Few-shot Image Classification for Breast Cancer Detection

An Advanced Approach in Medical Imaging

Qirun Dai, Jingzhi Sun,
Pengyu Chen, Xiaowei Zeng

UC Davis & Fudan University

March 2, 2024

Problem Statement

- **Background:** Breast cancer is a pressing concern worldwide, ranking as the second leading cause of cancer death in women. Early detection and identification of breast cancer can lead to timely treatment, effectively reducing the risk of further deterioration or death [Miller et al. 2022].
- **Motivation:**
  - Traditional medical image classification: Consumes a lot of manpower and time.
  - Pattern recognition and machine learning: Automates and streamlines medical image classification.
- **Objective:** Few-shot image classification for breast cancer detection.

Major Challenges and Guidelines

- **Data sparsity:** How to generalize the model and prevent overfitting in a few-shot training setting? [Varoquaux and Cheplygina 2022]
  - Utilize a special breast image dataset that aligns with a genuine few-shot training scenario.
- **Rapid development of new vision models:** How to compare and evaluate the traditional machine learning models and cutting-edge deep learning models?
  - **Parametric supervised learning:** Naïve Bayes, Logistic Regression.
  - **Non-parametric supervised learning:** Support Vector Machine (SVM).
  - **Ensembling methods:** Decision Tree, Random Forest.
  - **Deep learning models:** Convolutional Neural Networks (CNN), Vision Transformers (ViT).

Breast Ultrasound Images Dataset

- **Collection:** 2018 [Al-Dhabyani et al. 2020].
- **Observations:** 600 breast ultrasound images among women in ages between 25 and 75 years old.
- **Data size:** 780 images with an average image size of 500*500 pixels.
- **Labels:** Three classes, which are **normal, benign, and malignant**.
- **Format:** PNG, including original images and masked images.
- **Randomly split:** 80% for training and the remaining 20% for testing.
- This dataset not only vividly simulates the real-world scenario of medical data sparsity, but also poses a great challenge to the efficient training of machine learning models.

Masking in Image Classification

Masking is a technique in image processing that highlights specific areas, crucial in medical imaging for focusing on regions of interest. This approach offers several advantages:

- **Enhanced Focus:** Directs analysis to key regions, aiding in accurate diagnosis.
- **Accuracy:** Improves classification accuracy by focusing on relevant areas (Shape and Area Size).
- **Noise Reduction:** Reduces background noise and extraneous details.
- **Efficiency:** Simplifies image content for more targeted analysis.

Illustration of Masked Images

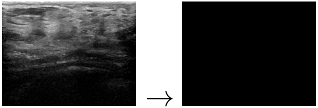

Figure: Normal and its Mask

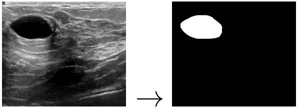

Figure: Benign and its Mask

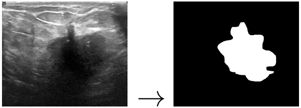

Figure: Malignant and its Mask

Methods and Models

1. **Naïve Bayes (Parametric)**
2. Logistic Regression (Parametric)
3. Support Vector Machine (Non-Parametric)
4. Decision Tree and Random Forest (Ensemble)
5. Deep Learning Methods

Naïve Bayes

- **"Naïve":** The strong assumption that the features are conditionally independent of one another given the class label, which enhances the algorithm's computational efficiency.
- **Performance in real-world scenarios:** While the independence assumption is often violated in practice, Naïve Bayes delivers competitive classification accuracy [Webb, Keogh, and Miikkulainen 2010].
- **Features:**
  - Computational efficiency.
  - Robustness in the face of noise.
  - Robustness in the face of missing values.

Naïve Bayes in Medical Imaging

- Nav̈e Bayes is really **suitable** for medical image classification tasks.
- **Resilience against noise:**
  - It uses **all attributes** for predictions, regardless of the presence of noisy or irrelevant attributes.
  - It estimates the probability based on the **overall likelihood**.
- **Alignment with independence assumption:**
  - Due to the randomness of medical events [Zaw, Maneerat, and Win 2019].
- Consequently, Naïve Bayes consistently delivered high predictive accuracy [Ramesh Kumar and Vijaya 2022].

Naïve Bayes on Original Images

Images were resampled to a 64x64 resolution. We trained the Naïve Bayes classifier with default hyperparameters as the baseline.

Confusion Matrix	Benign	Malignant	Normal
Benign	45	12	27
Malignant	8	25	10
Normal	12	6	11

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.69	0.54	0.60	84
Malignant	0.58	0.58	0.58	43
Normal	0.23	0.38	0.29	29

Accuracy: 0.52 Total Support: 156

Table: Classification Report

Naïve Bayes on Masked Images

We trained another Naïve Bayes classifier with default parameters on masked images. The following result is the baseline for later parameter tuning.

Confusion Matrix	Benign	Malignant	Normal
Benign	79	8	0
Malignant	30	16	0
Normal	0	0	27

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.72	0.91	0.81	87
Malignant	0.67	0.35	0.46	46
Normal	1.00	1.00	1.00	27

Accuracy: 0.76 Total Support: 160

Table: Classification Report

Parameter Tuning for Naïve Bayes

var_smoothing is the stability calculation to smooth the curve and therefore account for more samples that are further away from the distribution mean. After the log-scale search and cross validation, we chose 0.0032.

Confusion Matrix	Benign	Malignant	Normal
Benign	78	9	0
Malignant	24	22	0
Normal	0	0	27

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.76	0.90	0.83	87
Malignant	0.71	0.48	0.57	46
Normal	1.00	1.00	1.00	27
Accuracy: 0.79		Total Support: 160		

Table: Classification Report

Findings in Naïve Bayes Learning Process

During the parameter tuning and evaluation of the Naïve Bayes model, we made several key observations:

1. **Sampling Adequacy:** A 64x64 resampling was found to be sufficient for this problem. Increasing the sampling dimensions did not contribute to higher accuracy.
2. **Computational Efficiency:** The training time is **linear** with both the number of training examples and the number of attributes, and the classification time is **linear** with the number of attributes and unaffected by the number of training examples.
3. **Improvement After Parameter Tuning:** Higher **recall** rate for malignant tumor and higher overall accuracy.
4. **Classification Challenges:** Models sometimes misclassify benign and malignant tumors, requiring more effective identification

Methods and Models

1. Naïve Bayes (Parametric)
2. **Logistic Regression (Parametric)**
3. Support Vector Machine (Non-Parametric)
4. Decision Tree and Random Forest (Ensemble)
5. Deep Learning Methods

Logistic Regression

- Logistic regression is a a cornerstone in supervised learning method in classification with the following features:
  - **Applicability to Small Datasets:** Logistic regression exhibits proficiency with relatively small datasets, making it valuable in medical scenarios where acquiring labeled data brings privacy concerns and more cost.
  - **Linear Assumption:** Logistic regression assumes that the relationship between the log-odds of the variables should be approximately linear. While this assumption might be reasonable in certain cases, the complex nature of medical images introduces challenges to this linearity assumption.
- **Performance in medical image classification:** In a study byDinesh and Kalyanasundaram 2022, logistic regression demonstrated superior performance compared to SVM, KNN, Decision Tree using the Wisconsin dataset.

Logistic Regression on Original Images

Images were resampled to a 64x64 resolution. We trained the support vector machine with default hyperparameters and original images as the baseline.

Confusion Matrix	Benign	Malignant	Normal
Benign	71	8	5
Malignant	17	24	2
Normal	15	2	12

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.69	0.85	0.76	84
Malignant	0.71	0.56	0.62	43
Normal	0.63	0.41	0.50	29

Accuracy: 0.69 Total Support: 156

Table: Classification Report

Logistic Regression on Masked Images

We trained another Logistic regression classifier with default parameters on masked images. The following result is the baseline for later parameter tuning.

Confusion Matrix	Benign	Malignant	Normal
Benign	61	22	4
Malignant	27	19	0
Normal	0	0	27

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.69	0.70	0.70	87
Malignant	0.46	0.41	0.44	46
Normal	0.87	1.00	0.93	27

Accuracy: 0.67 Total Support: 160

Table: Classification Report

Parameter Tuning for Logistic Regression

Using grid search, we optimized the Logistic Regression classifier's hyperparameters.
C: 0.025, **penalty:** l2.

Confusion Matrix	Benign	Malignant	Normal
Benign	72	14	1
Malignant	23	23	0
Normal	0	0	27

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.76	0.83	0.79	87
Malignant	0.62	0.50	0.55	46
Normal	0.96	1.00	0.98	27
	Accuracy: 0.76		Total Support: 160	

Table: Classification Report

2-stage Logistic Regression

Introducing a mask image improved the classification accuracy between normal and tumors. Then train another Logistic Regression using original image to classify benign and malignant tumors

Confusion Matrix	Benign	Malignant	Normal
Benign	83	8	1
Malignant	17	22	0
Normal	0	0	27

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.90	0.83	0.86	87
Malignant	0.77	0.73	0.75	46
Normal	0.96	1.00	0.98	27
Accuracy: 0.84		Total Support: 160		

Table: Classification Report

Findings in Logistic Regression learning progress

During the parameter tuning and evaluation of Logistic Regression classifier, we made several key observations:

1. **Sampling Adequacy:** A 64x64 resampling was found to be sufficient for this problem. Increasing the sampling dimensions did not contribute to higher accuracy.
2. **Mask image:** Introducing a mask image significantly improved the classification accuracy between normal images and those with tumors, which led to a decrease in the classification accuracy between benign tumors and malignant tumors. So we introduce a two-stage Logistic Regression, which significantly improve the accuracy

Methods and Models

1. Naïve Bayes (Parametric)
2. Logistic Regression (Parametric)
3. **Support Vector Machine (Non-Parametric)**
4. Decision Tree and Random Forest (Ensemble)
5. Deep Learning Methods

Support Vector Machine

- **Theoretical Foundation:** Support Vector Machines (SVM) originated in the 1990s and, similar to an extension of neural networks, focuses on discovering an optimal hyperplane for classification.
- **Performance in medical images:** Chi, Feng, and Bruzzone 2008 emphasize SVM's efficiency in classifying small-sized training datasets, showcasing its ability to generalize well in scenarios with high-dimensional input spaces.
- **Key concepts:**
  - **Support Vector:** SVM relies on support vectors lying on class boundaries, which is useful in small datasets. This attribute proves especially beneficial in medical imaging.
  - **Kernel function:** SVM utilizes a kernel function to transform data into a higher-dimensional space, enabling the establishment of a decision plane to eparate distinct classes.

SVM on Original Images

Images were resampled to a 64x64 resolution. We trained the support vector machine with default hyperparameters and original images as the baseline.

Confusion Matrix	Benign	Malignant	Normal
Benign	78	6	0
Malignant	21	22	0
Normal	22	2	5

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.64	0.93	0.76	84
Malignant	0.73	0.51	0.60	43
Normal	1.00	0.17	0.29	29

Accuracy: 0.67 Total Support: 156

Table: Classification Report

SVM on Masked Images

We trained another support vector machine with default parameters on masked images. The following result is the baseline for later parameter tuning.

Confusion Matrix	Benign	Malignant	Normal
Benign	77	7	3
Malignant	16	30	0
Normal	0	0	27

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.83	0.89	0.86	87
Malignant	0.81	0.65	0.72	46
Normal	0.90	1.00	0.95	27

Accuracy: 0.84 Total Support: 160

Table: Classification Report

Parameter Tuning for SVM

Using grid search, we optimized the support vector machine classifier's hyperparameters.
C: 22.5, **gamma:** 0.003, **kernel:** rbf.

Confusion Matrix	Benign	Malignant	Normal
Benign	77	10	0
Malignant	14	32	0
Normal	0	0	27

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.85	0.89	0.87	87
Malignant	0.76	0.70	0.73	46
Normal	1.00	1.00	1.00	27
	Accuracy: 0.85		Total Support: 160	

Table: Classification Report

Findings in SVM learning progress

During the parameter tuning and evaluation of our support vector machine, we made several key observations:

1. **Sampling Adequacy:** A 64x64 resampling was found to be sufficient for this problem.
2. **Model performance:**
   - Improving accuracy from 0.67 to 0.84 through the inclusion of mask data is a significant enhancement, indicating that the additional information captured by the mask features is valuable for classification task.
   - Limited improvement of 0.01 through parameter tuning suggests that the model might already be performing close to its optimal capacity. Other factors are needed to improve performance.
   - Low False Negative Rate is critically important in medical image classification, especially when dealing with conditions like cancer.
3. **Classification Challenges:** SVM classifier sometimes misclassifies benign and malignant tumors, requiring more effective features.

Methods and Models

1. Naïve Bayes (Parametric)
2. Logistic Regression (Parametric)
3. Support Vector Machine (Non-Parametric)
4. **Decision Tree and Random Forest (Ensemble)**
5. Deep Learning Methods

Decision Tree and Random Forest

- **Decision Trees in Medical Imaging:** Kaganov, Ades, and Fraser 2018 highlighting the effectiveness of decision tree models in MRI signal intensity classification.
- **Advancement with Random Forests:** Criminisi et al. 2010 Describing the use of random forests for automatic detection and localization in three-dimensional CT scans. Increase accuracy and robust.
- **Study Aim:** Exploring the application of Random Forests in few-shot image classification for breast cancer detection - a three-class problem: normal, benign, or malignant.

Initial Method with Raw Images

In this initial approach, images were resampled to a 64x64 resolution, training the Random Forest classifier with default hyperparameters. The obtained results were:

Confusion Matrix	Benign	Malignant	Normal
Benign	82	2	0
Malignant	24	19	0
Normal	20	1	8

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.65	0.98	0.78	84
Malignant	0.86	0.44	0.58	43
Normal	1.00	0.28	0.43	29
Accuracy: 0.70		Total Support: 156		

Table: Classification Report

Results with Masked Images

After applying a 64x64 resampling and using default hyperparameters, the Random Forest classifier was trained on masked images. The results were as follows:

Confusion Matrix	Benign	Malignant	Normal
Benign	77	8	2
Malignant	22	24	0
Normal	0	0	27

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.78	0.89	0.83	87
Malignant	0.75	0.52	0.62	46
Normal	0.93	1.00	0.96	27
Accuracy: 0.80		Total Support: 160		

Table: Classification Report

Parameter Tuning with Grid Search

Using grid search, we optimized the Random Forest classifier's hyperparameters.

Max Depth: 9, **Min Sample Split:** 4, **Num of Estimators:** 70.

Confusion Matrix	Benign	Malignant	Normal
Benign	81	3	3
Malignant	17	29	0
Normal	0	0	27

Table: Confusion Matrix

Class	Precision	Recall	F1-score	Support
Benign	0.83	0.93	0.88	87
Malignant	0.91	0.63	0.74	46
Normal	0.90	1.00	0.95	27
Accuracy: 0.86		Total Support: 160		

Table: Classification Report

Findings in Random Forest Learning Process

During the parameter tuning and evaluation of our Random Forest model, we made several key observations:

1. **Sampling Adequacy:** A 64x64 resampling was found to be sufficient for this problem. Increasing the sampling dimensions did not contribute to higher accuracy.
2. **Algorithm Stability:** The model showed good stability across different random seeds, indicating low overfitting.
3. **Classification Challenges:** Our masked-image model excelled in distinguishing diseased from non-diseased cases, likely due to non-diseased masks being entirely black while diseased ones show increased white values. However, it faced challenges differentiating between benign and malignant cases.

These findings provide insights into the strengths and limitations of our approach, guiding future improvements.

Summary for Machine Learning Methods

The accuracy of the four machine learning methods are shown below.

Accuracy	Naïve Bayes	Logistic	SVM	Random Forest
Original	0.52	0.69	0.67	**0.70**
Masked	0.76	0.67	**0.84**	0.80
Masked + Tuning	0.79	0.76	0.85	**0.86**

Table: Summary of the Results

- Except the Logistic model, all the methods have higher accuracy when training with masked data than with original data.
- Random Forest and SVM performs better on breast cancer detection.
- The Logistic model has better performance in distinguishing between benign and malignant tumors when training with original data. While training with mask data, all models have better performance in distinguishing whether patient the has a cancer or not. (Normal v.s. Benign & Malignant).

Methods and Models

1. Naïve Bayes (Parametric)
2. Logistic Regression (Parametric)
3. Support Vector Machine (Non-Parametric)
4. Decision Tree and Random Forest (Ensemble)
5. **Deep Learning Methods**

References I

Al-Dhabyani, Walid et al. (2020). "Dataset of breast ultrasound images". In: *Data in brief* 28, p. 104863.

Chi, Mingmin, Rui Feng, and Lorenzo Bruzzone (2008). "Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem". In: *Advances in space research* 41.11, pp. 1793–1799.

Criminisi, A. et al. (2010). "Regression Forests for Efficient Anatomy Detection and Localization in CT Studies". In: pp. 106–117. DOI: 10.1007/978-3-642-18421-5_11.

Dinesh, Paidipati and P Kalyanasundaram (2022). "Medical image prediction for diagnosis of breast cancer disease comparing the machine learning algorithms: SVM, KNN, logistic regression, random forest, and decision tree to measure accuracy". In: *ECS Transactions* 107.1, p. 12681.

References II

Kaganov, Helen, A. Ades, and David S. Fraser (2018). "PREOPERATIVE MAGNETIC RESONANCE IMAGING DIAGNOSTIC FEATURES OF UTERINE LEIOMYOSARCOMAS: A SYSTEMATIC REVIEW". In: *International Journal of Technology Assessment in Health Care* 34, pp. 172 –179. DOI: 10.1017/S0266462318000168.

Miller, Kimberly D et al. (2022). "Cancer treatment and survivorship statistics, 2022". In: *CA: a cancer journal for clinicians* 72.5, pp. 409–436.

Ramesh Kumar, P and A Vijaya (2022). "Naïve Bayes machine learning model for image classification to assess the level of deformation of thin components". In: *Materials Today: Proceedings* 68. 4th International Conference on Advances in Mechanical Engineering, pp. 2265–2274. ISSN: 2214-7853. DOI: https://doi.org/10.1016/j.matpr.2022.08.489.

References III

Varoquaux, Gaël and Veronika Cheplygina (2022). "Machine learning for medical imaging: methodological failures and recommendations for the future". In: *NPJ digital medicine* 5.1, p. 48.

Webb, Geoffrey I, Eamonn Keogh, and Risto Miikkulainen (2010). "Naïve Bayes.". In: *Encyclopedia of machine learning* 15.1, pp. 713–714.

Zaw, Hein Tun, Noppadol Maneerat, and Khin Yadanar Win (2019). "Brain tumor detection based on Naïve Bayes Classification". In: *2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, pp. 1–4. DOI: 10.1109/ICEAST.2019.8802562.