Article

Explainable Ensemble Machine Learning for Breast Cancer Diagnosis Based on Ultrasound Image Texture Features

Alireza Rezazadeh [1,*], **Yasamin Jafarian** [2] **and Ali Kord** [3]

1   Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55414, USA
2   Department Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55414, USA; yasamin@umn.edu
3   Division of Interventional Radiology, Department of Radiology, University of Cincinnati, Cincinnati, OH 45221, USA; kordai@uc.edu
*   Correspondence: rezaz003@umn.edu

Abstract: Image classification is widely used to build predictive models for breast cancer diagnosis. Most existing approaches overwhelmingly rely on deep convolutional networks to build such diagnosis pipelines. These model architectures, although remarkable in performance, are black-box systems that provide minimal insight into the inner logic behind their predictions. This is a major drawback as the explainability of prediction is vital for applications such as cancer diagnosis. In this paper, we address this issue by proposing an explainable machine learning pipeline for breast cancer diagnosis based on ultrasound images. We extract first- and second-order texture features of the ultrasound images and use them to build a probabilistic ensemble of decision tree classifiers. Each decision tree learns to classify the input ultrasound image by learning a set of robust decision thresholds for texture features of the image. The decision path of the model predictions can then be interpreted by decomposing the learned decision trees. Our results show that our proposed framework achieves high predictive performance while being explainable.

Keywords: ultrasound image texture analysis; breast cancer prediction; explainable machine learning; ensemble classification; decision tree classification

1. Introduction

Ultrasound imaging is an effective method for breast cancer diagnosis [1,2] that, compared to alternative modalities, is more accessible and less costly. Several recent studies have explored building data-driven automated breast cancer diagnosis machine learning pipelines to detect the malignancy of tumors observed in ultrasound images [3–6]. These studies dominantly rely on deep convolutional neural network architectures to classify tumor images. Convolutional neural networks essentially learn to map the input image pixel information to a lower-dimensional feature space through a series of hidden layers.

Although notable in prediction performance, convolutional networks are largely black-box machine learning models that provide little to no insight into the logic behind their predictions [7,8]. In general, humans tend to be unwilling to rely on procedures that are not interpretable, explainable and transparent, especially for making critical predictions such as cancer diagnosis [8,9]. However, the explainability of machine learning models for cancer diagnosis falls short of the increasing demand for interpretable and reliable artificial intelligence [10].

In a recent study, Moon et al. [3] adopted standard deep convolutional neural network architectures (including VGG, ResNet, and DenseNet) to classify breast ultrasound images and detect the malignancy of tumors. The study reports a high predictive performance for these standard architectures. In a similar study, Masud et al. [11] evaluated pretrained convolutional models for ultrasound image classification. Other studies, such as [12,13], focused on semantic segmentation of the breast tumor from ultrasound images.

In this paper, we propose an explainable machine learning pipeline for probabilistic breast cancer diagnosis based on ultrasound images. We formulate this as a binary classification problem. First, through a comprehensive texture analysis, we extract first- and second-order texture features of the region of interest in the ultrasound image. We then use these features to train an ensemble of decision trees. Each decision tree in the model learns to classify the input image through a set of test conjunctions where each test compares a texture feature with a robust numerical decision threshold. We show that our proposed pipeline achieves a high predictive performance that is comparable to the existing black box convolution neural network architectures. More importantly, we demonstrate that our proposed model can be probed to accurately track and explain the decision path behind its prediction.

2. Materials and Methods

2.1. Data

We use a public dataset of breast ultrasound images [2]. In this dataset, a total number of 780 images are obtained from 600 female patients (age of 25–75 years old). This includes 133 normal cases with no mass, 210 cases with a benign mass, and 487 cases with a malignant mass. Images are obtained using a LOGIQ E9 ultrasound and LOGIQ E9 Agile ultrasound systems. These instruments produce DICOM images with 1280×1024 resolution using 1–5 MHz transducers on an ML6-15-D Matrix linear probe. The raw DICOM images are cropped, preprocessed, and converted to PNG format with an average resolution of 500×500 pixels. For each case with a mass, a ground-truth binary mask of the region of interest (ROI) is manually created. The dataset is split into an 80% train set and a 20% test set. The model performance metrics are reported based on a k-fold cross validation (k = 5).

2.2. Texture Analysis

Texture features are important quantifiable metrics to characterize and describe a region of interest in an image [14–16]. Texture feature analysis is typically measured using first-order and second-order statistical metrics [17].

2.2.1. First-Order Statistics' Texture Features

First-order texture features are computed based on first-order statistics of the one-dimensional gray level histogram of the image. Therefore, a first-order texture feature does not take into account the pixel neighborhood information [17,18]. In this study, we compute eight common first-order statics of the ROI pixels: Mean, Variance, RMS, Energy, Entropy, Kurtosis, Skewness, and Uniformity. These statistical metrics described in Table A1.

2.2.2. Second-Order Statistics' Texture Features

The second-order statistical texture features are computed based on the gray-level co-occurrence matrix (GLCM). GLCM elements are an estimation of the probability of transition from one gray level to another along a certain pixel distance and direction [19–21]. We measure five common statistics based on the GLCM computed for the ROI: Contrast, Dissimilarity, Homogeneity, Energy, and Correlation. A description of these statistics is included in Table A2.

2.3. Decision Tree Models

A decision tree (DT) is a predictive model that consists of a set of test conjunctions, where each test compares a feature of data with a numerical threshold [22]. Decision tree classification models are learned by recursively partitioning the feature space to discover a set of robust decision rules [23,24]. One major advantage of decision tree modeling is their interpretability. The set decision rules learned in a decision tree model can be directly used to explain the logic behind the prediction of the model. Explainable predictions, as opposed to black-box predictions, can be used more reliably in applications such as medical diagnosis.

2.3.1. Gradient Boosting Decision Tree

Gradient boosting decision tree (GBDT) is an ensemble of sequential decision tree models. GBDT is frequently used in a variety of machine learning tasks due to its accuracy and efficiency. At each boosting iteration, the ensemble learns a decision tree to predict the residual errors [22,25]. Specifically, for a classification problem, let $\{(X_i, y_i)\}_{i=1...N}$ be a dataset where for each entry, X and y, correspond to feature vector and the class label the entry belongs to. The goal is to approximate the function $\hat{F}(X) = y$ that learns to map feature vectors to their corresponding class label under an arbitrary differentiable loss function $L(y_i, F(X_i))$, cross-entropy loss, in our case.

With gradient boosting, at first, the model is initialized with a constant,

$$F_0(x) = \min_\lambda \sum_{i=1}^{n} L(y_i, \lambda) \tag{1}$$

Next, at each boosting iteration m, for each entry, the residuals are computed as

$$r_{i,m} = -\partial L(y, F) / \partial F |_{F=F_{m-1}} \tag{2}$$

A decision tree with J terminal nodes is fitted to the residuals where

$$\lambda_{j,m} = \min_\lambda \sum_{j} L(y, F_{m-1}(x) + \lambda) \tag{3}$$

Finally, the boosted model is updated as:

$$F_m(x) = F_{m-1}(x) + \nu \sum_{J} \lambda_{j,m} I(x) \tag{4}$$

Through grading boosting, GBDT combines multiple "weak" learner classifiers into an ensemble of a strong classification model.

2.3.2. LightGBM

LightGBM is an open-source GBDT framework [25]. LightGBM is based on a gradient-based one-side sampling to filter data instances and an exclusive feature bundling to encode features into less dense space. Specifically, LightGBM discretizes continuous features using a histogram-based algorithm for a faster training process and reduced memory consumption. In addition, LightGBM uses a leaf-wise strategy of growing decision trees by discovering a leaf with the highest gain of variance. This enables LightGBM to achieve state-of-the-art performance in a variety of applications [26–28].

2.4. Machine Learning Diagnosis Pipeline

In this paper, we propose an interpretable machine learning pipeline for breast cancer diagnosis based on ultrasound images. We formulate this as a binary classification problem where class labels belong to {benign, malignant}. As shown in Figure 1, the pipeline takes as input the ultrasound image of the mass and a binary mask of the ROI. We use this binary mask to crop and compute the GLCM of the ROI. This yields a total number of 60 texture features (5 GLCM statistics × 3 distances × 4 angles). These texture features are denoted as *{statistic}_d{distance}_a{angle}* (e.g., *energy_d3_a135* refers to the GLCM energy computed within a distance of three pixels and along a 135° direction).

We then use a LightGBM classification model with a gradient boosting decision tree strategy, 10 leaves per tree, and a maximum feature bin size of 512. The classification model is trained by minimizing a binary log loss, with a learning rate of 0.05, and for a total number of 500 boosting iterations. For a given ultrasound image input, the model outputs the probability of the mass in ROI being benign or malignant. Importantly, the decision tree ensemble can be decomposed to explain how the model comes up with a prediction.

The learned model is a set of decision trees with multiple test conjunctions that compare the texture features of the ROI with numerical thresholds inferred from the data.
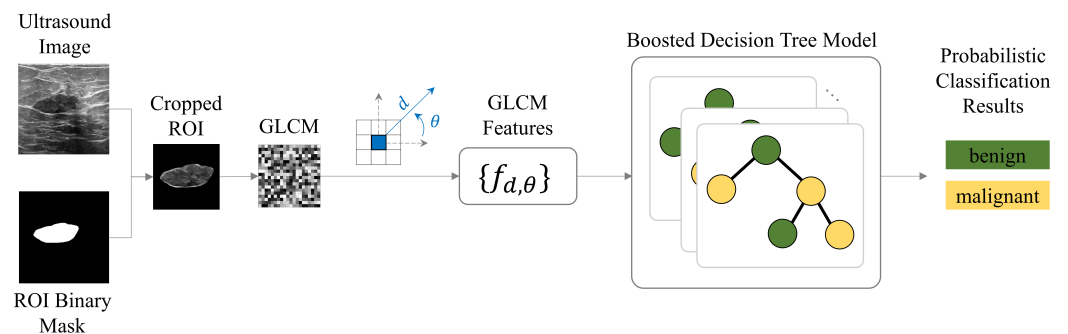

Figure 1. Our proposed pipeline uses the ultrasound image and ROI mask to extract GLCM texture features and learns a boosted decision tree model to make probabilistic diagnosis based on explainable decision trees.

3. Results

In this section, we first summarize the texture analysis results and then evaluate the performance of our purposed pipeline. Lastly, we highlight how our pipeline can be used as an explainable machine learning framework to understand the logic behind each of its diagnostic predictions.

3.1. Texture Analysis and Statistical Analysis

We perform texture analysis by computing the ROI first- and second-order statistics (see Tables A1 and A2 for mathematical descriptions). A standard *t*-test is used to compare each texture feature between the two groups of benign and malignant masses.

Most first-order statistic texture features are significantly different for the two sets of benign and malignant masses ($p < 0.001$). Mean ($p = 0.43$) and RMS ($p = 0.28$) first-order statistics, however, are not significantly different across the groups. Figure 2 demonstrates the first-order texture features comparison. All *t*-test results of this comparison are thoroughly reported in Table A1.
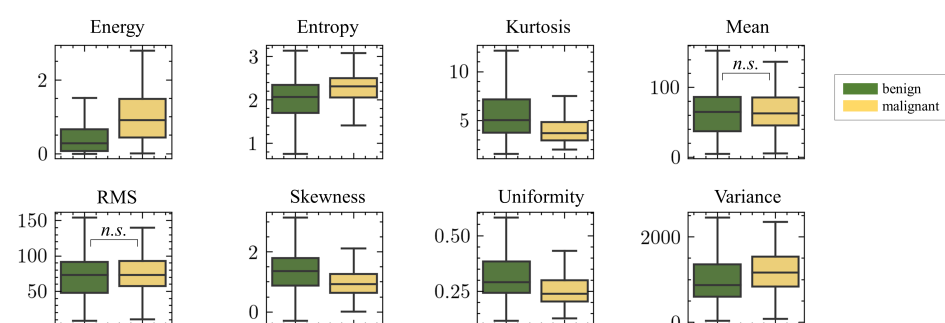

Figure 2. First-order texture features. All measured statistics, except Mean and RMS, are significantly different between the two groups of benign and malignant masses. Refer to Table A1 for a description of the metrics and *t*-test results.

Second-order statistics based on GLCM are computed across three distances $\{1, 3, 5\}$ pixels and four angles $\{0°, 45°, 90°, 135°\}$. All GLCM features are significantly different between benign and malignant groups ($p < 0.001$). See Tables A2–A7 for the complete *t*-test results. Interestingly, the difference between the two groups is consistent for various distances and angles. This further indicates that GLCM features are consistent for different orientations of the ROI. Figure 3 exemplifies the persistence of the difference in each GLCM feature across all angles for $d = 3$ pixels.
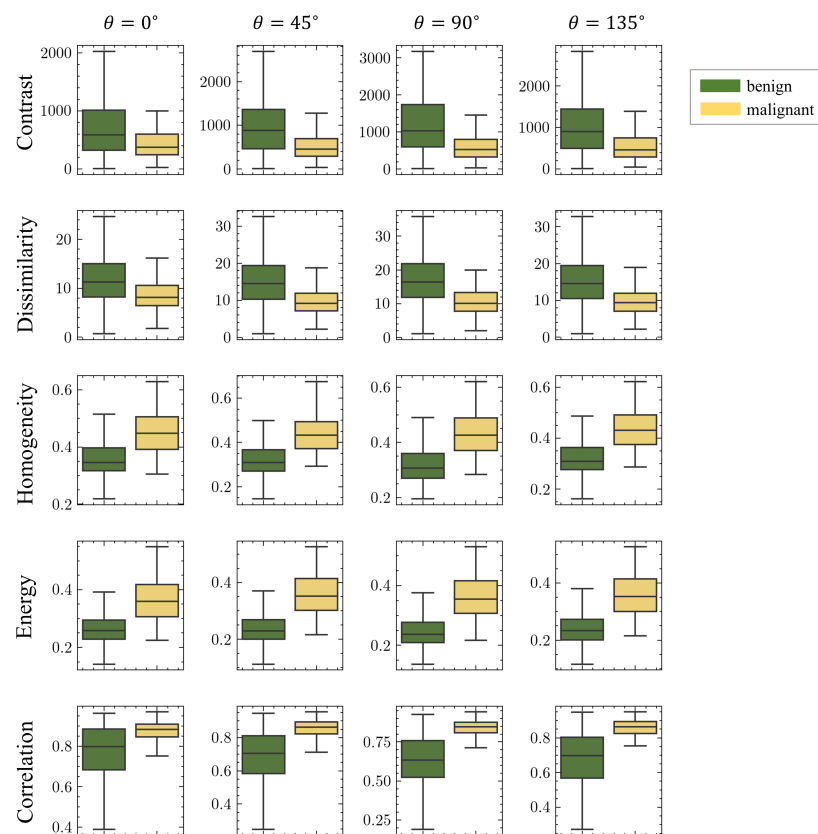
Figure 3. Second-order GLCM texture features. All measure statistics are significantly different between the benign and malignant groups. This significant difference is consistent across various pixel distances ($d = \{1, 3, 5\}$, only $d = 3$ results are demonstrated here) and angles ($\theta = \{0°, 45°, 90°, 135°\}$). Refer to Tables A2–A7 for the deception of each metric and detailed *t*-test results.

3.2. Machine Learning Diagnosis Pipeline Performance

In this section, we compare multiple variations of our framework. The comparison is made using common standard statistical metrics for evaluating classification performance on the test set. Specifically, we report: Accuracy, Precision, Recall, F1-score, and Area under the ROC Curve (AUC) (see [29] for mathematical descriptions of these metrics).

Using a LightGBM algorithm with 500 iterations of gradient boosting, our pipeline reaches its best performance with 0.91 accuracy, 0.94 precision, 0.93 recall, and 0.93 F1-score (Table 1). With a smaller number of boosting iterations, the accuracy slightly drops down. We compare our model with a simple DT and a random forest (RF) with 100 estimators and a maximum depth of two. Using LightGBM significantly improves the classification performance which further emphasizes the importance of gradient boosting approaches in our pipeline.

We also compare our model's performance with the standard convolutional neural network architectures from Moon et al. [3], which is also using a similar dataset to ours [2]. Our pipeline, based on decision tree ensembles, achieves comparable results to the convolutional network models (Table 2) while being explainable. Our model can be probed to trace the logic behind its predictions while the convolutional models do not provide any insight into the process behind their predictions. In comparison, our model achieves higher precision, recall, and F1-score. Note that the F1-score is more suitable to assess models for their classification performance on imbalanced datasets.

Table 1. Model evaluation. Standard classification performance metrics measured on the test set. The best performance is achieved with the LightGBM model with 500 gradient boosting iterations.

Model	Precision	Recall	F1-Score	AUC	Accuracy
DT	0.85	0.82	0.83	0.82	0.86
RF	0.87	0.84	0.86	0.85	0.88
LightGBM-10	0.90	0.80	0.83	0.79	0.87
LightGBM-50	0.88	0.87	0.88	0.90	0.87
LightGBM-100	0.93	0.91	0.92	0.92	0.90
LightGBM-500	**0.94**	**0.93**	**0.93**	**0.93**	**0.91**

Table 2. Model comparison with convolutional architectures. Standard classification performance metrics measured on the test set. Our explainable model based on decision trees achieves high predictive performance that is comparable to existing black box convolutional neural network architectures.

Model	Precision	Recall	F1-Score	AUC	Accuracy
VGG	0.75	0.76	0.76	0.87	0.85
ResNet	0.89	0.89	0.89	0.96	0.91
DenseNet	0.90	0.92	0.91	**0.97**	**0.94**
Decision Tree Ensemble (ours)	**0.94**	**0.93**	**0.93**	0.93	0.91

To further identify the most important texture features for the model, we quantify the feature importance with SHAP values [30,31]. The SHAP values are Shapley values from coalitional game theory and correspond to the magnitude of each feature's attribution on the output of the model. Intuitively, SHAP value of each feature singles out its importance on the model's prediction. With our dataset, the most important features are: GLCM correlation within three pixels along the 90° direction, GLCM energy within a 5-pixel distance along the 90° direction, GLCM energy within a 3-pixel distance along the 90° direction, and GLCM correlation within a 5-pixel distance along the 90° direction (see Figure 4). Interestingly, all top four important features are statistics measured in the 90° direction.
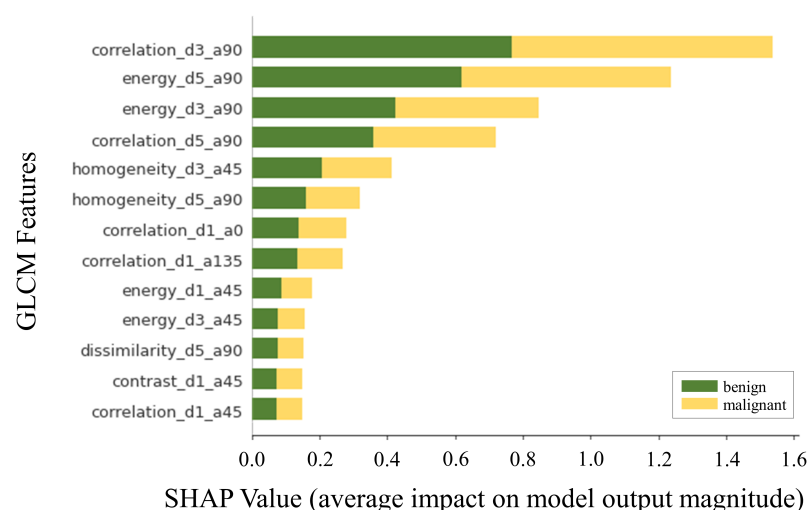

Figure 4. Feature importance. SHAP value quantifies the contribution of each feature to the prediction of each class. The features are denoted as *{statistic}_d{distance}_a{angle}*.

3.3. Explainability of the Machine Learning Diagnosis Pipeline

In this section, we give an overview of how the explainability of our pipeline enables tracking down the logic behind its predictions. First, we evaluate a benign case from the

test set (Figure 5). Given the ultrasound image and the cropped ROI, the learned model predicts the mass in ROI is benign with a likelihood of 0.97. We can further decompose the learned model to understand the decision path that leads to this prediction. The learned model is an ensemble of multiple decision trees ($T = 500$).

In this example (Figure 5), two decision trees from the ensemble are depicted. In *Tree Classifier 1*, the first nodes split the data based on *correlation_d3_a90*. This texture feature is measured as 0.36 in the cropped ROI, which, in comparison with the learned split threshold 0.78, determines the outcome of the split at this node. The histogram of nodes also compares the split threshold with the distribution of the texture feature for both malignant and benign cases in the train set. At the next node *energy_d5_a0* $= 0.23 < 0.26$, which leads to the final node *dissimilarity_d1_a0* $= 4.09 < 5.88$. Based on this decision path, the model predicts the ROI that belongs to the benign class. Other decision trees in the ensemble can also be interpreted in a similar way.

Next, we evaluate a malignant case from the test set (Figure 6). For this input, the model predicts that the mass in ROI is malignant with a likelihood of 0.93. In *Tree Classifier 1*, the model's decision path begins with a split based on *correlation_d3_a90* $= 0.86 > 0.78$. This is followed by the final node *energy_d5_a90* $= 0.31 > 0.30$, which predicts that the ROI mass is malignant.
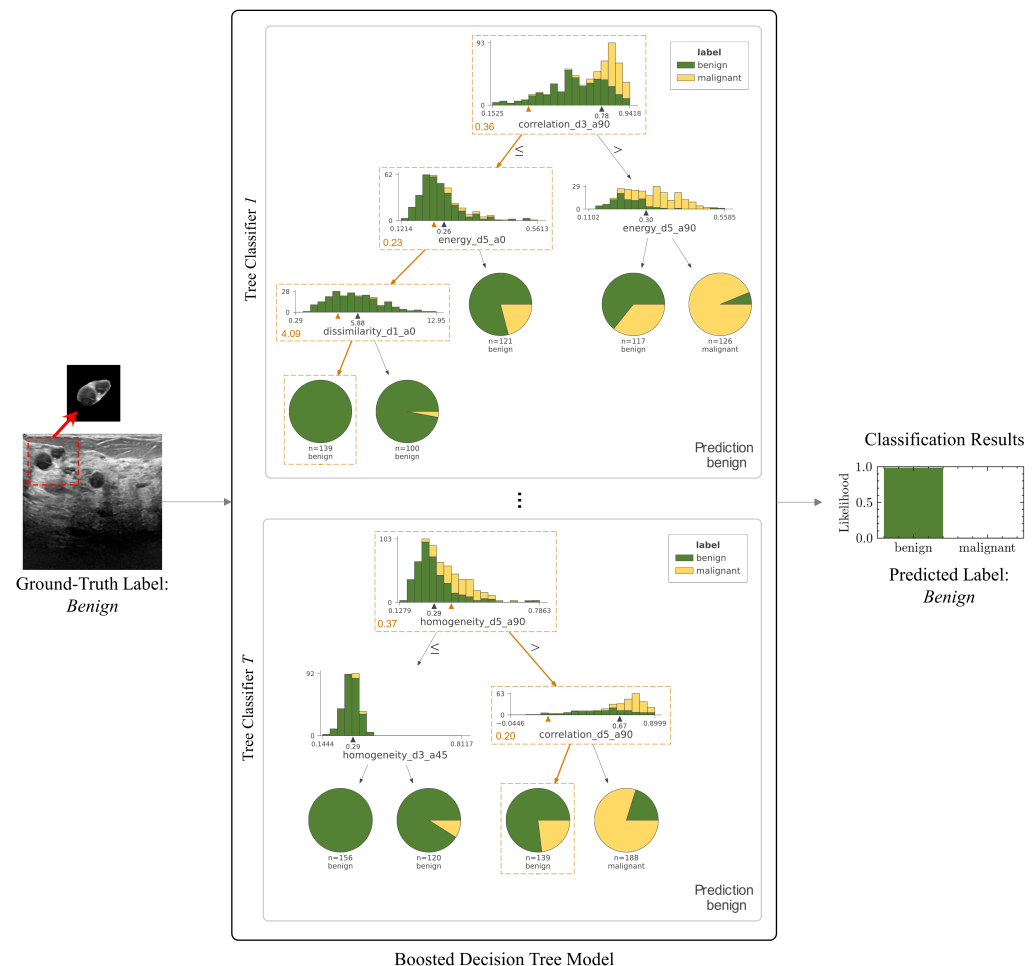

Figure 5. Qualitative results of a benign case. Our pipeline learns to infer a probabilistic diagnosis of the breast ultrasound images. The learned ensemble can be probed to obtain explainable decision paths in a set of learned decision trees. In each learned tree classifier, the orange arrows highlight the decision path. The model learns to compare the texture features obtained from the input image (orange numbers at the bottom of each dashed box) with the learned thresholds (black triangle on each histogram) at each node of the decision tree.
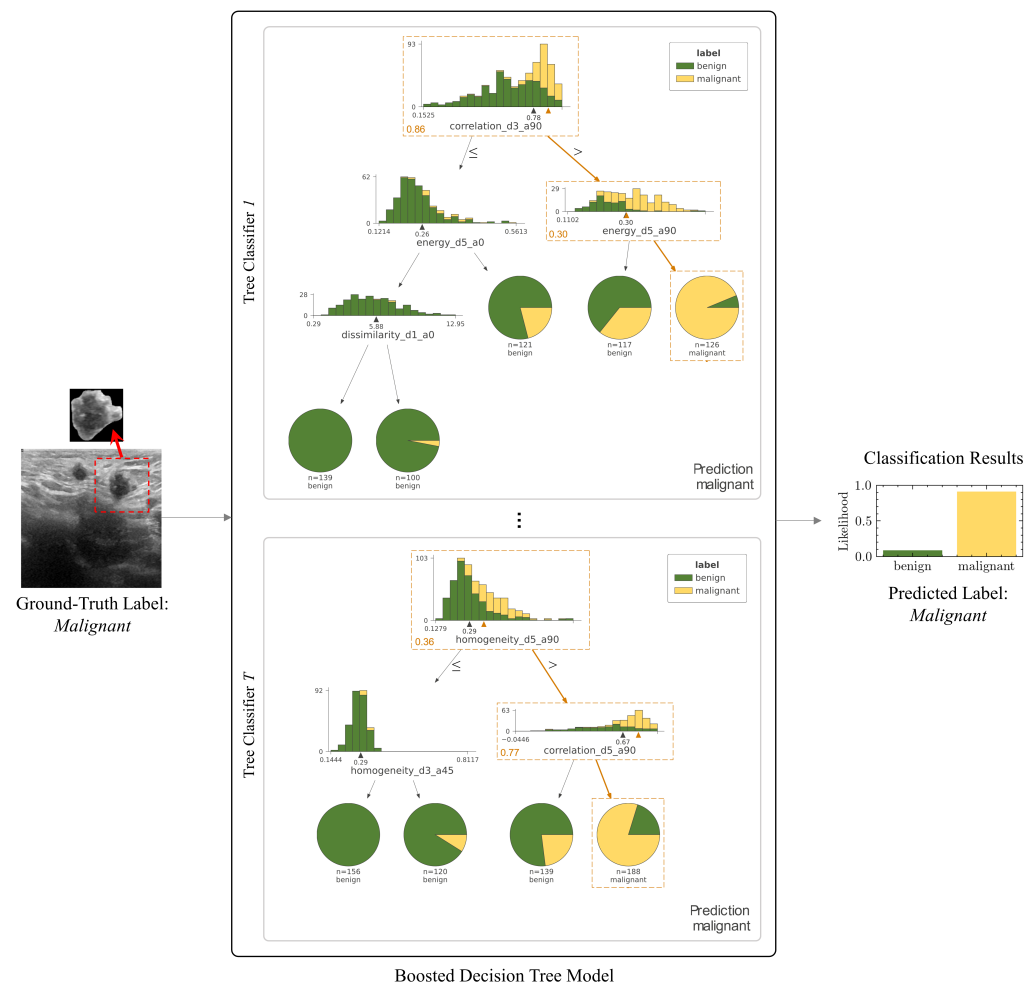
Figure 6. Qualitative results of a malignant case. Refer to the Figure 5 caption for more details.

4. Conclusions

Ultrasound imaging is an accessible and cost-effective imaging modality to diagnose breast cancer. Most recent work on building machine learning models for breast cancer diagnosis depends on convolutional neural network architectures. Although accurate in performance, convolutional networks are black-box models and cannot be interpreted in terms of the logic behind their predictions.

In this paper, we proposed a novel explainable machine learning pipeline for breast cancer diagnosis based on ultrasound images. Our pipeline uses texture analysis of the ultrasound images as its basis to learn an ensemble of decision trees to predict the likelihood of malignancy of breast tumors. Importantly, our model can be decomposed into its underlying decision trees to fully interpret the decision path behind its outputs by following test conjunctions in each decision tree.

We believe our work is a step towards building more practical and comprehensible machine learning tools for cancer diagnosis by increasing the transparency of the prediction process. An important limitation of our work is relying on manually extracted ROIs. This could be a drawback for scaling to larger datasets without ground truth ROI masks. An interesting future work is to combine convolutional networks with decision trees. Finally, we hope our approach in this work to inspire future research on data-driven medical diagnosis to devote more attention into increasing the explainability of their solutions.

Appendix A

This section gives an overview of the first- and second-order GLCM texture feature statistics along with a detailed report of the *t*-test of each statistic between the two groups of benign and malignant masses.

Table A1. First-order Texture Features. The metrics are computed for a set of N pixels inside of the ROI denoted as X. Benign and malignant groups are compared using a *t*-test.

Feature	Description	μ_{benign}	$\mu_{malignant}$	*p*-Value
energy	$\sum_{i=1}^{N} (\mathbf{X}(i) + c)^2$	6.66×10^7	1.17×10^8	<0.001
entropy	$-\sum_{i=1}^{N} p(i) \log_2 \left(p(i) + \epsilon \right)$	1.98	2.23	<0.001
kurtosis	$\dfrac{\frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}(i) - \bar{X})^4}{\left(\frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}(i) - \bar{X})^2 \right)^2}$	6.50	4.27	<0.001
mean	$\frac{1}{N} \sum_{i=1}^{N} \mathbf{X}(i)$	6.40×10^1	6.59×10^1	0.43
rms	$\sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}(i) + c)^2}$	7.26×10^1	7.51×10^1	0.28
skewness	$\dfrac{\frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}(i) - \bar{X})^3}{\left(\sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}(i) - \bar{X})^2} \right)^3}$	1.45	0.99	<0.001
uniformity	$\sum_{i=1}^{N} p(i)^2$	3.40×10^{-1}	2.69×10^{-1}	<0.001
variance	$\frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}(i) - \bar{X})^2$	1.04×10^3	1.22×10^3	<0.001

Table A2. Second-order Texture Features. The metrics are computed based on the gray-level co-occurrence matrix for a set of N gray level of pixels inside of the ROI.

GLCM Feature	Description
contrast	$\sum_{i=1}^{N} \sum_{j=1}^{N} (i - j)^2 p(i,j)$
energy	$\sum_{i=1}^{N} (\mathbf{X}(i) + c)^2$
correlation	$\dfrac{\sum_{i=1}^{N} \sum_{j=1}^{N} p(i,j) ij - \mu_x \mu_y}{\sigma_x(i) \sigma_y(j)}$
dissimilarity	$\sum_{i=1}^{N} \sum_{j=1}^{N} |i - j| p(i,j)$

Table A2. *Cont.*

GLCM Feature	Description
energy	$\sum_{i=1}^{N}\sum_{j=1}^{N}\left(p(i,j)\right)^2$
energy	$\sum_{i=1}^{N}\sum_{j=1}^{N}\left(p(i,j)\right)^2$
homogeneity	$\sum_{i=1}^{N}\sum_{j=1}^{N}\dfrac{p(i,j)}{1+\lvert i-j\rvert}$

Table A3. The gray level co-occurrence matrix contrast feature measured across three pixel distances $\{1,3,5\}$ and four angles $\{0°,45°,90°,135°\}$.

GLCM Contrast				
Distance	**Angle**	μ_{benign}	$\mu_{malignant}$	***p*-Value**
1	0°	184.86	114.20	<0.001
	45°	433.20	238.06	<0.001
	90°	304.20	154.26	<0.001
	135°	438.01	238.07	<0.001
3	0°	730.45	493.19	<0.001
	45°	1032.13	582.21	<0.001
	90°	1229.20	660.83	<0.001
	135°	1049.76	582.42	<0.001
5	0°	1136.95	830.26	<0.001
	45°	1768.76	1132.49	<0.001
	90°	1681.63	1040.57	<0.001
	135°	1801.47	1130.44	<0.001

Table A4. The gray level co-occurrence matrix correlation feature measured across three pixel distances $\{1,3,5\}$ and four angles $\{0°,45°,90°,135°\}$.

GLCM Correlation				
Distance	**Angle**	μ_{benign}	$\mu_{malignant}$	***p*-Value**
1	0°	0.94	0.97	<0.001
	45°	0.87	0.94	<0.001
	90°	0.91	0.96	<0.001
	135°	0.86	0.94	<0.001
3	0°	0.77	0.87	<0.001
	45°	0.68	0.85	<0.001
	90°	0.63	0.83	<0.001
	135°	0.68	0.85	<0.001
5	0°	0.64	0.78	<0.001
	45°	0.46	0.70	<0.001
	90°	0.47	0.73	<0.001
	135°	0.44	0.70	<0.001

Table A5. The gray level co-occurrence matrix dissimilarity feature measured across three pixel distances $\{1, 3, 5\}$ and four angles $\{0°, 45°, 90°, 135°\}$.

		GLCM Dissimilarity		
Distance	**Angle**	μ_{benign}	$\mu_{malignant}$	p-**Value**
1	0°	4.84	3.41	<0.001
	45°	8.67	5.75	<0.001
	90°	7.09	4.58	<0.001
	135°	8.69	5.74	<0.001
3	0°	12.10	8.84	<0.001
	45°	15.33	10.07	<0.001
	90°	17.04	10.97	<0.001
	135°	15.40	10.06	<0.001
5	0°	16.76	12.76	<0.001
	45°	23.10	15.95	<0.001
	90°	22.06	14.98	<0.001
	135°	23.18	15.90	<0.001

Table A6. The gray level co-occurrence matrix energy feature measured across three pixel distances $\{1, 3, 5\}$ and four angles $\{0°, 45°, 90°, 135°\}$.

		GLCM Energy		
Distance	**Angle**	μ_{benign}	$\mu_{malignant}$	p-**Value**
1	0°	0.30	0.38	<0.001
	45°	0.28	0.37	<0.001
	90°	0.29	0.38	<0.001
	135°	0.28	0.37	<0.001
3	0°	0.26	0.36	<0.001
	45°	0.24	0.35	<0.001
	90°	0.25	0.35	<0.001
	135°	0.24	0.35	<0.001
5	0°	0.24	0.34	<0.001
	45°	0.19	0.32	<0.001
	90°	0.22	0.33	<0.001
	135°	0.19	0.32	<0.001

Table A7. The gray level co-occurrence matrix homogeneity feature measured across three pixel distances $\{1, 3, 5\}$ and four angles $\{0°, 45°, 90°, 135°\}$.

		GLCM Homogeneity		
Distance	**Angle**	μ_{benign}	$\mu_{malignant}$	p-**Value**
1	0°	0.49	0.56	<0.001
	45°	0.41	0.49	<0.001
	90°	0.43	0.51	<0.001
	135°	0.41	0.50	<0.001
3	0°	0.37	0.45	<0.001
	45°	0.33	0.44	<0.001
	90°	0.33	0.43	<0.001
	135°	0.33	0.44	<0.001
5	0°	0.33	0.41	<0.001
	45°	0.26	0.32	<0.001
	90°	0.30	0.40	<0.001
	135°	0.27	0.39	<0.001

References

1. Kuhl, C.K.; Schrading, S.; Leutner, C.C.; Morakkabati-Spitz, N.; Wardelmann, E.; Fimmers, R.; Kuhn, W.; Schild, H.H. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *J. Clin. Oncol.* **2005**, *23*, 8469–8476. [CrossRef] [PubMed]
2. Al-Dhabyani, W.; Gomaa, M.; Khaled, H.; Fahmy, A. Dataset of breast ultrasound images. *Data Brief* **2020**, *28*, 104863. [CrossRef] [PubMed]
3. Moon, W.K.; Lee, Y.W.; Ke, H.H.; Lee, S.H.; Huang, C.S.; Chang, R.F. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput. Methods Programs Biomed.* **2020**, *190*, 105361. [CrossRef]
4. Samulski, M.; Hupse, R.; Boetes, C.; Mus, R.D.; den Heeten, G.J.; Karssemeijer, N. Using computer-aided detection in mammography as a decision support. *Eur. Radiol.* **2010**, *20*, 2323–2330. [CrossRef] [PubMed]
5. Sahiner, B.; Chan, H.P.; Roubidoux, M.A.; Hadjiiski, L.M.; Helvie, M.A.; Paramagul, C.; Bailey, J.; Nees, A.V.; Blane, C. Malignant and benign breast masses on 3D US volumetric images: Effect of computer-aided diagnosis on radiologist accuracy. *Radiology* **2007**, *242*, 716–724. [CrossRef]
6. Jiménez-Gaona, Y.; Rodríguez-Álvarez, M.J.; Lakshminarayanan, V. Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review. *Appl. Sci.* **2020**, *10*, 8298. [CrossRef]
7. Castelvecchi, D. Can we open the black box of AI? *Nat. News* **2016**, *538*, 20. [CrossRef]
8. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
9. Zhu, J.; Liapis, A.; Risi, S.; Bidarra, R.; Youngblood, G.M. Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. In Proceedings of the 2018 IEEE Conference on Computational Intelligence and Games (CIG), Maastricht, The Netherlands, 14–17 August 2018; pp. 1–8.
10. Preece, A.; Harborne, D.; Braines, D.; Tomsett, R.; Chakraborty, S. Stakeholders in explainable AI. *arXiv* **2018**, arXiv:1810.00184.
11. Masud, M.; Rashed, A.E.E.; Hossain, M.S. Convolutional neural network-based models for diagnosis of breast cancer. *Neural Comput. Appl.* **2020**, 1–12. [CrossRef]
12. Byra, M.; Jarosik, P.; Szubert, A.; Galperin, M.; Ojeda-Fournier, H.; Olson, L.; O'Boyle, M.; Comstock, C.; Andre, M. Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. *Biomed. Signal Process. Control* **2020**, *61*, 102027. [CrossRef] [PubMed]
13. Irfan, R.; Almazroi, A.A.; Rauf, H.T.; Damaševičius, R.; Nasr, E.A.; Abdelgawad, A.E. Dilated semantic segmentation for breast ultrasonic lesion detection using parallel feature fusion. *Diagnostics* **2021**, *11*, 1212. [CrossRef] [PubMed]
14. Tuceryan, M.; Jain, A.K. Texture analysis. *Handb. Pattern Recognit. Comput. Vis.* **1993**, 235–276.
15. Materka, A.; Strzelecki, M. Texture analysis methods—A review. *Tech. Univ. Lodz Inst. Electron. COST B11 Rep. Bruss.* **1998**, *10*, 4968.
16. Varghese, B.A.; Cen, S.Y.; Hwang, D.H.; Duddalwar, V.A. Texture analysis of imaging: What radiologists need to know. *Am. J. Roentgenol.* **2019**, *212*, 520–528. [CrossRef] [PubMed]
17. Srinivasan, G.; Shobha, G. Statistical texture analysis. *World Acad. Sci. Eng. Technol.* **2008**, *36*, 1264–1269.
18. Kim, N.D.; Amin, V.; Wilson, D.; Rouse, G.; Udpa, S. Ultrasound image texture analysis for characterizing intramuscular fat content of live beef cattle. *Ultrason. Imaging* **1998**, *20*, 191–205. [CrossRef]
19. Sebastian V, B.; Unnikrishnan, A.; Balakrishnan, K. Gray level co-occurrence matrices: Generalisation and some new features. *arXiv* **2012**, arXiv:1205.4831.
20. Iqbal, F.; Pallewatte, A.S.; Wansapura, J.P. Texture analysis of ultrasound images of chronic kidney disease. In Proceedings of the 2017 Seventeenth International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, 6–9 September 2017; pp. 1–5.
21. Xu, S.S.D.; Chang, C.C.; Su, C.T.; Phu, P.Q. Classification of liver diseases based on ultrasound image texture features. *Appl. Sci.* **2019**, *9*, 342. [CrossRef]
22. Sharma, H.; Kumar, S. A survey on decision tree algorithms of classification in data mining. *Int. J. Sci. Res. (IJSR)* **2016**, *5*, 2094–2097.
23. Myles, A.J.; Feudale, R.N.; Liu, Y.; Woody, N.A.; Brown, S.D. An introduction to decision tree modeling. *J. Chemom. A J. Chemom. Soc.* **2004**, *18*, 275–285. [CrossRef]
24. Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **1991**, *21*, 660–674. [CrossRef]
25. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
26. Rezazadeh, A. A Generalized Flow for B2B Sales Predictive Modeling: An Azure Machine-Learning Approach. *Forecasting* **2020**, *2*, 267–283. [CrossRef]
27. Chen, C.; Zhang, Q.; Ma, Q.; Yu, B. LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. *Chemom. Intell. Lab. Syst.* **2019**, *191*, 54–64. [CrossRef]
28. Sun, X.; Liu, M.; Sima, Z. A novel cryptocurrency price trend forecasting model based on LightGBM. *Financ. Res. Lett.* **2020**, *32*, 101084. [CrossRef]

29. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1.
30. Lundberg, S.M.; Erion, G.G.; Lee, S.I. Consistent individualized feature attribution for tree ensembles. *arXiv* **2018**, arXiv:1802.03888.
31. Sundararajan, M.; Najmi, A. The many Shapley values for model explanation. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 9269–9278.