Cancer Detection

Introduction

Cancer is the second most common cause of death in the US, according to the American Cancer Society; specifically, there is a 1 in 8 chance for a woman in the US will develop breast cancer. But not all cancers cause death. Statics provided by the American Society of Clinical Oncology indicates that the average 5-year survival rate for women with non-metastatic invasive breast cancer is 90 percent. Early cancer detection can increase the probability of successful treatments and the survival rate. Due to the importance of early cancer detection, we need to improve the diagnoses with the least possible noise and bias and be faster, easier, and more publicly available. As a result, with the help of artificial intelligence and machine learning techniques, Computer Aided Diagnosis systems (CADs) can be developed.

There can be four reformulations to solve the cancer detection problem using the Breakhis dataset, according to Benhammou [1]: Magnification-Specific Binary (MSB), Magnification-Independent Binary (MIB), Magnification-Specific Multi-category (MSM) and Magnification-Independent Multi-category (MIM) classifications. We decided to focus on MIM reformulation for the following advantages, which Benhammou discusses. First, multi-category models provide more information which leads the expert to choose the most suitable treatment among various treatments and also take actions to prevent the risk of cancer development. Secondly, developing magnification-specific models means a model for each magnification which is costly and requires intense effort, while magnification-independent models seem a more efficient choice. In addition, magnificent-dependent models are trained with additional data and extra information from various features of histological tissue images, which increase the generalization capacity of the model and cause better results.

Last but not least, a general model independent of magnification and not limited to a specific setting practically is more likely to be used in industry. In fact, not all laboratories are equipped with microscopes with all magnification levels. Furthermore, magnificent independent models have a considerable advantage. All mentioned advantages of MIM reformulation motivated us to investigate this topic and develop a model with higher accuracy than our baseline [1] and state-of-the-art [2].

Breakhis [3] is a publicly available breast cancer database of 82 patients in various magnification scales with eight labels half of them belong to Malignant, and the other half belongs to Benign classes. This comprehensive dataset feeds our model and compares the results with the other researchers. The proposed model includes three main modules: a pre-trained transfer model, which returns all features of a given image input, a feature selection phase, and a classifier.

This report is organized as follows: Dataset and data distribution in detail, Metrics of evaluation of the model, Background and previous must-known information and details of the selected pre-trained model and feature selector, Architecture of the

model including feature representations, classifier selection, and hyperparameters, Results of the comparison with the other researches and finally a visualization of feature embedding vectors after a dimension reduction.

Database

BreaKHis [3] is a public dataset that contains 7909 microscopic biopsy images, with a three-channel RGB belonging to two classes, Benign (B) and Malignant (M). In addition, each class includes four subclasses. Benign breast tumors can be divided into adenosis(A), fibroadenoma (F), phyllodes tumor (PT), tubular adenoma (TA), and malignant tumors are separated into ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC) subclasses. Figure 1 shows examples of images belonging to each binary category or eight subcategories. The histological tissue images are gathered from a clinical study for almost a year, from January 2014 to December 2014, from 82 patients. Also, images can be grouped by magnification level. Figure 2 shows each example of four magnification levels, 40X, 100X, 200X, and 400X.

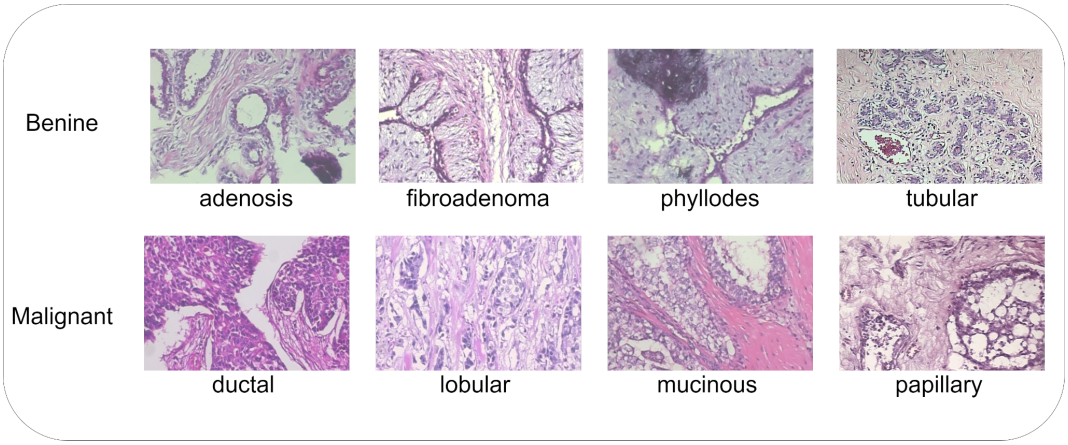


Figure 1: The microscopic biopsy images are divided into two main classes and eight subclasses.

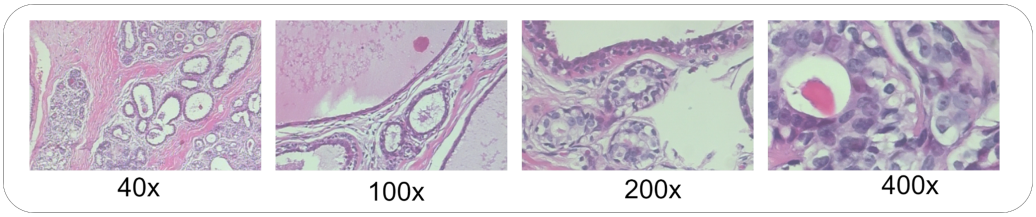


Figure 2: The microscopic biopsy images and four magnification levels.

Figure 3 shows the distribution of data for both binary classes and multi-category classes. The images are not uniformly distributed in two classes or eight subclasses; therefore, An experiment was done to reveal the impact of under-sampling on classification accuracy, but there was not an impressive improvement in classification accuracy, so we decided to continue with the same original distribution of data.

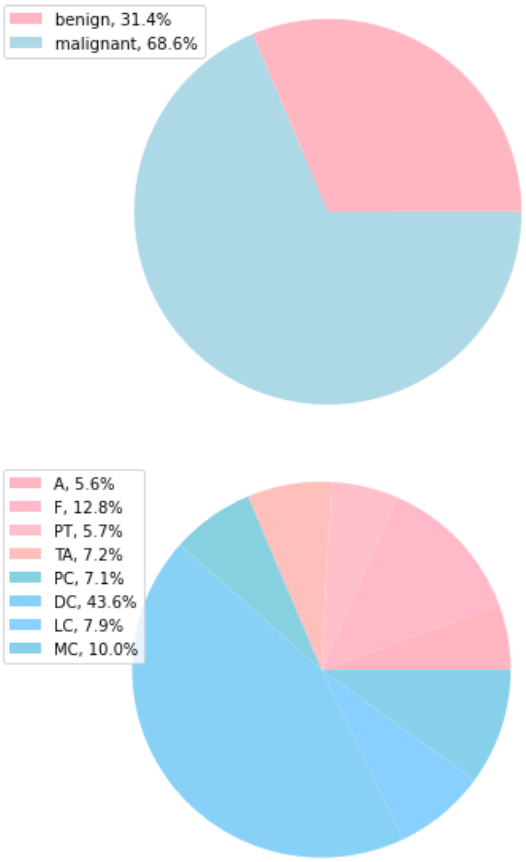


Figure 3: The data distribution in each class for both benign and malignant classes and eight subcategories (A/F/PT/TA/PC/DC/LC/MC).

Metrics

The baseline introduces two evaluation metrics for a fair comparison, image-level accuracy(ILA) and patient-level accuracy(PLA) or recognition rate. The image level accuracy does not consider any information about the patient and focuses on the number of correctly classified images per total image. Besides, patient-level accuracy calculates an individual score for each patient, which describes the accuracy of predictions per patient. The average of all the scores is considered patient-level accuracy.

Background

The model architecture consists of three modules the encoder, feature selector, and classifier. Figure 4 shows how the modules are well-connected and classify each image input. In the following section, we go through the details of each model module.

First, an encoder is needed to encode an image to an embedding vector representing the image details. Transform-based modes are selected for this module. Transformer-based models have achieved as high accuracy as Cnn-based models in image processing; indeed, depending on the task, they even can perform better. The vision transformer is a transformer-based model that relies on self-attention over patches of images. The input images go through a feature extractor that rescales and normalizes them, then split into a fixed-size non-overlapping sequence. Then, the linear embeddings of the patches are calculated and fed to the encoder.

The Vision Transformer (ViT) model, trained on Image-net21k, is the base model used for feature extraction. The pre-trained used model is google/vit-base-patch16-224-in21k, which is trained on 14 million images and 21,843 classes where every single input is a sequence of fixed-size patches (16x16). Then the model returns an encoding with the shape of (1, 197, 768) for a given image, representing the image features in detail.

Secondly, a feature selection module should be selected because it can reduce the computational complexity and cost; Indeed, a lower number of features chosen can quickly achieve higher accuracy in the classification task because noise and unnecessary features are removed.

The ANOVA (Analysis of Variance) method, a type of F-statistics, is chosen for feature selection. It is suitable for either categorical target variables or numerical data input. According to the effect of each feature on the distance between classes and compactness of classes, a score for each feature is calculated, and then just the top K features with the highest score are selected. Some experiments led us to a proper k-value selection which is explained in detail in the following section.

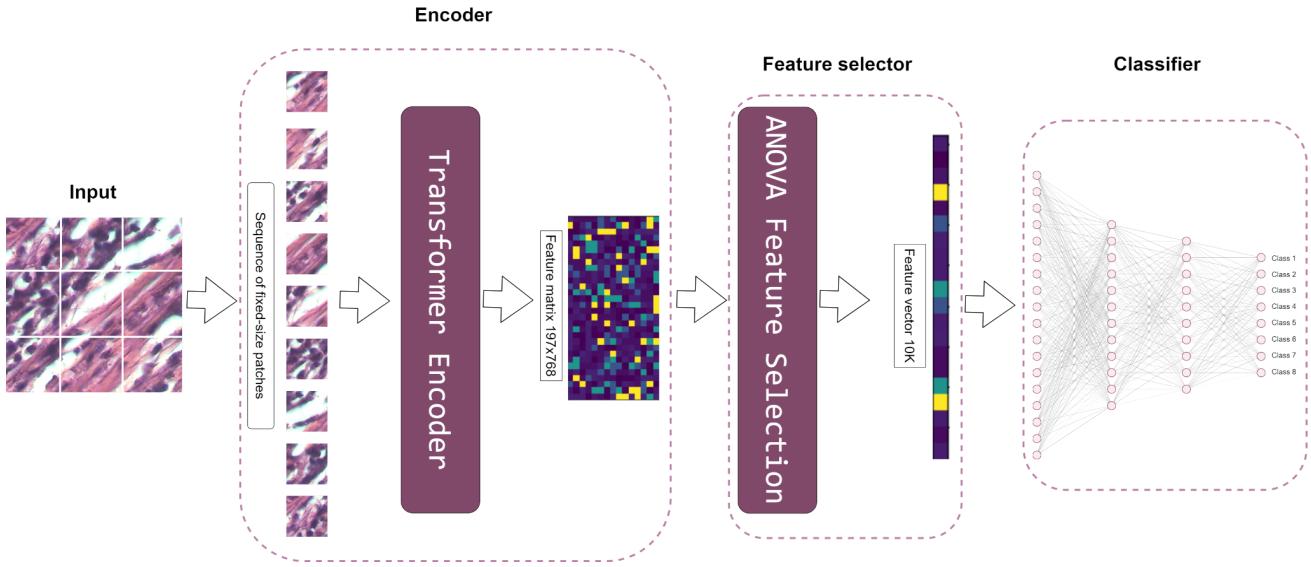


Figure 4: The visualization of the proposed architecture for the MIM task

Architecture

Feature representations

The Breakhis dataset contains just about 7k histological tissue images; in fact, lack of data is a common issue in medical image processing. Therefore, the transfer learning technique is a tricky choice in this problem. The selected model is Vision Transformer (ViT), which returns an encoding with the shape of (1, 197, 768) for a given input representing the input features. This high-dimension output embedding includes excellent image details, while the task is just classifying each image into eight classes. Thus, a dimension reduction on extracted features can significantly reduce the computational complexity. A feature selection technique is needed for simplifying extracted features for more accurate classification and lower computational costs. The ANOVA method is the feature selector. According to recent experiments, a K of 10K is a reasonable choice for the number of selected features.

A list of experiments, which include different K values followed by different classifiers, is done to assign a proper value to K. Figure 5 demonstrates that a higher k value results in higher accuracy in the classification, whether the classifier is an SVM, logistic regression, or deep neural network. In search of an appropriate amount for k, two factors are considered, computational capacity and information convention to be used in the following classifier. Therefore, a value of 10K is selected for K.

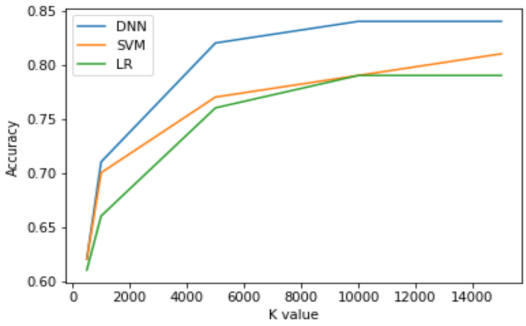


Figure 5: The accuracy of classifiers trained on a range of K values

Classification

The third section of the proposed architecture is a classifier. The type and hyperparameters of the classifiers are the two main issues to be selected and tuned. In the first step of a better classifier selection, a comparison has been made, which shows how deep neural networks (DNN) perform much better than the other classifiers: a support vector machine (SVM), logistic regression, K nearest neighbor, extreme gradient boosting (XGB), decision tree, and gaussian naive base classifiers. Figure 6 is a chart bar that illustrates a sorted list of classifiers according to their accuracy in the MIM task.

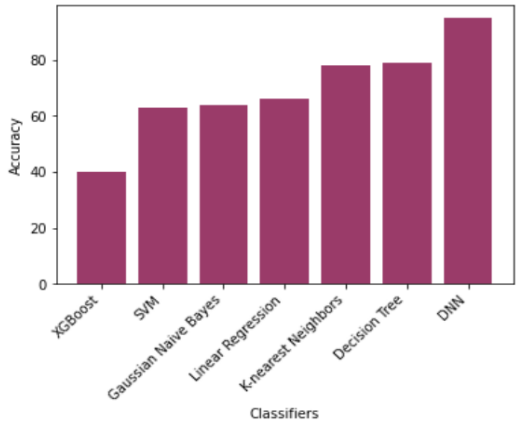


Figure 6: Sorted accuracy values of different classifiers.

Then in the second step, the hyperparameters of the DNN should be tuned. The designed model is a fully connected neural network with about 10 million parameters. To be more specific, seven dense layers create the DNN. Also, three dropout layers and a gaussian nose layer are added to the model for higher generalization and to avoid overfitting. The Keras tuner optimization framework, which manages the hyperparameters search process, is used to find the optimum number of nodes per layer. The details of the architecture of the model are illustrated in figure 7.

Figure 7: The architecture of the designed classifier

The proposed model is significantly low-cost and highly efficient computationally. The low number of selected features and a light classifier have achieved high accuracy on a system with 16GB of RAM and a GPU with 4GB of memory. Further configuration of the hardware is provided in table 1.

OS	Windows 11, 64-bit
CPU	Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz, 8 core
GPU	Nvidia GeForce GTX 1050 with Max-Q design 4GB
RAM	16GB

Table 1: Hardware configuration.

Results

In this section, the results of the proposed model are compared to both the baseline and the state-of-the-art. The comparison frameworks of the two related works are different, so two different experiments are done. Specifically, the two experiments differ in train and test split portion, data augmentation, and normalization.

First of all, As the MIM reformulation is discussed by Benhammou [1], the results of the proposed model are compared to it as the baseline model; furthermore, the raw dataset, without data augmentation or normalization, is fed to the model.

The model has trained five trials with different folds, which are split with 70-30 distribution each time with a batch size of 16. The comparison between our work and the baseline model is provided in table 2, which provides the mean value and standard derivation of classification accuracy at image level (ILA) over five trials on MIM reformulation. For a better comparison, the ILA is reported separately for the two M/B classes.

	Malignant	Benign
Baseline	57.4 ± 2.0	47.6 ± 1.6
Our work	**84.6 ± 2.3**	**80.2 ± 1.3**

Table 2: The comparison of the proposed model and the baseline for MIM reformulation reported by the accuracy

Table 2 demonstrates that the proposed transformer-based architecture achieves significantly higher results than the fine-tuned CNN model.

Secondly, the results are compared to this the state of the art model [2]. Accordingly, the dataset is randomly split into 20-80 portions with a batch size of 8. Then as data augmentation, the training data is doubled with its vertically flipped value. Also, images are normalized by torchstain, a stain normalization tool that is designed for histological analysis and computational pathology. Figure 8 gives detailed results of classifying histological tissue images into eight classes, including a classification report and confusion matrix.

	precision	recall	f1-score	support
DC	0.90	0.97	0.94	685
F	0.92	0.89	0.90	190
C	0.93	0.88	0.91	162
LC	0.99	0.79	0.88	106
TA	0.94	0.92	0.93	127
PC	0.98	0.88	0.93	115
PT	0.91	0.90	0.91	101
A	0.93	0.91	0.92	96
accuracy			0.92	1582
macro avg	0.94	0.89	0.91	1582
weighted avg	0.92	0.92	0.92	1582

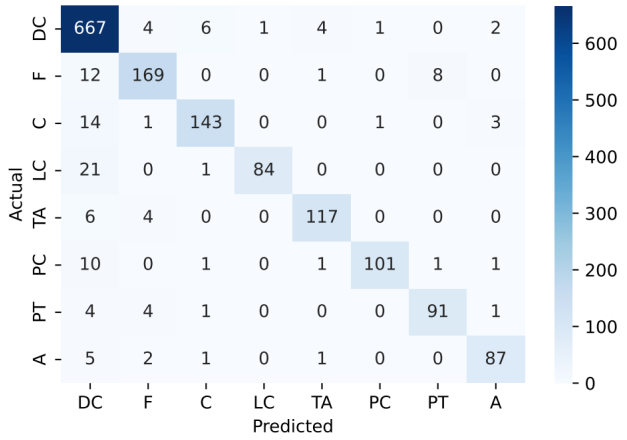


Figure 8: The detailed result of the second experiment reported per class and the confusion matrix

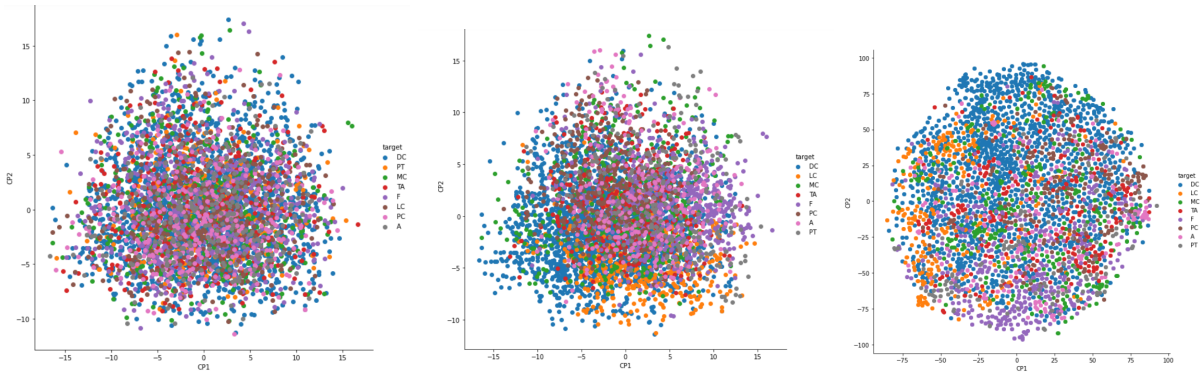
Table 3 demonstrates how the proposed model improves the state-of-the-art model in MIM reformulation. In addition, another reformulation of histological tissue image classification is MIB which classifies images into benign or malignant classes independent of their magnification factor. For this binary classification, just the last layer of the proposed model is modified. The final results of this task are included in table 3, which indicates the accuracy of the state-of-the-art and proposed model in MIB reformulation are tightly close to each other.

	Classification	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
State of the art	MIB	**98.42**	**98.75**	**99.01**	**98.88**
	MIM	92.03	91.39	90.28	90.77
Our work	MIB	97.47	97.59	98.68	97.46
	MIM	**92.22**	**92.43**	**92.22**	**92.16**

Table 3: The comparison of the proposed model and the state-of-the-art for both MIM and MIB reformulation reported by accuracy, precision, recall, and f-measure.

Visualization

A dimension reduction and visualization are done to compare the distribution of the classes in both histological tissue images and feature embeddings with a shape of 10k.



a b c

Figure 9: Visualizing images and feature embeddings after dimension reduction.
a) 2-dimensional PCA applied on histological tissue images. b) 2-dimensional PCA applied on feature embeddings. c) 2-dimensional TSN-E applied on feature embeddings.

First, 5k random images with the shape (460, 700, 3) are selected, then the dimension of the images is reduced from 966000 to 2 using PCA. The below diagram shows the distribution of the dimension-redacted images per class. Secondly, 5k random embedding vectors, which represent the key features of the images with the shape of 10k, are selected, and their dimension is reduced to 2 by PCA. Third, the TSN-E technique is used as another method for dimension reduction on the same feature embeddings. The below diagram shows how classifying the 10k embedding vectors is easier than the original images, even after dimension reduction. Figure 10 indicates that feature embeddings (b and c) are better distributed per class than histological tissue images (a). Therefore, feature embeddings convey prominent main features of images with the slightest noise, while noise and irrelevant features in original images negatively affect the classification.

References

1. Yassir Benhammou, Boujemâa Achchab, Francisco Herrera, Siham Tabik, “BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights”, Neurocomputing, Volume 375,2020, Pages 9-24
2. Said Boumaraf, Xiabi Liu, Zhongshu Zheng, Xiaohong Ma, Chokri Ferkous, “A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images”, Biomedical Signal Processing and Control, Volume 63, 2021,102192
3. F. A. Spanhol, L. S. Oliveira, C. Petitjean and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," in IEEE Transactions on Biomedical Engineering, vol. 63, no. 7, pp. 1455-1462, July 2016, doi: 10.1109/TBME.2015.2496264.