Few-shot Image Classification for Breast Cancer Detection

Qirun Dai **Jingzhi Sun** **Pengyu Chen** **Xiaowei Zeng**

1 Problem Statement

Breast cancer is a pressing concern worldwide, ranking as the second leading cause of cancer death in women [Miller et al., 2022]. Fortunately, early detection and identification of breast cancer can lead to timely treatment, effectively reducing the risk of further deterioration or death. Breast ultrasound image classification serves as a primary method for such detection. However, traditional medical image classification often demands considerable human expertise and time, making it impractical for many underdeveloped countries and regions. Hence, there has been a shift towards pattern recognition and machine learning approaches to automate and streamline medical image classification and breast cancer detection. Numerous learning-based methods, ranging from traditional machine learning to modern deep learning, have been proposed.

However, two major challenges remain relatively unaddressed. Firstly, due to the inherent data sparsity of real-life medical image data [Varoquaux and Cheplygina, 2022], current machine learning methods that thrive on extensive datasets struggle to generalize and prevent overfitting in a few-shot training setting. Secondly, the swift advancements in the deep learning community mean a constant influx of new vision models, complicating comprehensive comparisons and evaluations between traditional machine learning models and cutting-edge deep learning models. Our project aims to address these challenges with the following guidelines:

1. Our research seeks to mimic the data-scarce real-world environment of the medical image classification domain. To achieve this, we will utilize a special breast ultrasound image dataset comprising only 780 training and testing images combined. By doing so, our project aligns with a genuine few-shot training scenario, adding significant real-world application value.
2. We also aim to perform a comprehensive comparison and evaluation between traditional machine learning models and state-of-the-art deep learning models. Specifically, we plan to experiment on five traditional machine learning models representing three distinct learning paradigms: for **parametric supervised learning**, we plan to use Logistic Regression and Naive Bayes; for **non-parametric supervised learning**, we plan to use Support Vector Machine; for **ensembling methods**, we plan to use Decision Tree and Random Forest. For the deep learning models, we intend to employ two leading vision architectures: **Convolutional Neural Networks (CNN)** and **Vision Transformers (ViT)**. Through this approach, our project endeavors to create a comprehensive evaluation framework for few-shot medical image classification, encompassing both statistical learning and deep learning paradigms indicative of prior research.

2 Dataset Information

The dataset we use for our classification task includes breast ultrasound images collected from 600 female patients aged between 25 and 75 years old [Al-Dhabyani et al., 2020]. The dataset consists of 780 black-and-white images, each of which containing 500*500 pixels, and is categorized into three classes: normal, benign, and malignant. One of the most intriguing features of this dataset is the exceptionally small size. With a 80%/20% training/testing split, only about 600 images can be used to train a model for three-class classification. Such a characteristic not only vividly simulates the

Preprint. Under review.

real-world scenario of medical data sparsity but also poses a great challenge to the efficient training of machine learning models.

3 Methods

As is stated above, we propose to explore the few-shot medical image classification performance of five statistical learning models encompassing three distinct learning paradigms: Logistic Regression, Naive Bayes, Support Vector Machine, Decision Tree and Random Forest. These models are widely applied in both academic and industrial practice, due to their low implementation difficulty and rich interpretability. Besides, we also intend to compare the performance of traditional machine learning models with two other state-of-the-art deep learning architectures: Convolutional Neural Networks (e.g. ResNet [He et al., 2016]) and Vision Transformers (e.g. Swin Transformer [Liu et al., 2021]). It is noteworthy that our research motivation is to not only compare the few-shot classification performance of various architectures, but also conduct a grounded analysis on the possible latent factors contributing to the differences in final performance, thus establishing both an empirical impression and theoretical interpretation of various models.

3.1 Logistic Regression

Logistic regression, a fundamental supervised learning method rooted in probability functions, is mainly utilized for classification tasks. Differing from traditional regression, it gauges event likelihood by leveraging the sigmoid function, enabling the transformation of output into a range between 0 and 1. This shift aids in confining the continuous input space, effectively assigning probabilities to different classes in a more refined manner.

An advantage of logistic regression is its competence with relatively small datasets, which proves beneficial in scenarios where obtaining labeled medical imaging data is restricted. This advantage is particularly pertinent in medical domains, where challenges like data privacy and the cost of obtaining labeled medical images limit dataset sizes. In a study by [Dinesh and Kalyanasundaram, 2022], logistic regression exhibited superior performance compared to SVM, KNN, Decision Tree, and Random Forest in breast cancer detection using the Wisconsin dataset.

3.2 Naïve Bayes

The Naïve Bayes algorithm calculates the posterior probability for each class and predicts the class with the highest probability. The term 'naïve' comes from a strong assumption that the features are conditionally independent of one another given the class label, which enhances the algorithm's computational efficiency. While this assumption may not always hold in real-world scenarios, it actually demonstrates competitive classification accuracy in most cases [Webb et al., 2010].

Naïve Bayes is suitable for image classification tasks due to its resilience against common noise often encountered in image datasets. It embraces all available attributes when making predictions, regardless of the presence of noisy or irrelevant attributes. Moreover, it leverages probability estimates rooted in the overall likelihood, thereby reducing the adverse impact of noisy data points on model fitting. Consequently, even in scenarios rife with noise, Naïve Bayes can make informed decisions.

Compared to other tasks, the independence assumption of Naïve Bayes is less susceptible to violations in the context of breast cancer classification, which can be attributed to the randomness of medical events. [Zaw et al., 2019] employed Naïve Bayes for brain tumor detection and observed that this method proved significantly effective due to independence on the occurrence of the extracted features. In scenarios with high-dimension features, [Ramesh Kumar and Vijaya, 2022] introduced feature ranking techniques and Principal Component Analysis (PCA) to enhance the Naïve Bayes Classifier, reducing feature dimensionality and computational modeling costs while increasing interpretability at minimized information loss. The results showed that Naïve Bayes consistently delivered high predictive accuracy.

3.3 Support Vector Machine

Originating in the 1990s by Vapnik and his team, SVM theory, considered as an extension of neural networks, focuses on discovering an optimal hyperplane for classification. By utilizing a kernel function to transform data into a higher-dimensional space, SVM establishes a decision plane that effectively separates distinct classes.

A significant aspect of SVM is its reliance on support vectors lying on class boundaries for classification. This attribute proves especially beneficial in medical imaging, where acquiring training data is often costly and limited due to factors such as expert annotation or specific imaging techniques. In a study by [Chi et al., 2008], the authors highlight SVM's efficacy in classifying small-sized training datasets. They emphasize its large-margin classification and strong generalization capacity in scenarios with high-dimensional input spaces, addressing challenges associated with limited training data.

3.4 Decision Tree and Random Forest

In the domain of 'Breast cancer image' classification, where images are categorized as 'normal,' 'benign,' or 'malignant,' Decision Trees (DTs) have proven invaluable. DTs are adept at organizing diverse image data, ensuring accurate categorization.

Each DT assesses image features, including pixel brightness, texture changes, color patterns, and breast tissue-related details. By iteratively partitioning data based on these features, DTs differentiate 'normal,' 'benign,' and 'malignant' images, ensuring precision.

However, it's crucial to control the level of detail in DTs to avoid overfitting, especially in medical imaging. Excessive focus on irrelevant details can lead to misclassification. Thus, regular adjustments and refinements are essential, as demonstrated by studies:

[Riri et al., 2016] introduced an algorithm using a decision tree to classify 19 types of dental images, highlighting the method's effectiveness.Moreover, [Kaganov et al., 2018] found that a decision tree model based on MRI signal intensities aids in diagnosing uterine leiomyosarcomas, revealing a significant relationship between histopathological type and T1 and T2 intensity signals.

In RF, each tree evaluates images independently, scrutinizing various features like texture, color patterns, and brightness to detect breast tissue changes. Since each tree is trained with different data and features, they offer unique perspectives. This diversity yields a broader and more detailed understanding of breast cancer images.

The use of multiple trees in RF results in significant variability, preventing overfitting to the training data and ensuring adaptability, a crucial aspect in medical imaging.When applied to 'Breast cancer image' classification, the Random Forests model, with its numerous trees, provides a comprehensive assessment, aiding in distinguishing 'normal,' 'benign,' and 'malignant' images. The consensus of all trees in the final decision-making process enhances reliability and accuracy, effectively addressing overfitting issues associated with decision trees.

[Criminisi et al., 2010] proposed a paper based on multi-class random regression forests as an algorithm for the efficient, automatic detection and localization of anatomical structures within three-dimensional CT scans. Quantitative validation is performed on a database of 100 highly variable CT scans. Localization errors are shown to be lower (and more stable) than those from global affine registration approaches. The regressor's parallelism and the simplicity of its context-rich visual features yield typical runtimes of only 1s. Applications include semantic visual navigation, image tagging for retrieval, and initializing organ-specific processing.

[Mirmohammadi et al., 2021] proposed multi-class random regression forests as an algorithm for the efficient, automatic detection and localization of anatomical structures within three-dimensional CT scans.A single pass of our probabilistic algorithm enables the direct mapping from voxels to organ location and size; with training focusing on maximizing the confidence of output predictions.

3.5 Deep Learning Methods

CNN and ViT. Since our focus is on performing a comprehensive comparison between traditional machine learning models and state-of-the-art deep learning models, we plan to employ Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) as the base architectures for the deep learning evaluation, as they are two most capable and pioneering architectures in the current field of computer vision. At their core, CNNs utilize convolution kernels to recognize patterns in images. These kernels operate with two primary insights: firstly, they detect features or patterns that are typically smaller than the complete input image, and secondly, they acknowledge that these features can emerge multiple times in various parts of the image. This design inherently understands the spatial hierarchies in visual data compared to traditional Multi-layer Perceptrons (MLPs), effectively recognizing intricate structures within images. In contrast, Vision Transformers (ViTs) dissect an image into fixed-size patches and then transform them into a series of vectors using a transformer encoder. This approach is rooted in harnessing the potent long-range modelling capability of the attention mechanism present in transformer architectures. By doing so, ViTs evade the inductive bias inherently introduced by CNNs' convolution design, and push back the frontiers of long-range dependency modelling in computer vision. In essence, while CNNs bank on localized pattern recognition, ViTs exploit global attention, emphasizing fine-grained relationships across the entire image. Both architectures, though conceptually different, highlight the diverse ways in which machines can be taught to 'see'.

Transfer Learning. Despite the proven capabilities of CNN and ViT in general vision tasks, it is not guaranteed that they will consistently show strong performance on medical image classification, mainly due to the severe training data scarcity of medical images. In order to tackle the data insufficiency problem, we plan to employ a new training paradigm called Transfer Learning (TL), which has been trending in various computer vision tasks in recent years. The core idea of transfer learning is to pretrain a vision model on a very large dataset (e.g. ImageNet [Deng et al., 2009], which contains 1.2 million images with 1000 categories), and then use the pretrained model either as an initialization or a fixed feature extractor for the task of interest. [Kim et al., 2022] conducted a comprehensive survey of transfer learning in the field of medical image classification, demonstrating its efficacy with deep convolutional neural networks (e.g. ResNet [He et al., 2016] and GoogLeNet [Szegedy et al., 2015]) as the pretrained backbone model. Moreover, with the advent of Vision Transformers, a large proportion of recent research also concentrates on the application of ViT under the transfer learning framework. [Matsoukas et al., 2022] studied the working mechanism of two data-efficient ViT models (DeiT [Touvron et al., 2021] and Swin Transformers [Liu et al., 2021]) in transfer learning for medical image classification, and found that the benefits from transfer learning increase with reduced data size and models with fewer inductive biases. However, their work did not focus on the extreme data-insufficient scenario like ours, and mainly concentrated on the domain discrepancy between the natural image domain (i.e. ImageNet that the backbone model was trained on) and the medical image domain.

References

Kimberly D Miller, Leticia Nogueira, Theresa Devasia, Angela B Mariotto, K Robin Yabroff, Ahmedin Jemal, Joan Kramer, and Rebecca L Siegel. Cancer treatment and survivorship statistics, 2022. *CA: a cancer journal for clinicians*, 72(5):409–436, 2022.

Gaël Varoquaux and Veronika Cheplygina. Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ digital medicine*, 5(1):48, 2022.

Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in brief*, 28:104863, 2020.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

Paidipati Dinesh and P Kalyanasundaram. 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. *ECS Transactions*, 107(1):12681, 2022.

Geoffrey I Webb, Eamonn Keogh, and Risto Miikkulainen. Naïve bayes. *Encyclopedia of machine learning*, 15(1):713–714, 2010.

Hein Tun Zaw, Noppadol Maneerat, and Khin Yadanar Win. Brain tumor detection based on naïve bayes classification. In *2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, pages 1–4, 2019. doi: 10.1109/ICEAST.2019.8802562.

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

Mingmin Chi, Rui Feng, and Lorenzo Bruzzone. Classification of hyperspectral remote-sensing data with primal svm for small-sized training dataset problem. *Advances in space research*, 41(11): 1793–1799, 2008.

Hicham Riri, A. Elmoutaouakkil, A. Beni-hssane, and Farid Bourezgui. Classification and recognition of dental images using a decisional tree. *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, pages 390–393, 2016. doi: 10.1109/CGIV.2016.82.

Helen Kaganov, A. Ades, and David S. Fraser. Preoperative magnetic resonance imaging diagnostic features of uterine leiomyosarcomas: A systematic review. *International Journal of Technology Assessment in Health Care*, 34:172 – 179, 2018. doi: 10.1017/S0266462318000168.

A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu. Regression forests for efficient anatomy detection and localization in ct studies. pages 106–117, 2010. doi: 10.1007/978-3-642-18421-5_11.

P. Mirmohammadi, Marjan Ameri, and Ahmad Shalbaf. Recognition of acute lymphoblastic leukemia and lymphocytes cell subtypes in microscopic images using random forest classifier. *Physical and Engineering Sciences in Medicine*, 44:433 – 441, 2021. doi: 10.1007/s13246-021-00993-5.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Hee E Kim, Alejandro Cosa-Linan, Nandhini Santhanam, Mahboubeh Jannesari, Mate E Maros, and Thomas Ganslandt. Transfer learning for medical image classification: a literature review. *BMC medical imaging*, 22(1):69, 2022.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015. URL `http://arxiv.org/abs/1409.4842`.

Christos Matsoukas, Johan Fredin Haslum, Moein Sorkhei, Magnus Söderberg, and Kevin Smith. What makes transfer learning work for medical images: Feature reuse & other factors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9225–9234, June 2022.

Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International conference on machine learning*, pages 10347–10357. PMLR, 2021.