

Artificial Intelligence Models for the Detection of Breast Cancer in Medical Imaging

Ece Akaydın¹, Ayşe Rüya Arslan² and Gülay Çiçek³

^{1,2}Department of Software Engineering, ³Department of Biomedical Engineering
Istanbul Beykent University, Faculty of Engineering and Architecture, Sariyer, Istanbul, Turkey

¹2103013295@student.beykent.edu.tr, ²2203013050@student.beykent.edu.tr

³gulaycicek@beykent.edu.tr

Abstract—This study compares the performance of deep learning (ResNet50, VGG16) and traditional machine learning (Random Forest, SVM) algorithms in breast cancer diagnosis. Convolutional neural networks (CNN) were used for feature extraction, and the impact of Principal Component Analysis (PCA) and dimensionality reduction on model performance was analyzed. In experiments on the BreakHis dataset, the effect of PCA on AUC values was evaluated, and comparisons were made between models. The results show that the ResNet50 model using PCA achieved the highest performance, reaching 99.21% AUC. While PCA improved performance in some models, it negatively impacted performance in others. This study is unique in that it compares deep learning and traditional machine learning methods under the same dataset and experimental conditions, with and without PCA. This study allows for the integration of the obtained results not only into academic but also clinical decision support systems. The comprehensive analysis provides valuable findings for both researchers and healthcare professionals...

Keywords—Breast Cancer Diagnosis, Deep Learning, Machine Learning, PCA, BreakHis Dataset

I. INTRODUCTION

Breast cancer is the most common malignant tumor in women, accounting for approximately 30% of all female cancer cases. Early diagnosis is critical to treatment success. Imaging methods such as mammography, ultrasound, and MRI are widely used in the diagnostic process; however, due to limitations such as cost, access, and margin of error, these methods can be inadequate in some cases. Manual evaluations by pathologists are also time-consuming and error-prone. In this context, the use of artificial intelligence, and particularly deep learning methods, has great potential to increase accuracy and speed up the process of breast cancer diagnosis. Image processing techniques and CNN-based structures can detect abnormal tissues at an early stage. Significant artificial intelligence and deep learning-based research on breast cancer conducted in recent years is summarized. The reviewed studies were evaluated in terms of the datasets used, methods, results obtained, shortcomings, and future contributions. Dembrower et al. (2023) reported that artificial intelligence was not as effective as a second radiologist in a study of 55,581 data points, indicating that the technology is still developing. [1]. Ibrahim et al. (2020) reported that the CNN model they developed on mammography images obtained from different

hospitals demonstrated higher AUC performance than doctors; however, the lack of dataset size limits the generalizability of the study. [2]. Alom et al. (2019) reported accuracies of 97.95% and 99.05% using the IRRCNN model with the Kaggle and BreakHis datasets, respectively; however, the study did not include machine learning methods, which is limiting in terms of methodological diversity [3].

The unique contributions of this study to the literature are as follows:

- While many studies in the literature focus solely on deep learning or machine learning methods, this study considers both approaches together and systematically examines the effects of PCA implementation/non-implementation on both classical algorithms and deep network architectures.
- By evaluating the performance of different model groups on breast cancer images under the same dataset and conditions, the study provides a fair and comparable analysis. In this respect, it serves as an important reference for model selection in the field of medical image processing.
- Designing the findings in a way that can be integrated not only into academic settings but also into clinical applications adds practical value to the study and increases its applicability for decision support systems.

II. Method

A. Dataset and Preprocessing

This study used the BreakHis dataset, obtained from the Kaggle platform, which contains histopathological images for breast cancer detection. [4]. The dataset comprises a total of 7,909 images collected from 82 patients, with 2,480 benign and 5,429 malignant samples, highlighting a class imbalance that could potentially affect model training. Images were captured at magnification levels of 40X, 100X, 200X, and 400X, each with a resolution of 700×460 pixels and a 3-channel RGB format. During preprocessing, the accuracy of the labels was checked and corrected if necessary to ensure data quality. The dataset was then divided using a 5-fold stratified cross-validation strategy, allocating approximately 70% of the data for training and 30% for testing in each fold. This systematic preprocessing and splitting approach ensures that both training and evaluation phases are robust and reproducible.



B. Feature Extraction

In this study, Convolutional Neural Networks (CNN) will be used for feature extraction. In CNN, image pixels are processed with convolutional layers to extract features.

C. Feature Selection

Features extracted from visual data are often high-dimensional, which can increase model complexity and negatively impact the learning process. Therefore, dimensionality reduction techniques were employed to improve classification performance and reduce the risk of overfitting. In this study, PCA was used to retain only the most informative features, providing more efficient input to the classification models.

D. Model Selection

- **Machine Learning Algorithms:** Machine Learning is a process that utilizes mathematical models to enable a computer to learn without requiring explicit instructions. In recent years, machine learning methods have become frequently preferred because they have significantly increased classification accuracy. Two machine learning algorithms, Random Forest and Support Vector Machines, were implemented in this study.
 - **Random Forest:** A Random Forest is an algorithm consisting of multiple decision trees. It trains each decision tree with different pieces of data. This way, each tree provides a different perspective, and the overall accuracy of the model increases [5]
 - **Support Vector Machines (SVM):** This algorithm is generally used for data classification. It is used to classify both linearly separable and non-linearly separable data [6].
- **Deep Learning Architectures:** Deep learning is a more advanced version of artificial neural networks. It is used specifically for classification, recognition, and detection [7]. The fundamental architecture of Deep Learning is considered the CNN architecture.
 - **ResNet:** ResNet is a specialized CNN architecture consisting only of convolutional layers. Essentially, it's a structure that allows layer blocks to be skipped via shortcuts. Unlike previous architectures, it's a deeper architecture designed with 152 layers. It also won the 2015 ImageNet competition with a 3.6% error rate.
 - **VGGNet:** VGG is a CNN architecture developed for image processing on large datasets. It consists of a structure consisting of 11, 16, and 19 layers, called VGG-11, VGG-16, and VGG-19, respectively.
 - **Convolutional Neural Networks (CNNs)** were chosen for model selection due to their success, particularly in image processing and classification. CNNs automatically extract features from images, reducing the

need for manual feature engineering. Furthermore, when used in conjunction with techniques such as data augmentation, they reduce the risk of overfitting and increase the model's generalization ability. This increases the model's accuracy and reliability.

E. Model Training

Dropout was applied after fully connected layers in model training to prevent overfitting. [8]. Training continued for 50 epochs, and hyperparameters such as learning rate and batch size were determined to increase model performance [9]. Model performance was evaluated using metrics such as accuracy, sensitivity, specificity, precision, and F-score. [10]. These metrics are calculated based on the TP, TN, FP, and FN values in the complexity matrix. Additionally, the model's classification performance is visualized with the ROC curve, and the overall success is measured with the AUC (area under the curve). The AUC value reflects the model's ability to distinguish between positive and negative classes.

III. Experimental Studies

This section presents experimental studies conducted with different deep learning and machine learning models for the classification of breast cancer images. PCA was applied to the obtained features for dimensionality reduction, and their classification performance was evaluated. All models were tested on the BreakHis dataset and analyzed using metrics such as accuracy, precision, sensitivity, and F-measure.

Confusion Matrix: It is one of the main metrics used to evaluate the performance of classification algorithms. This matrix, whose rows represent the classes predicted by the model and whose columns represent the true classes, numerically represents the model's correct and incorrect predictions. Each cell represents the number of examples at the intersection of a given class prediction and the true class.

A. PCA On - Off Applications and Metrics

This section evaluates the performance impact of applying PCA to different classification models. Table I presents a comparison of the accuracy, precision, sensitivity, and F-measure values obtained for Support Vector Machines (SVM), Random Forest (RF), ResNet50, and VGG16 models before (OFF) and after (ON) PCA. This examines the contribution of dimensionality reduction to the classification performance of deep learning-based feature extractors, along with traditional machine learning algorithms. The graph below provides a comparative graph of the Machine and Deep Learning models. Table I presents a class-by-class comparison of the performance of four different models before and after PCA application in terms of metrics such as accuracy, precision, sensitivity, and F1 score. This table allows us to numerically observe the impact of PCA on each model,

and clearly demonstrates which models experience performance gains or losses.

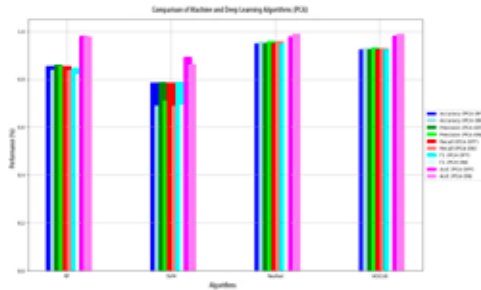


Fig. 1: Comparison of Machine and Deep Learning Models

When the results presented in Table I are evaluated in general, the performance differences between the models before and after PCA can be summarized as follows: Analyzing the obtained results, the accuracy rate for the VGG16 model before PCA was applied was calculated as 92.58%. The area under the ROC curve (AUC) value for this structure was determined as 0.9693. With PCA applied, the accuracy rate increased to 93.16%, and the AUC value reached 0.9781. Small improvements were also observed in metrics such as sensitivity, specificity, and F1-measure. The accuracy rate achieved without PCA in the Random Forest (RF) algorithm was 85.57%. In this case, the AUC value was measured as 0.9315. However, when PCA was applied, the accuracy dropped to 83.99%, and the AUC value was recorded as 0.9307. The reason for this negative effect is that the RF algorithm creates decision trees based on raw features. When PCA loses some discriminative features, the model loses some of the distinction between classes and performance declines. Similar decreases were observed in metrics such as macro-average and F1-score. These data reveal that PCA has a negative impact on the RF model in terms of both overall accuracy and discrimination power. The ResNet model achieved an accuracy rate of 94.98% without PCA, and the AUC value for this structure was calculated as 0.9889, a very high level. With PCA applied, the accuracy increased to 95.85%, and the AUC value increased even further, reaching 0.9921. The reason for this decline is that PCA loses some of the discriminative information from the high-dimensional CNN features. In particular, the fine details that distinguish malignant and benign cells are not sufficiently represented after dimensionality reduction, causing the model performance to decrease somewhat. Significant improvements were also observed in the precision, recall, and F1-score metrics. These results show that PCA provides a significant performance increase over the ResNet architecture, and that the model can distinguish the positive class with higher accuracy. The accuracy rate of the SVM algorithm before PCA was applied was determined to be 78.77%. The AUC value for this structure was 0.8354. However, when PCA was applied, the accuracy rate decreased to 69.04%, and

the AUC value decreased to 0.7274. Significant decreases were also observed in the precision, recall, and F1-score metrics. This indicates that PCA significantly negatively impacted the classification success of the SVM model. This negative effect can be explained by the fact that SVM is highly sensitive to the distribution and scaling of features. PCA reduces dimensionality by selecting principal components that explain the most variance, but some features that are discriminative for SVM classification may be discarded during this process. Overall, while PCA increased the classification success and discrimination ability (AUC) in the ResNet and VGG16 models, it decreased the performance in terms of both accuracy and area under the ROC curve in the Random Forest and SVM algorithms. However, the sensitivity (recall) values for the malignant class were generally high in all models, indicating that the models were generally successful in detecting malignant cases.

Figure 2 compares the average AUC values of four different models (VGG16, SVM, ResNet50, and Random Forest) without (PCA OFF) and after (PCA ON). The results show that PCA provides small improvements in AUC values for some models (e.g., VGG16 and ResNet50), but performance decreases for the SVM model.

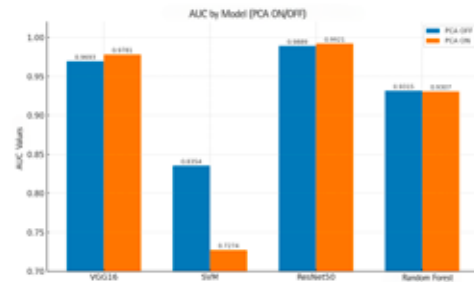


Fig. 2: Comparison of Machine and Deep Learning Models

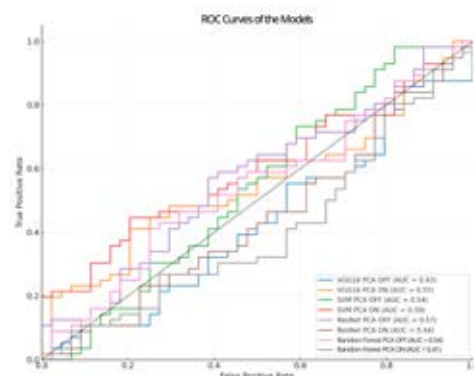


Fig. 3: Comparison of Machine and Deep Learning Models

Figure 3 presents the ROC curves obtained for all models with PCA ON and PCA OFF. The ROC curves visually demonstrate the classification performance of each model at different threshold values and are supported by AUC

TABLE I: Experimental Results of Classification Algorithms (PCA)

Algorithm	PCA	Accuracy	Precision		Recall		F-Skor		AUC
			Class 0	Class 1	Class 0	Class 1	Class 0	Class 1	
VGG16	OFF	0.93	0.90	0.94	0.86	0.96	0.88	0.95	0.9693
	ON	0.93	0.90	0.94	0.88	0.96	0.89	0.95	0.9781
RF	OFF	0.86	0.89	0.85	0.62	0.97	0.73	0.90	0.9315
	ON	0.84	0.94	0.82	0.52	0.98	0.67	0.89	0.9307
ResNet	OFF	0.95	0.92	0.96	0.92	0.97	0.92	0.96	0.9889
	ON	0.96	0.95	0.96	0.92	0.98	0.93	0.97	0.9921
SVM	OFF	0.79	0.66	0.85	0.67	0.84	0.66	0.84	0.8354
	ON	0.69	0.51	0.81	0.62	0.72	0.56	0.76	0.7274

scores. These graphs provide a visual comparison of the models' accuracy-performance balance.

IV. Discussion

The effectiveness of Principal Component Analysis (PCA) in machine learning varies significantly based on the model. In traditional models like Random Forest and SVM, applying PCA generally led to a decrease in performance. For Random Forest, accuracy dropped from 85.57% to 83.99%, while the drop was more substantial for SVM (from 78.77% to 69.04%), suggesting PCA can be detrimental to some tree-based and non-linear models. Conversely, PCA had a positive effect on deep learning models. For VGG16, accuracy saw a slight increase from 92.58% to 93.16%, and the improvement was more pronounced with ResNet, where accuracy rose from 94.98% to 95.85%, along with significant gains in other metrics. This indicates that PCA can be a beneficial preprocessing step in deep learning architectures, especially with complex, high-dimensional data. Ultimately, these findings show that dimensionality reduction techniques like PCA are not universally effective; their success is directly tied to the compatibility between the specific algorithm and the preprocessing technique used.

V. CONCLUSION

This study systematically evaluated the effect of Principal Component Analysis (PCA) on both traditional machine learning and deep learning models for breast cancer detection. The findings reveal that PCA's impact is model-dependent: it improved performance in deep learning architectures, achieving the highest accuracy of 95.85% with ResNet and also boosting VGG16. Conversely, PCA resulted in a decrease in metrics for traditional algorithms like Random Forest and SVM. This disparity underscores that dimensionality reduction techniques are not universally effective. The results clearly demonstrate that PCA can be particularly beneficial for deep learning models, highlighting the significant potential of AI-based systems for accelerating and improving the accuracy of early breast cancer diagnosis. The joint, systematic analysis of both model families under the influence of PCA

constitutes a unique contribution to medical image processing research. Future work should focus on enhancing model generalization using larger, balanced datasets and exploring next-generation Transformer-based approaches.

REFERENCES

- [1] Karin Dembrower, Alessio Crippa, Eugenia Colón, Martin Eklund, Fredrik Strand, and ScreenTrustCAD Trial Consortium. Artificial intelligence for breast cancer detection in screening mammography in sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit. Health*, 5(10):e703–e711, October 2023.
- [2] Asmaa Ibrahim, Paul Gamble, Ronnachai Jaroensri, Mohammed M Abdelsamea, Craig H Mermel, Po-Hsuan Cameron Chen, and Emad A Rakha. Artificial intelligence in digital breast pathology: Techniques and applications. *Breast*, 49:267–273, February 2020.
- [3] Md Zahangir Alom, Chris Yakopcic, Mst Shamima Nasrin, Tarek M Taha, and Vijayan K Asari. Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. *J. Digit. Imaging*, 32(4):605–617, August 2019.
- [4] BreakHis Dataset. BreakHis – Breast Cancer Histopathological Image Classification. <https://www.kaggle.com/datasets/ambarrish/breakhis/data>, 2025. Access date: August 8, 2025.
- [5] Özlem Akar and Oğuz Güngör. Classification of multispectral images using random forest algorithm. *Journal of Geodesy and Geoinformation*, (106):139–146, 2012. [in Turkish].
- [6] Özlem Akar and Esra Tunç Görmüş. Land use maps produced from göktürk-2 and hyperion eo-1 satellite images using random forest classifier and support vector machines. *Geomatics*, 4(1):68–81, 2019. [in Turkish].
- [7] Ferdi Doğan and İbrahim Türkoğlu. Comparison of deep learning algorithms' success in leaf classification. *Sakarya University Journal of Computer and Information Sciences*, 1(1):10–21, 2018. [in Turkish].
- [8] Numan Halit Güldemir and Ahmet Alkan. Classification of optical coherence tomography images using deep learning. *Firat University Journal of Engineering Sciences*, 33(2):607–615, 2021. [in Turkish].
- [9] Murat Karakoyun and Mehmet Hacıbeyoğlu. Statistical comparison of machine learning classification algorithms with biomedical datasets. *Dokuz Eylül University Faculty of Engineering Journal of Science and Engineering*, 16(48):30–42, 2014. [in Turkish].
- [10] Gülay Gürgen and Soydan Serttaş. Performance comparison of machine learning algorithms for early diagnosis of heart failure. *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences*, 10(28):165–174, 2023. [in Turkish].