

Comparative Evaluation of Vision-Based Machine Learning Models for Histopathological Breast Cancer Classification

Project Leaders: Russell Day, Andrew Zhou, Andre Gala-Garza

Introduction & Objectives

- Breast cancer diagnosis through histopathological image analysis is critical for early detection and treatment planning. While machine learning is commonly used in this domain, most existing approaches rely on text-based data or traditional model comparisons. This has left a notable gap in the application and evaluation of vision-based machine learning models specifically trained on histopathological images.
- Recent advancements in computer vision offer a promising opportunity to address this gap. Models such as Convolutional Neural Networks (CNN), Bag of Visual Words (BoW), and Vision Transformers (ViT) are capable of learning complex visual patterns directly from image data. These models can be trained to classify breast tissue samples from the BreakHis dataset as either benign or malignant, or to recognize more granular cancer subtypes.
- In trying to reach our objective of exceeding 90% classification accuracy, We explore whether learning finer-grained visual features can enhance a model's sensitivity and accuracy in a clinically relevant context.
- Hypothesis:** Training models to first classify cancer subtypes improves their ability to predict whether a tumor is benign or malignant compared to training directly on benign/malignant labels.

Materials & Methods

Dataset

The BreakHis (Breast Cancer Histopathological Database) public dataset, comprising 9,109 microscopic images of breast tumor tissues, was utilized for this study. The dataset includes 2,480 benign and 5,429 malignant images, classified into eight distinct histological types:

- Benign tumors: Adenosis (A), Fibroadenoma (F), Phyllodes Tumor (PT), Tubular Adenoma (TA)
- Malignant tumors: Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), Papillary Carcinoma (PC)

Data Collection Method

Samples in the BreakHis dataset were collected through surgical open biopsy (SOB), also known as partial mastectomy or excisional biopsy. This procedure involves removing larger tissue samples compared to needle biopsies and is performed under general anesthesia in a hospital setting.

Model Training

Models were trained exclusively on histopathological images using TensorFlow due to its efficiency and minimal hardware demands. Two training approaches were implemented:

- Binary Classification: Models trained to classify images strictly as malignant or benign.
- Cancer-Type Classification: Models trained to identify specific cancer subtypes, inherently categorizing each as either benign or malignant based on subtype labels.

Models were trained and a snapshot of the model was taken at its best epoch.

Training and Testing Protocol

An 80/20 training-testing data split was applied to ensure robust evaluation. Each model's performance was assessed separately on both the training and testing datasets.

AI Models Evaluated

Four vision-based machine learning architectures were selected for comparative evaluation:

- Convolutional Neural Network (CNN)
- Bag of Words (BoW)
- Vision Transformer (ViT)

All models were trained and evaluated using identical image-based inputs without additional metadata.

Figures

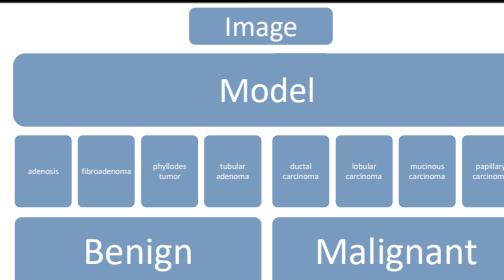


Figure 1. Training the model on the subtype of the cancer first then mapping to benign/malignant

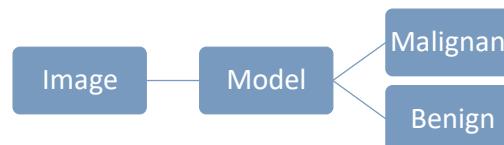


Figure 2. Training the model to recognize the image directly to malignant/benign

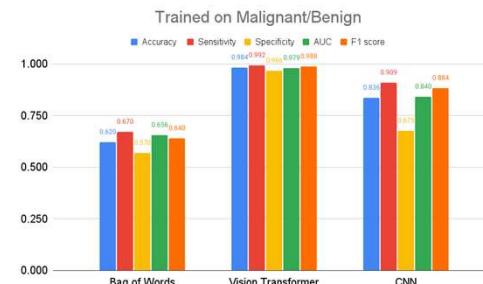


Table 1. Results of models trained on recognizing malignant/benign directly on test data

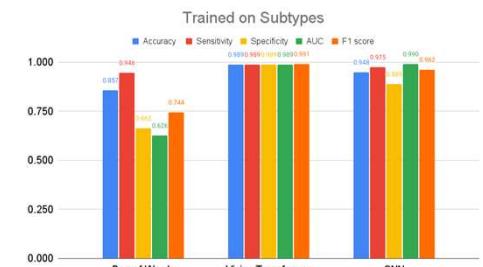


Table 2. Models trained to distinguish between subtypes of cancers and the results are based on if the models were able to correctly state if the image was malignant/benign on test data

Results

- Vision Transformer with subtype training achieved the highest overall performance**, with accuracy, sensitivity, specificity, and AUC all around **98.9–99.1%**, and an F1 score of **0.991**.

Subtype training consistently outperformed direct benign/malignant training across all models.

- BoW improved the most**, increasing its accuracy from **62.0%** to **85.7%** — a **+23.7% gain**.
- CNN also improved**, from **83.6%** to **94.8% accuracy**, with a higher sensitivity (from 90.9% to 97.5%).
- ViT gained 0.5% accuracy**, reinforcing its strong baseline performance.

Conclusions

- Our results confirm that training models on cancer subtypes before classifying tumors as benign or malignant improves overall diagnostic performance.
- All three models — Bag of Visual Words, CNN, and Vision Transformer — showed measurable gains in accuracy, sensitivity, and F1 score when subtype training was used.
- Importantly, the Vision Transformer model with subtype training achieved the highest performance across all metrics, validating it as a highly effective approach for histopathological image classification.

Recommendations

We recommend using a Vision Transformer (ViT) architecture for histopathological breast cancer classification tasks, particularly when early and accurate identification of malignancy is critical.

To maximize performance:

- Train the model first to distinguish cancer subtypes, allowing it to learn detailed morphological features.
- Then map subtype predictions to benign or malignant labels as a second step.

Future Research

- Looking for other ways to improve these models would be very impactful as it would speed up not only accuracy of diagnosis, but also reducing time needed to diagnosis malignance or if it is benign.
- Explore other medical datasets to validate generalizability of subtype-first training approach

Credits/References

- F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016, doi: 10.1109/TBME.2015.2496264.
- A. Heikal, A. El-Ghamry, S. Elmougy, and M. Z. Rashad, "Fine tuning deep learning models for breast tumor classification," *Sci. Rep.*, vol. 14, no. 1, Art. 10753, May 10, 2024, doi: 10.1038/s41598-024-60245-w.
- A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv preprint arXiv:2010.11929*, 2020.
- S. Gidaris, A. Bursuc, N. Komodakis, P. Pérez, and M. Cord, "Learning Representations by Predicting Bags of Visual Words," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 6926–6936, doi: 10.1109/CVPR42600.2020.00696.
- K. O’Shea and R. Nash, "An Introduction to Convolutional Neural Networks," *arXiv preprint arXiv:1511.08458*, 2015.