Machine Learning Augmented Breast Tumors Classification using Magnetic Resonance Imaging Histograms

Ahmed M. Sayed

Biomedical Engineering Department, Helwan University, Helwan, Cairo, Egypt
EECS Department, MSOE University, Milwaukee, WI, USA

Abstract—**At present, breast cancer survival rate significantly varies with the stage at which it was first detected. It is crucial to achieve early detection of malignant tumors to reduce their negative effects. Magnetic resonance imaging (MRI) is currently an important imaging modality in the detection of breast tumors. A need exists to develop computer aided methods to provide early diagnosis of malignancy. In this study, I present machine learning models utilizing new image histogram features using the pixels least significant bit. The models were first trained on an MRI breast dataset that included 227 images captured using the short TI inversion recovery (STIR) sequence and diagnosed as either benign or malignant. Three data classification methods were utilized to differentiate between the tumor's classes. The examined classification methods were the Discriminant Analysis, K-Nearest Neighborhood, and the Random Forest. Algorithms' testing was performed on a completely different dataset that included another 186 MRI STIR images showing breast tumors with verified biopsy diagnostics. A significant tumor classification efficiency was found, as judged by the pathological diagnosis. Classification's accuracy was calculated as 94.1% for the DA, 94.6% for the KNN and 80.6% for the RF algorithm. Receiver operating curves also showed significant classification performances. The proposed tumor classification techniques can be used as non-invasive and fast diagnostic tools for breast tumors, with the capability of significantly reducing false errors associated with common MRI imaging-based diagnosis.**

Keywords—*Tumor classification; histogram analysis; magnetic resonance imaging; breast cancer; machine learning*

I. INTRODUCTION

Breast cancer is the most common cancer type in women worldwide. It is the fifth cause of female deaths due to cancer [1]. Around 300,000 new female cases is estimated to occur each year in the United States alone [2]. The survival rate for breast cancer have generally improved over the past few years, as diagnosis at an early and localized stage is now possible, because of the progressive improvement in treatment strategies [3]. Early diagnosis of malignant tumors is crucial to avoid tumor metastasis and subsequently elevate the survival rate of diseased cases. If the tumor was not diagnosed early, it may spread beyond the original breast organ to other distant organs. Currently the routine method for diagnosing suspected breast tumors is imaging using Mammography, the main imaging modality for the breast organ that is then followed by pathological diagnosis through extraction of a biopsy sample from the tumor invasively. Mammography breast cancer

detection sensitivity is generally high [4], however, this sensitivity goes down to near 62% when imaging females with dense breasts [5]. Additionally, the costly biopsy procedure, the gold standard for diagnosis, is routinely performed under ultrasound guidance. However, about 75% of the performed biopsy procedures yield a benign diagnosis [6, 7], which is considered an unnecessary, costly, and time consuming and painful procedure to patients. In order to reduce the wasted biopsy procedures, other imaging modalities were proposed, such as magnetic resonance imaging (MRI) and ultrasound elastography [8-13].

Lately, MRI has become a useful and important modality to visualize and detect breast tumors in today's clinical practice [8, 14, 15]. This imaging modality is becoming increasingly in use to preoperatively evaluate DCIS tumors and define their extent [16, 17]. MRI has the advantages of not producing ionizing radiation, exhibiting high imaging contrast, good sensitivity rate, ability to show auxiliary nodes, and enjoying 3D imaging capabilities [18]. Additionally, Short inversion time Inversion Recovery (STIR) MRI scanning sequence provides a means for suppressing fat and inflammatory tissue from the normal tissue in the resultant images [19, 20]. If MRI is used in daily routine examinations, specific types of cancer would have been significantly diagnosed with higher sensitivity rates at an earlier stage [8], but cost remains a major impediment. Nevertheless, breast imaging using MRI exhibit relatively moderate specificity rates (down to 79%) that increase the erroneous false positive diagnostic percentages [15, 21-23].

One of the MRI imaging characterization methods is histogram analysis that is usually used to distinguish different anatomical and morphological regions, in addition to its more fundamental usage as an image enhancement tool [24, 25]. Some previous studies used histogram methods to illustrate the relation between the tumors physiological changes and their associated histogram parameters to achieve improved utilization of these histogram parameters as substitutive and representative markers describing heterogeneity of the tumor compositions [15, 26, 27]. In the past years, several studies have exploited histogram approaches in various imaging modalities [24, 27-32] with a growing emphasis on different MRI techniques and imaging sequences. Histogram processing methods showed its value for investigating various tumor parameter distributions, for example, in dynamic contrast-enhanced MRI (DCE MRI) it was possible to differentiate

between responder and non-responder groups in brain tumors radiotherapy [33], and in apparent diffusion coefficient (ADC) using diffusion MRI, it was also possible to detect specific types of cervical cancer [34] and endometrial cancer [35]. Despite that, information related to tumor's heterogeneities remain not fully scrutinized [15]. Along with the advances in high resolution MRI and its associated signal processing methods, histogram analysis of cancer tumors scanned using MRI will be used to a greater extent.

In this research, breast tumor's heterogeneity was investigated and described by the least significant byte histogram parameters calculated from STIR MRI imaging sequences for a number of clinically and pathologically verified patients diagnosis. The aim of this study is to differentiate between the two main breast tumors' classes; benign and malignant, with higher accuracy rates. Computer aided diagnosis was achieved using three classification algorithms to categorize the acquired data. Following that, the classification efficiency was calculated and compared with the outcomes of pathological tumor's diagnosis; consequently, the classification errors and receiver operating curves were calculated using two MRI data sets; one dataset for training the classifiers and the other dataset was utilized for testing. Throughout this study, it will be demonstrated that the proposed breast tumor classification technique has the potential as a noninvasive early diagnosis tool. This may lead to earlier and faster tumors characterization, and also may reduce the number of unnecessary biopsies performed pathologically to determine benignancy or malignancy; the applicable criteria that follow.

II. MATERIALS AND METHODS

In this research, the used training dataset was an online imaging dataset made available for scientific studies. It was published by the Cancer Imaging Archive (TCIA) [36] under the Breast-Diagnosis collection [37]. Table I lists the mass types included in this training dataset and their pathologic diagnosis, as published in the clinical, pathology, and radiologist reports [37]. A different MRI dataset was used for classification testing purposes. This testing dataset was previously acquired from different health care faculties, where tumors diagnosis was also verified with histopathology, as it was used in a previously published study of the research team [38]. The dataset included 186 tumor images, as listed in Table I, and their biopsy results were also available.

TABLE I. PATHOLOGIC DIAGNOSIS OF THE EXAMINED CASES

Case Diagnosis	Pathological Diagnosis (count)	
	Training Dataset	**Testing Dataset**
Benign	Fibroadenoma (27) Fibrocystic Change (22) Fibrosis (25) Stromal Hyperplasia (8)	Fibroadenoma (43) Fibrocystic Change (15) Cystic Lesion (14)
Malignant	Invasive Ductal Carcinoma (91) Ductal Carcinoma In Situ (19) Invasive lobular Carcinoma (35)	Invasive Ductal Carcinoma (114)
Total Numbers	**Training Dataset**	**Testing Dataset**
Benign	82	72
Malignant	145	114

A. Generating Histograms

The presented analyses in this study focus on obtaining tumor's histograms and identify the important classification features. Apparently, obtaining the whole image's histogram will degrade the classification overall accuracy by including imaging features that represent the surrounding non-tumorous tissue and healthy organs. Selection of the region of interest (ROI); i.e., delineating only the tumor, is a common practice in the routine radiology analysis. This is a particularly necessary step in this study to exclude non-tumorous features from the classification process.

Tumors' locations and pathological diagnosis were already determined in the training dataset's pathological reports. This information was used to manually select ROIs for all independent breast tumors in the dataset. Fig. 1(a) and Fig. 1(c) show selected ROIs for malignant and benign breast tumors, respectively. Following that the histogram for the selected image ROIs was generated for the least significant byte (LSB) only; the reason for that will be explained shortly. The corresponding histograms are presented in Fig. 1(b) and Fig. 1(d), in which the horizontal axes represent gray scale level variations, and the vertical axes represent the number of pixels for a specific gray scale level.

After generating all tumors' ROI imaging histograms, their classifying features were then computed. The classifying features were chosen to be ten histogram parameters. Those parameters were used to describe the shape and profile of a histogram. The ten histogram parameters were generally used in similar studies found in the literature [15, 26, 34, 38, 39] that aimed to differentiate tumors or identify various morphological regions based on imaging data. The used histogram features were: maximum, median, mean, mode, entropy, standard deviation, kurtosis, skewness, 75 percentile and the 25 percentile values. Entropy measures the degree of uniformity of a histogram. Kurtosis represents a measure of the histogram general shape. Skewness represents a histogram's data asymmetry about the mean value. Percentile values represent a value below a specified limit of the calculated histogram data.
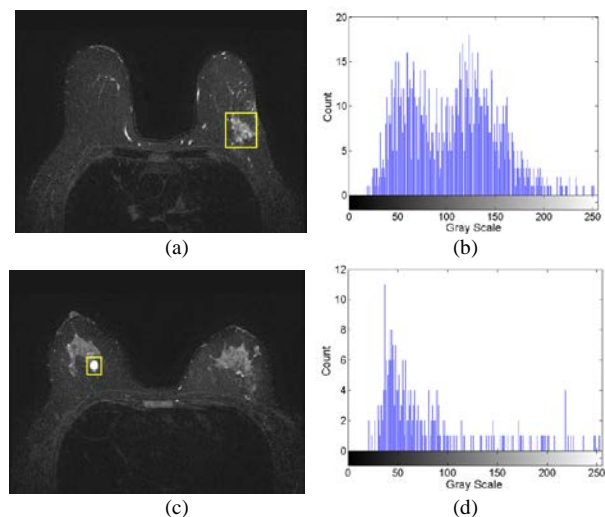

Fig. 1. MRI Imaging Examples of Breast Tumors' ROI and their Least Significant Byte (LSB) Associated Histograms: (a, b) Malignant Tumor; (c, d) Benign Tumor.

Statistically speaking, this study is an observational study, with no control over the classification features. Statistical testing was used to examine the classification significance of the chosen histogram parameters/features. First normality test had to be applied to select a suitable significance test. The Jarque–Bera normality test was utilized to examine the features normality. All features failed the normality test; therefore, the parametric t-test could not be used and the non-parametric Wilcoxon rank sum test was used instead. This test showed that some features had a significant classification power, provided that the alternative testing hypothesis of different features distribution medians for the two classes (benign and malignant) were different. Statistical significance level α of 0.05 was considered throughout this study.

For both training and testing data sets, only Short TI Inversion Recovery (STIR) MRI imaging were examined, which is an imaging procedure that aims at highlighting the breast mass's morphology and facilitates visualization of tumor's heterogeneity. Each pixel of the studied MRI images consists of two bytes; least significant bytes (LSB) and most significant bytes (MSB). Two types of histograms were generated: histograms based on the whole pixel size; LSB and MSB, and histograms based using LSB only. As shown in Table II, the number of significant histogram features using the full pixel size were only 6, compared to 8 significant features when using the LSB. Kurtosis and skewness were the additional significant features in the second case. For the full-length histograms, the mode was the most significant feature with P value of 0.0026, while for the LSB case, skewness was the most significant feature with 8.35E-05 P value.

Evidently, features based on the LSB histograms would provide more classification power between the two tumor types. The LSB's histogram information may have magnified the tumor's image heterogeneity and adherence pattern with the surrounding normal tissue. Fig. 1 shows benign and malignant examples along with their LSB histograms. The malignant LSB histogram has a greater content of low pixel values, in contrast to the benign LSB histogram that has larger content of high pixel values that has been truncated in the calculation of the LSB histograms. Therefore, this interesting and useful effect was encouraging to proceed the classification process using LSB histograms rather than full pixel length histograms, which according to my knowledge, has not been reported before.

B. Data Classifiers

Three classifiers were exploited to automatically categorize the examined images as either malignant or benign, according to the corresponding histogram features. The used classifiers were the discriminant analysis (DA), K-Nearest Neighbor (KNN), and Random Forest (RF) classifiers. The 227 data points were utilized to train and validate the three classifiers, judged by their consequent resubstitution error, and the leave-one-out analysis. The three classifiers were chosen for their popularity, implementation simplicity, and prior use in similar applications [40-42].

The DA classifier tries to find a combination of the classifying features that divides the two disease main classes; benign and malignant. The discriminant analysis as a parametric method, attempts to estimate a categorical or grouping dependent variable based on a number of continuous independent variables; i.e. predictor variables using a preselected discriminant function. The dependent variable in our application was the tumor diagnosis outcome, while the independent variables were the MRI imaging features. Previous studies show that this classification method has shown an acceptable classification performance, even with inappropriate features selections [41, 43].

KNN is a nonparametric classification method. A data point is classified according to the distance between it and its neighbors in the feature space, with the point being assigned to a class that is closest to its K nearest neighbors. The main parameter controlling the performance of such classifier is the number of neighbors, K. A common tradeoff in selecting the right value of the parameter K exists, where larger K values make classification outcomes less vulnerable to the effect of noise or data outliers but results in less distinct classification boundaries. On the other hand, lower values of K produce uneven and irregular classification boundaries [44]. Therefore, the classification analysis was repeated in the training phase for different values of the K parameter, and the resultant classification error was reported accordingly.

The third used classifier in this study was the RF which is also a nonparametric classification method. In this method, a group learning model is constructed using a large number of decision trees. Classification is performed according to the mode of the decision trees. Classification trees have the advantage of making a good fit to the training data [45, 46]. The main parameter in this method is the number of trees T used to build the classification model. Therefore, the classification analysis was repeated in the training phase for different values of the T parameter, and results were compared at each selected value. It is worth mentioning that it has been reported by Lin and Jeon [47], that a relationship exists between RF and KNN methods, where both belong to the weighted neighborhoods schemes.

TABLE II.   STATISTICAL TESTING OF HISTOGRAM CLASSIFICATION FEATURES USING TWO IMAGING PIXELS SIZES

P-values based on LSB+MSB										# Significant features
Entropy	Max	Median	Mean	STD	Mode	Kurtosis	skewness	prctile75	prctile25	
0.1077	0.0135	0.013	0.0034	0.59	0.0026	0.1077	0.4112	0.0125	0.0063	6
P-values based on LSB										
Entropy	Max	Median	Mean	STD	Mode	Kurtosis	skewness	prctile75	prctile25	
0.847	0.0468	1.34E-04	0.0071	0.327	0.0251	1.47E-04	8.35E-05	1.71E-04	0.0011	8

After training each classifier using the training MRI dataset, the associated resubstitution error was calculated. Additionally, leave one out analysis (LOO) was performed as a validation step. In leave one out analysis each classifier was trained on the whole data set except for one data point, and classification was then predicted for this data point. This process is repeated until all points were diagnosed based on the model generated using the other trained data points. Both resubstitution and LOO analysis data were used to select the classifiers' parameters that generate the lowest false negative error with a high level of accuracy. Achieving low false negative errors is very crucial, as misclassifications of positive malignant tumors are so severe, as they lose the early detection and treatment privileges.

Following training, testing of the classification model follows using another independent dataset that consists of 186 testing images. Classification accuracy and receiver operating characteristic (ROC) curves were plotted for each classification model and analyzed accordingly. A flowchart summarizing the tumor's classification process is shown in Fig. 2.
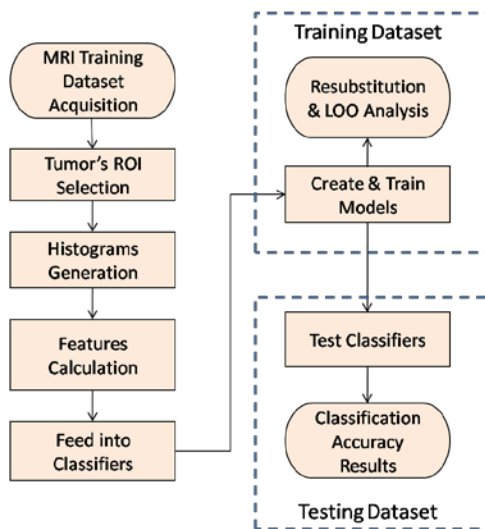

Fig. 2. Flow Chart of the Breast Tumor's Classification Process.

III. RESULTS

Application of the described approach resulted in labeled histogram data points. The points were used for training and evaluation of the three classifiers. The classifiers where then tested, and classification was found to be significant, with different efficiencies according to the used classifier, as will be shown in the following subsections.

A. Classifiers Training Evaluation

1) DA classifier: Five DA discriminant functions were evaluated and the results of resubstitution error and leave one analysis are shown in Table III. The table lists the True Negative (TN), True Positive (TP), False Negative (FN) and False Positive (FP) values for each discriminant function. Sensitivity, specificity, and accuracy values are used to determine the classifier's performance. The sensitivity value provides a very important indication, as higher sensitivity values point out the classifier's ability to identify malignant

tumors as malignant. It is ideal to have a false negative value of zero (sensitivity of 1). The Positive and Negative Likelihood Ratios (PLR and NLR respectively) are also reported in Table III. It is desirable to have a classifier with higher positive likelihood ratio > 2 and NLR < 0.5 for a better discriminatory classification [48].

From the table, it is evident that the Mahalanobis discriminant function provides satisfactory classification results with high sensitivity and low NLR reflecting the low possibility of producing false negative errors. Although the other functions provided a lower FP error, yet the severity of FN errors is much higher than FP errors.

2) KNN classifier: The KNN classifier was trained using the standard Euclidean distance metric. This classifier requires determination of the number of neighboring points; K parameter, to be included while creating the model decision boundary. A good value of K would provide a suitable compromise between the classifier's sensitivity to noise at low values of K form one hand, and the reduced classification accuracy at high values of K on the other hand. Therefore, different K values were examined, as shown in Fig. 3.

In Fig. 3, both the resubstitution and LOO analysis were performed for K values ranging from 1 to 100, to explore any potential useful values of K. Only the classification accuracy is being graphed to show the overall performance of the classifiers, but specificity and sensitivity are also reported and analyzed in Table IV. The resubstitution accuracy profile in Fig. 3(a) shows a rapid accuracy decline as K increases with a small peak that appears at K=15, then the curve declines again until it settles at an accuracy of 79% approximately. The LOO accuracy profile in Fig. 3(b) shows two peaks at K=5 and K=15, then the curve settles at about 79 % accuracy level. From both curves, it is rational to select the value of 15 rather than 5, to make the model more generalized and less sensitive to data outliers and data irregularities. For this value of K = 15, an overall training classification sensitivity of about 85.5% was achieved with a low NLR; which is an indication of a low possibility of producing false negative diagnosis.

3) Random Forest Classifier: In this classifier, the main parameter is the number of trees (T) composing the forest. The resubstitution and LOO analyses were performed for T values from 5 to 300, to explore potentially useful T values, as illustrated in Fig. 4. The resubstitution accuracy profile shown in Fig. 4(a), exhibits a steady accuracy of 100% for T values larger than 40 trees. The LOO profile in Fig. 4(b) shows accuracy fluctuations around the 79% accuracy level for almost all values of T. It can be noticed from the figure that the resubstitution error is infinitesimal and only occurs for small tree numbers. Nevertheless, the LOO analysis reveals the actual performance of the RF algorithm when data points not included in the training are tested, which indicates a model overfitting effect. Based on these results, a classifier with a decision trees number T of 100 was chosen for further testing, and the corresponding training evaluation calculations are listed in the second section of Table IV.

TABLE III.   TRAINING AND EVALUATION OF THE DISCRIMINANT ANALYSIS CLASSIFIER USING DIFFERENT DISCRIMINANT FUNCTIONS

A- Resubstitution Analysis										
Discriminant function	**TN**	**TP**	**FN**	**FP**	**Count**	**Sensitivity**	**Specificity**	**Accuracy**	**PLR**	**NLR**
Mahalanobis	52	137	8	30	227	0.945	0.634	0.832	2.583	0.087
Linear	56	134	11	26	227	0.924	0.683	0.837	2.914	0.111
Diagonal Linear	68	106	39	14	227	0.731	0.829	0.767	4.282	0.324
Quadratic	73	107	38	9	227	0.738	0.89	0.793	6.723	0.294
Diagonal Quadratic	72	81	64	10	227	0.559	0.878	0.674	4.58	0.503
B- Leave One Out Analysis										
Discriminant function	**TN**	**TP**	**FN**	**FP**	**Count**	**Sensitivity**	**Specificity**	**Accuracy**	**PLR**	**NLR**
Mahalanobis	52	136	9	30	227	0.938	0.634	0.828	2.564	0.098
Linear	54	133	12	28	227	0.917	0.659	0.824	2.686	0.126
Diagonal Linear	68	104	41	14	227	0.717	0.829	0.758	4.20	0.340
Quadratic	69	107	38	13	227	0.738	0.841	0.775	4.655	0.311
Diagonal Quadratic	72	81	64	10	227	0.559	0.878	0.674	4.580	0.503

TABLE IV.   EVALUATION OF THE KNN AND RF CLASSIFIERS' TRAINING PERFORMANCE

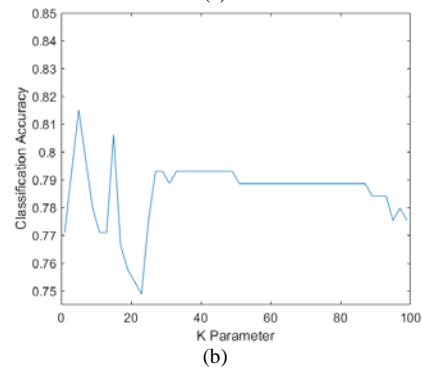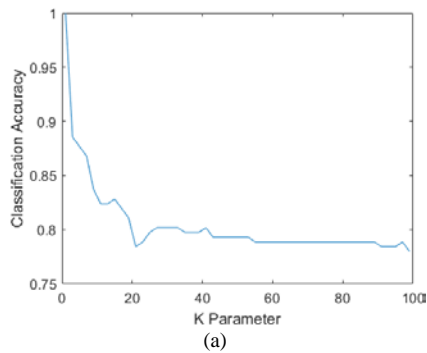|	**TN**	**TP**	**FN**	**FP**	**Count**	**Sensitivity**	**Specificity**	**Accuracy**	**PLR**	**NLR**
KNN Resubstitution Analysis										
K = 15	62	126	19	20	227	0.869	0.756	0.828	3.563	0.173
KNN Leave One Out Analysis										
K = 15	61	122	23	21	227	0.841	0.744	0.806	3.285	0.213
RF Resubstitution Analysis										
T = 100	82	145	0	0	227	1	1	1	Inf	0
RF Leave One Out Analysis										
T = 100	58	127	18	24	227	0.876	0.707	0.815	2.993	0.176

(a)

(b)

Fig. 3.   KNN Classifier: (a) Training Accuracy for different Values of K Parameter using Resubstitution Analysis. (b) Training Accuracy for different Values of K Parameter using LOO Analysis.
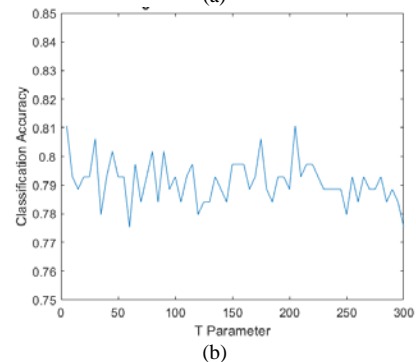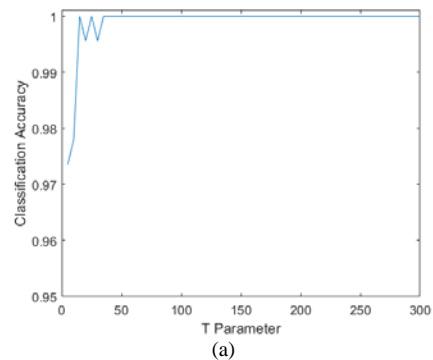

(a)

(b)

Fig. 4.   Random Forest Training Evaluation: (a) Training Accuracy for different Values of T Parameter using Resubstitution Analysis. (b) Training Accuracy for different Values of T Parameter using LOO Analysis.

B. Classifiers Testing Evaluation

One contribution of this study is exploitation of trained classification models to classify an entirely different testing dataset that was not included in the models training. No further tuning or post processing was applied on the models, and testing was performed directly resulting in the following tumor diagnosis outcomes.

1) Testing DA classifie: A summary of DA testing outcomes is presented in Table V. The DA classified the tumors with a sensitivity of 99.0%, specificity of 87.8%, and accuracy of 94.1% with a very low NLR ratio.

2) Testing KNN classifier: The KNN classifier testing results are summarized in Table V for the same K value used in the training process. The testing accuracy of the algorithm was calculated for all values of K; from 1 to 100, as shown in Fig. 5(a). Two accuracy peaks appear at 15 and 23 and giving the same exact classification accuracy.

It is apparent that the KNN has a very close performance to the DA classifier, yet the sensitivity measure was found to be better using the KNN, as the calculated NLR was almost zero with no FN errors.
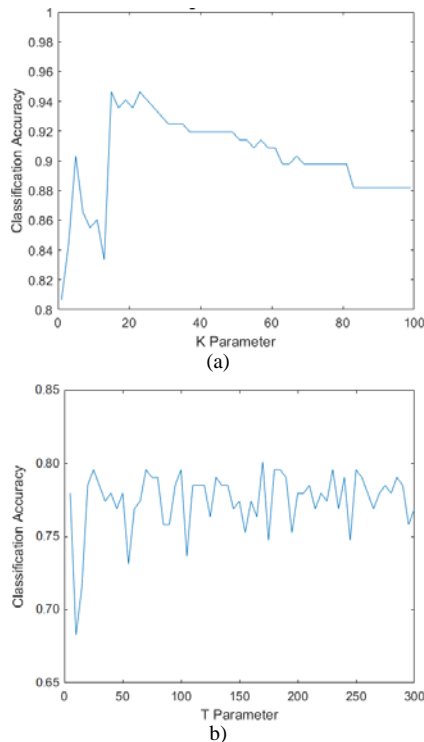

(a)

b)

Fig. 5. (a) Accuracy of Classification Testing for all Values of K Parameter for the KNN Classifier. (b) Accuracy of Classification for all Values of T Parameter for the RF Classifier.

3) Testing random forest classifier: The RF algorithm tumor categorization outcomes were summarized in Table V for the same T value used in the training process. Once more, the testing accuracy of the algorithm was calculated for all values of T; from 5 to 300, as shown in Fig. 5(b). The accuracy profile does not show specific peaks or range of T values with higher accuracy levels. A fluctuating, yet steady

performance is noticed for T values of more than 25, around the 77.5 % accuracy level. Yet, a sensitivity level of only 76% was noted, as the classifier failed to correctly categorize 25 malignant tumors. The RF testing profile is very similar to the RF LOO training profile. This observation is interesting, as it shows that the RF classifier performance can be accurately predicted based on the LOO training curve profile.

To compare the three classifiers' performance at the selected classification parameters, a combined ROC curve was plotted in Fig. 6. It is clear that the DA and KNN classifiers are superior to the RF classifier, in terms of sensitivity and specificity. The areas under the curves (AUC) were 0.956 and 0.953 for the DA and KNN, respectively, indicating a very good and significant classification performance. The AUC value for the RF model was calculated to be only 0.845 reflecting a moderate classification performance.
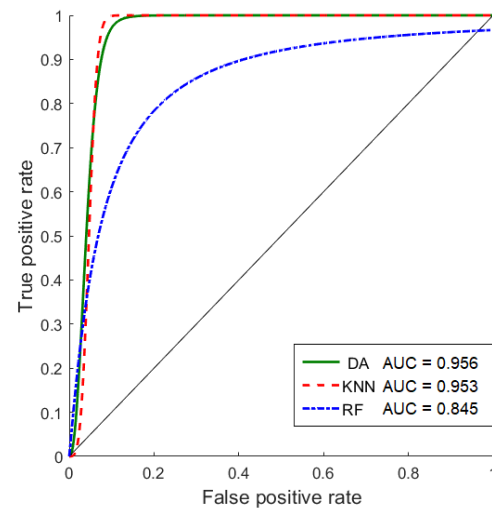

Fig. 6. ROC Curves showing Performance of the Three Breast Tumor Classifiers.

IV. DISCUSSION

In this article, breast tumor classification methods based on MRI LSB histogram parameters were demonstrated. The selected histogram features were used to train three different machine learning methods as tumor classifiers, and their corresponding performances were compared. Previous studies in the literature used histogram parameters to characterize breast cancer and its response to therapy [27, 41, 49-51]. Their main goal was finding the statistical significance of histogram features to categorize the examined breast tumors, yet the usage of machine learning techniques for such a purpose was limited in the literature. The study reported by Vidić and coworkers presented a support victor machine algorithm to evaluate breast tumors classifications [50]. Although the authors showed an overall classification accuracy of up to 0.96, they reported only accuracy values without considering specificity nor sensitivity ratios, therefore, no information were provided about false positive or negative classification errors. Lee and associates reported five machine learning algorithms to predict prognostic biomarkers of breast cancer [51]. The authors reported an AUC value of 0.8 using a random forest

model, which is quite close to the RF AUC value we report herein, although higher AUC values were achieved using other algorithms, as demonstrated in the paper in hand. Additionally, usage of histogram features calculated using only the LSB imaging pixels has never been reported, according to my best knowledge, which have shown more significant differences between benign and malignant tumors. This effect can be explained as follows: important information pertaining to the tumor's heterogeneity, adherence to the surrounding normal tissue, and response to magnetic excitations may have been emphasized in the image's low gray scale values, due to the scanning nature of STIR MRI sequence that suppresses the fatty normal tissues, i.e., it nulls the signal from fat.

Herewith, two different breasts MRI dataset were exploited: one for training the classifiers and the other for independently testing them. One main contribution in this study is the application of a trained machine learning method on a totally different dataset from a different source. This strengthens the hypothesis that the described methods are generalized classification methods that could be efficiently used to classify any given STIR MRI breast tumor images. Furthermore, the described methods can be easily repeated and validated by other research groups on their own datasets.

Statistical analysis of the selected histogram features showed skewness as the most significantly different parameter between benign and malignant tumors, with a P value of 8.35E-05. In general, skewness represents the shape and asymmetry of a given histogram. Based on the training MRI dataset, the average skewness value for the benign images was 5.866, while for the malignant tumors the value was 2.001; approximately 3 folds. This indicates that benign histograms were quite asymmetric around the mean and more right-skewed towards the higher image pixel values as compared to malignant histograms.

As mentioned earlier the selection of the three classifiers was based on their inherent implementation simplicity and prior use in similar applications [40-42]. The aid of machine learning algorithms to improve diagnostic accuracy is of significant interest and utility, as human interpretation of MRI breast data is neither 100% sensitive nor 100% specific. Even though the DA algorithm assumes a multivariate normal distribution between the used features, which is not the case here, yet it was successful in categorizing the testing tumors data. It has been reported that violations of the normality assumption can be permitted in certain cases, and the algorithm outcomes can still be considered reliable, given that the non-normality violations are not caused by data outliers [52]. The other two algorithms; KNN and RF, do not assume normality for the input data points.

The KNN classifier was also very successful and specific in classifying the examined breast tumors. Selection of a certain K parameter was a compromise between good classification performance and robustness against noise and data outliers. Optimization techniques can be used to find the optimum KNN

parameters by minimizing the cross-validation loss error. Readily available hyperparameter optimization methods (MATLAB, The MathWorks Inc., Natick, Massachusetts, US) was attempted using the training data and tested as well using the testing dataset. The optimization process recommended using the Spearman distance function (instead of the standard Euclidean method used throughout this study) and a K value of 38. In this case, the classification error occurred only for 1 FP and 1 FN data points out of the 186 testing points (sensitivity of 99% and specificity of 98.8%), which is a remarkable classification performance. Yet, the goal of this paper is to demonstrate the feasibility of using different machine learning techniques to categorize breast tumors and compare between their performances in a pilot study. The task of finding an optimal classifier for that purpose would need more investigation and testing using larger datasets, which is considered future work.

The DA and KNN classification performance metrics showed significantly better outcomes in categorizing testing data over the training data. Training outcomes showed an accuracy of approximately 83%, while testing data showed about 94% accuracy; more than 10% of accuracy increase. This effect can be explained by the fact that the training dataset was larger and more diverse than the testing dataset. As demonstrated in Table I, the training dataset included 227 independent images with 7 different tumor types, while the testing dataset included 186 images showing 4 types of common breast tumors. The training dataset included more tumor types, however, some of them were uncommon and rare tumor types, such as stromal hyperplasia and Ductal Carcinoma in Situ [53, 54]. The trained algorithms used a more generalized data than the testing data, which was the main cause behind the accuracy differences between the two situations. The lack of a more generalized testing dataset is considered another limitation of this study. Despite the encouraging results that have been shown in this study, testing the developed classification methods on a larger and more diverse MRI dataset is an ongoing work.

The RF algorithm had though a moderate classification performance with an utmost accuracy ratio of 80%. It has also the disadvantage of being expensive regarding the computational time. RF LOO analysis was completed in approximately 70 minutes to run using MATALB (The MathWorks Inc., Natick, Massachusetts, US) on a modern computer (Windows 10, 10th Generation Core I5, 2.11 GHz, 16GB RAM), while DA LOO and KNN LOO were completed in 2 and 6 seconds, respectively. Nevertheless, this method's testing performance may be directly predicted from the training LOO data analysis, as the RF algorithm behaved in a very similar and consistent way in both cases, with slight accuracy degradation under the testing mode. This was clear from the demonstrations in Fig. 4(b) and 5(b). Another interesting observation is that the three unoptimized algorithms had very close specificity ratios though; 87.8 for DA, 87.8 for KNN and 85.4 for RF, respectively.

TABLE V. TESTING RESULTS FOR THE BREAST TUMOR'S CLASSIFICATION USING THE THREE ALGORITHMS

Algorithm Parameter	TN	TP	FN	FP	Count	Sensitivity	Specificity	Accuracy	PLR	NLR
DA Algorithm										
Mahalanobis	72	103	1	10	186	0.990	0.878	0.941	8.121	0.011
KNN Algorithm										
K = 15	72	104	0	10	186	1	0.878	0.946	8.2	0
RF Algorithm										
T = 100	70	79	25	12	186	0.76	0.854	0.801	5.191	0.282

V. CONCLUSION AND FUTURE WORK

New breast tumors' classification methods based on MRI imaging were presented. The methods showed the potential to provide more accurate tumor's diagnosis non-invasively and timely efficient. This method may provide an alternative approach to the unnecessary biopsy procedures routinely performed to verify a breast tumor preliminary diagnosis. From the demonstrated results, it has been shown that the discriminant analysis and K nearest neighborhood methods can provide good tumor categorization performance with a significant sensitivity and accuracy levels. The random forest method proved to provide a moderate degree of classification accuracy, however, it showed consistent outcomes in both the training and testing data. The reported least significant byte histogram-based algorithms may be applied on other tumor types, but this requires further investigation to prove being valid. Future research projects include applying the described methods on larger and diverse STIR MRI imaging datasets to find an optimized tumors classification scheme.

REFERENCES

[1] J. Ferlay et al., "Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012," Int J Cancer, vol. 136, no. 5, pp. E359-86, Mar 2015.

[2] "American Cancer Society. Breast Cancer Facts & Figures 2015-2016," 2015.

[3] "Diet, Nutrition, Physical Activity and breast cancer," 2017.

[4] N. Houssami, S. J. Lord, and S. Ciatto, "Breast cancer screening: emerging role of new imaging techniques as adjuncts to mammography," Medical Journal of Australia, vol. 190, no. 9, pp. 493-498, May 2009.

[5] P. A. Carney et al., "Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography," Ann Intern Med, vol. 138, no. 3, pp. 168-175, Feb 2003.

[6] S. P. Poplack, P. A. Carney, J. E. Weiss, L. Titus-Ernstoff, M. E. Goodrich, and A. N. A. Tosteson, "Screening mammography: Costs and use of screening-related services," Radiology, vol. 234, no. 1, pp. 79-85, Jan 2005.

[7] S. Goenezen et al., "Linear and nonlinear elastic modulus imaging: an application to breast cancer diagnosis," IEEE Trans Med Imaging, vol. 31, no. 8, pp. 1628-37, Aug 2012.

[8] C. K. Kuhl et al., "Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer," J Clin Oncol, vol. 23, no. 33, pp. 8469-76, Nov 20 2005.

[9] A. Sayed, G. Layne, J. Abraham, and O. M. Mukdadi, "Nonlinear Characterization of Breast Cancer using Multi-Compression 3D Ultrasound Elastography In Vivo," Ultrasonics vol. 53, no. 5, pp. 979-991, 2013.

[10] A. M. Sayed, G. Layne, J. Abraham, and O. M. Mukdadi, "3-D visualization and non-linear tissue classification of breast tumors using ultrasound elastography in vivo," Ultrasound Med Biol, vol. 40, no. 7, pp. 1490-502, Jul 2014.

[11] E. Warner et al., "Comparison of breast magnetic resonance imaging, mammography, and ultrasound for surveillance of women at high risk for hereditary breast cancer," J Clin Oncol, vol. 19, no. 15, pp. 3524-31, Aug 2001.

[12] C. D. Lehman et al., "Screening women at high risk for breast cancer with mammography and magnetic resonance imaging," Cancer, vol. 103, no. 9, pp. 1898-905, May 2005.

[13] A. M. Sayed, M. A. Naser, A. A. Wahba, and M. A. A. Eldosoky, "Breast Tumor Diagnosis Using Finite-Element Modeling Based on Clinical in vivo Elastographic Data,", J Ultrasound Med, vol. 39, no. 12, pp. 2351-2363, Dec 2020.

[14] C. H. Lee et al., "Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer," J Am Coll Radiol, vol. 7, no. 1, pp. 18-27, Jan 2010.

[15] N. Just, "Improving tumour heterogeneity MRI assessment with histograms," (in eng), Br J Cancer, vol. 111, no. 12, pp. 2205-13, Dec 2014.

[16] C. J. Allegra et al., "National Institutes of Health State-of-the-Science Conference Statement: Diagnosis and Management of Ductal Carcinoma In Situ September 22-24, 2009," J Natl Cancer I, vol. 102, no. 3, pp. 161-169, Feb 2010.

[17] M. Pilewskie et al., "Effect of MRI on the Management of Ductal Carcinoma In Situ (DCIS) of the Breast," Ann Surg Oncol, vol. 19, pp. S11-S11, Feb 2012.

[18] J. E. Joy, E. E. Penhoet, D. B. Petitti, National Cancer Policy Board (U.S.). Committee on New Approaches to Early Detection and Diagnosis of Breast Cancer., National Research Council (U.S.). Policy and Global Affairs., and National Research Council (U.S.). Board on Science Technology and Economic Policy., Saving women's lives : strategies for improving breast cancer detection and diagnosis. Washington, D.C.: National Academies Press, 2005.

[19] R. Golfieri, H. Baddeley, J. S. Pringle, and R. Souhami, "The role of the STIR sequence in magnetic resonance imaging examination of bone tumours," Br J Radiol, vol. 63, no. 748, pp. 251-6, Apr 1990.

[20] E. M. Delfaut, J. Beltran, G. Johnson, J. Rousseau, X. Marchandise, and A. Cotten, "Fat suppression in MR imaging: techniques and pitfalls," Radiographics, vol. 19, no. 2, pp. 373-82, Mar-Apr 1999.

[21] E. Othman et al., "Comparison of false positive rates for screening breast magnetic resonance imaging (MRI) in high risk women performed on stacked versus alternating schedules," Springerplus, vol. 4, p. 77, 2015.

[22] C. J. Allegra et al., "National Institutes of Health State-of-the-Science Conference statement: Diagnosis and Management of Ductal Carcinoma In Situ September 22-24, 2009," J Natl Cancer Inst, vol. 102, no. 3, pp. 161-9, Feb 2010.

[23] W. Chen, M. L. Giger, U. Bick, and G. M. Newstead, "Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI," Med Phys, vol. 33, no. 8, pp. 2878-87, Aug 2006.

[24] J. S. H.D. Cheng, Wen Ju, Yanhui Guo, Ling Zhang, "Automated breast cancer detection and classification using ultrasound images: A survey," Pattern Recognition, vol. 43, pp. 299-317, 2010.

[25] H. Satake, S. Ishigaki, R. Ito, and S. Naganawa, "Radiomics in breast MRI: current progress toward clinical application in the era of artificial intelligence, Radiol Med, 2021.

[26] J. S. Carter et al., "Quantitative multiparametric MRI of ovarian cancer," (in eng), J Magn Reson Imaging, vol. 38, no. 6, pp. 1501-9, Dec 2013.

[27] Y. C. Chang, C. S. Huang, Y. J. Liu, J. H. Chen, Y. S. Lu, and W. Y. Tseng, "Angiogenic response of locally advanced breast cancer to neoadjuvant chemotherapy evaluated with parametric histogram from dynamic contrast-enhanced MRI," Phys Med Biol, vol. 49, no. 16, pp. 3593-602, Aug 21 2004.

[28] I. Christoyianni, A. Koutras, E. Dermatas, and G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms," Comput Med Imaging Graph, vol. 26, no. 5, pp. 309-19, Sep-Oct 2002.

[29] K. Holli et al., "Characterization of breast cancer types by texture analysis of magnetic resonance images," Acad Radiol, vol. 17, no. 2, pp. 135-41, Feb 2010.

[30] D. R. Chen, R. F. Chang, W. J. Kuo, M. C. Chen, and Y. L. Huang, "Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks," Ultrasound Med Biol, vol. 28, no. 10, pp. 1301-10, Oct 2002.

[31] D. Checkley, J. J. Tessier, J. Kendrew, J. C. Waterton, and S. R. Wedge, "Use of dynamic contrast-enhanced MRI to evaluate acute treatment with ZD6474, a VEGF signalling inhibitor, in PC-3 prostate tumours," Br J Cancer, vol. 89, no. 10, pp. 1889-95, Nov 17 2003.

[32] G. Y. Cho et al., "Evaluation of breast cancer using intravoxel incoherent motion (IVIM) histogram analysis: comparison with malignant status, histological subtype, and molecular prognostic factors," Eur Radiol, vol. 26, no. 8, pp. 2547-58, Aug 2016.

[33] S. L. Peng et al., "Analysis of parametric histogram from dynamic contrast-enhanced MRI: application in evaluating brain tumor response to radiotherapy," NMR Biomed, vol. 26, no. 4, pp. 443-50, Apr 2013.

[34] K. Downey et al., "Relationship between imaging biomarkers of stage I cervical cancer and poor-prognosis histologic features: quantitative histogram analysis of diffusion-weighted MR images," AJR Am J Roentgenol, vol. 200, no. 2, pp. 314-20, Feb 2013.

[35] S. Ytre-Hauge et al., "Preoperative tumor texture analysis on MRI predicts high-risk disease and reduced survival in endometrial cancer," J Magn Reson Imaging, vol. 48, no. 6, pp. 1637-1647, 2018.

[36] V. B. Clark K, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," Journal of Digital Imaging, vol. 26, no. 6, pp. 1045-1057, 2013.

[37] B. N. Bloch, A. Jain, and C. C. Jaffe. Data From BREAST-DIAGNOSIS, The Cancer Imaging Archive, 2015a.

[38] Ahmed M Sayed, Eman Zaghloul, and T. M. Nassef, "Automatic Classification of Breast Tumors Using Features Extracted from Magnetic Resonance Images," Procedia Computer Science, vol. 95, pp. 392 – 398, 2016.

[39] H. Chandarana et al., "Histogram analysis of whole-lesion enhancement in differentiating clear cell from papillary subtype of renal cell cancer," Radiology, vol. 265, no. 3, pp. 790-8, Dec 2012.

[40] E. I. Zacharaki et al., "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," Magn Reson Med, vol. 62, no. 6, pp. 1609-18, Dec 2009.

[41] D. J. Tozer, G. R. Davies, D. R. Altmann, D. H. Miller, and P. S. Tofts, "Principal component and linear discriminant analysis of T1 histograms of white and grey matter in multiple sclerosis," Magn Reson Imaging, vol. 24, no. 6, pp. 793-800, Jul 2006.

[42] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Comput Struct Biotechnol J, vol. 13, pp. 8-17, 2015.

[43] R. Wolz et al., "Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease," PLoS One, vol. 6, no. 10, p. e25446, 2011.

[44] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," IEEE Trans Neural Netw, vol. 12, no. 2, pp. 181-201, 2001.

[45] B. A. Goldstein, A. E. Hubbard, A. Cutler, and L. F. Barcellos, "An application of Random Forests to a genome-wide association dataset: methodological considerations & new findings," BMC Genet, vol. 11, p. 49, Jun 2010.

[46] L. Breiman, "Random forests," Mach Learn, vol. 45, pp. 5-32, 2001.

[47] Y. Lin and Y. Jeon, "Random forests and adaptive nearest neighbors," Journal of the American Statistical Association, vol. 101, no. 474, 2006.

[48] S. McGee, "Simplifying likelihood ratios," J Gen Intern Med, vol. 17, no. 8, pp. 646-9, Aug 2002.

[49] R. Johansen et al., "Predicting survival and early clinical response to primary chemotherapy for patients with locally advanced breast cancer using DCE-MRI," J Magn Reson Imaging, vol. 29, no. 6, pp. 1300-7, Jun 2009.

[50] I. Vidić et al., "Support vector machine for breast cancer classification using diffusion-weighted MRI histogram features: Preliminary study," J Magn Reson Imaging, vol. 47, no. 5, pp. 1205-1216, 2018.

[51] J. Y. Lee et al., "Radiomic machine learning for predicting prognostic biomarkers and molecular subtypes of breast cancer using tumor heterogeneity and angiogenesis properties on MRI," Eur Radiol, vol. 32, no. 1, pp. 650-660, 2022.

[52] B. G. Tabachnick and L. S. Fidell, Using multivariate statistics, 6th ed. Boston: Pearson Education, 2013.

[53] S. D. Raj et al., "Pseudoangiomatous Stromal Hyperplasia of the Breast: Multimodality Review With Pathologic Correlation," Curr Probl Diagn Radiol, vol. 46, no. 2, pp. 130-135, 2017 Mar - Apr 2017.

[54] S. S. Badve and Y. Gökmen-Polar, "Ductal carcinoma in situ of breast: update 2019," Pathology, vol. 51, no. 6, pp. 563-569, Oct 2019.