Utilization of Machine Learning and Deep Learning Techniques for Image Classification in BreakHis

Ahmed Tamer, Pınar Yılmaz

***Abstract*—Breast cancer is one of the leading causes of cancer death in women worldwide. Although histopathological image analysis is the diagnostic gold standard, they are time-consuming and prone to subjectivity. The aim of this study is to accelerate breast cancer diagnosis by evaluating and comparing the performance of machine learning and deep learning models in classifying the BreaKHis database. The database contains a wide range of histopathological images taken at different magnification levels. Deep learning architectures were evaluated including InceptionResNetV2,ResNet152V2, InceptionV3,VGG16, and Xception. Then, two different feature extraction methods were applied to the segmented images and their performances were evaluated via machine learning models such as SVM, Random Forest, K-NN and XGBoost. The accuracy rate varies between 80% and 90%, indicating that there is room for improvement. The results showed that InceptionResNetV2 outperformed all other models, achieving 90% accuracy at 200x magnification. Among the ML models, XGBoost with PFTAS achieved a competitive accuracy rate of 87%, outperforming all other ML configurations. These findings highlight the importance of deep learning in interpreting histopathological images and suggest that traditional machine learning approaches may also be useful when fine-tuned. The study contributes to the advancement of computer-aided diagnostic systems by accelerating diagnosis.**

***Index Terms*—Breast cancer, histopathology, image classification, ML, DL.**

I. INTRODUCTION

CANCER is one of the major health problems of the 21st century. According to the International Agency for Research on Cancer of the World Health Organization,cancer is one of the leading causes of death worldwide, causing approximately 10 million deaths in 2020. The study shows that breast cancer in particular, is the second most common cancer in terms of incidence with more than 2.2 million new cases. It also ranks fourth in terms of mortality, causing more than 666,000 deaths, according to the study conducted by WHO [1]. Early detection significantly increases survival rates.

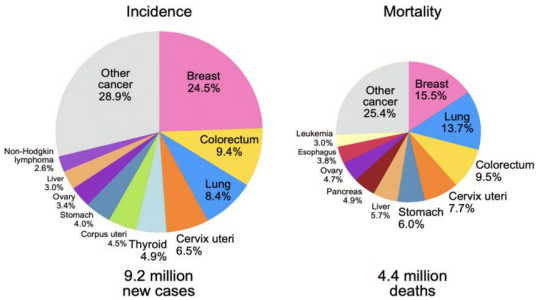

Fig. 1. Distribution of global cancer incidence and mortality,

Currently, techniques such as digital mammography and MRI are widely used in the diagnosis of breast cancer. Digital Mammography detects early tumors, However, its efficiency decreases in dense breasts. As such, MRI provides detailed screening for high-risk patients, such as those with a strong family history of breast cancer, but is expensive and has higher rate of false positives, leading to unnecessary biopsies. While these techniques are important for early detection, a more definitive diagnosis of breast cancer is essential. This is where histopathology becomes important. Among biopsy techniques, the most common are fine needle aspiration and surgical open biopsy (SOB)[2]. It involves looking at a small sample of tissue under a microscope to see the shape of cells, growth pattern and the state of spread to the surrounding tissue in these images, which helps doctors decide the best treatment. In benign tumors, the cells are uniform and the boundaries are obvious. On the other hand, malignant tumors are characterized by irregular cells, enlargement of nuclei, excessive cell division, and tumor infiltration into surrounding tissue (Elston and Ellis, 1991). Furthermore, manual inspection of histopathology images is time-consuming and requires a lot of experience. This is exactly why computer-aided diagnosis (CAD) and machine learning (ML) systems have advanced significantly in cancer detection in recent years. In particular, tumor classification is carried out using convolutional neural networks (CNNs) on histopathological images. Rather than replacing doctors, ML aims to perform standard tasks to ease their work and reduce variability in diagnosis. However, the development and validation of these systems critically depend on access to large, publicly available, and well-organized datasets—which, unfortunately, are still hard to come by for researchers. Veta et al. [3] highlight that the primary obstacle to the development of new histopathology image analysis methods is the absence of large, publicly available, annotated datasets.Without enough data, it's hard to make sure these systems work well in real-world medical settings. That's why the development of imaging techniques in medicine has greatly aided medical diagnosis because they give researchers a reliable way to develop and compare different methods. One good example is the BreakHis dataset. It's a large collection of histopathology images specifically focused on breast cancer, and it has become a valuable resource for training and evaluating classification algorithms.

In conclusion, the main purpose of this paper is to examine the efficiency and identify the best-suited model of ML and DL in the early detection of breast cancer using the BreaKHis

dataset. Our ML pipeline begins with image preprocessing through clustering-based and threshold-based segmentation, followed by the extraction of morphological and texture features. Unlike many previous studies that analyze raw RGB histology images, we employ fuzzy c-means (FCM) clustering to enhance tissue structure representation. These segmented images serve as the basis for feature extraction, which is then used to train a gradient boosting classifier (XGBoost) with stratified K-fold cross-validation. Additionally, we trained four DL models with various fine-tuning hyperparameters to identify the best-performing architecture. Both ML and DL approaches were used to classify tumors as benign or malignant and to distinguish between the four subtypes within each category.

II. BREAKHIS DATASET

BreakHis (Breast Cancer Histopathologic Database) is a publicly available database containing microscopic images of breast tumor tissue collected from 82 patients. The images taken at 40X, 100X, 200X an and 400X magnifications, as shown in Figure 1. The database contains total of 7,909 images in three-channel RGB format with 8-bit depth and 700 × 460 pixels in each channel. The database was collected within the scope of a clinical study conducted from January to December 2014 in collaboration with the P&D Laboratory and the Pathological Anatomy and Cytopathology Laboratory in Paraná, Brazil.

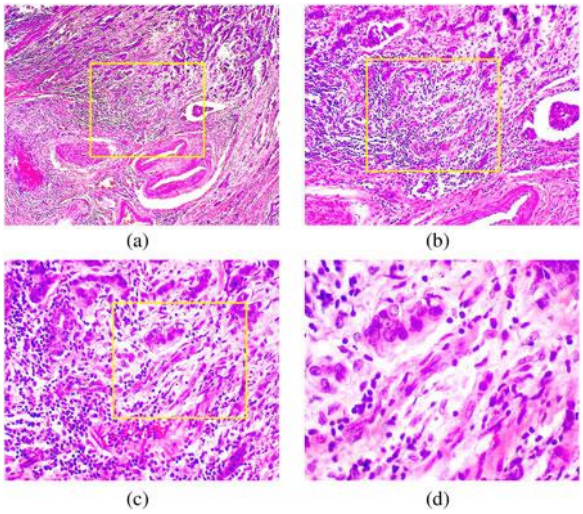

Fig. 2. Slide of breast malignant tumor (stained with HE) seen in different magnification factors: (a) 40X,(b) 100X,(c) 200X, and (d) 400X . Highlighted rectangle (manually added for illustrative purposes only) is the area of interest selected by pathologist to be detailed in the next higher magnification factor.

The data was collected using the SOB method. This procedure removes a larger tissue sample than other methods of needle biopsy and is performed in a hospital under general anesthesia. The two main classes of the BreakHis dataset are benign and malignant tumors. Benign tumors lack malignant features, such as metastasis, are slow-growing, and remain localized. In contrast, malignant tumors, or cancers, are invasive, destroying adjacent tissues and potentially metastasizing to distant sites. Both benign and malignant tumors are distributed into four different subcategories based on the tumor's appearance under a microscope.

The database contains four benign tumor types: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA). Malignant: Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), and Papillary Carcinoma (PC). The dataset contains 2,480 benign samples and 5,429 malignant samples at all magnifications. Each image is encoded with important information such as file name, biopsy processing method, patient ID, tumor classification (benign/malignant), specific tumor type, and magnification factor. A detailed summary of the distribution of images and patients across these classes and subcategories is presented in Table 1.

TABLE I
DISTRIBUTION OF BREAKHIS DATABASE

Classes	Sub-classes	Number of samples per Magnification factors					
		40X	100X	200X	400X	Total	# Patients
Benign	A	114	113	111	106	444	4
	F	253	260	264	237	1014	10
	TA	109	121	108	115	453	3
	PT	149	150	140	130	569	7
	total	**625**	**644**	**623**	**588**	**2480**	**24**
Malign	DC	864	903	896	788	3451	38
	LC	156	170	163	137	626	5
	MC	205	222	196	169	792	9
	PC	145	142	135	138	560	6
	total	**1370**	**1437**	**1390**	**1232**	**5429**	**58**
Total		**1995**	**2081**	**2013**	**1820**	**7909**	**82**

III. SEGMENTATION

With a comprehensive dataset like this, it is important to be able to differentiate cancerous and noncancerous areas from other normal tissues. To identify areas as benign or malignant tumors, segmentation is the key. The segmentation process for the BreakHis dataset was designed to isolate regions of interest (ROIs) while preserving morphological and textural features critical for classification.In this section, we briefly describe the two segmentation we have used to train the classifiers. One of them is the fuzzy C-means (FCM) method and the other is Otsu morphological operation with PFTAS.

A. Fuzzy C-Means(FCM)

Fuzzy C-Means (FCM) is an unsupervised clustering technique that assigns membership values to each pixel, allowing it to belong to multiple clusters with varying degrees of membership. This method is particularly suitable for histopathological image analysis due to the inherent ambiguity in tissue structures.Each image was reshaped into a feature matrix based on RGB pixel values, and FCM was executed with key parameters: the number of clusters set to 5, the fuzziness coefficient set to 1.8 to allow some separation between clusters, and a maximum of 100 iterations. By applying the FCM algorithm, the images were separated into distinct clusters representing different tissue regions.

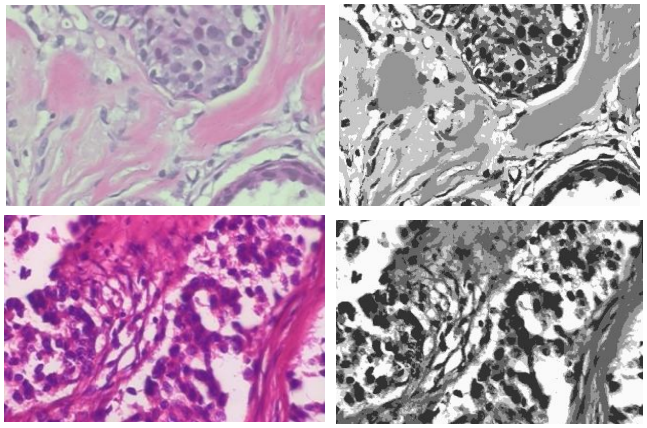

Fig. 3. Examples of original and segmented images. The top row shows a benign histopathological image alongside its segmented version, while the bottom row presents a malignant image with its corresponding segmentation. These segmented outputs highlight regions of interest used for feature extraction.

B. Otsu's Threshold

Also, Otsu's threshold technique was used for segmentation. This method finds the ideal threshold value from the grayscale histogram of the image by reducing the intra-class variance. Therefore, it successfully separates the foreground (texture) from the background. In this implementation, Otsu's method was used to create binary masks that highlight important texture areas in each RGB channel after processing each channel separately. These segmented masks were used to improve feature extraction. By focusing on these segmented tissue regions, the method minimizes the influence of background noise and improves the ability of the classifier to distinguish between benign and malignant samples.

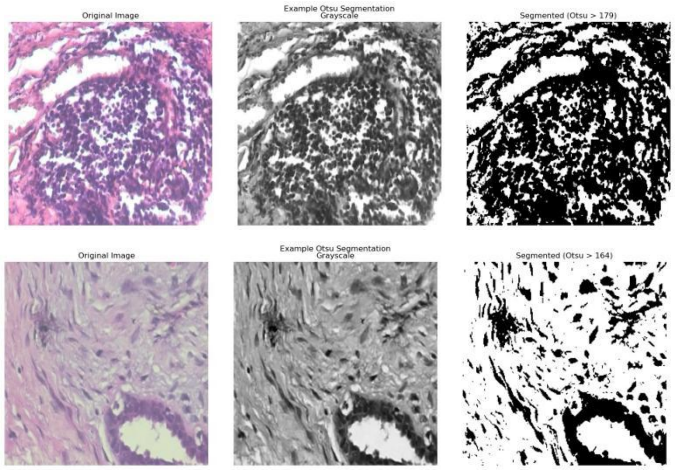

Fig. 4. Examples of original and segmented images. The top row shows a benign histopathological image alongside its segmented version and the wanted mask. the bottom row presents a malignant image with its corresponding segmentation and the wanted mask.

II. FEATURE EXTRACTORS

Following the segmentation process, the next critical step is feature extraction. The segmented images are converted into numerical features that can be used for classification. Different feature extractors were used according to different segmentation methods. For the basic segmentation with Otsu threshold, a technique known as Parameter-Free Threshold Adjacency Statistics (PFTAS) was applied. In contrast, for the segmented image via FCM, radiomic features were extracted. Radiomic features capture the essential patterns, textures, shapes, and intensity distributions in medical images. By aligning specific feature extraction methods with each segmentation technique, it enhances the classifier's ability to distinguish between cancerous and noncancerous tissues.

A. Parameter-Free Threshold Adjacency Statistics (PFTAS)

PFTAS is a texture-based method that captures local spatial structure without predefined parameters. Basically, two PFTAS variants were employed as 'Internal' and 'External'. Internal PFTAS features were calculated using adaptive thresholds derived from the Otsu method on each RGB channel separately, producing 162 features per image. To enhanced local texture analysis by segmentation-based features incorporated which external PFTAS. Specifically, we applied Otsu's threshold to the grayscale image to generate tissue/background masks. These masks allowed us to calculate how often different colors appeared next to each other, resulting in 324 new features (3 color channels × 3 masks × 9 bins × 2).We extracted 486 features from each image, combining color-based texture with segmentation-guided spatial information.

B. Radiomic Features

Radiomic features are a set of quantitative descriptors that capture the underlying patterns, textures, shapes, and intensity distributions within medical images. These features are widely used in biomedical image analysis because they can reveal subtle tissue characteristics that may not be visible to the human eye. In this study, radiomic-inspired features were utilized through the Parameter-Free Threshold Adjacency Statistics (PFTAS) method, which focuses on capturing texture information without the need for predefined thresholds. PFTAS features fall under the category of texture-based radiomic features, as they describe how pixel intensities are spatially distributed and related to their neighbors across different regions of the image. Additionally, segmentation-based extraction helped isolate relevant tissue areas, allowing for more targeted and meaningful radiomic feature computation. By leveraging these features, the classification model gains deeper insight into the heterogeneity and structural differences between benign and malignant histopathological images. 94 feature were extracted and the best 20 of them used for classification.

III. CLASSIFIERS

Two main approach is employed to classification of BreakHis database, which are machine learning based and deep learning based approaches. Four different machine learning based classifiers were used to evaluate the above-mentioned feature sets: Support Vector Machines (SVM), XGBoost, and Random Forests (RF), and K-NN.

A. Machine Learning Based Classifiers

Support Vector Machines (SVM) are supervised machine learning algorithms that are primarily used for classification tasks. SVM works by determining the optimal boundary called hyperplane that divides two classes by the largest amount. XGBoost is a powerful method that builds a series of decision trees, where each new tree corrects the mistakes of the previous ones, leading to high accuracy. Random Forest is another tree-based method that creates many decision trees using different parts of the data and combines their results, which helps improve reliability and reduce overfitting.

B. Deep Learning Based Classifiers

CNNs are designed to automatically learn and extract spatial patterns and textures from images through layer-by-layer convolution, pooling, and nonlinear activation. This makes them particularly effective for image analysis tasks such as histopathological classification. Unlike machine learning models that rely on hand-extracted features, such as PFTAS, CNNs learn features directly from raw image pixels during training. However, they require huge computing power, making their training demanding on equipment. That is why we tried both ML and DL to observe if the required heavy computing power actually makes a difference and pays off. The study used five deep learning architectures, including VGG19, ResNet-152v2, InceptionV3, Inception-ResNetV2, and Xception. These models were pre-trained on the large-scale ImageNet dataset and fine-tuned for histopathological image classification. The framework offers robust feature extraction, high classification accuracy, and efficient learning on smaller, domain-specific datasets.

IV. RESULT AND DISCUSSION

A. Process

In this study, histopathological breast tissue images from the BreaKHis database at four different magnifications were used. BreaKHis database is divided into training (70%) and testing (30%) sets at each magnification level according to the CSV file that is provided in the database. This approach prevents data leakage and produces a more realistic estimate of model performance in clinical settings. We used Support Vector Machines (SVM), Random Forests (RF), and Extreme Gradient Boosting (XGBoost) models for their success in previous histopathological imaging research. The discussed ML and DL models were applied to each magnification level separately. We used Otsu's threshold technique and FCM on each RGB channel of the images for segmentation in order to improve the input data. By capturing local texture patterns from both the original and segmented regions, the Parameter-Free Threshold Adjacency Statistics (PFTAS) method guided feature extraction using the segmented masks. The combined feature set was then standardized and used to train machine learning classifiers.

B. Results

We evaluated various traditional machine learning and deep learning models at different magnification levels of the BreaKHis dataset. Although the masks of segmented images may appear visually similar across methods, the resulting accuracies vary significantly due to differences in the discriminative power of the feature extraction techniques used.

The performance of all machine learning models is presented in Table VI. Among the ML methods using PFTAS features, XGBoost achieved the highest performance, with an accuracy of 0.87 at 200x magnification. Across magnification levels, models generally performed best at 200x, indicating that this level captures the most informative features for classification. Similarly, K-NN and RF achieved notable accuracies of 0.86 and 0.83 respectively at this magnification, reinforcing the effectiveness of PFTAS based texture descriptors.

In contrast, ML models using radiomics with FCM segmentation showed significantly lower performance. For example, XGBoost reached only 0.32 at 200x, which highlights the variability in feature quality depending on the extraction method. This stark difference demonstrates how the choice of feature extraction directly influences classification performance.

Deep learning model results are also summarized in Table VI. Overall, CNN architectures outperformed traditional ML models, especially at higher magnifications. InceptionResNetV2 achieved the best result, with 0.90 accuracy at 200x magnification, along with high precision (0.87), recall (0.88), and F1-score (0.89). Other models such as InceptionV3 and Xception also performed strongly, with accuracies of 0.87 and 0.84 at 200x, respectively.

These results confirm that 200x magnification consistently provides the most discriminative information across both ML and DL models. As such, deeper evaluations were focused on this magnification level. The findings highlight the importance of selecting both an effective magnification level and a compatible feature extraction strategy or model architecture to achieve optimal classification performance.

C. Discussion

As can be understood from the results, there is distinct difference in performance based on the feature extraction technique. ML models trained with PFTAS texture features performed significantly better than models trained with radiomics features and FCM segmentation. For example, when using PFTAS, XGBoost obtained an accuracy of 0.87 at 200x magnification, whereas when using radiomics with FCM, the same model only obtained an accuracy of 0.32. This difference indicates that, at least for the purposes of this investigation, texture-based descriptors—like PFTAS—are better to radiomics-based methods in terms of identifying discriminative features from histopathological images.

TABLE II

MACHINE LEARNING METHODS PERFORMANCE BY MAGNIFICATION FACTOR

Otsu's threshold + PFTAS	Magnification Factor			
	40X	100X	200X	400X
SVM	0.79	0.78	0.8	0.79
RF	0.8	0.81	0.83	0.81
K-NN	0.8	0.79	**0.86**	0.83
XGBoost	0.82	0.81	**0.87**	0.85

TABLE III

DEEP LEARNING METHODS PERFORMANCE BY MAGNIFICATION FACTOR

FCM + Radiomic	Magnification Factor			
	40X	100X	200X	400X
SVM	0.63	0.71	0.70	0.69
RF	**0.76**	0.65	0.54	0.69
K-NN	0.71	0.69	0.58	0.68
XGBoost	0.69	0.58	0.32	0.71

TABLE IV

DEEP LEARNING METHODS PERFORMANCE AT 200X

DL Methods	Magnification Factor			
	40X	100X	200X	400X
VGG16	0.59	0.61	0.64	0.63
ResNet152V2	0.78	0.75	0.8	0.78
InceptionV3	0.85	0.84	**0.87**	0.86
InceptionResNetV2	0.86	0.86	**0.9**	0.89
Xception	0.81	0.8	**0.84**	0.83

V. CONCLUSION

In this study, we investigated and compared machine learning (ML) and deep learning (DL) methods to classify histopathological breast tissue images from the BreaKHis database. Two different segmentation methods were evaluated. Depending on the segmentation approach used, relevant feature extraction techniques were applied to the segmented images. Result of segmented and feature-selected data tested through machine learning models. There is room for improvement but we achieved very good results especially for malignant recognition. In parallel, we evaluated advanced Convolutional Neural Network (CNN) architectures on the raw data. The best classification performance was always achieved with 200x magnification in all models. 200x magnification provides a balance between fine-grained cellular features and larger tissue structure.We want to send an important message with this study. Carefully constructed, interpretable machine learning pipelines, when combined with advanced deep learning models, can be useful for robustly and accurately detecting breast cancer because they speed up the manual process. Our findings suggest that, especially in cases with limited data or computational resources, traditional classifiers can compete with deep models by combining preprocessing with various feature extractions. In conclusion, the proposed system can help speed up the diagnosis process and reduce the workload of pathologists by combining the strengths of both ML and DL approaches.

REFERENCES

[1] World - IARC (no date) International Agency for Research on Cance. Available at: https://gco.iarc.fr/today/data/factsheets/populations/900-world-fact-sheets.pdf (Accessed: 28 August 2023).

[2] Breastcancer.org. (2012) Biopsy. [Online]. Available: http://www.breastcancer.org/symptoms/testing/types/biopsy

[3] Veta, M., Pluim, J. P., van Diest, P. J., & Viergever, M. A. (2014). Breast cancer histopathology image analysis: a review. IEEE transactions on biomedical engineering, 61(5), 1400–1411. https://doi.org/10.1109/TBME.2014.2303852