extractor		Classifier	Performance (%)		
			Accuracy	Specificity	Sensitivity
Transfer learning (TL)	ResNet50	SVM	94.89	94.73	95.29
		KNN	90.09	84.94	95.57
		ANN	99.38	99.55	98.57
	Xception	SVM	92.88	91.85	94.30
		KNN	86.07	76.11	96.13
		ANN	98.77	98.84	98.20
	DenseNet201	SVM	93.96	94.01	94.32
		KNN	93.50	90.85	96.57
		ANN	**100**	**100**	**100**
	Inception V3	SVM	90.71	90.06	91.89
		KNN	90.25	86.78	93.87
		ANN	97.38	98.46	95.71
HandCrafted	LBP	SVM	78.79	76.79	80.97
		KNN	83.58	85.68	81.80
		ANN	86.58	96.23	75.66
	HOG	SVM	85.91	85.23	78.72
		KNN	78.48	61.49	96.02
		ANN	98.46	98.18	97.50

of 97.50%.

4.2.3 Comparison with Published Works

Comparison with other works, in the context of the classification of US breast cancer images, is very difficult because of the difference in protocols and databases used for evaluation. To enable a fair comparison, we made a comparison with works evaluated on the same datasets. According to Table 5, our method compares favorably, or even outperforms the results obtained by other approaches using the same databases considered in this study. This performance is mainly due to the use of the CNN models Resnet50, Inception and DenseNet201 as feature extractors, combined with the ANN classifier whose learning is regulated by the ES criterion. This regulation permits to avoid overfitting of the ANN classifier, which helps to improve the performance of the proposed procedure. In perspective, the emphasis will be on categorizing ultrasound images based on the multiclass BI-RADS (Breast Imaging-Reporting and Data System) [48] [49] level by leveraging other open-source datasets.

4.2.4 Discussions

The two experiments carried out in this work allowed us to make the following observations:

- The results of the first experiment showed that the ANN classifier significantly outperformed SVM and KNN, regardless of the feature extraction method used. This observation underlines the effectiveness of the approach based on artificial neural networks, regulated by the ES criterion.
- The quality and representativeness of the data in a dataset are critical to classification and feature extraction performance.
- It was found that dataset-4 requires in-depth analysis by an expert to eliminate outliers to improve its quality.

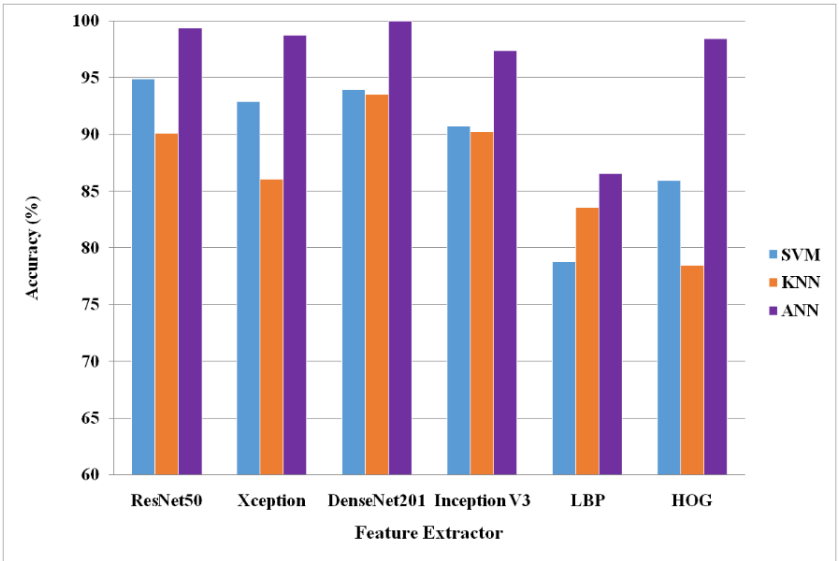

Figure 7: The accuracy of each classifier on the merged datasets (Experiment 2)

Table 5: Comparison of the proposed procedure with state-of-the-art ultrasound breast cancer classification methods

Paper	**FE methods**	**Classification methods**	**Image Dataset**	**Average Accuracy**
Wilding R and al. [24] (2022)	CNN	End- to-end CNN	BUSI + Dataset B	96 %
Labcharoenwongs P et al. [25] (2023)	CNN	End-to-end CNN (Yolo7)	Thammasat	95.07%
Sirjani N et al. [26] (2023)	CNN	End-to-end CNN (InceptionV3)	BUSI + Dataset B + Thammasat	81%
Zhuang Z et al. [16] (2021)	CNN	ANN	BUSI + OMI + Dataset B	95.48%
The proposed procedure	CNN-Resnet50	ANN	BUSI	100%
	CNN-InceptionV3		Thammasat	100%
	CNN-DenseNet201		BUSI + Dataset B + Thammasat	100%

- In the second experiment, we merged datasets-1, datasets-2 and datasets-3 to evaluate the adopted procedure. The results showed that in this context, the DenseNet201 model adapted as a feature extractor and combined with the ANN classifier gave the best accuracy.

- Using the HOG extractor with an ANN classifier resulted in high and balanced performance. This highlights that handcrafted methods may be appropriate in ultrasound image processing approaches for breast tumor classification.

- Our procedure compares favorably or even surpasses other methods for the same ultrasound image datasets, mainly through the use of CNN-FE models as feature extractors together with an ANN classifier regulated by the ES criterion.

5 Conclusion

In this paper, we consider the implementation of a procedure for classifying breast tumors in ultrasound imaging. We have used six CNN-FE models pre-trained by transfer learning and two handcrafted methods to extract features, combined with three classifiers, namely ANN, SVM and KNN. Features are first extracted from breast tumor US images after being denoised and equalized. Then, SVM, KNN and ANN classifiers are respectively utilized to learn these extracted features. The obtained findings show an improvement in classification performance, whether in the case without fusion or when the datasets are merged. The ANN classifier stands out for its superior performance in terms of generalization, while maintaining low complexity architecture compared to linear SVM and KNN classifiers. The results clearly demonstrate the effectiveness of using pre-trained Resnet50, InceptionV3 and DenseNet201 CNN models as feature extractors compared to handcrafted extractors, namely LBP and HOG. We compared our approach to other state-of-the-art methods for classifying breast cancer from US images using the same datasets 1, 2 and 3. Two series of experiments were carried out: one without fusion of ultrasound images and the other with fusion (except for dataset-4). The results showed outstanding performance of the ANN classifier in both cases, outperforming KNN and SVM with 100% accuracy. Among the studied CNN-FE models, ResNet50, DensNet201, and InceptionV3 achieved 100% accuracy without fusion, while DenseNet201 achieved 100% accuracy with ultrasound image fusion.
That said, implementing a neural network, whether for CNN models or for classification with an artificial neural network, remains a complex task. For instance, in the case of the ANN classifier, the number of neurons in the hidden layer was determined heuristically. Conducting experiments to identify the optimal number of neurons for our application would be more effective. This optimization could significantly enhance model performance. Additionally, other hyperparameters, such as the learning rate and regularization, could also benefit from systematic evaluation to maximize the neural network's efficiency.
For future research, the focus will also be on categorizing ultrasound images according to the BI-RADS [48] [49] multiclass level by leveraging other open-source datasets. This prompts us to consider the multiclass scenario for classifier design.

References

[1] American Cancer Society, "Cancer Facts and Figures", 2022. http://cancerstatisticscenter.cancer.org

[2] M. Seely and T. Alhassan,"Screening for Breast Cancer in 2018—What Should We Be Doing Today?", Current Oncology (Toronto, Ont.), 25(11):115–24, 2018. https://doi.org/10.3747/co.25.3770

[3] R. Guo, G. Lu, B. Qin, B. Fei, "Ultrasound Imaging Technologies for Breast Cancer Detection and Management: A Review", Ultrasound in medicine and biology, 44(1):37–70, 2018. https://doi.org/10.1016/j.ultrasmedbio.2017.09.012

[4] M. Bentoumi, M. Daoud, M. Benaouali, A. Taleb Ahmed, "Improvement of Emotion Recognition from Facial Images Using Deep Learning and Early Stopping Cross Validation", Multimedia Tools and Applications, 81(21):29887–29917, 2022. https://doi.org/10.1007/s11042-022-12058-0

[5] Y. Jiménez-Gaona, M. J. Rodríguez-Álvarez, V. Lakshminarayanan, "Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review", Applied Sciences, 10(22):8298, 2020. https://doi.org/10.3390/app10228298

[6] Z. He, Z. Chen, M. Tan, S. Elingarami, Y. Liu, T. Li, W. Li, "A Review on Methods for Diagnosis of Breast Cancer Cells and Tissues", Cell Proliferation, 53(7):e12822, 2020. https://doi.org/10.1111/cpr.12822

[7] H. Zhi, B. Ou, B. M. Luo, X. Feng, Y. L. Wen, H. Y. Yang, "Comparison of Ultrasound Elastography, Mammography, and Sonography in the Diagnosis of Solid Breast Lesions", Journal of Ultrasound in Medicine: Official Journal of the American Institute of Ultrasound in Medicine, 26(6):807–815, 2007. https://doi.org/10.7863/jum.2007.26.6.807

[8] M. Benaouali, M. Bentoumi, M. Touati, A. Taleb Ahmed, M. Mimi, "Segmentation and Classification of Benign and Malignant Breast Tumors via Texture Characterization from Ultrasound Images", in 2022 7th International Conference on Image and Signal Processing and Their Applications (ISPA), Mostaganem, Algeria, May 2022. https://doi.org/10.1109/ISPA54004.2022.9786350

[9] Y. LeCun, Y. Bengio, G. Hinton, "Deep Learning", Nature, 521:436–444, 2015. https://doi.org/10.1038/nature14539

[10] T. Sadad, A. Hussain, A. Munir, M. Habib, S. A. Khan, S. Hussain, M. Alawairdhi, "Identification of Breast Malignancy by Marker-Controlled Watershed Transformation and Hybrid Feature Set for Healthcare", Applied Sciences, 10(6):1900, 2020. https://doi.org/10.3390/app10061900

[11] A. K. Mishra, P. Roy, S. Bandyopadhyay, S. K. Das, "Breast Ultrasound Tumor Classification: A Machine Learning—Radiomics Based Approach", Expert Systems, 38(7):e12713, 2021. https://doi.org/10.1111/exsy.12713

[12] M. Wei, Y. Du, X. Wu, Q. Su, J. Zhu, L. Zheng, J. Zhuang, "A Benign and Malignant Breast Tumor Classification Method via Efficiently Combining Texture and Morphological Features on Ultrasound Images", Computational and Mathematical Methods in Medicine,5894010, 2020. https://doi.org/10.1155/2020/5894010

[13] W. Liu, M. Guo, P. Liu, Y. Du, "MfdcModel: A Novel Classification Model for Classification of Benign and Malignant Breast Tumors in Ultrasound Images", Electronics, 11(16):2583, 2022. https://doi.org/10.3390/electronics11162583

[14] K. Jabeen, M. A. Khan, M. Alhaisoni, U. Tariq, Y. D. Zhang, A. Hamza, R. Damaševičius, "Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion", Sensors, 22(3):807, 2022. https://doi.org/10.3390/s22030807

[15] M. Byra, "Breast Mass Classification with Transfer Learning Based on Scaling of Deep Representations", Biomedical Signal Processing and Control, 69:102828, 2021. https://doi.org/10.1016/j.bspc.2021.102828

[16] Z. Zhuang, Z. Yang, A. N. J. Raj, C. Wei, P. Jin, S. Zhuang, "Breast Ultrasound Tumor Image Classification Using Image Decomposition and Fusion Based on Adaptive Multi-Model Spatial Feature Fusion", Computer Methods and Programs in Biomedicine, 208:106221, 2021. https://doi.org/10.1016/j.cmpb.2021.106221

[17] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, "Deep Learning Approaches for Data Augmentation and Classification of Breast Masses Using Ultrasound Images", International Journal of Advanced Computer Science and Applications, 10(5), 2019. https://doi.org/10.14569/ijacsa.2019.0100579

[18] T. Pang, J. H. D. Wong, W. L. Ng, C. S. Chan, "Semi-Supervised GAN-Based Radiomics Model for Data Augmentation in Breast Ultrasound Mass Classification", Computer Methods and Programs in Biomedicine, 203:106018, 2021. https://doi.org/10.1016/j.cmpb.2021.106018

[19] A. E. Ilesanmi, U. Chaumrattanakul, S. S. Makhanov, "Methods for the Segmentation and Classification of Breast Ultrasound Images: A Review", Journal of Ultrasound, 24(4):367–382, 2021. https://doi.org/10.1007/s40477-020-00557-5

[20] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, "Dataset of Breast Ultrasound Images", Data in Brief, 28:104863, 2020. https://doi.org/10.1016/j.dib.2019.104863

[21] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Senti, R. Zwiggelaar, R. Marti, "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks", IEEE Journal of Biomedical and Health Informatics, 22(4):1218–1226, 2017. https://doi.org/10.1109/JBHI.2017.2731873

[22] M. Lin, Q. Chen, S. Yan, "Network in Network", arXiv:1312.4400v3, 2014. https://arxiv.org/abs/1312.4400

[23] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, "Learning Deep Features for Discriminative Localization", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. https://doi.org/10.1109/cvpr.2016.319

[24] R. Wilding, V. M. Sheraton, L. Soto, N. Chotai, E. Y. Tan, "Deep Learning Applied to Breast Imaging Classification and Segmentation with Human Expert Intervention", Journal of Ultrasound, 25:659–666, 2022. https://doi.org/10.1007/s40477-021-00642-3

[25] P. Labcharoenwongs, S. Vonganansup, O. Chunhapran, D. Noolek, T. Yampaka, "An Automatic Breast Tumor Detection and Classification Including Automatic Tumor Volume Estimation Using Deep Learning Technique", Asian Pacific Journal of Cancer Prevention, 24(3):1081–1088, 2023. https://doi.org/10.31557/APJCP.2023.24.3.1081

[26] N. Sirjani, M. G. Oghli, M. K. Tarzamni, M. Gity, A. Shabanzadeh, P. Ghaderi, M. Taghipour, "A Novel Deep Learning Model for Breast Lesion Classification Using Ultrasound Images: A Multicenter Data Evaluation", Physica Medica, 107:102560, 2023. https://doi.org/10.1016/j.ejmp.2023.102560

[27] B. Shareef, M. Xian, S. Sun, A. Vakanski, J. Ding, C. Ning, H. Cheng, "A Benchmark for Breast Ultrasound Image Classification", Available at SSRN: https://ssrn.com/abstract=4339660 or http://dx.doi.org/10.2139/ssrn.4339660

[28] Y. Yu, S. T. Acton, "Speckle Reducing Anisotropic Diffusion", IEEE Transactions on Image Processing, 11(11):1260–1270, 2002. https://doi.org/10.1109/TIP.2002.804276

[29] K. Drukker, C. A. Sennett, M. L. Giger, "Automated Method for Improving System Performance of Computer-Aided Diagnosis in Breast Ultrasound", IEEE Transactions on Medical Imaging, 28(1):122–128, 2009. https://doi.org/10.1109/TMI.2008.928178

[30] R. V. Menon, P. Raha, S. Kothari, S. Chakraborty, I. Chakrabarti, R. Karim, "Automated Detection and Classification of Mass from Breast Ultrasound Images", in 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 1:1-4, IEEE, 2015. https://doi.org/10.1109/NCVPRIPG.2015.7490070

[31] W. K. Moon, I. L. Chen, J. M. Chang, S. U. Shin, C. M. Lo, R. F. Chang, "The Adaptive Computer-Aided Diagnosis System Based on Tumor Sizes for the Classification of Breast Tumors Detected at Screening Ultrasound", Ultrasonics, 76:70–77, 2017. http://dx.doi.org/10.1016/j.ultras.2016.12.017

[32] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1:770-778, 2016. https://doi.org/10.1109/CVPR.2016.90

[33] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1:1800–1807, 2017. https://doi.org/10.1109/cvpr.2017.195

[34] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, "Densely Connected Convolutional Networks", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1:2261-2269, 2017. https://doi.org/10.1109/cvpr.2017.243

[35] M. Tan, Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", International Conference on Machine Learning, 1:6105-6114, May 2019. https://doi.org/10.48550/arXiv.1905.11946

[36] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the Inception Architecture for Computer Vision", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1:2818–2826, 2016. https://doi.org/10.1109/cvpr.2016.308

[37] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv preprint arXiv:1409.1556, 2014. https://doi.org/10.48550/arXiv.1409.155

[38] T. Ojala, M. Pietikäinen, D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions", Pattern Recognition, 29(1):51–59, 1996. https://doi.org/10.1016/0031-3203(95)00067-4

[39] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection", in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1:886-893, IEEE, 2005. https://doi.org/10.1109/CVPR.2005.177

[40] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer New York, 2000. https://doi.org/10.1007/978-1-4757-3264-1

[41] R. Min, D. A. Stanley, Z. Yuan, A. Bonner, Z. Zhang, "A Deep Non-linear Feature Mapping for Large-Margin kNN Classification", 2009 Ninth IEEE International Conference on Data Mining, 1:357–366, 2009. https://doi.org/10.1109/icdm.2009.27

[42] F. X. Diebold, R. S. Mariano, "Comparing Predictive Accuracy", Journal of Business and Economic Statistics, 20(1):134–144, 2002. https://doi.org/10.1198/073500102753410444

[43] M. Buckland, F. Gey, "The Relationship Between Recall and Precision", Journal of the American Society for Information Science, 45(1):12–19, 1994. https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L

[44] Stanislav Makhanov, "Ultrasound Images 2012", Online Medical Images, 2012. http://www.onlinemedicalimages.com/index.php/en/sitemap

[45] T. Geertsma, "Ultrasoundcases.info, FujiFilm", Ultrasoundcases.info, 2014. https://www.ultrasoundcases.info/cases/breast-andaxilla/

[46] J. Kriti Virmani, R. Agarwal, "Deep Feature Extraction and Classification of Breast Ultrasound Images", Multimedia Tools and Applications, 79:27257–27292, 2020. https://doi.org/10.1007/s11042-020-09337-z

[47] Klaus Nordhausen, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Second Edition by Trevor Hastie, Robert Tibshirani, Jerome Friedman, International Statistical Review, 77(3):482, 2009. https://doi.org/10.1111/j.1751-5823.2009.00095_18.x

[48] S. Ackerman, "ACR BI-RADS for Breast Ultrasound", Ultrasound in Medicine & Biology, 41(4):S95, 2015. https://doi.org/10.1016/j.ultrasmedbio.2014.12.391

[49] Sedgwick EL., "Chapter 12: Imaging Analysis: Ultrasonography", in Diseases of the Breast, 5th ed., J.R. Harris, M.E. Lippman, M. Morrow, and C.K. Osborne, Eds. Philadelphia, Pa: Lippincott Williams & Wilkins, 2014.