2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

NGMMs: Neutrosophic Gaussian Mixture Models for Breast Ultrasound Image Classification

Kuan Huang, Meng Xu*, and Xiaojun Qi

Abstract— **Ultrasound imaging is commonly used for diagnosing breast cancers since it is non-invasive and inexpensive. Breast ultrasound (BUS) image classification is still a challenging task due to the poor image quality and lack of public datasets. In this paper, we propose novel Neutrosophic Gaussian Mixture Models (NGMMs) to more accurately classify BUS images. Specifically, we first employ a Deep Neural Network (DNN) to extract features from BUS images and apply principal component analysis to condense extracted features. We then adopt neutrosophic logic to compute three probability functions to estimate the truth, indeterminacy, and falsity of an image and design a new likelihood function by using the neutrosophic logic components. Finally, we propose an improved Expectation Maximization (EM) algorithm to incorporate neutrosophic logic to reduce the weights of images with high indeterminacy and falsity when estimating parameters of each NGMM to better fit these images to Gaussian distributions. We compare the performance of the proposed NGMMs, its two peer GMMs, and three DNN-based methods in terms of six metrics on a new dataset combining two public datasets. Our experimental results show that NGMMs achieve the highest classification results for all metrics.**

I. Introduction

Breast cancer is one of the most common cancers among U.S. women. It alone accounts for 30% of female cancers [1]. About 13% of U.S. women develop breast cancer in their lifetime. Estimated 330,840 new women cases and 43,600 women death cases are reported in U.S. in 2020 [2]. Breast cancer can be classified into two classes: benign and malignant. In general, benign tumors have smooth shape while malignant tumors tend to have irregular border [3].

Early diagnosis and treatment of breast cancer are essential to reduce mortality [4]. Ultrasound imaging is a non-invasive, inexpensive, and effective diagnostic tool that has been commonly used to detect women breast cancers at early stages. The Breast Ultrasound (BUS) image classification is an effective tool to assist radiologists to make correct decisions [5]. It has been extensively studied for many years. However, it is still a challenging task due to the poor quality of BUS images and lack of public datasets.

Over recent decades, many methods have been proposed for breast cancer classification. Support Vector Machine (SVM), K-Nearest Neighbors (KNN), random forest, Multiple-Instance Learning (MIL), and Deep Neural Network (DNN) based methods have been well studied. Wang *et al.* [6] propose to extract a local descriptor named Phased Congruency-based Binary Pattern (PCBP) and feed it to the SVM for breast cancer classification. Liu *et al.* [7] employ the SVM on three edge-based features (e.g., sum of maximum curvature, sum of maximum curvature and peak, and sum of maximum curvature and standard deviation) extracted from BUS images for breast cancer classification. Cong *et al.* [8] propose to capture texture features in the Gray-Level Co-occurrence Matrix (GLCM) and apply a selective ensemble method that incorporates KNN, SVM, and Naive Bayes for classification. Ding *et al.* [9] propose a MIL algorithm to combine local distance and sparseness features and use KNN for classification. Bing *et al.* [10] propose a sparse representation based-MIL method that uses concentric circle to extract global and local features and employ a Relevance Vector Machine (RVM) for classification. To address speckle noise and low contrast issues, Abdel-Nasser *et al.* [11] propose to use a super-resolution computation method to reconstruct a high-resolution image from a set of input BUS images and then compute regions of interest and texture features to better represent a BUS image. Random forests are finally deployed on these features for breast cancer classification. Shi *et al.* [12] propose a stacked deep polynomial network that generates high-level texture features and achieves good classification accuracy on small ultrasound datasets. Virmani *et al.* [13] show that features extracted by DNNs (e.g., VGG [14] and ResNet [15]) are efficient for BUS image classification.

Gaussian Mixture Models (GMMs) are commonly used in BUS image segmentation to smooth images and remove noise [16]. However, they have not been widely used for BUS image classification. In this paper, we propose novel Neutrosophic Gaussian Mixture Models (NGMMs) to more accurately classify BUS images with tumors. Specifically, we train one NGMM with five Gaussian distributions for benign tumors and one NGMM with four Gaussian distributions for malignant tumors. Each NGMM uses probability functions to estimate three neutrosophic logic components (i.e., the truth, indeterminacy, and falsity) of an image. These components are then incorporated into an improved Expectation-Maximization (EM) algorithm to train the two NGMMs to represent benign and malignant tumors for more accurate tumor classification.

Our major contributions are: (1) Incorporating neutrosophic logic into GMMs to build NGMMs for BUS image classification; (2) Employing a DNN to extract features, which are further processed by Principal Component Analysis (PCA) to obtain condensed features; (3) Adopting neutrosophic logic to compute three probability functions to estimate the truth, indeterminacy, and falsity of an image to be-

*Corresponding author: Meng Xu meng.xu0201@gmail.com
Kuan Huang, Meng Xu, and Xiaojun Qi are with the Department of Computer Science, Utah State University, Logan, Utah, USA, 84322

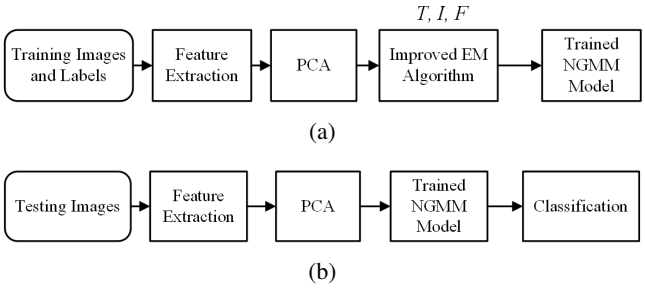

Fig. 1: Overview of NGMMs in (a) Training; (b) Testing.

long to an NGMM; (4) Designing a new likelihood function using the neutrosophic logic components; (5) Proposing an improved EM algorithm to employ neutrosophic logic-based probability functions to reduce weights of images with high indeterminacy and falsity when estimating parameters of an NGMM, which leads to better fit these images to Gaussian distributions in their respective NGMM; (6) Comparing the proposed NGMMs with its two peers, namely, conventional GMMs and Type-2 Fuzzy GMMs (T2FGMMs) [17], and three DNN-based methods on a new dataset to demonstrate its superior performance in tumor classification.

II. The Proposed Method

The proposed NGMMs has three stages: (1) feature extraction, (2) NGMMs construction, and (3) NGMMs classification. The first stage uses a DNN to extract features from input BUS images and then applies PCA to reduce dimensions. The second stage accomplishes three tasks: (1) designing three neutrosophic logic-based probability functions to estimate the truth, indeterminacy, and falsity of an image; (2) proposing a new likelihood function by incorporating neutrosophic logic components; (3) employing an improved EM algorithm to reduce the weights of images with high indeterminacy and falsity to better fit them to Gaussian distributions in their respective NGMM. The third stage uses two trained NGMMs to classify breast tumors.

In this section, we first introduce the overview of proposed NGMMs from both training and testing perspectives. We then present three neutrosophic logic components of NGMMs and how to use them to estimate appropriate weights of training images with different characteristics. Finally, we describe the training algorithm (e.g., improved EM) in detail.

A. Overview

Fig. 1 presents the block diagram of proposed NGMMs, where Fig. 1(a) and 1(b) show overview from training and testing perspectives, respectively.

As illustrated in Fig. 1(a), BUS images of arbitrary sizes are resized to 224×224 and then fed into a DNN for feature extraction. PCA is then employed on the reshaped feature map in the form of a vector to reduce its high dimension to 1×25 to keep important principal components that could lead to convergence of NGMMs. An improved EM algorithm, which incorporates the weights computed from neutrosophic logic for each training image, is applied on these preprocessed features to train two NGMMs to better capture characteristics of benign and malignant tumors. Specifically, the NGMM representing benign tumors consists of five Gaussian distributions and the NGMM representing malignant tumors consists of four Gaussian distributions. During the training procedure, three neutrosophic logic components, T, I and F, are calculated for each input image based on all Gaussian distributions in both NGMMs. These neutrosophic components are used in the improved EM algorithm to estimate the weights of each input image to more accurately train two NGMMs.

As illustrated in Fig. 1(b), the testing procedure follows the same flow. Each testing image goes through feature extraction and PAC process to obtain a preprocessed feature vector. This vector is then fed into two trained NGMMs to calculate its probability in each NGMM. It is finally classified to the category with the highest probability.

B. Neutrosophic Logic Components of NGMMs

We propose to incorporate neutrosophic logic into GMMs to build NGMMs to improve their classification power. Specifically, we use five Gaussian distributions to construct an NGMM to classify benign tumors and four Gaussian distributions to construct an NGMM to classify malignant tumors. In this subsection, we discuss three probability functions in each NGMM to compute neutrosophic components for estimating the weights of images with high indeterminacy and falsity. These weights facilitate each Gaussian distribution to pay less attention to images with high indeterminacy and falsity and therefore lead to better NGMM models to represent all training images.

For each training BUS image, we use T, I, and F to represent its truth, indeterminacy, and falsity membership in a Gaussian distribution, respectively. In the following, we explain the formulas of computing these three neutrosophic logic components.

Let C denote the number of categories (e.g., $C = 2$ for breast tumor classification), X denote the set of all training images containing either benign or malignant tumors, X^c denote the set of all training images in category c, and N_c denote the number of training images in category c. For an NGMM corresponding to category c, where $c = 1$ represents the benign category and $c = 2$ represents the malignant category, there are K_c Gaussian distributions. The probability of a BUS image x_i belonging to the k_c^{th} Gaussian distribution in category c is calculated by:

$$p_{k_c}(x_i) = \omega_{k_c} \mathcal{N}(x_i | \mu_{k_c}, \Sigma_{k_c}) \tag{1}$$

where k_c is the index of Gaussian distributions in the NGMM corresponding to category c (k_c ranges from 1 to 5 when $c = 1$ and k_c ranges from 1 to 4 when $c = 2$), and ω_{k_c}, μ_{k_c}, and Σ_{k_c} are the weights, mean vector, and co-variance matrix of the k_c^{th} Gaussian distribution, respectively. We then define $T_{k_c}(x_i)$, the truth membership for x_i to be in the k_c^{th} Gaussian distribution, as follows:

$$T_{k_c}(x_i) = \frac{p_{k_c}(x_i)}{\sum_{j=1}^{K_c} p_j(x_i)} \tag{2}$$

where $T_{k_c}(x_i)$ is the normalization of $p_{k_c}(x_i)$ and $\sum T_{k_c} = 1$. K_c is the number of Gaussian distributions in the NGMM corresponding to category c, where $K_c = 5$ when $c = 1$ and $K_c = 4$ when $c = 2$. The smaller the value of $T_{k_c}(x_i)$, the lower probability of x_i belonging to the k_c^{th} Gaussian distribution. An image x_i has the highest probability to belong to a Gaussian distribution in category c when its $T_{k_c}(x_i)$ corresponding to that Gaussian distribution is the highest. We then define $T^c(x_i)$, the truth membership for x_i to be in the c^{th} category, as follows:

$$T^c(x_i) = \max_{k_c=1,2,...,K_c} T_{k_c}(x_i) \tag{3}$$

If the value of $T_{k_c}(x_i)$ is small for all k_c's, it indicates that the probability of x_i belonging to category c is small. In other words, the falsity membership for x_i to belong to category c is high. To this end, we define the falsity membership for x_i to be in the c^{th} category as follows:

$$F^c(x_i) = 1 - T^c(x_i) \tag{4}$$

where $T^c(x_i)$ measures the highest probability for x_i to belong to category c. The smaller the $T^c(x_i)$ value, the higher the $F^c(x_i)$ value. The high $F^c(x_i)$ value indicates the high probability for x_i to falsely belong to category c.

We adopt the concept of entropy [18] to define the indeterminacy for an image x_i to belong to the c^{th} category as follows:

$$I^c(x_i) = -\frac{1}{\log K_c} \times \sum_{k_c=1}^{K_c} T_{k_c}(x_i) \log T_{k_c}(x_i) \tag{5}$$

when $T_{k_c}(x_i) = \frac{1}{K_c}$ for $k_c = 1, 2, ..., K_c$ (i.e., the probability of x_i belongs to each Gaussian distribution in category c is the same), the degree of indeterminacy $I^c(x_i)$ achieves the highest value of 1. This indicates x_i is at the most chaotic (i.e., indeterminacy) status.

By incorporating both indeterminacy and falsity, we define the likelihood of all images $x_i \in X^c$ to belong to the c^{th} category as follows:

$$\mathrm{L}^c(X^c) = \prod_{i=1}^{N_c} \left(\sum_{k_c=1}^{K_c} p_{k_c}(x_i) \right)^{G^c(x_i)} \tag{6}$$

where $G^c(x_i) = (1-I^c(x_i))(1-F^c(x_i))$. $I^c(x_i)$ and $F^c(x_i)$ measure the indeterminacy and falsity memberships for x_i in category c, respectively. This likelihood function utilizes neutrosophic logic to estimate the contributions of each training image in all Gaussian distributions of its associated GMM and adjust its likelihood based on its contributions.

C. Improved EM Algorithm

We propose an improved EM algorithm to train two NGMMs to represent two categories using benign and malignant training images. Specifically, we incorporate three neutrosophic logic components, namely, T, I, and F, into the conventional EM algorithm to assign weights to images with different characteristics when computing the log-likelihood of each training image in the c^{th} category. The log-likelihood instead of likelihood is employed in the EM algorithm to make it easier to compute the partial derivatives with regard to the weights, mean vector, and co-variance matrix of each Gaussian distribution. By employing the logarithmic operation on Eq. (6), we convert the likelihood to its log-likelihood by:

$$\ell^c(X^c) = \sum_{i=1}^{N_c} G^c(x_i) \log(\sum_{k_c=1}^{K_c} p_{k_c}(x_i)) \tag{7}$$

This conversion makes $G^c(x_i)$ become a positive scalar of the conventional log-likelihood employed in the EM algorithm. $G^c(x_i)$ takes into account of both $I^c(x_i)$ and $F^c(x_i)$ of an image x_i in category c and assigns a weight in the range of $(0, 1)$ based on the contribution of x_i in category c. Using the log-likelihood expression, the values of μ_{k_c}, Σ_{k_c}, and ω_{k_c} are respectively updated by taking a partial derivative with regard to μ_{k_c}, Σ_{k_c}, and ω_{k_c}, as follows:

$$\begin{aligned}
\mu_{k_c} &= \frac{\sum_{i=1}^{N_c} G^c(x_i) \cdot T_{k_c}(x_i) \cdot x_i}{\sum_{i=1}^{N_c} G^c(x_i) \cdot T_{k_c}(x_i)} \\
\Sigma_{k_c} &= \frac{\sum_{i=1}^{N_c} G^c(x_i) \cdot T_{k_c}(x_i) \cdot (x_i - \mu_{k_c})(x_i - \mu_{k_c})^{\mathrm{T}}}{\sum_{i=1}^{N_c} G^c(x_i) \cdot T_{k_c}(x_i)} \\
\omega_{k_c} &= \frac{\sum_{i=1}^{N_c} G^c(x_i) \cdot T_{k_c}(x_i)}{\sum_{i=1}^{N_c} G^c(x_i)}
\end{aligned} \tag{8}$$

The improved EM algorithm is summarized in **Algorithm 1**. It should be noted that the improved EM algorithm is individually employed on BUS benign training images and BUS malignant training images to train their corresponding NGMMs. For each category, we calculate a set of T, I and F values for each training image and incorporate them into the likelihood function to consider the contribution of each training image to all Gaussian distributions.

Algorithm 1 Improved EM Algorithm

Input: X^c containing a set of benign BUS images or malignant BUS images

Output: The parameters μ_{k_c}, Σ_{k_c} and ω_{k_c} for the c^{th} category's NGMMs

Initialization: Initialize the parameters μ_{k_c}, Σ_{k_c} and ω_{k_c} by k-means clustering where $k = K_c$.

while not converged **do**
  E-Step for each $x_i \in X^c$ **do**
    Calculate $p_{k_c}(x_i)$ using Eq. (1) for all k_c's.
    Calculate $T_{k_c}(x_i)$ using Eq. (2) for all k_c's.
    Calculate $F^c(x_i)$ using Eq. (4).
    Calculate $I^c(x_i)$ using Eq. (5).
  end E-Step for each $x_i \in X^c$
  M-Step **do**
    Re-estimate μ_{k_c}, Σ_{k_c} and ω_{k_c} using Eq. (8).
  end M-Step
  Evaluate log-likelihood using Eq. (7)
end while

When using the improved EM algorithm to train the NGMM for each category, $G^c(x_i)$ reduces the weights of

TABLE I: Summary of Classification Results ($\pm$ Standard Deviation)

Methods	TPR	FPR	ACC	PRE	F_1-score
LeNet	0.9270 $\pm$ 0.0608	0.3962 $\pm$ 0.0572	0.8217 $\pm$ 0.0409	0.8289 $\pm$ 0.0223	0.8743 $\pm$ 0.0330
VGG16	0.9417 $\pm$ 0.0517	0.4510 $\pm$ 0.3764	0.8137 $\pm$ 0.0964	0.8339 $\pm$ 0.1147	0.8769 $\pm$ 0.0516
ResNet-101	0.9450 $\pm$ 0.0431	0.1618 $\pm$ 0.0664	0.9101 $\pm$ 0.0277	0.9248 $\pm$ 0.0282	0.9338 $\pm$ 0.0215
VGG16 + GMMs	0.9531 $\pm$ 0.0331	0.4167 $\pm$ 0.2150	0.8329 $\pm$ 0.0712	0.8321 $\pm$ 0.0764	0.8866 $\pm$ 0.0447
VGG16 + T2FGMMs	0.9267 $\pm$ 0.0580	0.2536 $\pm$ 0.1787	0.8679 $\pm$ 0.0461	0.8906 $\pm$ 0.0711	0.9049 $\pm$ 0.0317
VGG16 + NGMMs	0.9219 $\pm$ 0.0700	0.1628 $\pm$ 0.1038	0.8943 $\pm$ 0.0583	0.9229 $\pm$ 0.0446	0.9209 $\pm$ 0.0454
ResNet-101 + GMMs	0.9457 $\pm$ 0.0343	0.3121 $\pm$ 0.1517	0.8617 $\pm$ 0.0565	0.8653 $\pm$ 0.0561	0.9028 $\pm$ 0.0375
ResNet-101 + T2FGMMs	0.9287 $\pm$ 0.0592	0.1700 $\pm$ 0.1477	0.8965 $\pm$ 0.0706	0.9213 $\pm$ 0.0633	0.9239 $\pm$ 0.0521
ResNet-101 + NGMMs	**0.9587 $\pm$ 0.0370**	**0.1249 $\pm$ 0.1085**	**0.9315 $\pm$ 0.0472**	**0.9424 $\pm$ 0.0458**	**0.9499 $\pm$ 0.0341**

images with either high indeterminacy or high falsity or both. For example, for an image with either high $I^c(x_i)$ or $F^c(x_i)$ value, the value of $G^c(x_i)$ is small and leads to a small value of $G^c(x_i) \cdot T_{k_c}(x_i)$ for three learnable parameters μ_{kc}, Σ_{kc} and ω_{kc}. In other words, the small value of $G^c(x_i)$ reduces the influence of image x_i when estimating the value of μ_{kc}, Σ_{kc}, and ω_{kc}. As a result, our improved EM algorithm is able to guide NGMMs to pay less attention to images with either high indeterminacy or falsity or both during the training procedure. It can better train the NGMM to represent all training images by considering their contributions in each Gaussian distribution.

III. Experiments

In this section, we describe the BUS datasets, explain the evaluation metrics, and present the experimental results of three DNN-based methods, the proposed NGMMs, and its two peer methods (conventional GMMs and T2GMMs).

Datasets. There are two publicly available BUS datasets: one is Data B [19] and the other is Data BUSI [20]. Data B contains 110 images with benign tumors and 53 images with malignant tumors (163 images in total) with an average size of 760×570 pixels. Data BUSI contains 487 images with benign tumors, 210 images with malignant tumors, and 133 images without tumors (830 images in total) with an average size of 500×500 pixels. Ethical approval is not required as confirmed by the license attached with the open access data.

Evaluation Metrics. We combine the images with benign and malignant tumors in both Data B and Data BUSI datasets (860 images in total) to get a new dataset. We perform 10-fold cross-validation to evaluate the performance of each compared classification method on the new dataset in terms of six evaluation metrics. These metrics are True Positive Ratio (TPR), False Positive Ratio (FPR), Classification Accuracy (ACC), Precision (PRE), F_1-score, and Receiver Operating Characteristic (ROC) curve. All classification results are presented by their mean values achieved over 10 runs $\pm$ their Standard Deviation (SD).

Experimental Results. Since we use the DNN to extract features to build the NGMMs, we compare the classification results of three common DNN-based methods (e.g., LeNet [21], VGG16 [14], and ResNet-101 [15]) on our new dataset. Their classification results in terms of five evaluation metrics are listed in the first three rows of Table I. It clearly shows all DNN-based methods achieve good BUS classification results, which are consistent with the findings in [13] that features extracted by DNNs are efficient for BUS image classification. ResNet-101 outperforms both VGG16 and LeNet and achieves the best classification result in terms of all metrics.

We then implement the proposed NGMMs using the best DNN for the BUS classification task (i.e., ResNet-101) to extract features and implement its peers (GMMs and T2FGMMs [17]) using ResNet-101 to extract features. We also implement their counterparts using VGG16 to extract features due to its simple network architecture. The classification results for these six methods are listed in the last six rows of Table I. Based on the features extracted by ResNet-101, NGMMs achieve the best classification results in terms of all five metrics and T2FGMMs outperform GMMs in terms of four metrics: FPR, ACC, PRE and F_1-score. Based on the features extracted by VGG16, it also shows that NGMMs achieve the best classification results in terms of all five metrics except for TPR. T2FGMMs achieve the second best classification results in terms of four metrics: FPR, ACC, PRE and F_1-score and GMMs have the highest TPR. Among all methods shown in Table I, the proposed ResNet-101+NGMMs achieve the highest TPR value of 0.9587, the lowest FPR value of 0.1249, the highest ACC value of 0.9315, the highest PRE value of 0.9424, and the highest F_1-score of 0.9499.

Fig. 2 shows ROC curves with values of Area Under the ROC Curve (AUC) for each method listed in table I. Among three DNNs, ResNet-101 yields the highest AUC value of 0.9618; VGG16 and LeNet yield significantly smaller AUC values of 0.8582 and 0.8456, respectively. Using the features extracted by VGG16, NGMM achieves the highest AUC value of 0.9212, which is marginally larger than the AUC value of 0.9150 obtained by T2FGMMs and significantly larger than the AUC value of 0.8792 obtained by GMMs. Using the features extracted by ResNet-101, NGMMs achieve the highest AUC value of 0.9631, T2FGMMs achieve the second highest AUC value of 0.9447, and GMMs achieve the smallest AUC value of 0.9080. Overall, the proposed ResNet-101+NGMMs achieve the best classification accuracy with the highest AUC of 0.9631.

Both Table I and Fig. 2 show the similar classification trends. The proposed NGMMs perform the best classification results using features extracted from the DNN when comparing with their peers (GMMs and T2FGMMs) in terms of all six evaluation metrics. In addition, it has the smallest SD values, which indicates that its classification results are stable

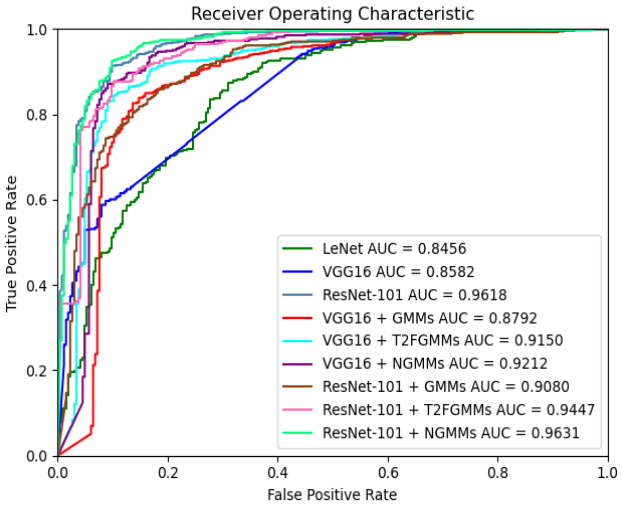

Fig. 2: ROC curves of nine compared classification methods.

across multiple runs. The proposed NGMMs also significantly improve their corresponding DNN-based classification methods (i.e., without using GMMs). Specifically, ResNet-101+NGMMs improves the ResNet-101 by 1.45% in TPR, 22.81% in FPR, 2.35% in ACC, 1.90% in PRE, and 1.72% in F1-score, respectively. The VGG16+NGMMs improves the VGG16 by 63.90% in FPR, 9.91% in ACC, 10.67% in PRE, and 5.02% in F1-score, respectively.

Implementation Details. All experiments are conducted on Ubuntu 18.04 system, Intel(R) Xeon(R) CPU E5-2620 2.00 GHz, and two NVIDIA GeForce 1080Ti graphics cards. For the training of convolutional neural networks, Adam optimizer is used with a learning rate of 0.0001, momentum parameters β_1 of 0.9, β_2 of 0.99, a weight decay of 0.0005, a batch size of 20, and the number of training epochs of 80. Cross-entropy is employed in the loss function. The feature map from the VGG16 network is extracted by its fifth convolutional block followed by a global average pooling, which is 1×512. The same operation is conducted to extract the feature map from the ResNet-101, which is 1×2048. Both reshaped features can be processed by PCA to reduce its dimension to 1×25. Features after PCA are utilized in the EM algorithm to train GMMs to represent benign and malignant tumors in two categories. The number of PCA components is 25. The number of Gaussian distributions in the NGMM for benign and malignant classes are empirically set to be 5 and 4, respectively.

IV. Conclusions

We propose novel NGMMs for BUS image classification. It employs a DNN for feature extraction to train two NGMMs to represent benign and malignant tumors in two categories. We first utilize the neutrosophic logic-based probability functions to estimate the truth, indeterminacy, and falsity membership of an image. We then incorporate these neutrosophic components to not only design a novel likelihood function but also reduce the weights of images with high indeterminacy and falsity when estimating the parameters of an NGMM in an improved EM algorithm. Overall, the proposed NGMMs outperform three state-of-the-art DNN-based methods and its two peer GMMs in terms of six metrics on the new dataset.

References

[1] R. L. Siegel, K. D. Miller, A. Goding Sauer, S. A. Fedewa, L. F. Butterly, J. C. Anderson, A. Cercek, R. A. Smith, and A. Jemal, "Colorectal cancer statistics, 2020," *CA: a cancer journal for clinicians*, vol. 70, no. 3, pp. 145–164, 2020.

[2] "U.s. breast cancer statistics," Feb 2021. [Online]. Available: https://www.breastcancer.org/symptoms/understand_bc/statistics

[3] R.-F. Chang, W.-J. Wu, W. K. Moon, and D.-R. Chen, "Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors," *Breast cancer research and treatment*, vol. 89, no. 2, pp. 179–185, 2005.

[4] X. Qi, L. Zhang, Y. Chen, Y. Pi, Y. Chen, Q. Lv, and Z. Yi, "Automated diagnosis of breast ultrasonography images using deep neural networks," *Medical image analysis*, vol. 52, pp. 185–198, 2019.

[5] Q. Huang, X. Huang, L. Liu, Y. Lin, X. Long, and X. Li, "A case-oriented web-based training system for breast cancer diagnosis," *Computer methods and programs in biomedicine*, vol. 156, pp. 73–83, 2018.

[6] L. Cai, X. Wang, Y. Wang, Y. Guo, J. Yu, and Y. Wang, "Robust phase-based texture descriptor for classification of breast ultrasound images," *Biomedical engineering online*, vol. 14, no. 1, pp. 1–21, 2015.

[7] Y. Liu, L. Ren, X. Cao, and Y. Tong, "Breast tumors recognition based on edge feature extraction using support vector machine," *Biomedical Signal Processing and Control*, vol. 58, p. 101825, 2020.

[8] J. Cong, B. Wei, Y. He, Y. Yin, and Y. Zheng, "A selective ensemble classification method combining mammography images with ultrasound images for breast cancer diagnosis," *Computational and mathematical methods in medicine*, vol. 2017, 2017.

[9] J. Ding, H. Cheng, M. Xian, Y. Zhang, and F. Xu, "Local-weighted citation-knn algorithm for breast ultrasound image classification," *Optik*, vol. 126, no. 24, pp. 5188–5193, 2015.

[10] L. Bing and W. Wang, "Sparse representation based multi-instance learning for breast ultrasound image classification," *Computational and mathematical methods in medicine*, vol. 2017, 2017.

[11] M. Abdel-Nasser, J. Melendez, A. Moreno, O. A. Omer, and D. Puig, "Breast tumor classification in ultrasound images using texture analysis and super-resolution methods," *Engineering Applications of Artificial Intelligence*, vol. 59, pp. 84–92, 2017.

[12] J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, and T. Wang, "Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset," *Neurocomputing*, vol. 194, pp. 87–94, 2016.

[13] J. Virmani, R. Agarwal *et al.*, "Deep feature extraction and classification of breast ultrasound images," *Multimedia Tools and Applications*, vol. 79, no. 37, pp. 27 257–27 292, 2020.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[16] A. E. Ilesanmi, U. Chaumrattanakul, and S. S. Makhanov, "Methods for the segmentation and classification of breast ultrasound images: a review," *Journal of Ultrasound*, pp. 1–16, 2021.

[17] J. Zeng, L. Xie, and Z.-Q. Liu, "Type-2 fuzzy gaussian mixture models," *Pattern Recognition*, vol. 41, no. 12, pp. 3636–3643, 2008.

[18] C. Li, D. Lin, B. Feng, J. Lü, and F. Hao, "Cryptanalysis of a chaotic image encryption algorithm based on information entropy," *Ieee Access*, vol. 6, pp. 75 834–75 842, 2018.

[19] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Marti, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE journal of biomedical and health informatics*, vol. 22, no. 4, pp. 1218–1226, 2017.

[20] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in brief*, vol. 28, p. 104863, 2020.

[21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.