



Accurate Breast Tumor Identification Using Computational Ultrasound Image Features

Yongqing Li¹ and Wei Zhao^{1,2}(✉)

¹ School of Physics, Beihang University, Beijing, China
zhaow85@buaa.edu.cn

² The Beihang Hangzhou Innovation Institute, Yuhang, Xixi Octagon City, Hangzhou, China

Abstract. Breast cancer ranks the first noncutaneous malignancy incidence and mortality in women worldwide, and seriously endangers the health and life of women. Ultrasound plays a key role and yet provides an economical solution for breast cancer screening. While valuable, ultrasound is still suffered from limited specificity, and its accuracy is highly related to the clinicians, resulting in inconsistent diagnosis. To address the challenge of limited specificity and inconsistent diagnosis, in this retrospective study, we first develop a learning model based on the computational ultrasound image features and identified a set of clinically relevant features. Then, the abstract spatial interaction patterns of the ultrasound images together with the extracted features were employed for breast malignancy diagnosis. We evaluate the proposed algorithm on the Breast Ultrasound Images Dataset (BUSI). The proposed algorithm achieved a diagnostic accuracy of 89.32% and a significant area under curve (AUC) of 0.9473 with the repeated cross-validation scheme. In conclusion, our algorithm shows superior performance over the existing classical methods and can be potentially applied to breast cancer screening.

Keywords: Computational features · Ultrasound · Breast cancer

1 Introduction

Breast cancer is the most commonly diagnosed cancer and causes the most deaths for women diagnosed with cancers. [6] Early diagnosis plays an important role in both treatment and prognosis for breast cancer. It has been extensively reported that patients diagnosed with smaller primary breast tumors had a significantly higher disease-free survival and overall survival, compared to patients with locally advanced breast tumors. Early detection and diagnosis of breast cancer are therefore of interest. Various imaging modalities have been applied to breast cancer diagnosis. Among these, ultrasound (US) imaging which employs sound waves to generate images of the internal morphology of the breast is the

most widely used method due to its safety and painlessness. The US is able to help diagnose breast lumps and other abnormalities in a noninvasive way.

Despite its usefulness and wide applicability, breast US has suffered from limited specificity and interobserver variability, both of which contribute to a high rate of false-positive and false-negative. The misdiagnoses cause either a number of unnecessary biopsies and surgeries, or missed cases. To address the challenge of limited specificity and interobserver variability, there has been a growing interest in the application of machine learning technology for automatic US breast tumor identification [4].

Different from conventional US diagnosis, the machine learning approaches make decisions based on extracted computational features. The features extraction procedure can be performed using either deep neural networks [2] or spatial and texture computational tools. While the deep neural network-based features are usually illusive and lack interpretability, the spatial and texture computational tools extract features that are directly related to tumor size and shape, image intensity histogram, and relationships between image voxels from radiologic images. The mathematical definitions of these features are explicit and easy to reproduce. Some of these features, such as tumor texture, have been demonstrated to be useful for differentiating malignant from benign tumors in breast cancer. In this study, we aimed to develop a learning model based on the computational ultrasound image features and applied the model to breast tumor identification. Clinically relevant features were used to differentiate breast tumor malignancy.

2 Method

Radiomics researches have a rather clear pipeline [3] which we adopted. First, we prepared the data, where the segmentation of region of interest (ROI) had been already available. Next, we extract features from ROIs with PyRadiomics package. Then, we selected and eliminated features and prepared them for modeling. At last, we built our model and evaluated the model by common metrics. The adopted pipeline is shown in Fig. 1.

2.1 Data Preparation

The BUSI dataset [1] was collected from 600 female patients and divided into three categories: benign, malignant, and normal. Both ultrasound images and segmentation masks are stored as 8-bit pngs. A sample of a malignant ultrasound image, a benign ultrasound image, and their corresponding masks are shown in Fig. 2.

Since the radiomics extract information from the region of interest (ROI) instead of the entire image, an ultrasound image with more than one tumor will result in the situation that the number of tumors ROIs is greater than that of the ultrasound image. Through the pairing of the ultrasound images and the masks, 454 benign tumor ROIs and 211 malignant tumor ROIs were finally obtained.

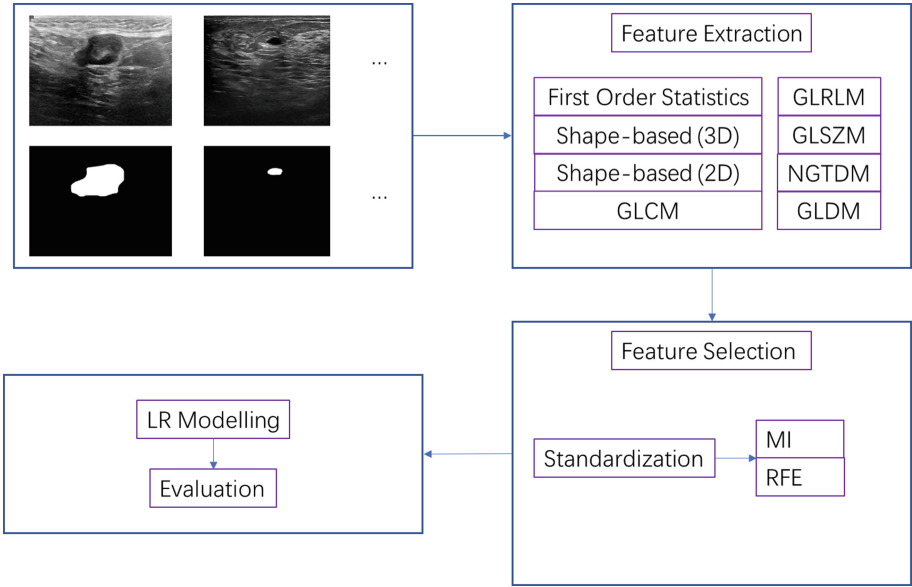


Fig. 1. Adopted pipeline of the research.

2.2 Feature Extraction

PyRadiomics [7] is an open-source Python library for radiomics feature extraction. With PyRadiomics, we extracted 1318 image-related features, which consist of eight classes:

- First Order Statistics
- Shape-based (2D)
- Shape-based (3D)
- Gray Level Cooccurrence Matrix (GLCM)
- Gray Level Run Length Matrix (GLRLM)
- Gray Level Size Zone Matrix (GLSZM)
- Neighbouring Gray Tone Difference Matrix (NGTDM)
- Gray Level Dependence Matrix (GLDM).

2.3 Feature Selection

Features with too high dimension hinder the implementation of classification algorithms, so feature selection is required. After the following steps, the number of features is controlled in an appropriate range.

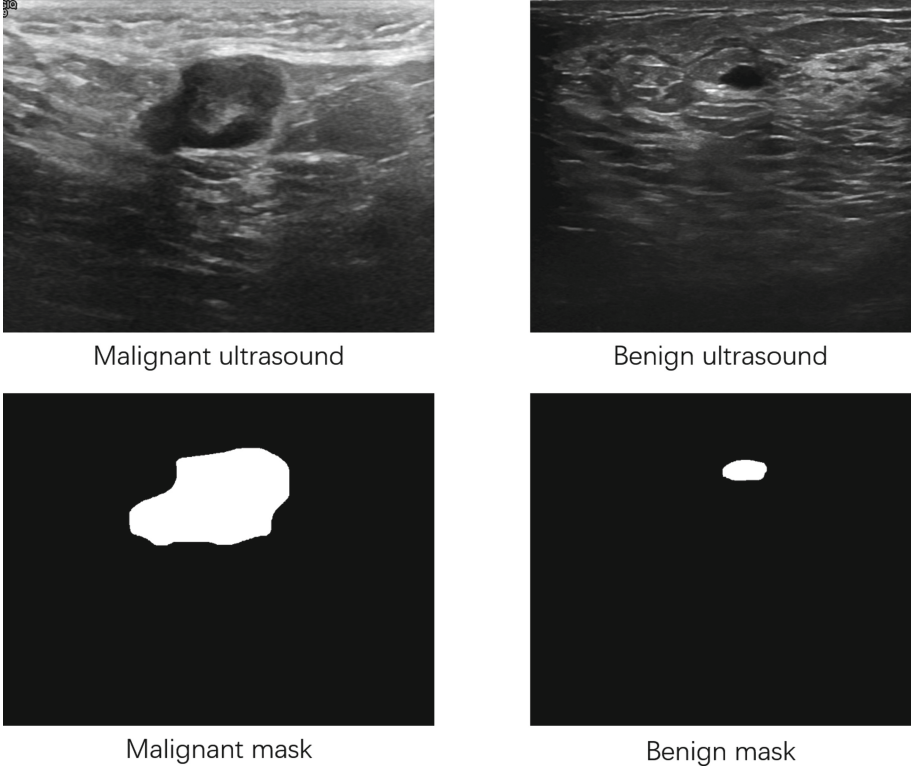


Fig. 2. Sample of a malignant ultrasound image, a benign ultrasound image, and their corresponding masks.

Data Standardization. The standardization process unifies the dimensions of the features and prevents the effect of the different magnitude order during the selection and modeling process.

We standardized the data by the formula

$$\hat{x} = \frac{x - \mu}{\sigma}, \quad (1)$$

where x represents the original data and \hat{x} represents the standardized data. μ represents the mean of the data, and σ represents the standard deviation of the data.

Mutual Information Filtering The mutual information (MI) of a chosen feature X and label Y is defined as

$$I(X; Y) = E [I(x_i; y_j)] = \sum_{x_i \in X} \sum_{y_j \in Y} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i) p(y_j)}, \quad (2)$$

where x_i represents the chosen feature of i -th sample, and y_j represents the binary label of j -th sample.

For a chosen feature, the less mutual information it has with the label, the less information it provides for classification. Based on this principle, we performed feature filtering based on the MI, and the features whose MI with the label was lower than the threshold of 0.1 was eliminated.

Recursive Feature Elimination. Recursive feature elimination (RFE) method works with predictive models. The feature which contributes the least to the result is determined by the model during each recursion and then eliminated. The recursive process goes on until the number of remaining features does not exceed the threshold we set.

In our implementation, we used random forest as the predictive model during the RFE process, where 25 decision trees were ensembled. 30 features were selected.

It is worth mentioning that the above steps of feature selection are not quite clear at the initial stage. Instead, they are determined by trying applying common feature selection methods(including filters, wrappers and embedded ones) by following the principles that through one single selection process, an appropriate number of features can be eliminated. Removing too many or too few features in one process are avoided because the extreme threshold of the former extremizes the training data distribution, and the latter fails the selection process.

2.4 Modeling and Evaluation

We chose linear regression, a simple machine learning model for the purpose of classification, with L_1 norm as the penalty, and liblinear as the solver. The max iteration was set to 10^4 .

For evaluation, we used common metrics, including:

- F1-score
- Accuracy
- Sensitivity
- Specificity
- Precision
- ROC curve [5] and area under curve (AUC).

Each metrics were calculated with respect to the 30% test data for 50 random splits of the dataset.

3 Results and Discussions

3.1 Metrics Performance

The performance of the LR model on the selected metrics is listed in Table 1, and visualized in Fig. 3. The error bar indicates the 95% confidence interval (95% CI).

It can be seen from the figure that the model is robust to different split of training and test sets. Thus the metrics have a small interval of 95% CI.

The sensitivity is relatively low compared with other metrics. As the BUSI dataset suffers from data imbalance, where the number of available benign ROIs is nearly twice as that of malignant ones. Considering the definition of the sensitivity metric, it may be improved by properly oversampling the positive samples, i.e. malignant ROIs.

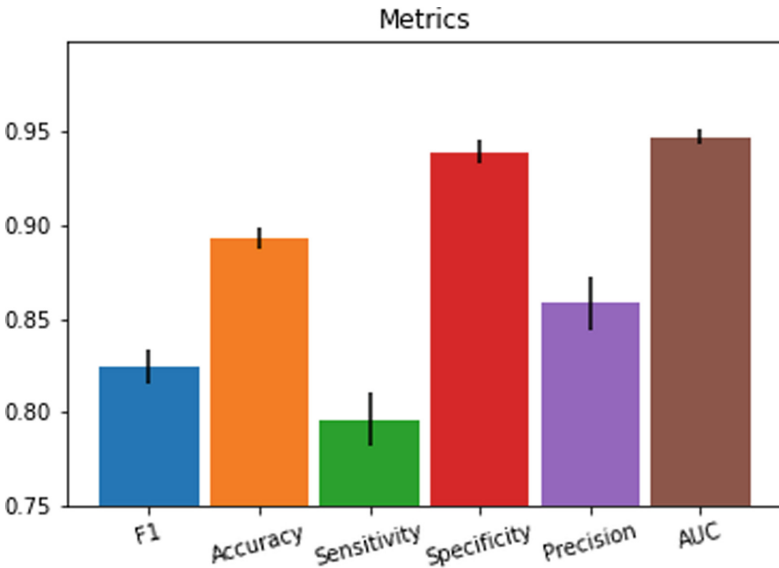


Fig. 3. Numerical metrics

Table 1. Metrics

F1-score	Accuracy	Sensitivity	Specificity	Precision	AUC
0.8246 ± 0.0089	0.8932 ± 0.0053	0.797 ± 0.014	0.9389 ± 0.0063	0.8584 ± 0.014	0.9473 ± 0.004

3.2 ROC Curve

The ROC curve of the model on a random split of the dataset is shown in Fig. 4. The corresponding AUC is 0.9469.

The ROC curve and corresponding AUC reveal that the model has a relatively high predictive value from an overall perspective, especially considering the imbalance of the dataset in this study.

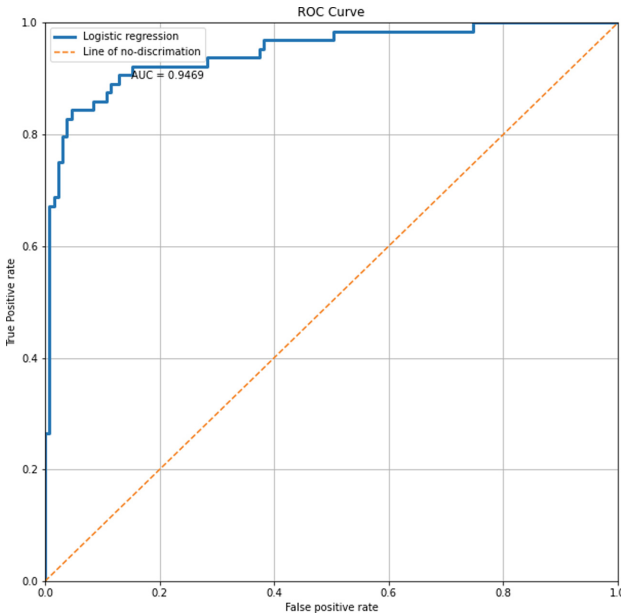


Fig. 4. The blue line is the ROC curve of our model on a random split of the dataset. (Color figure online)

3.3 Calibration Curve

The calibration curve corresponding to the model with the ROC Curve above is shown below in Fig. 5. As can be seen from the figure, when the predicted value is at lower (<0.3) and higher (>0.7) values, the calibration curve of the model is close to the perfectly calibrated curve. The deviation on the interval around 0.5 indicates that the model has much room for improvement. Attaching attention technologies or simply put more weight on the training samples whose predicted value falls in the interval around 0.5 may lead to the calibration curve approaching to the perfectly calibrated one and improve the performance of the model.

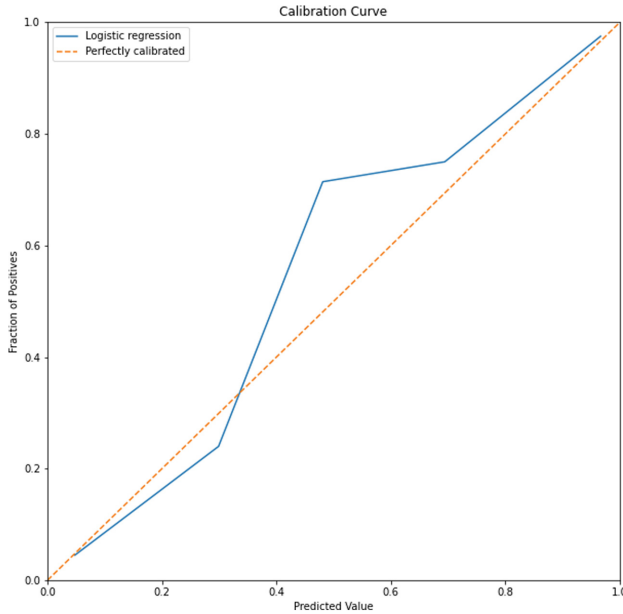


Fig. 5. The calibration curve of a random split.

4 Conclusion

We present a computational US image modeling algorithm to accurately identify breast tumors. The algorithm is able to extract reproducible and interpretable features to differentiate breast tumor malignancy. Using these clinically relevant features, the proposed classification model achieves promising results based on clinical US images from public BUSI dataset. We anticipate that the proposed tumor identification and feature extraction and selection scheme can adapt to a broader category of cancers.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China (No. 12175012).

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. *Data Brief* **28**, 104863 (2020)
2. Becker, A.S., Mueller, M., Stoffel, E., Marcon, M., Ghafoor, S., Boss, A.: Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study. *Br. J. Radiol.* **91**(1083), 20170576 (2018)
3. Bibault, J.E., et al.: Radiomics: a primer for the radiation oncologist. *Cancer/Radiothérapie* **24**(5), 403–410 (2020)
4. Cole-Beuglet, C., Beique, R.A.: Continuous ultrasound B-scanning of palpable breast masses. *Radiology* **117**(1), 123–128 (1975)

5. Cook, N.R.: Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin. Chem.* **54**(1), 17–23 (2008)
6. Ferlay, J., et al.: Global cancer observatory: cancer today. International Agency for Research on Cancer, Lyon, France (2020)
7. Van Griethuysen, J.J., et al.: Computational radiomics system to decode the radiographic phenotype. *Can. Res.* **77**(21), e104–e107 (2017)