Deep Multiple Instance Learning for Automatic Breast Cancer Assessment Using Digital Mammography

Abdelali Elmoufidi, *Member, IEEE*

Abstract—Breast cancer causes serious public health problems; it is the most common cancer among women worldwide. Screening and early detection of signs of breast cancer increase the chance of survival. Early diagnosis is a crucial task for radiologists and physicians. Therefore, many computer-aided detection and diagnosis (CADx) systems are being developed to ensure the survival of radiologists' decisions. In this article, we describe a framework to automate assessment of suspicious regions, detected in screening mammography, without having carried out additional examinations, especially unnecessary biopsies in the case where the suspect regions are benign tumors. The setup of the proposed framework is ordered as follows: regions of interest (ROIs) have been segmented using a modified K-means algorithm; the bidimensional empirical mode decomposition (BEMD) algorithm is applied to derive many layers [bidimensional intrinsic mode function (BIMF)] from ROIs. Then, textural features are extracted from the obtained ROIs. First, directly from segmented ROI, second from the ROI and its sublayers (BIMFs + Residue). The features extracted in the second time have been grouped into a bag descriptive of the ROI under consideration. This bag is the input parameter of the classification algorithm based on the support vector machine (which has been confirmed to be beneficial for the classification of breast cancer). The average obtained sensitivity, specificity, accuracy, and area under the receiver operating characteristic (ROC) curve (AUC) rates, are, respectively, 98.60%, 98.65%, 98.62%, and 98.23%. Generally, the experimental results in INbreast, digital database of screening mammography (DDSM), and Mammography Image Analysis Society (MIAS) datasets demonstrate the robustness and the efficiency of the developed framework compared to previous works in the literature and have shown a significant advance.

Index Terms—Bidimensional empirical mode decomposition (BEMD), breast cancer, computer-aided diagnosis, deep learning, machine learning, mammography, texture feature.

I. INTRODUCTION

BREAST cancer causes serious problems of public health; it is the most frequently diagnosed cancer among women worldwide. According to the World Health Organization (WHO),[1] one in eight women is currently affected by breast cancer [1], [2]. Breast cancer develops in three-quarters of cases in women over 40 years, this disease is the most common and leading cause of death among women aged between 40 to 74 years.[1]

Actors in the medical field consider that increasing cure chances, decreasing the rate of morbidity and mortality caused by this disease are possible only if it is taken care of in the early stages of its appearance. Therefore, specialists have recommended early identification through a screening system. In fact, the most widely used screening tool is mammography, to ensure early detection of such breast changes. Consequently, radiologists have been led to increase the frequency of mammograms, especially for the most affected age groups. Generally, for each examination, four mammography images have been performed, two views [Cranio Caudal (CC) and medio lateral oblique MLO] per breast. This screening resulted in an exponential increase in mammograms number performed.

Seeing the high number of mammograms, the diagnosis task has become difficult for radiologists to handle. Mammogram evaluation is a critical task, subjective, and highly dependent on the radiologist's expertise. For that, the rate of breast cancer detection increases by about (15%–20%) with a second reading [3]–[5]. Except that the number of radiologists required for this job increases, the wait time is much longer, and the related charge are increased. However, approximately 10%–30% of mammary lesions are missing in screening due to limitations of human observation [6], [7]. Therefore, computer-aided detection (CAD) systems have emerged that introduce relatively novel technologies. These CADs have introduced some novel functions into the mammography device that allow for simple interpretation of mammograms. The systems also assist radiologists in increasing sensitivity to (21%–23%) in the detection of breast cancer [3], [6]. Therefore, they are helpful when there is a swell between observers, observation lack, or inability for double reading with two or more radiologists. Seeing exponential growth of new cases of breast cancer throughout the validation of the relevance of the last decades and after the CAD system's relevance validation for early breast cancer diagnosis. Much attention is devoted to this topic as a scientific, bioinformatics, and biomedical investigation axis. The goal of CAD systems is to assist radiologists in evaluating breast abnormalities, to see an automatic diagnosis based on the mammogram [8]–[11]. They have improved the diagnosis of breast abnormalities (masses, calcifications, microcalcification, architectural distortions, etc.). Many studies have been investigated to develop the CAD system, [12]–[16].

In fact, many CAD systems are based on artificial intelligence to automate the diagnosis of breast abnormalities using mammograms. The evaluation mammogram plays an important role in the prognosis of breast cancer. Therefore, the mammograms must be classified with maximum accuracy. Given the major difficulty of this task, as it is subjective and highly dependent on professional and human expertise. Therefore, in the last years, several scientific research has focused to develop automated classification methods, for the goal of helping radiologists to augment diagnosis accuracy, minimize errors and efforts [17]–[25]. These studies relate to the lives of human beings, so a slight improvement in performance has a major impact on the prognosis and research. Among the limitations of such methods, are stability and precision, as they which are still under discussion. On our side, in this article, we implement a framework to participate in automating regions of interest (ROI) diagnosis based only on mammography screening, and without the need for additional examinations, to know unnecessary biopsies in the case that suspect regions are benign. Multilayers based on bidimensional empirical mode decomposition (BEMD) is a method variant that can be used to solve the problems of feature extraction and selection. The BEMD algorithm gives an optimal solution to describe exactly ROIs and returns the classification step very easily using machine learning. The rest of the article is arranged as follows: Section II describes an overview of some previous related work. Section III introduces the materials and methods used. In Sections I–V, we present the experimental results of the proposed method. The discussion is given in Section V. Section VI summarizes the work of this article.

II. RELATED PREVIOUS WORKS

A. Bidimensional Empirical Mode Decomposition

Empirical mode decomposition (EMD) [26] is an approach adapted to nonstationary time frequency analysis and has been applied in various fields, i.e., biological [27], marine environment [28], structural diagnosis [29], mechanical diagnosis faults [30], [31] and other fields based on 1-D signal. In addition, Nunes *et al.* [2], [32] have introduced EMD in image processing (2-D signal) and they have developed the algorithm for 2-D decomposition in empirical mode (BEMD). This algorithm has attracted the attention of several researchers. Has been applied in image denoising [33]; image compression [34], [35]; image texture segmentation [36], [37], scaling image [38]; extraction of image characteristics [39], [40]; texture synthesis [41]; and classification of image texture [17].

In this article, we have implemented the BEMD to describe ROI in order to make the classification phase very easy given the use of several layers in the same region. Section II-B introduces some related previously developed methods of breast abnormalities assessment. Numerous promising approaches are emerging to solve the problems of assessing breast abnormalities in mammograms.

There are many methods of decomposition signals and images (discrete cosine transform (DCT), Wavelets, Fourier, etc.) which assume decomposition on a basis given *a priori*. The major advantage of empirical modal

decomposition (EMD) is the ability to define decompositions of the images, which do not depend on the choice of a particular base. In addition, the EMD is particularly well suited for the study of nonstationary signals. We will present the original EMD algorithm whose efficiency is recognized by the applications it allows processing. Similar to the EMD of the 1-D signal, the BEMD for a 2-D image based on the extrema that exist in the original image or obtained from the first derivative of the original or the higher-order derivative, to achieve the decomposition of the image signal. Distances between extrema may provide information to describe the image intrinsic length scales. The EMD method is an adaptive decomposition that allows the decomposition of any signal into a large set of intrinsic mode functions (IMFs) of the signals, denoted bidimensional IMF (BIMF) [42].

Given a 2-D image denoted by $I(m, n)$, the procedure of the BEMD can be described as follows [2], [27], and [42].

1- Initialization: $r_0(m, n) = I(m, n)$ where $(m, n) \in [0, M-1] \times [0, N-1]$, M and N represent numbers of the rank on the discrete image plane and $j = 1$ (index number of BIMF);

2- Extract the $j^{th} IMF$:

  a) Initialization parameters, $h_{j,0}(m, n) = r_{j-1}(m, n)$; $i = 1$;

  b) Obtain the local minima/maxima of $h_{j,i-1}(m, n)$;

  c) Compute upper envelope and lower envelope funtions by interpolate between local minima and between maxima, respectively, to get two envelope surfaces, $\text{Env}_{min,i-1}(m, n)$ and $\text{Env}_{max,i-1}(m, n)$ of $h_{j,i-1}(m, n)$.

  d) Compute the mean envelope surface of these two envelope surfaces: $Env_{moy,i-1}(m, n) = (Env_{min,i-1}(m, n) + Env_{max,i-1}(m, n))/2$;

  e) Update the original signal $h_{j,i}(m, n) = h_{j,i-1} - \text{Env}_{moy,i-1}$, $i = i + 1$;

  f) Calculate stopping criterion SD_{ij}; (standard deviation) given by (1); $SD_{ij} = \sum_{m=0}^{M} \sum_{n=0}^{N} \left[\frac{| h_{j,i-1}(m, n) - h_{j,i}(m, n) |^2}{h_{j,i-1}^2(m, n)} \right]$

  g) Repeat steps b) to f) until the calculated SD_{ij} is less than a predetermined criterion standard deviation (generally taken as 0.2-0.3) [43] and stop the iteration. Therefore, represents an IMF; that is, $imf_j(m, n) = h_{i,j}(m, n)$;

3- Update this signal and obtain the residual signal $r_j(m, n) = r_{j-1}(m, n) - imf_j(m, n)$;

4- If r_j is not monotonic, go to step 2 with $j = j + 1$ otherwise, stop the process of BEMD the decomposition is complete, finally obtain all IMF components.

5- When the decomposition is complet and the IMFs are extracted through the sifting process, we can write the original image $I(m, n)$ as the sum of all IMFs and the residual r_j, given by: $I(m, n) = \sum_{k=1}^{j} imf_k(m, n) + r_j(m, n)$.

B. Breast Abnormalities Classification

Currently, many methods based on supervised learning algorithms have been developed to assess breast abnormalities [1], [20], [22], [25]. In our previous work [1], we have proposed a CAD system to automate the diagnosis of breast cancer using mammography images and multiple instances learning algorithms. The proposed method contains the next steps: in order, automatically segment ROIs, textural and shape features are selected from detected ROIs, then classify ROIs as "benign," or "malignant," by implementing multiple instance learning (MIL) algorithms. Kayode *et al.* [22] have developed a method to automate the classification of mammograms based on a modified support vector machine. In this study, the authors have extracted textural features from mammograms using a gray level cooccurrence matrix in four different angular directions: $\theta = 0°, 45°, 90°, 135°$, with two distances: $D = 1, 2$. Then support vector machine (SVM) has been implemented to classify mammograms as normal, benign, or malignant. Limits of this study that the authors have implemented traditional techniques to extract features and no improvement was observed in the quantitative results obtained. Danala *et al.* [19] have implemented a method to classify breast masses as malignant or benign using a contrast-enhanced mammogram of the CAD scheme. In this study, the authors have selected 109 features separated by features class, which are tumor background density, tumor density related, shape, and wavelet. Then to classify suspicious breast masses, they have used a multilayer perceptron (MLP) based artificial neural network (ANN). The study novelties if compared with previous CAD schemes of full field digital mammography (FFDM) images are 1) segmentation technique used; 2) added lesion density heterogeneity features; 3) used average image features computed from CC and MLO views; and 4) implement an interactive visual aid tool for CAD scheme of contrast-enhanced digital mammography (CEDM) images. The limitations of this study are that the results obtained have a modest performance area under the receiver operating characteristic (ROC) curve (AUC = 0.848 ± 0.038) compared to the CAD scheme of low energy (LE) images with AUC = 0.753 ± 0.047. KULKARNI, Kulkarni and Stranieri [20] in this work, the authors have developed an approach to classify mammographic lesions in order to diagnose breast cancer. They have used the three most common classifiers, MLP, SVM, and K-nearest neighbors (KNNs). Rouhi *et al.* [25] in this article, the authors have presented a method to diagnose tumor benign or malignant using mammograms. First, they have segmented ROIs. In the next stage, they have extracted features from segmented tumors and applied genetic algorithms (GAs) to select the appropriate ones. ANNs have then been implemented to classify mammograms as benign or malignant. The authors have achieved good performance: the sensitivity, specificity, and accuracy rates are, respectively, 96.87%, 95.94%, and 96.47%. The obtained results in the previous works demonstrate some limits of the proposed CAD system. Abdelrahman *et al.* [44] have discussed in a survey which could serve as a road map for developing convolutional neural network (CNN)-based solutions to further improve mammographic detection of breast cancer. In this new work, we have obtained improvements in performance. These improvements have been obtained due to the feature extraction techniques that are used, by implementing the BEMD algorithm.

III. MATERIALS AND METHODS

In this article, we propose an approach to automate assessment abnormalities using mammograms, to suggest breast cancer diagnosis. This approach contains the following steps: we start by detecting ROIs, second, we apply the BEMD algorithm to derive the sublayers BIMF from the ROIs in question, and each BIMF represents a sublayer of the detected ROI. The third step is to extract the features of each layer and then combine them into a matrix. The last step is to classify the input using the support vector machine. In this section, we introduce the materials and methods used. First, we present the used datasets, next by implementing algorithms.

A. Used Dataset

To train, test, and validate the proposed framework, we need reference datasets. These datasets should be complete and include all possible study cases. Among these, we have found the three most well-known and public available databases, which are INbreast [45] database, digital database of screening mammography (DDSM) database [46], and Mammography Image Analysis Society (MIAS) database [47], which are the reference databases of most research work in the literature. Below we present the three databases INbreast, DDSM, and MIAS.

1) INbreast Database: INbreast [46] was collected at the Breast Center in the Centro Hospitalar de São João, Porto, under permission of both institutions: the National Committee of Data Protection and the Hospital's Ethics Committee, between April 2008 and July 2010. This study concerned 115 cases (examination records) studies, 90 of them contain two views (CC and Mediolateral oblique) of the left and right breast separately. The 25 cases that stay are from women who had a mastectomy and two views of only one breast were included. The database contains 410 mammography images. They were digitized at 70 μm and 14-bit depth pixels. The digitizer used is the SIEMENS Novation FFDM Digital Mammography. In addition, contextual description (digitization dates, the American College of Radiology (ACR), Bi-Rads, type abnormalities, etc.). Information on types of suspicious regions (a cluster of MCs or a mass). Fig. 1 shows an example of mammography image from *INbreast* database. More information regarding the INbreast database is provided in [45].

2) DDSM Database: The DDSM [47] was collected by cooperation between three institutions, which are the University of South Florida, the Massachusetts General Hospital (MGH), and the Sandia National laboratories. This database contains 2620 examination records (complete screening mammograms) collected at four institutions: Louis Medical Center (WU), MGH, Washington University of St and Sacred Heart Hospital (SH), and Wake Forest University School of Medicine (WFU), between October of 1988 and February of 1999. For
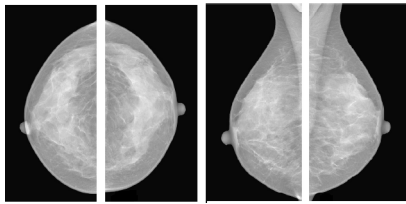
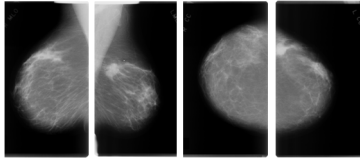Fig. 1. Example of mammography study from INbreast database.

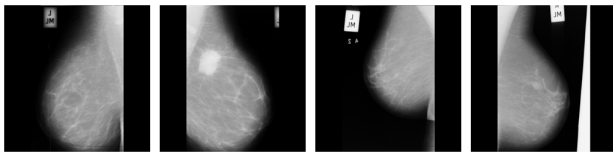Fig. 2. Example of mammography study from DDSM database.

Fig. 3. Example of mammography study from MIAS database.

each patient in the participant, the examination contains two mammogram views: The mediolateral oblique (MLO) views and the CC views of the right and left breasts separately. Mammograms were digitized at 50 μm and 12-bit depth pixels. The digitizers used are DBA, HOWTEK, and LUMISYS. In addition, the database contains contextual descriptions (digitizer, digitization dates, acquisition, patient age, etc.) and radiologist report. The mammograms containing abnormalities (have associated "ground truth" pixel-level) come with a manual segmentation by radiologist and abnormalities assessment (a mass or cluster of MC, benign or malignant). Fig. 2 shows an example of mammography image from DDSM database. More information regarding the DDSM database is provided in [46].

3) MIAS Database: The MIAS [43] is an organization of research groups in the U.K. interested in understanding and studying mammograms. It has collected a dataset of digital mammograms, from 161 patients taken from the U.K. National Breast Screening Program (U.K.) have been participated in this study. The database contains 322 screening mammograms. This dataset contains only MLO views of the left and right breasts. Mammography was digitized with a scanning micro densitometer at a resolution of 50 μm $\times$ 50 μm and 8-bit depth pixel. Mammography images have been reduced to 200-μm pixels. The database also includes radiologists' "ground truth-markings" at the locations of any abnormalities that may be present. Fig. 3 shows an example of mammography image from MIAS database. More detail about the MIAS database is provided in [47].

B. K-Fold Cross-Validation

Training and testing the parameters of a prediction function on the same dataset is a methodological error: a model that

would repeat the labels of the samples we have just seen would have a perfect score but could not predict anything useful on a new dataset. This case is called overfitting. Therefore, to avoid this problem, it is common during the building of a machine learning model (supervised) to keep a part of the dataset for tests. Therefore, in this work, we have used a ten-fold cross-validation technique to avoid overfitting.

C. Proposed Methodology

When suspicious lesions have been detected in a breast using a mammogram, additional examinations are necessary to determine whether they are cancer or not. For this, doctors perform a sample of the lesion, that is to say, a biopsy. Generally, it is very difficult to know the basis of mammograms whether a lesion is cancerous or not. For this, it is necessary to study the tissues of the lesion directly under the microscope, which makes it possible to see whether cancer cells are present or not. Therefore, it is necessary to take a tissue sample, different techniques exist (samples through the skin; ultrasound-guided biopsy; stereotaxic biopsy), and the choice of one or another depends mainly on the location of the lesion and its size. Although breast biopsies have generally been considered safe, they have certain risks, including bleeding; infection; swelling, and the appearance of bruises on the breast; a scar may remain visible after the breast has healed; also, a change in the appearance of the breast may be noticed, depending on the amount of tissue removed by doctors. Additionally, there is a slight risk that the biopsy result is falsely negative (that is, the test result is negative, but it is an error and there is really cancer). A false-negative result is inversely proportional to the amount of tissue removed. That is to say, in order to minimize the false-negative rate, if the amount of tissue removed is important. This implies an additional doubt about the exciting ones. Faced with these drawbacks, several research studies have focused on automating breast cancer diagnosis. On our part, in this work, we propose a very important new technique, which allows examining the suspect tissue in the mammography image and without having to resort to the biopsy. The strong point of the proposed approach implemented is the BEMD algorithm. The BEMD technique allows us to derive several layers of the same ROI; this allows us to analyze the 2-D image as being a 3-D image. The principle of the BEMD technique makes it possible to extract the different levels of BIMFs in the region in question; each level of BIMFs contains an additional texture detail of the detected ROI. Formally, each BIMF represents a layer and the residue is the last layer. Textural features selected from the obtained layers of the same region then are grouped together in a single vector. These vectors are the SVM-based input parameters of the SVM-based classification algorithm to classify them as "normal" or "abnormal."

1) Defining ROI: The aim of this work is to automate the assessment of the suspect regions detected. Therefore, first, we extract the ROIs based on our previous modified K-means algorithm [1] in order to classify them later. Figs. 6, 7, and 9 show three examples of suspect regions detected. Traditional classification techniques (by extracting features of the region under consideration and implementing supervised
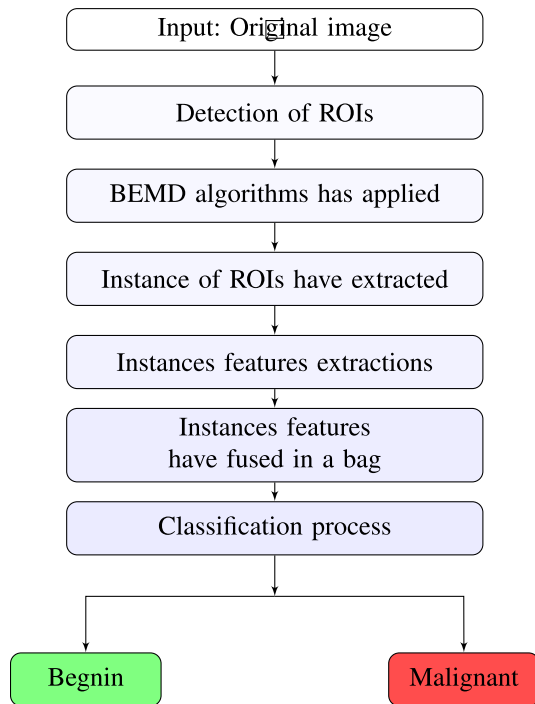
Fig. 4. Steps of the proposed method.

learning algorithms) have given good results, except that these results are too sensitive to the features extracted from detected ROIs, especially in the case where there are details hidden behind layers with lower or higher intensity levels. In this article, we have proposed a technique to solve this drawback, by decomposing the ROIs into several sublayers based on the BEMD algorithm in order to be able to describe these ROIs as much as possible. See four example listed in Figs. 9–12. Therefore, instead of classifying the ROI directly, we derive it into several layers (BIMFs); each layer contains specific detail. The combination of features layers in a matrix has improved the description of the region in question. Fig. 4 shows the steps of the proposed approach and Fig. 5 shows the architecture of our proposed approach.

2) Feature Extraction: After the ROIs have been detected and BEMD applied to derive sublayers (BIMFs levels) from the ROI into consideration. The next step is to extract the textural features from each layer using one of the following four well-known families. The Haralick features based on those matrices (Haralick), Gray-Level Cooccurrence Matrices (GLCM) [48], Gray-Level Run-Length Matrices (GLRLM) [49] or histograms of local binary patterns (LBPH) [50]. The features have been extracted from the layers, each layer represented by a single vector. The vectors of the same ROI have combined in a matrix (V1, V2, . . . , Vn) corresponding the layers (IMF1, IMF2, . . . , IMFn, Residue) describing a single ROI. These matrices are the input parameters of various classifiers.

3) Classifier Implimented: Mammography image classification plays an important role in breast cancer screening and diagnosis. Therefore, the images must be classified with maximum accuracy. Given the major difficulty of this task, as it is subjective and highly dependent on professional and

human expertise. Many types of research have focused on developing automated classification methods. In this article, we propose a framework based on SVM classifier types, with a new technique for extracting features of ROI using the BEMD algorithm to describe in detail ROI, and then obtain best classification accuracies. The traditional SVM is a nonprobability, linear or nonlinear kernel. In practice, we are going to implement an SVM with a nonlinear kernel that is the radial basis function (RBF), also tuning the two important hyperparameters of SVMs, C and gamma. The classification step goes through two phases: a learning phase and a testing phase. The learning phase constructs a model that will classify future unknown data. In the test phase, we will use the structure created previously to classify our unknown samples, and finally, we can calculate the accuracy of the classifier.

IV. EXPERIMENTAL RESULT ANALYSIS ON USED DATASET

Accurate breast tumor assessment of malignant or benign remains a challenging task actually. Although many CAD systems based on mammograms have been developed in order to assist radiologists. The classification stage represents a very crucial task in the CAD system development chain. Classification accuracy, using supervised classification algorithms strongly depends on using descriptors. Extracting features directly from the ROIs in question then applying a supervised classifier has almost become stationary and gives too close results in several previous works. Therefore, in this work, we applied the BEMD algorithm to derive several intensity levels (BIMFs) from ROIs. Consequently, each BIMF (layer) contains an additional detail. The next step is to extract the descriptors from each layer and then combine them get a matrix (V1, V2, . . . , Vn) that describes the ROI in consideration. These matrices are the input parameters of our classifier later. On the other hand, highly supervised learning algorithms were widely used in the field of classification of mammographic images. For this, and after studying the advantages and the disadvantages of these algorithms; we have used the SVM classifier as the best one. To evaluate the effectiveness of the proposed framework, we should apply it to large-scale datasets. Therefore, we have applied it to three publicly available datasets that are frequently used in breast cancer studies. The first one is INbreast [45], the second one is Digital Database for Screening Mammography [46], and the third one is MIAS database [47]. From the INbreast dataset, we have used 108 mammograms containing benign and malignant masses; from the DDSM database, we have used 1700 mammograms containing benign and malignant masses; and from the MIAS database, we have used 115 mammograms containing benign or malignant masses. Tables I and II include statistical information about using datasets.

A. Training Results

To validate the proposed method, we have built three models, which are trained in three different databases. As the used databases are not divided into subsets for learning and subsets for testing, we have used the cross-validation technique ten-fold to minimize over-learning. The following are the three proposed models.
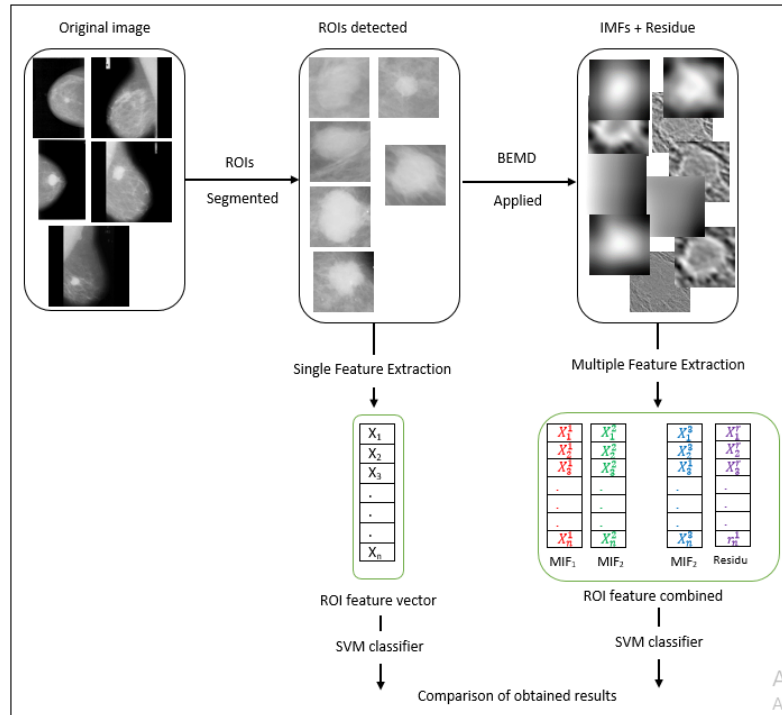
Fig. 5.   Proposed framework architecture for classifying breast abnormalities using mammography images.

TABLE I
STATISTICS ABOUT USING DATASETS

Dataset	Benign images	Malignant Images	Total
INbreast dataset [45]	59	49	108
DDSM dataset [46]	850	850	1700
MIAS dataset [47]	63	52	115
Total	972	951	1923

TABLE II
STATISTICS ABOUT USING ROIS EXTRACTED FROM USED DATASETS

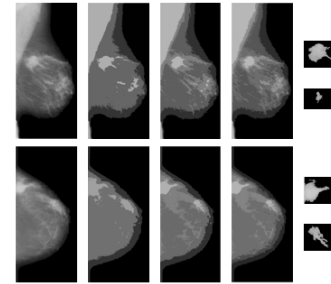| Dataset	Benign ROIS	Malignant ROIs	Total
INbreast dataset [45]	59	49	108
DDSM dataset [46]	1802	1844	3646
MIAS dataset [47]	71	65	136
Total	1932	1958	3890

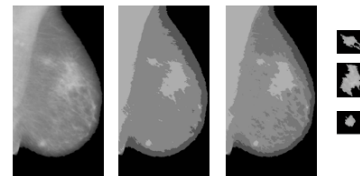Fig. 6.   Example 1: ROIs detected using DDSM database.

Fig. 7.   Example 2: ROIs detected using MIAS database.

1) Results on INbreast Dataset: This section describes a classifier model built using the INbreast database [45], which contains 115 patient records. In this study, we are interested in benign and malignant cases. Therefore, we have used 59 benign cases and 49 malignant cases. Knowing that in the INbreast database, we have two views CC and Mediolateral oblique (MLO) for each breast. Therefore, to classify a mammogram, we use either CC view or MLO view for the same breast. In addition, some patients contain tumors in both breasts; in this case, we treat each breast independently of the other, because the goal is to diagnose each breast separately using a mammogram. Table III below shows the quantitative

results obtained in detail. To summarize, the model achieved an AUC accuracy of 0.9832 ± 0.0039 using BEMD (multiple features), or an AUC accuracy of 0.9463 ± 0.0049 using single features. Fig. 13 shows the obtained ROC curves averaged over ten folds using the INbreast dataset and multiple features extracted based on the BEMD algorithm against single features extracted directly from ROI and SVM classifiers.

2) Results on DDSM Dataset: This section describes a classifier model built using the DDSM database [45], which contains 2620 patient records. In this study, we are interested in benign and malignant cases. Therefore, we have used
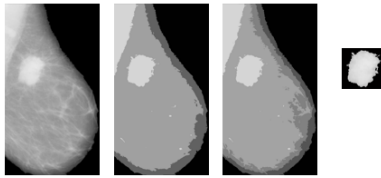
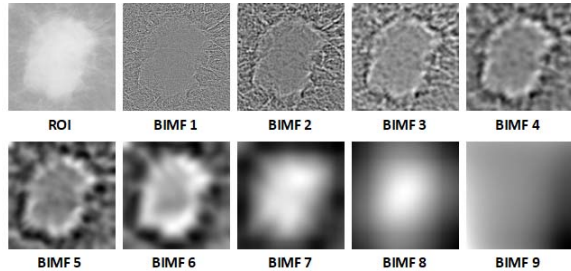Fig. 8. Example 3: ROIs detected using MIAS database.

Fig. 9. Example 3: decomposed BEMD components from malignant ROI segmented from the mammography image named mdb184 in MIAS dataset in the order of high to low-level intensities from left to right in each row.
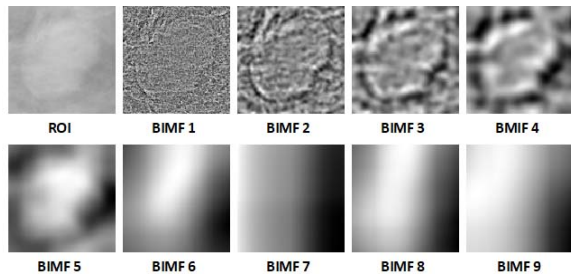

Fig. 10. Example 2: decomposed BEMD components from benign ROI segmented from the mammography image named mdb019 in the MIAS dataset, in the order of high to low-level intensities from left to right in each row.

Fig. 11. Example 3: decomposed BEMD components from malignant ROI segmented from the mammography image named case0089 in the DDSM dataset, in the order of high to low-level intensities from left to right in each row.
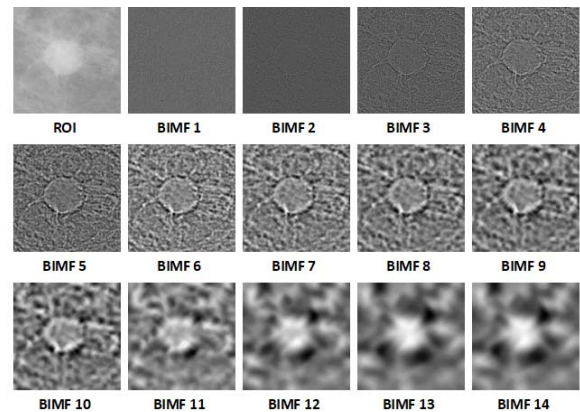
Fig. 12. Example 4: decomposed BEMD components from benign ROI segmented from the mammography image named case4121 in the DDSM dataset, in the order of high to low-level intensities from left to right in each row.

850 benign cases and 850 malignant cases. Knowing that in the DDSM database, we have both views: CC and Mediolateral oblique (MLO) for each breast. Therefore, to classify a mammogram, we have used a CC view or an MLO view for the same breast. In addition, some patients contain tumors in both breasts; in this case, we treat each breast independently of the other, because the goal is to diagnose each breast separately using a mammogram. Tables I–V shows the quantitative results obtained in detail. To summarize, the model achieved an AUC accuracy of 0.9820 ± 0.0040 using BEMD (multiple features), or an AUC accuracy of 0.9428 ± 0.0045 using single features. Fig. 14 shows the obtained ROC curves averaged over ten folds using the DDSM dataset and multiple features extracted based on the BEMD algorithm against single features extracted directly from ROI and SVM classifiers.

3) Results on MIAS Dataset: In this section, we have built a model trained using the MIAS dataset [47]. In this dataset, we have 322 mammography images, including 63 benign mammography images and 52 malignant mammography images. To classify a mammogram, we have used the profile view because in the MIAS database, we only have the profile view for each breast. Table V below shows the quantitative

results obtained in detail. To summarize, this model achieved an AUC of 0.9803 ± 0.0035 using BEMD, against an AUC of 0.9404 ± 0.0041 using single features. Fig. 15 shows obtained ROC curves averaged over ten folds by using the MIAS dataset and multiple features extracted based on the BEMD algorithm against single features extracted directly from ROI and SVM classifiers.

B. Testing Results

Tables VI–VIII below show the obtained results in the testing phase using the three training models formed, which are, respectively, the model trained on the INbreast dataset; the model trained on the DDSM dataset, and the model trained on the MIAS dataset. The ROI descriptors have been selected using both modalities mentioned previously: the first one, by using simple features, which are directly extracted from the ROIs detected. The second one, by using multiple features, which are extracted based on BEMD. Then, in the last stage, the SVM was applied as a classifier. An observable improvement in the results obtained results using the new technique implemented in this article can be stated, in which the features

TABLE III

MASS DIAGNOSIS PERFORMANCE AVERAGE AUC (95% CI[1]) AND STANDARD ERROR, USING INBREAST DATASET

Folds	Haralick MF[2]	Haralick SF[3]	GLCM MF	GLCM SF	GLRLM MF	GLRLM SF	LBPH MF	LBPH SF
Fold1	0.9834 ± 0.0047	0.9456 ± 0.0049	0.9826 ± 0.003	0.9457 ± 0.0043	0.9821 ± 0.0028	0.9442 ± 0.0056	0.984 ± 0.0033	0.9363 ± 0.0032
Fold2	0.9845 ± 0.0027	0.9484 ± 0.0056	0.9833 ± 0.004	0.9488 ± 0.0073	0.9792 ± 0.0034	0.9462 ± 0.004	0.9824 ± 0.0028	0.9401 ± 0.0046
Fold3	0.984 ± 0.0026	0.9432 ± 0.0048	0.9849 ± 0.0041	0.9459 ± 0.0059	0.9824 ± 0.0032	0.9411 ± 0.0069	0.9801 ± 0.0038	0.941 ± 0.0028
Fold4	0.984 ± 0.0031	0.9436 ± 0.0058	0.9816 ± 0.0047	0.9456 ± 0.0037	0.9816 ± 0.0034	0.9421 ± 0.0066	0.9805 ± 0.0042	0.9384 ± 0.0047
Fold5	0.9825 ± 0.0023	0.9487 ± 0.0063	0.9838 ± 0.0045	0.9436 ± 0.0035	0.9839 ± 0.0036	0.9425 ± 0.0061	0.9807 ± 0.0036	0.9404 ± 0.0039
Fold6	0.9816 ± 0.0043	0.945 ± 0.0061	0.9827 ± 0.0037	0.943 ± 0.0049	0.9805 ± 0.0047	0.946 ± 0.0043	0.9835 ± 0.0047	0.9376 ± 0.0024
Fold7	0.9821 ± 0.0038	0.9432 ± 0.0043	0.9843 ± 0.0041	0.946 ± 0.0043	0.9796 ± 0.0031	0.9461 ± 0.0056	0.9812 ± 0.0035	0.9352 ± 0.0045
Fold8	0.9811 ± 0.0034	0.945 ± 0.0066	0.9812 ± 0.0037	0.9494 ± 0.0041	0.9835 ± 0.0038	0.9417 ± 0.0039	0.9826 ± 0.0039	0.9358 ± 0.0027
Fold9	0.9843 ± 0.0029	0.9423 ± 0.0057	0.9847 ± 0.0047	0.9465 ± 0.0041	0.9838 ± 0.0047	0.9475 ± 0.0034	0.9832 ± 0.0046	0.9355 ± 0.0032
Fold10	0.9813 ± 0.0029	0.9475 ± 0.007	0.9833 ± 0.0031	0.949 ± 0.005	0.9797 ± 0.0043	0.9426 ± 0.0064	0.9838 ± 0.0044	0.9354 ± 0.0025
AVG	**0.9828 ± 0.0032**	**0.9452 ± 0.0056**	**0.9832 ± 0.0039**	**0.9463 ± 0.0049**	**0.9816 ± 0.0036**	**0.944 ± 0.0054**	**0.9822 ± 0.0038**	**0.9375 ± 0.0034**

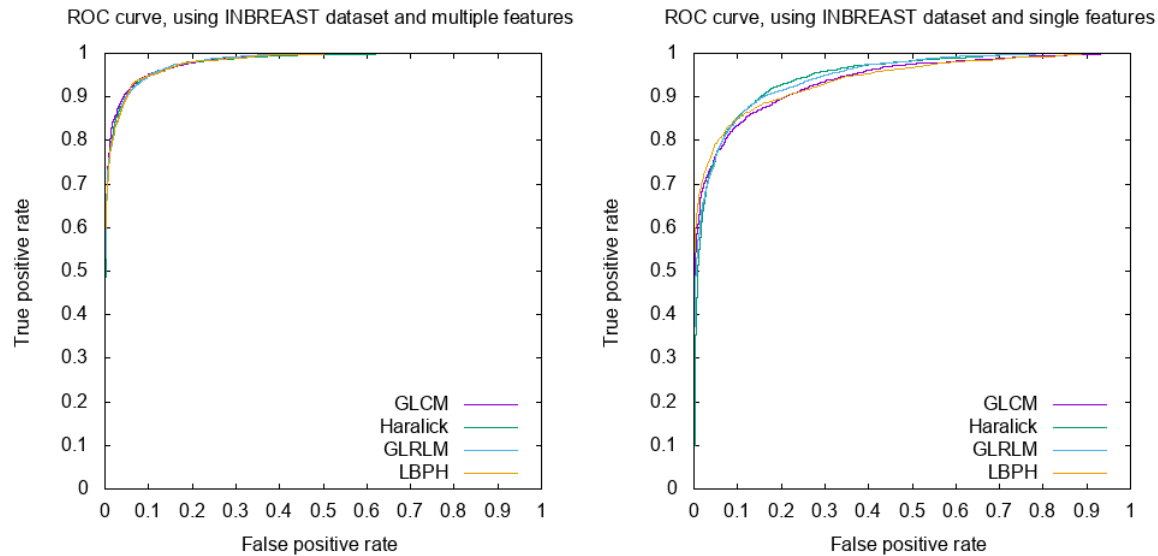CI[1]: Confidential Interval;   MF[2]: Multiple Features;   SF[3]: Single features

Fig. 13.   Obtained ROC curves averaged over ten folds using INbreast dataset and: 1) multiple features extracted based on BEMD (left) and 2) single features extracted directly from ROIs (right).

TABLE IV

MASS DIAGNOSIS PERFORMANCE–AVERAGE AUC (95% CI) AND STANDARD ERROR, USING DDSM DATASET

Folds	Haralick MF	Haralick SF	GLCM MF	GLCM SF	GLRLM MF	GLRLM SF	LBPH MF	LBPH SF
Fold1	0.9816 ± 0.0038	0.9410 ± 0.0040	0.9810 ± 0.0031	0.9397 ± 0.0052	0.9811 ± 0.0034	0.9401 ± 0.0033	0.9801 ± 0.0048	0.9407 ± 0.0048
Fold2	0.9792 ± 0.0045	0.9441 ± 0.0044	0.9817 ± 0.0039	0.9401 ± 0.0050	0.9827 ± 0.0025	0.9408 ± 0.0062	0.9818 ± 0.0032	0.9393 ± 0.0050
Fold3	0.9824 ± 0.0029	0.9415 ± 0.0037	0.9822 ± 0.0042	0.9424 ± 0.0040	0.9807 ± 0.0046	0.9427 ± 0.0047	0.9800 ± 0.0043	0.9407 ± 0.0059
Fold4	0.9813 ± 0.0040	0.9459 ± 0.0044	0.9805 ± 0.0042	0.9420 ± 0.0072	0.9813 ± 0.0032	0.9424 ± 0.0035	0.9806 ± 0.0034	0.9398 ± 0.0038
Fold5	0.9829 ± 0.0026	0.9434 ± 0.0048	0.9838 ± 0.0032	0.9413 ± 0.0044	0.9804 ± 0.0030	0.9408 ± 0.0035	0.9806 ± 0.0041	0.9384 ± 0.0047
Fold6	0.9792 ± 0.0043	0.9405 ± 0.0035	0.9828 ± 0.0049	0.9408 ± 0.0052	0.9809 ± 0.0045	0.9412 ± 0.0043	0.9815 ± 0.0027	0.9388 ± 0.0053
Fold7	0.9801 ± 0.0035	0.9403 ± 0.0062	0.9808 ± 0.004	0.9397 ± 0.0061	0.9819 ± 0.0042	0.9412 ± 0.0064	0.9819 ± 0.0036	0.9403 ± 0.0056
Fold8	0.9796 ± 0.0035	0.9421 ± 0.0063	0.9811 ± 0.0043	0.9425 ± 0.0069	0.9828 ± 0.0041	0.9399 ± 0.0044	0.9814 ± 0.0046	0.9401 ± 0.0064
Fold9	0.9792 ± 0.0047	0.9458 ± 0.0042	0.9828 ± 0.0044	0.9409 ± 0.0057	0.9828 ± 0.0047	0.9392 ± 0.007	0.9812 ± 0.0028	0.9385 ± 0.9385
Fold10	0.9825 ± 0.0046	0.9442 ± 0.0042	0.9836 ± 0.0038	0.9404 ± 0.0068	0.9821 ± 0.0042	0.9422 ± 0.0041	0.9804 ± 0.0046	0.9416 ± 0.0045
AVG	**0.9808 ± 0.0038**	**0.9428 ± 0.0045**	**0.9820 ± 0.0040**	**SF 0.9409 ± 0.0056**	**0.9816 ± 0.0038**	**0.9411 ± 0.0047**	**0.9809 ± 0.0038**	**0.9398 ± 0.0051**

have been extracted based on BEMD, compared to the classic method in which the features were extracted directly from the suspicious region in question. We distinguish between two kinds of obtained results: on the one hand, results obtained using the classical method of extracting characteristics directly from the ROIs, and the new method based on BEMD.

Table VI reveals the results obtained using the model trained on the DDSM dataset, then testing on INbreast and MIAS datasets. First, in INbreast: using single features (classical method), we obtained sensitivity, specificity, accuracy, and an AUC, which are, respectively, 0.9437, 0.9461, 0.9428, and 0.9419. Using multiple features (based on BEMD), we obtained a specificity, sensitivity, precision, and an AUC, which are, respectively, 0.9825, 0.9830, 0.9815, and 0.9803. Second, in the MIAS dataset: using the single features we obtained sensitivity, specificity, accuracy, and an AUC, which

TABLE V

MASS DIAGNOSIS PERFORMANCE–AVERAGE AUC (95% CI) AND STANDARD ERROR, USING MIAS DATASETS

| Folds	Haralick MF	Haralick SF	GLCM MF	GLCM SF	GLRLM MF	GLRLM SF	LBPH MF	LBPH SF
Fold1	0.9792 ± 0.0031	0.9408 ± 0.0035	0.9804 ± 0.0043	0.9401 ± 0.0041	0.9802 ± 0.0049	0.9407 ± 0.0022	0.9781 ± 0.0035	0.9400 ± 0.0054
Fold2	0.9786 ± 0.0020	0.9404 ± 0.0053	0.9804 ± 0.0029	0.9397 ± 0.0036	0.9801 ± 0.0024	0.9400 ± 0.0034	0.9789 ± 0.0027	0.9396 ± 0.0043
Fold3	0.9788 ± 0.0029	0.9405 ± 0.0032	0.9793 ± 0.0039	0.9402 ± 0.0033	0.9796 ± 0.0025	0.9406 ± 0.0030	0.9786 ± 0.0048	0.9394 ± 0.0054
Fold4	0.9808 ± 0.0045	0.9408 ± 0.0032	0.9807 ± 0.0021	0.9399 ± 0.0047	0.9810 ± 0.0050	0.9403 ± 0.0039	0.9783 ± 0.0042	0.9392 ± 0.0044
Fold5	0.9789 ± 0.0025	0.9405 ± 0.0062	0.9787 ± 0.0035	0.9396 ± 0.0053	0.9805 ± 0.0048	0.9411 ± 0.0037	0.9784 ± 0.0053	0.9390 ± 0.0059
Fold6	0.9789 ± 0.0033	0.9404 ± 0.0048	0.9803 ± 0.0044	0.9396 ± 0.0051	0.9800 ± 0.0024	0.9407 ± 0.005	0.9787 ± 0.0041	0.9403 ± 0.0025
Fold7	0.9789 ± 0.0041	0.9395 ± 0.004	0.9788 ± 0.0022	0.9408 ± 0.0056	0.981 ± 0.0031	0.9404 ± 0.0047	0.9799 ± 0.0044	0.9399 ± 0.0047
Fold8	0.9806 ± 0.0047	0.9404 ± 0.0056	0.9796 ± 0.0029	0.9409 ± 0.0041	0.9806 ± 0.0028	0.9399 ± 0.0053	0.9784 ± 0.0049	0.9391 ± 0.0049
Fold9	0.9801 ± 0.0036	0.9398 ± 0.0063	0.9812 ± 0.0042	0.9402 ± 0.0053	0.9803 ± 0.0047	0.9398 ± 0.0047	0.9793 ± 0.0026	0.9396 ± 0.0054
Fold10	0.9796 ± 0.0021	0.9408 ± 0.0030	0.9801 ± 0.0057	0.9402 ± 0.0053	0.9801 ± 0.0026	0.9411 ± 0.0054	0.9784 ± 0.0026	0.9405 ± 0.0046
AVG	0.9794 ± 0.0032	0.9403 ± 0.0045	0.9799 ± 0.0036	0.9401 ± 0.0046	0.9803 ± 0.0035	0.9404 ± 0.0041	0.9787 ± 0.0039	0.9396 ± 0.0047
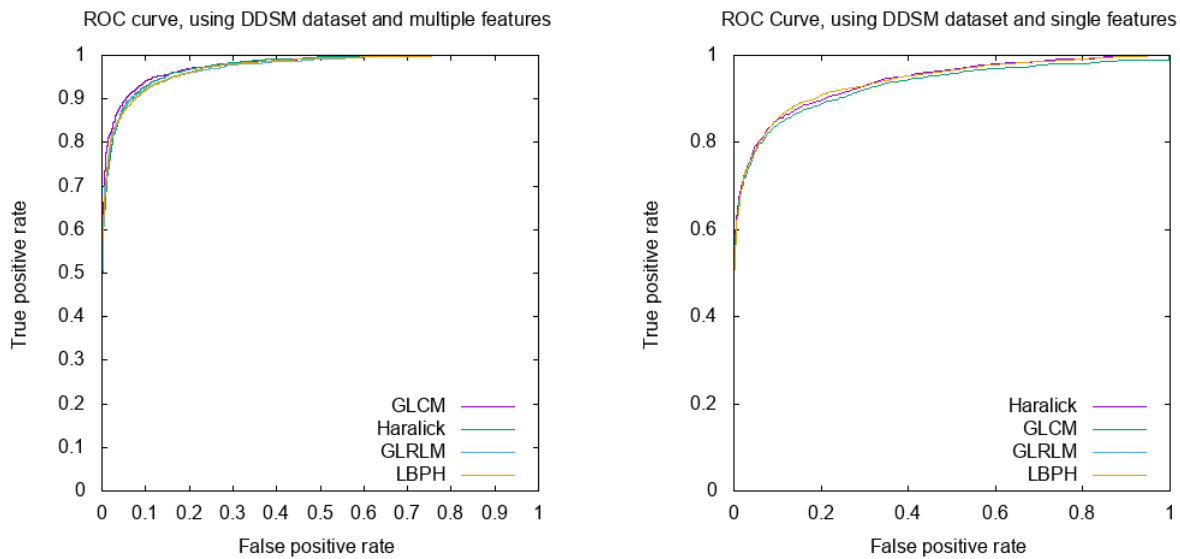
Fig. 14. ROC curves averaged over ten folds using DDSM dataset and: 1) multiple features extracted based on BEMD (left) and 2) single features extracted directly from ROIs (right).
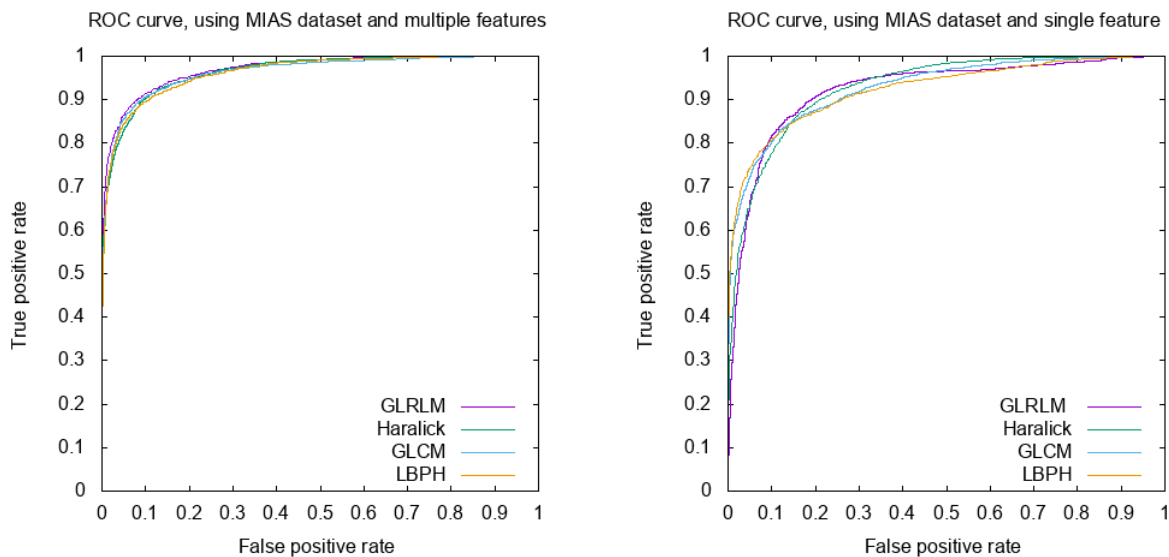

Fig. 15. Obtained ROC curves averaged over ten folds using MIAS dataset and: 1) multiple features extracted based on BEMD (left) and 2) single features extracted directly from ROIs (right).

TABLE VI

RESULTS OBTAINED IN INBREAST AND MIAS DATASET, USING THE MODEL TRAINED ON THE DDSM DATASET

Testing Set		Haralick		GLCM		GLRLM		LBPH	
		MF	SF	MF	SF	MF	SF	MF	SF
INbreast	Sen	1.00	0.9492	1.00	0.9492	0.9830	0.9492	0.9830	0.9322
	Spe	0.9796	0.9388	1.00	0.9388	0.9796	0.9592	0.9796	0.9388
	Acc	0.9907	0.9444	1.00	0.9444	0.9814	0.9537	0.9814	0.9351
	AUC	0.9833 ± 0.0039	0.9417 ± 0.0030	0.9827 ± 0.0044	0.9435 ± 0.0048	0.9843 ± 0.0052	0.9419 ± 0.0060	0.9813 ± 0.0042	0.9407 ± 0.0051
MIAS	Sen	0.9841	0.9365	0.9841	0.9523	1.00	0.9365	0.9841	0.9365
	Spe	0.9807	0.9411	1.00	0.9411	0.9803	0.9607	0.9803	0.9411
	Acc	0.9826	0.9304	0.9913	0.9391	0.9826	0.9391	0.9739	0.9304
	AUC	0.9818 ± 0.0055	0.9403 ± 0.0037	0.9819 ± 0.0053	0.9409 ± 0.0062	0.9833 ± 0.0051	0.9414 ± 0.0059	0.9807 ± 0.0058	0.9401 ± 0.0046

MF: Multiple Features;   SF: Single features;   AUC: Area Under the ROC Curve;   Sen: Sensitivity;   Spe: Specificity;   Acc: Accuracy

TABLE VII

RESULTS OBTAINED IN INBREAST AND DDSM DATASET, USING THE MODEL TRAINED ON THE MIAS DATASET

Testing Set		Haralick		GLCM		GLRLM		LBPH	
		MF	SF	MF	SF	MF	SF	MF	SF
INbreast	Sen	0.9661	0.9322	0.9831	0.9492	0.9796	0.9322	0.9661	0.9322
	Spe	0.9796	0.9388	0.9592	0.9388	1.00	0.9592	0.9796	0.9184
	Acc	0.9722	0.9352	0.9722	0.9444	0.9907	0.9444	0.9722	0.9259
	AUC	0.9813 ± 0.0050	0.9402 ± 0.0044	0.9822 ± 0.0072	0.9394 ± 0.0034	0.9818 ± 0.0053	0.9415 ± 0.0056	0.9815 ± 0.0069	0.9389 ± 0.0061
DDSM	Sen	0.9894	0.9482	0.9871	0.9458	0.9882	0.9424	0.9824	0.9471
	Spe	0.9847	0.9435	0.9906	0.9411	0.9835	0.9494	0.9859	0.9447
	Acc	0.9871	0.9459	0.9888	0.9435	0.9859	0.9459	0.9841	0.9459
	AUC	0.9814 ± 0.0046	0.9408 ± 0.0033	0.9810 ± 0.0059	0.9402 ± 0.0060	0.9817 ± 0.0071	0.9421 ± 0.0058	0.9811 ± 0.0054	0.9404 ± 0.0062

TABLE VIII

RESULTS OBTAINED IN MIAS AND DDSM DATASETS, USING THE MODEL TRAINED ON THE INBREAST DATASET

Testing Set		Haralick		GLCM		GLRLM		LBPH	
		MF	SF	MF	SF	MF	SF	MF	SF
DDSM	Sen	0.9824	0.9435	0.9882	0.9412	0.9859	0.9482	0.9847	0.9400
	Spe	0.9871	0.9400	0.9847	0.9447	0.9918	0.9459	0.9835	0.9424
	Acc	0.9841	0.9418	0.9865	0.9429	0.9888	0.9471	0.9841	0.9412
	AUC	0.9834 ± 0.0035	0.9399 ± 0.0057	0.9829 ± 0.0070	0.9393 ± 0.0045	0.9822 ± 0.0065	0.9413 ± 0.0054	0.9805 ± 0.0048	0.9400 ± 0.0074
MIAS	Sen	0.9683	0.9365	0.9841	0.9206	0.9841	0.9365	0.9683	0.9206
	Spe	0.9808	0.9231	0.9808	0.9423	1.00	0.9231	0.9615	0.9231
	Acc	0.9739	0.9304	0.9826	0.9304	0.9913	0.9304	0.9652	0.9217
	AUC	0.9819 ± 0.0034	0.9397 ± 0.0051	0.9813 ± 0.0053	0.9399 ± 0.0060	0.9817 ± 0.0047	0.9401 ± 0.0044	0.9809 ± 0.0050	0.9388 ± 0.0067

are, respectively, 0.9437, 0.9461, 0.9428, and 0.9419. With multiple features, we obtained specificity, sensitivity, accuracy, and an AUC, which are, respectively, 0.9825, 0.9830, 0.9815, and 0.9803.

Table VII reveals the results obtained using the model trained on the MIAS dataset, then testing on INbreast and DDSM datasets. First, in INbreast: using the single features we have obtained sensitivity, specificity, accuracy, and AUC, which are, respectively, 0.9437, 0.9461, 0.9428, and 0.9419. Using multiple features, we have obtained a specificity, sensitivity, accuracy, and an AUC, which are, respectively, 0.9825, 0.9830, 0.9815, and 0.9803. Second, in the DDSM dataset: using the single features we have obtained a sensitivity, specificity, accuracy, and an AUC, which are, respectively, 0.9437, 0.9461, 0.9428, and 0.9419. With multiple features, we have obtained a specificity, sensitivity, accuracy, and an AUC, which are, respectively, 0.9825, 0.9830, 0.9815, and 0.9803.

Table VIII presents the results obtained using the model trained on the INbreast dataset, and then tested on the DDSM and MIAS datasets. First, in DDSM: using the single features we have obtained a sensitivity, specificity, accuracy, and an AUC, which are, respectively, 0.9437, 0.9461, 0.9428, and 0.9419. Using multiple features, we have obtained a specificity,

TABLE IX

OBTAINED PERFORMANCE, USING THE PROPOSED METHOD IN THE DIFFERENT DATASETS

Performance		INbreast		DDSM		MIAS	
		MF	SF	MF	SF	MF	SF
Sen	Max	1.00	0.9492	0.9894	0.9492	1.00	0.9523
	Min	0.9661	0.9322	0.9824	0.9400	0.9683	0.9206
	Avg	0.9760	0.9407	0.9860	0.9446	0.9812	0.9345
Spe	Max	1.00	0.9388	0.9918	0.9494	1.00	0.9607
	Min	0.9592	0.9184	0.9835	0.9400	0.9615	0.9231
	Avg	0.9821	0.9413	0.9865	0.9440	0.9831	0.9371
Acc	Max	1.00	0.9537	0.9888	0.9471	0.9913	0.9391
	Min	0.9722	0.9259	0.9841	0.9192	0.9652	0.9217
	Avg	0.9826	0.9409	0.9862	0.9443	0.9804	0.9315
AUC	Max	0.9843	0.9435	0.9834	0.9421	0.9835	0.9414
	Min	0.9813	0.9389	0.9805	0.9393	0.9807	0.9388
	Avg	0.9823	0.9409	0.9818	0.9405	0.9817	0.9402

Max: Maximum;   Min: Minimum;   Avg: Average

sensitivity, accuracy, and an AUC, which are, respectively, 0.9825, 0.9830, 0.9815, and 0.9803. Second, in the MIAS dataset: using the single features we have obtained a sensitivity, specificity, accuracy, and an AUC, which are, respectively, 0.9437, 0.9461, 0.9428, and 0.9419. With multiple features, we have obtained a specificity, sensitivity, accuracy,

TABLE X

QUANTITATIVE OBTAINED RESULTS OF THE PROPOSED METHOD COMPARED TO RELATED PREVIOUS WORKS

Methodology	Dataset	Classifier	Sensitivity	Specificity	Accuracy	AUC
KAYODE, A, A et al. [22]	MIAS	SVM	0.944	0.913	0.9268	N/A
Our previous work [1]	DDSM	MI-SVM	0.9084	0.9017	0.9033	0.8840 ± 0.0097
	MIAS	MI-SVM	0.9008	0.9041	0.9019	0.8719 ± 0.0090
Rouhi R, et al. [25]	DDSM	MLP	0.9687	0.9594	0.9647	0.9510
	MIAS	MLP	0.9270	0.9054	0.9016	0.9558
Wang et al. [56]	DDSM	SVM	N/A	N/A	0.9274	0.9650
Liu et al. [53]	DDSM	SVM	0.92	0.93	0.93	0.9439
Dhungel, et al. [54]	INbreast	AlexNet+ MIL	N/A	N/A	0.91 ± 0.02	0.76 ± 0.23
Zhu, et al. [55]	INbreast	CNN+RF	N/A	N/A	0.90 ± 0.02	0.89 ± 0.04
Chakraborty, et al. [57]	DDSM	RF	0.85	N/A	0.833	0.92
Al-Antari, et al. [58]	INbreast	CNN	N/A	N/A	95.64	94.78
The proposed approach	DDSM	SVM	**0.9860**	**0.9865**	**0.9862**	**0.9818 ± 0.0039**
	MIAS	SVM	**0.9812**	**0.9831**	**0.9804**	**0.9817 ± 0.0050**
	INbreast	SVM	**0.9760**	**0.9821**	**0.9826**	**0.9823 ± 0.0045**

MIL: **M**ultiple **I**nstance **L**earning; **CNN**: **C**onvolutional **N**eural **N**etwork ; RF: Random Forest

and an AUC, which are, respectively, 0.9825, 0.9830, 0.9815, and 0.9803.

V. DISCUSSION

In this article, a novel computer-aided diagnosis system for the evaluation of abnormalities on mammograms has been presented. A novel strategy was proposed to drive sublayers from the ROIs according to the intensity of the level. The proposed study has shown that BEMD and SVM algorithms can be used to automate classifying breast abnormalities. It also represents an effective diagnostic tool to help radiologists make a good diagnosis. Detected ROIs in the early stages due to screening mammography have similar appearances with normal tissue; it is not like the obvious messes, with heterogeneous appearances and well-defined margins, whether they are malignant or benign. This similarity leads to major difficulties to make a precise diagnosis. Supervised learning algorithms are currently known in various classification and recognition operations. Except that, automating feature extraction to well describe ROIs (which are not limited to those conceived by humans) is a crucial step. Allows us to precisely classify various radiological appearances of cancer. For this, applying the BEMD algorithm makes it possible to observe several levels of intensity and textures, each level (BIMF) was considered as a layer, the cooperation of the layers and the study of the dependencies between them allow giving a good description of ROIs. Therefore, the classification step becomes obvious. SVM classifier and BEMD algorithm could contribute to the diagnosis of breast cancer in mammography by increasing the classification performance of tumors and by reducing unnecessary biopsies. To train, test, and validate the proposed framework, three public databases are used, which are called INbreast, DDSM, and MIAS. The use of diversified datasets (datasets have been scanned with different scanners and with different resolutions) is important to validate. The robustness of the proposed method. Obtained quantitative results in the three datasets are presented in Table IX below:

Table IX, summarizes the obtained performance. We have found that the proposed framework gives high performance and it is stable in the case where we have used BEMD. Whatever data has used either for training or for the test,

as well as whatever the technique used (Haralick, GLCM, GLRLM, LBPH) for ROIs feature extraction. Due to the diversified data, the proposed method has shown excellent performance in various trainings, testing, and validation datasets. Therefore, the robustness of the proposed method lie in the implemented BEMD technique, which allows one to have as much detail for each region to make the classification task as easy as the data. The proposed framework has the following characteristics: First, our method has shown significant performance in the classification of ROIs. This means that it can be approved as a technique to automate classify suspect regions in mammography screening, without having to resort to additional examinations and, especially, unnecessary biopsies. Significantly increasing the diagnostic accuracy of radiologists by overcoming the problems of traditional CAD. Second, the proposed method has shown that detecting and diagnosing breast cancer only by mammography screening could help reduce breast cancer and improve diagnostic results. Third, the diagnostic performance of the method used has less affected by breast density than that of radiologists, thanks to the implemented BEMD algorithm, which allows multiple layers of the intensity of the same ROIs to be exploited. Generally, radiologists' precision decreased with dense breasts, because the dense parenchymal tissue is more likely to mask suspect regions during mammography examinations [51]. The difference in radiological sensitivity between dense and fatty breasts was much greater than that of our method, so it can lead to a significant improvement in diagnostic performance assisted by radiologists in general and dense breasts in particular. Finally, automatic diagnosis instead of surgical biopsy for breast lesions leads to a reduction in the overall costs of treatment [52]. This reduction has been estimated to be between 40% and 85% [52]. Many similar studies on a CAD system for breast cancer diagnosis has recently been published [22], [25], [53]–[55] these studies allow us to make a comparison in order to assess the performance of our Framework. Table X below shows a comparison of the obtained results in this work and the results of previous works. Our approach is slightly better. In another word, a comparison between obtaining results using single features selected directly from ROI, and multiple features selected based on BEMD shows the

relevance of the BEMD paradigm for the task. In the proposed framework, the classification performances based on BEMD and SVM paradigm better results than a single feature with SVM classifier.

VI. Conclusion

In this proposal, we addressed the problem of breast cancer diagnosis using mammography imaging screening images. The objective is to automate the diagnosis based on mammography screening alone and without having to resort to additional examinations, especially unnecessary biopsies in cases where the detected ROIs are benign. The proposed framework, which is based on the BEMD algorithm to decompose ROI considered into layers (BIMFs) and the SVM algorithm for the classification task, has shown robustness despite the diversity of the dataset used for training or testing. As well as the results obtained have compared with others recently published in the literature and they showed a significant improvement in the diagnostic performance of breast cancer.

Declarations

References

[1] A. Elmoufidi, K. El Fahssi, S. Jai-Andaloussi, A. Sekkaki, Q. Gwenole, and M. Lamard, "Anomaly classification in digital mammography based on multiple-instance learning," *IET Image Process.*, vol. 12, no. 3, pp. 320–328, Mar. 2018.

[2] J. C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, and P. Bunel, "Image analysis by bidimensional empirical mode decomposition," *Image Vis. Comput.*, vol. 21, no. 12, pp. 1019–1026, 2003.

[3] J. M. Seely and T. Alhassan, "Screening for breast cancer in 2018—What should we be doing today?" *Current Oncol.*, vol. 25, no. 11, pp. 115–124, Jun. 2018.

[4] S. Ciatto *et al.*, "Second reading of screening mammograms increases cancer detection and recall rates. Results in the Florence screening programme," *J. Med. Screening*, vol. 12, no. 2, pp. 103–106, Jun. 2005.

[5] P. C. Brennan *et al.*, "Benefits of independent double reading in digital mammography: A theoretical evaluation of all possible pairing methodologies," *Acad. Radiol.*, vol. 26, no. 6, pp. 717–723, 2019.

[6] M. J. G. Calas, B. Gutfilen, and W. C. A. de Pereira, "CAD and mammography: Why use this tool?" *Radiol. Bras.*, vol. 45, no. 1, pp. 46–52, 2012.

[7] H.-E. Kim *et al.*, "Changes in cancer detection and false-positive recall in mammography using artificial intelligence: A retrospective, multireader study," *Lancet Digit. Health*, vol. 2, no. 3, pp. e138–e148, Mar. 2020.

[8] L. J. W. Burhenne *et al.*, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology*, vol. 215, no. 2, pp. 554–562, May 2000.

[9] G. Quellec, M. Lamard, M. Cozic, G. Coatrieux, and G. Cazuguel, "Multiple-instance learning for anomaly detection in digital mammography," *IEEE Trans. Med. Imag.*, vol. 35, no. 7, pp. 1604–1614, Jul. 2016.

[10] A. Elmoufidi *et al.*, "Automatic detection of suspicious lesions in digital X-ray mammograms," in *Proc. Int. Symp. Ubiquitous Netw.* Singapore: Springer, 2016, pp. 375–385.

[11] A. Elmoufidi, K. El Fahssi, S. Jai-Andaloussi, N. Madrane, and A. Sekkaki, "Detection of regions of interest's in mammograms by using local binary pattern, dynamic k-means algorithm and gray level co-occurrence matrix," in *Proc. Int. Conf. Next Gener. Netw. Services (NGNS)*, May 2014, pp. 118–123.

[12] R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," *Radiology*, vol. 219, no. 1, pp. 192–202, Apr. 2001.

[13] R. F. Brem *et al.*, "Improvement in sensitivity of screening mammography with computer-aided detection: A multiinstitutional trial," *Amer. J. Roentgenol.*, vol. 181, no. 3, pp. 687–693, Sep. 2003.

[14] A. Malich *et al.*, "Influence of breast lesion size and histologic findings on tumor detection rate of a computer-aided detection system," *Radiology*, vol. 228, no. 3, pp. 851–856, Sep. 2003.

[15] A. Malich, S. Schmidt, D. R. Fischer, M. Facius, and W. A. Kaiser, "The performance of computer-aided detection when analyzing prior mammograms of newly detected breast cancers with special focus on the time interval from initial imaging to detection," *Eur. J. Radiol.*, vol. 69, no. 3, pp. 574–578, Mar. 2009.

[16] A. Elmoufidi *et al.*, "Automatic diagnosing of suspicious lesions in digital mammograms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 5, pp. 510–518, 2016.

[17] Z. He, Q. Wang, Y. Shen, J. Jin, and Y. Wang, "Multivariate gray model-based BEMD for hyperspectral image classification," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 889–904, May 2013.

[18] T. Onega *et al.*, "Radiologists' perceptions of computer aided detection versus double reading for mammography interpretation," *Acad. Radiol.*, vol. 17, no. 10, pp. 1217–1226, Oct. 2010.

[19] G. Danala *et al.*, "Classification of breast masses using a computer-aided diagnosis scheme of contrast enhanced digital mammograms," *Ann. Biomed. Eng.*, vol. 46, no. 9, pp. 1419–1431, Sep. 2018.

[20] P. Kulkarni and A. Stranieri, "Pixel *N*-grams representation for medical image classification," in *Hybrid Machine Intelligence for Medical Image Analysis*. Singapore: Springer, 2020, pp. 21–40.

[21] P. J. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, "Multiple instance learning for histopathological breast cancer image classification," *Expert Syst. Appl.*, vol. 117, pp. 103–111, Mar. 2018.

[22] A. A. Kayode, N. O. Akande, A. A. Adegun, and M. O. Adebiyi, "An automated mammogram classification system using modified support vector machine," *Med. Devices*, vol. 12, pp. 275–284, Aug. 2019.

[23] J. Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artif. Intell.*, vol. 201, pp. 81–105, Aug. 2013.

[24] J. Foulds and E. Frank, "A review of multi-instance learning assumptions," *Knowl. Eng. Rev.*, vol. 25, no. 1, pp. 1–25, 2010.

[25] R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, "Benign and malignant breast tumors classification based on region growing and CNN segmentation," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 990–1002, 2015.

[26] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 460, no. 2046, pp. 1597–1611, 2004.

[27] W. Huang, Z. Shen, N. E. Huang, and Y. C. Fung, "Use of intrinsic modes in biology: Examples of indicial response of pulmonary blood pressure to ± step hypoxia," *Proc. Nat. Acad. Sci. USA*, vol. 95, no. 22, pp. 12766–12771, 1998.

[28] H. Song, Y. Bai, L. Pinheiro, C. Dong, X. Huang, and B. Liu, "Analysis of ocean internal waves imaged by multichannel reflection seismics, using ensemble empirical mode decomposition," *J. Geophys. Eng.*, vol. 9, no. 3, pp. 302–311, Jun. 2012.

[29] A. Garcia-Perez, J. P. Amezquita-Sanchez, A. Dominguez-Gonzalez, R. Sedaghati, R. Osornio-Rios, and R. J. Romero-Troncoso, "Fused empirical mode decomposition and wavelets for locating combined damage in a truss-type structure through vibration analysis," *J. Zhejiang Univ. Sci. A*, vol. 14, no. 9, pp. 615–630, Sep. 2013.

[30] J. Zheng, J. Cheng, and Y. Yang, "Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 40, no. 1, pp. 136–153, Oct. 2013.

[31] K. Zhu, X. Song, and D. Xue, "Incipient fault diagnosis of roller bearings using empirical mode decomposition and correlation coefficient," *J. Vibroeng.*, vol. 15, no. 2, pp. 597–603, 2013.

[32] J. C. Nunes, S. Guyot, and E. Deléchelle, "Texture analysis based on local analysis of the bidimensional empirical mode decomposition," *Mach. Vis. Appl.*, vol. 16, no. 3, pp. 177–188, 2005.

[33] Y. Zhou and H. Li, "Adaptive noise reduction method for DSPI fringes based on bi-dimensional ensemble empirical mode decomposition," *Opt. Exp.*, vol. 19, no. 19, pp. 18207–18215, 2011.

[34] A. Linderhed, "2D empirical mode decompositions in the spirit of image compression," *Proc. SPIE*, vol. 4738, no. 7, pp. 1–8, 2002.

[35] A. Linderhed, "Compression by image empirical mode decomposition," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, Sep. 2005, p. I-553.

[36] L. Qiao, K. Niu, N. Wang, and L. Peng, "Perfect reconstruction image modulation based on BEMD and quaternionic analytic signals," *Sci. China Inf. Sci.*, vol. 54, no. 12, pp. 2602–2614, Dec. 2011.

[37] Y. Chen, L. Wang, Z. Sun, Y. Jiang, and G. Zhai, "Fusion of color microscopic images based on bidimensional empirical mode decomposition," *Opt. Exp.*, vol. 18, no. 21, pp. 21757–21769, 2010.

[38] A. Linderhed, "Adaptive image compression with wavelet packets and empirical mode decomposition," Ph.D. dissertation, Linköpings Universitet, Linköping, Sweden, 2004.

[39] V. T. Tran, B.-S. Yang, F. Gu, and A. Ball, "Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 38, no. 2, pp. 601–614, Jul. 2013.

[40] A. Elmoufidi, A. Skouta, S. Jai-Andaloussi, and O. Ouchetto, "CNN with multiple inputs for automatic glaucoma assessment using fundus images," *Int. J. Image Graph.*, Jan. 2022, Art. no. 2350012.

[41] Z. Liu and S. Peng, "Boundary processing of bidimensional EMD using texture synthesis," *IEEE Signal Process. Lett.*, vol. 12, no. 1, pp. 33–36, Jan. 2005.

[42] N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[43] F.-P. An, D.-C. Lin, X.-W. Zhou, and Z. Sun, "Enhancing image denoising performance of bidimensional empirical mode decomposition by improving the edge effect," *Int. J. Antennas Propag.*, vol. 2015, pp. 1–12, Jan. 2015.

[44] L. Abdelrahman, M. Al Ghamdi, F. Collado-Mesa, and M. Abdel-Mottaleb, "Convolutional neural networks for breast cancer detection in mammography: A survey," *Comput. Biol. Med.*, vol. 131, Apr. 2021, Art. no. 104248.

[45] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a full-field digital mammographic database," *Acad. Radiol.*, vol. 19, no. 2, pp. 236–248, 2012.

[46] M. Heath et al., "Current status of the digital database for screening mammography," in *Digital Mammography*. Dordrecht, The Netherlands: Springer, 1998, pp. 457–460.

[47] J. Suckling et al., "The mammographic image analysis society digital mammogram database," in *Exerpta Medica: International Congress Series*, vol. 1069. Amsterdam, The Netherlands: Elsevier, 1994, pp. 375–378. Accessed: Oct. 28, 2016. [Online]. Available: http://peipa.essex.ac.uk/info/mias.html

[48] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[49] M. M. Galloway, "Texture analysis using grey level run lengths," NASA STI/Recon Tech. Rep. N 1974, vol. 75, p. 18555.

[50] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.

[51] P. E. Freer, "Mammographic breast density: Impact on breast cancer risk and implications for screening," *RadioGraphics*, vol. 35, no. 2, pp. 302–315, Mar. 2015.

[52] L. Liberman et al., "Impact of stereotaxic core breast biopsy on cost of diagnosis," *Radiology*, vol. 195, no. 3, pp. 633–637, Jun. 1995.

[53] X. Liu and J. Tang, "Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method," *IEEE Syst. J.*, vol. 8, no. 3, pp. 910–920, Sep. 2014.

[54] N. Dhungel, G. Carneiro, and A. P. Bradley, "The automated learning of deep features for breast mass classification from mammograms," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 106–114.

[55] W. Zhu, Q. Lou, Y. S. Vang, and X. Xie, "Deep multi-instance networks with sparse label assignment for whole mammogram classification," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 603–611.

[56] Y. Wang, J. Li, and X. Gao, "Latent feature mining of spatial and marginal characteristics for mammographic mass classification," *Neurocomputing*, vol. 144, pp. 107–118, Nov. 2014.

[57] J. Chakraborty, A. Midya, and R. Rabidas, "Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns," *Expert Syst. Appl.*, vol. 99, pp. 168–179, Jun. 2018.

[58] M. A. Al-Antari, M. A. Al-Masni, M.-T. Choi, S.-M. Han, and T.-S. Kim, "A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification," *Int. J. Med. Inform.*, vol. 117, pp. 44–54, Sep. 2018.

Abdelali Elmoufidi (Member, IEEE) received the master's and Ph.D. degrees in mathematics and computer science from the Faculty of Sciences Ain Chock, Hassan II University, Casablanca, Morocco, in 2012 and 2018, respectively.

Since 2019, he has been with the Faculty of Sciences and Techniques, Sultan Moulay Slimane University, Beni-Mellal, Morocco, as an Associate Professor. He is currently with the Department of Computer Science, Faculty of Sciences and Techniques, Sultan Moulay Slimane University. His research interests are related to based image retrieval (CBIR) and artificial intelligence for medical applications.