
PREDICTION OF BREAST CANCER USING HISTOPATHOLOGY IMAGES

P6-01

K V L BHAVYA - HU21CSEN0101141

CH HARIKA - HU21CSEN0101163

K PRANATHI - HU21CSEN0100954

K SURYA TEJA - HU21CSEN0101927

UNDER THE GUIDANCE OF: DR. FIGLU MOHANTY



PROBLEM STATEMENT

Breast cancer is a leading cause of cancer-related deaths among women, making early and accurate detection crucial for improving survival rates. Histopathological image analysis is a standard diagnostic method, but is time-consuming and prone to human errors, affecting the reliability. This project aims to develop an automated system to classify breast cancer histopathological images into benign and malignant categories using the BreakHis dataset.

The system combines deep learning models like CNN and pretrained VGG16 for feature extraction, with machine learning classifiers such as Random Forest (RF), XGBoost, and K-Nearest Neighbors (KNN) for classification. Principal Component Analysis (PCA) is applied to reduce dimensionality, and metaheuristic JAYA Algorithm is used to optimize hyperparameters. This approach aims to provide a reliable decision-support tool for early breast cancer diagnosis, improving accuracy and reducing diagnostic subjectivity.

Dataset Collection

Collected the BreakHis dataset containing histopathology images of benign and malignant breast tumors



Feature Extraction

Extracted deep features using a custom CNN and the pre-trained VGG16 model

Optimization

Employed the JAYA algorithm for feature selection and hyperparameter tuning to enhance classifier performance

Image Preprocessing

Resized, normalized, and applied data augmentation techniques like rotation and flipping to enhance model generalization

Feature Reduction

Applied PCA to reduce dimensionality while preserving essential information for classification

Classification

Classified tumors using Random Forest (RF), K-Nearest Neighbors (KNN), and XGBoost (XGB) classifiers

ARCHITECTURE OF CNN

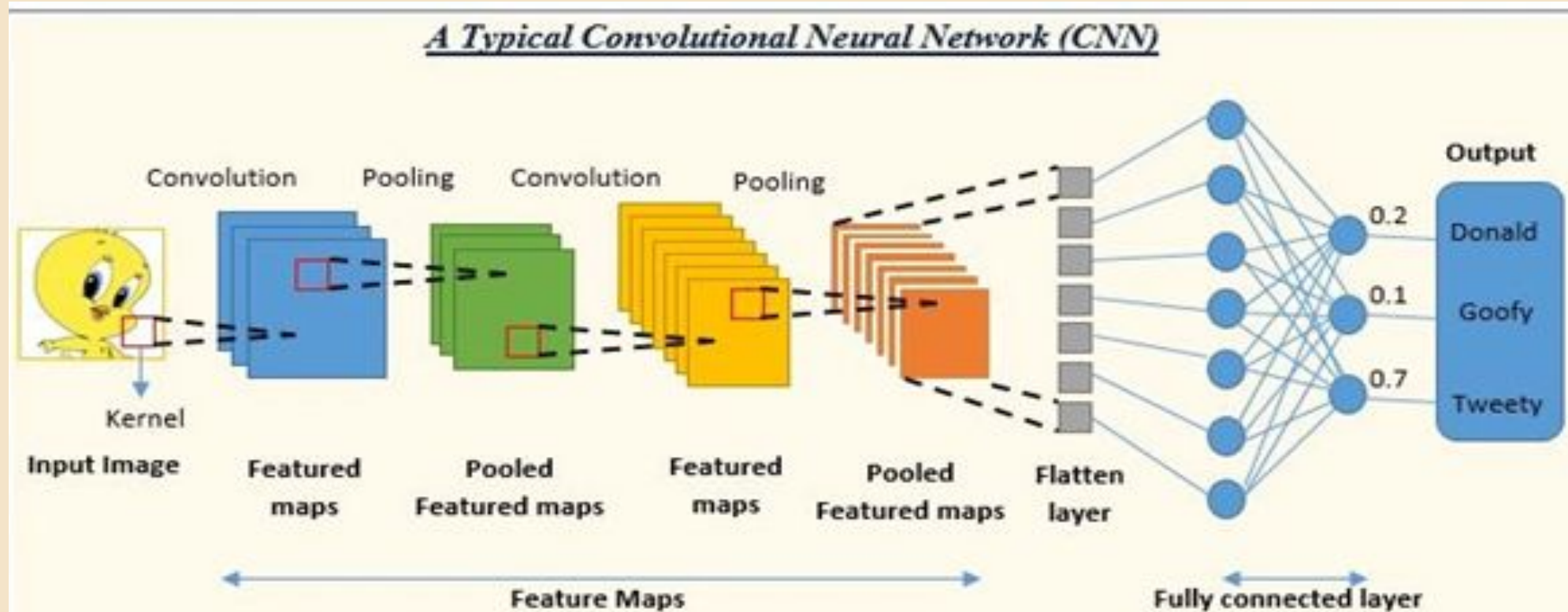


Fig. 1: CNN Architecture

ARCHITECTURE OF VGG16

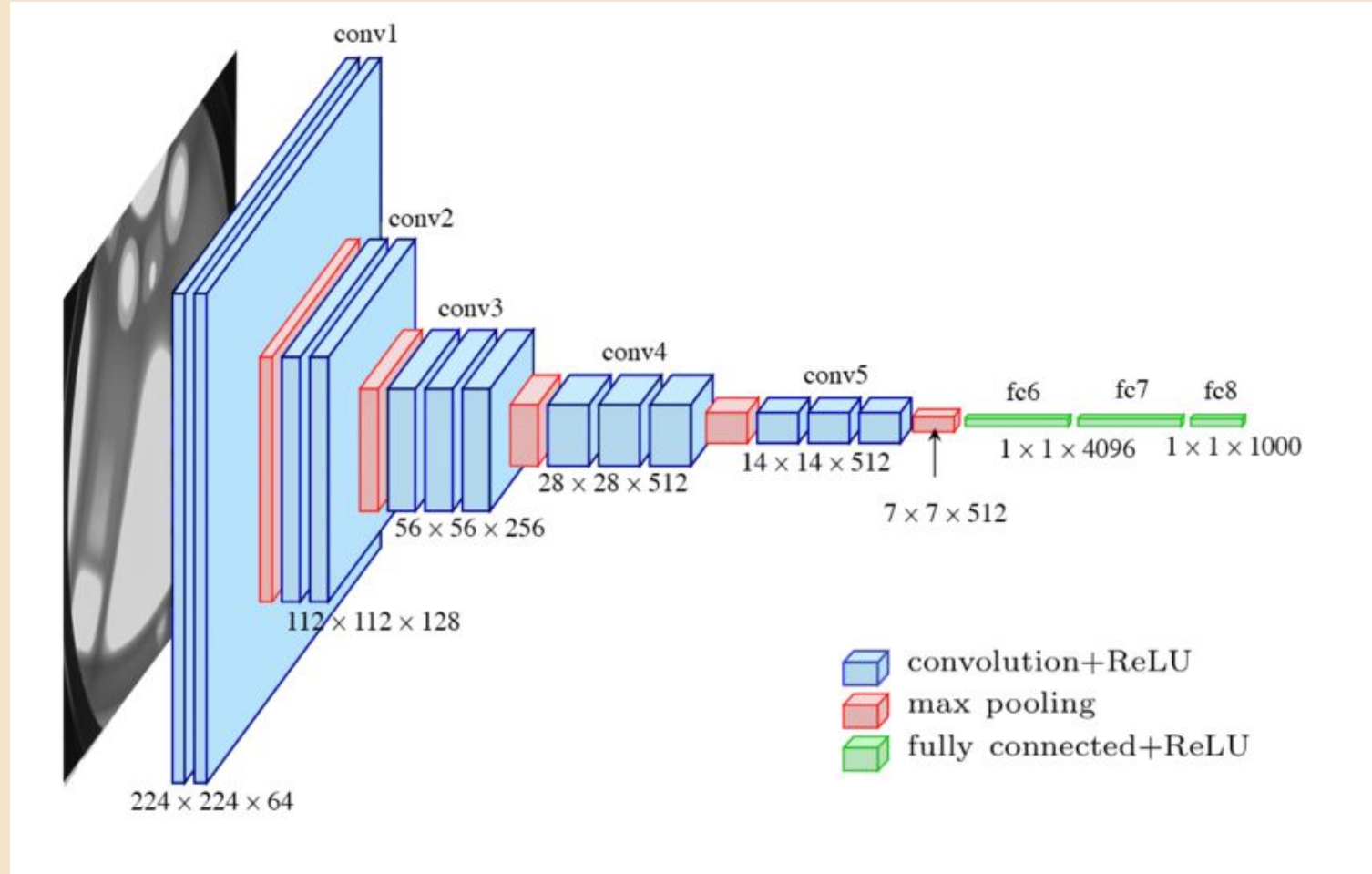


Fig. 2: VGG16 Architecture

PRINCIPAL COMPONENT ANALYSIS

- PCA is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space while preserving most of the variance.
- It identifies directions (principal components) with the most variation in data and projects data onto these directions.
- PCA removes redundant and less-informative features, reducing computational complexity and noise.
- By focusing on essential features, PCA prevents overfitting and improves model performance.
- In this project, PCA significantly reduced feature dimensions, leading to efficient processing and improved accuracy for certain model combinations.

CLASSIFIERS

- **Random Forest (RF):**

An ensemble method that combines multiple decision trees for robust predictions. Reduces overfitting and handles high-dimensional data efficiently. Applied with CNN and VGG16 features for tumor classification.

- **Extreme Gradient Boost (XGB):**

An efficient and scalable gradient boosting algorithm. Works well with structured data and is known for handling imbalanced datasets effectively. Used in this project to classify extracted features from histopathological images.

- **K-Nearest Neighbors (KNN):**

A simple instance-based learning algorithm. It classifies new data based on the majority class of its nearest neighbors. Though simple, it can work well when combined with good feature extraction like CNN.

RESULTS TABLE

MODEL	ACCURACY(%)
CNN+RF+PCA	85.9
CNN+XGB+PCA	86.83
CNN+KNN+PCA	85.61
VGG16+RF+PCA	68.62
VGG16+XGB+PCA	81.32
VGG16+KNN+PCA	72.1

Table 1: Results Table

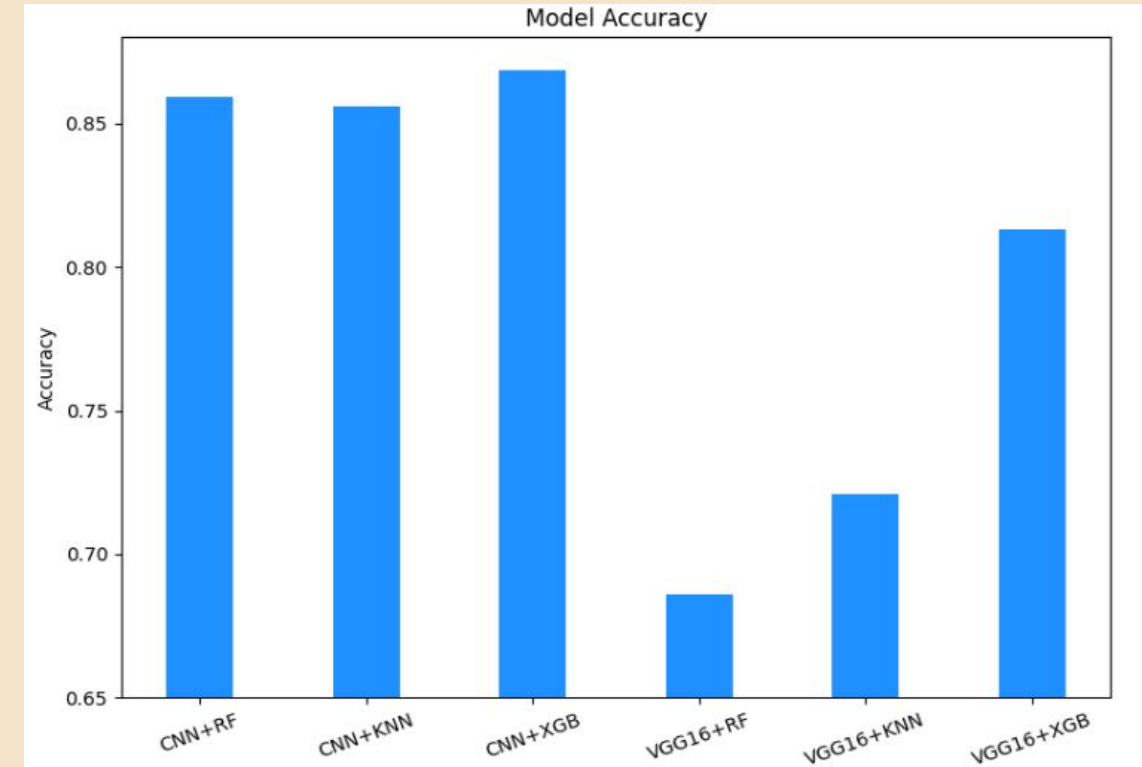


Fig. 3: Model Accuracy

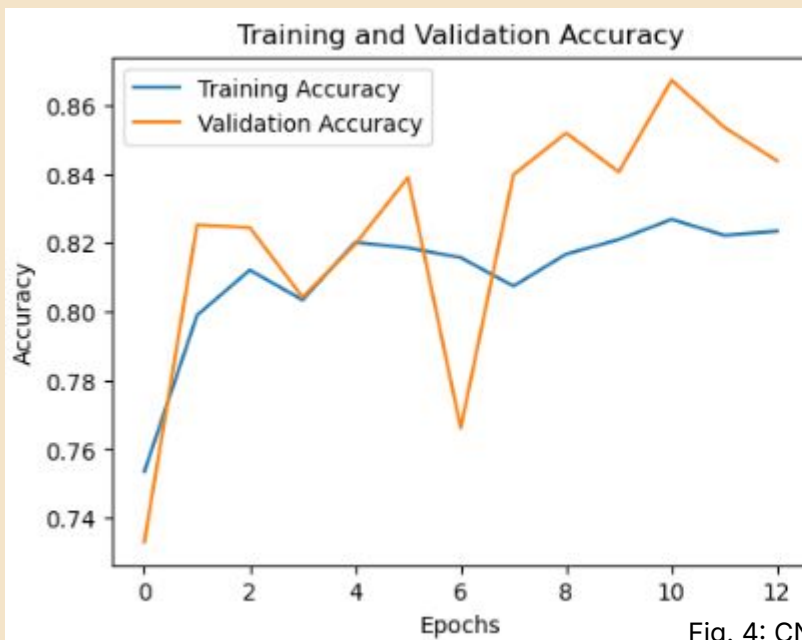
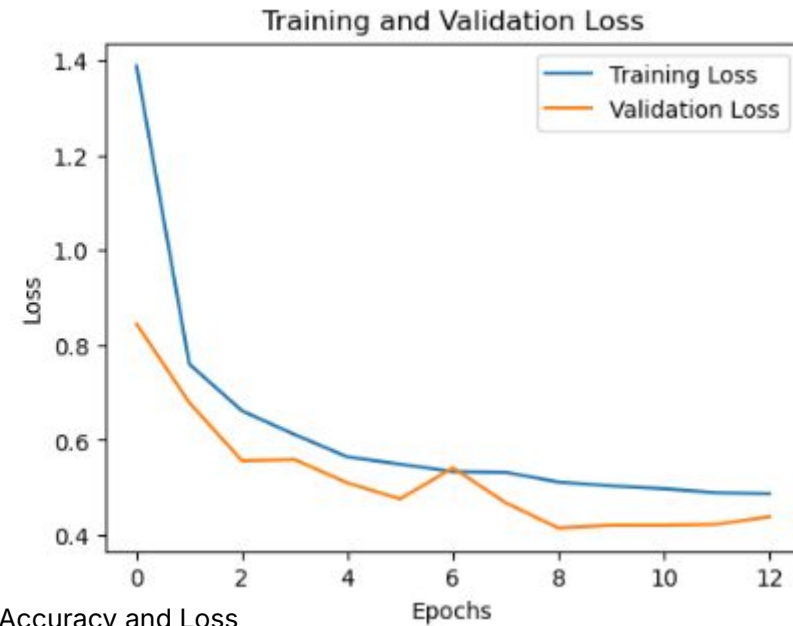


Fig. 4: CNN Accuracy and Loss



CNN Accuracy: 84.51%

VGG16 Accuracy: 78.24%

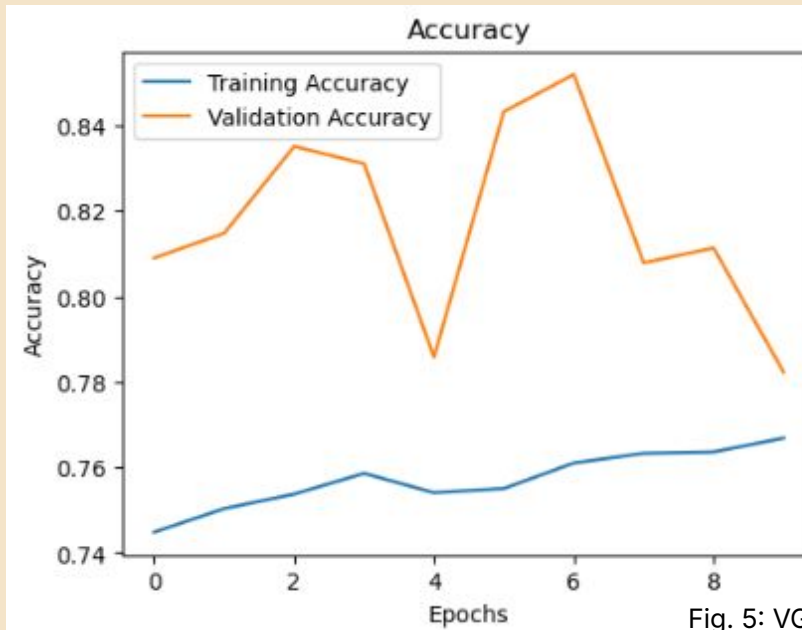
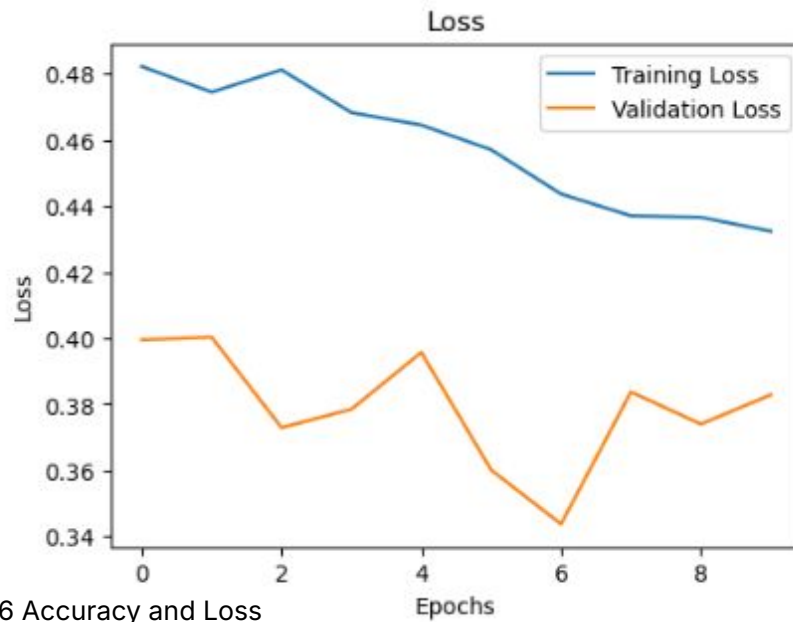


Fig. 5: VGG16 Accuracy and Loss



JAYA OPTIMIZATION ALGORITHM

- JAYA Algorithm was used to optimize feature selection and improve classifier performance for breast cancer classification.
- CNN and VGG16 were utilized for feature extraction from histopathological images.
- Hyperparameter tuning for RF, XGBoost, and KNN significantly influenced final classification accuracy.
- Traditional tuning methods like manual selection and grid search are computationally expensive and time-consuming.
- JAYA a metaheuristic optimization algorithm, automatically selects optimal hyperparameters without requiring algorithm-specific parameters.

RESULTS AFTER JAYA ALGORITHM

MODEL	ACCURACY BEFORE JAYA(%)	ACCURACY AFTER JAYA(%)
CNN+RF+PCA	85.9	86.37
CNN+XGB+PCA	86.83	87.18
CNN+KNN+PCA	85.61	85.85
VGG16+RF+PCA	68.62	75.06
VGG16+XGB+PCA	81.32	83.82
VGG16+KNN+PCA	72.1	80.68

Table 2: Results Table After Jaya algorithm

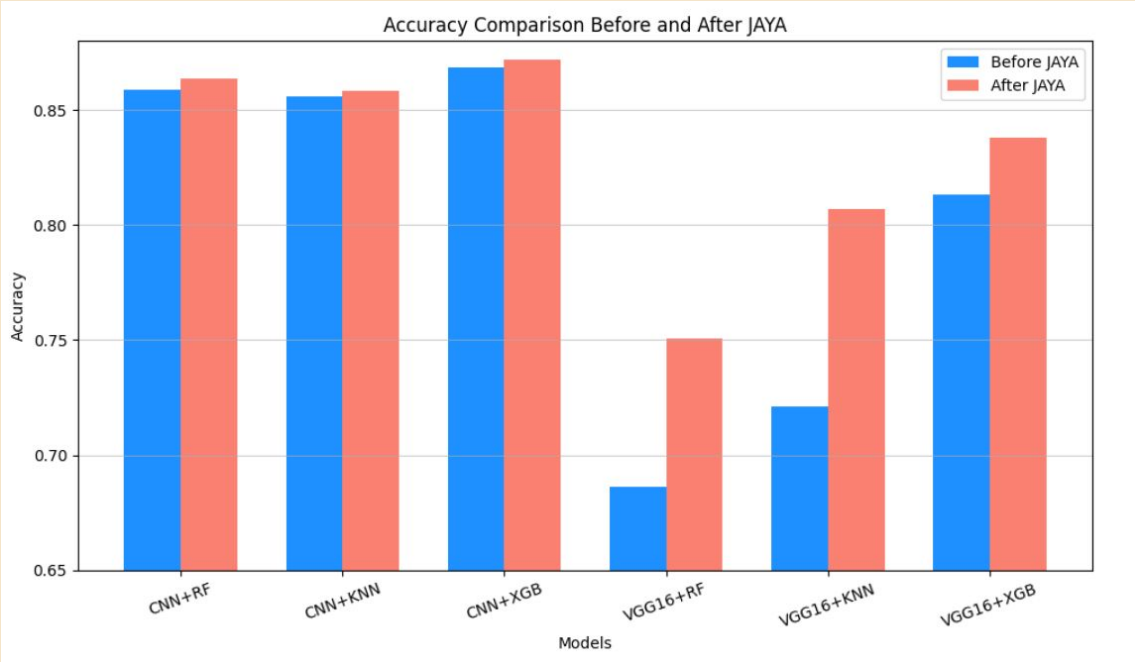


Fig.4: Model Accuracy After Jaya algorithm

CONCLUSION

- This project successfully implemented and evaluated deep learning and machine learning models for classifying breast cancer histopathological images into benign and malignant using the BreakHis dataset.
- The CNN+XGB with JAYA optimization achieved the highest test accuracy of 87.18%, demonstrating the effectiveness of combining deep learning-based feature extraction with optimized machine learning classifiers.
- PCA helped in dimensionality reduction while retaining important features, improving computational efficiency in certain models.
- The JAYA algorithm significantly enhanced accuracy by optimizing feature selection and hyperparameters for RF, KNN, and XGBoost classifiers.
- This study highlights that hybrid models integrating deep feature extraction with optimized machine learning classifiers provide a robust and reliable system to assist pathologists in early breast cancer detection.

REFERENCES

- Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., & Campilho, A. (2017). Classification of Breast Cancer Histology Images Using Convolutional Neural Networks. PLoS One, 12(6), e0177544. <https://doi.org/10.1371/journal.pone.0177544>
- Bayramoglu, N., Kannala, J., & Heikkilä, J. (2016). Deep Learning for Magnification Independent Breast Cancer Histopathology Image Classification. Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), 2440-2445. <https://doi.org/10.1109/ICPR.2016.7900001>
- Gupta, K., & Chawla, N. (2020). Analysis of Histopathological Images for Prediction of Breast Cancer Using Traditional Classifiers with Pre-Trained CNN. Procedia Computer Science, 167, 878-889. <https://doi.org/10.1016/j.procs.2020.03.427>
- Hoang, D.-T., et al. (2024). A Deep-Learning Framework to Predict Cancer Treatment Response from Histopathology Images through Imputed <https://doi.org/10.1038/s43018-024-00793-2> Transcriptomics. Nature Cancer, 5, 1-13.
- Li, X., Li, H., Cui, W., Cai, Z., & Jia, M. (2021). Classification on Digital Pathological Images of Breast Cancer Based on Deep Features of Different Levels. Mathematical Problems in Engineering, 2021, 1-13. <https://doi.org/10.1155/2021/8403025>
- Mondol, R. K., Millar, E. K. A., Graham, P. H., Browne, L., Sowmya, A., & Meijering, E. (2023). hist2RNA: An Efficient Deep Learning Architecture to Predict Gene Expression from Breast Cancer Histopathology Images. Cancers, 15(9), 2569. <https://doi.org/10.3390/cancers15092569>
- Rana, M., & Bhushan, M. (2023). Classifying Breast Cancer Using Transfer Learning Models Based on Histopathological Images. <https://doi.org/10.1007/s00521-023-08484-2> Neural Computing and Applications.
- Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks. Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN). <https://doi.org/10.1109/ijcnn.2016.7727519>
- Sudharshan, P. J., Petitjean, C., Spanhol, F., Oliveira, L. E., Heutte, L., & Honeine, P. (2019). Multiple Instance Learning for Histopathological Breast Cancer Image Classification. Expert Systems with Applications, 117, 103-111. <https://doi.org/10.1016/j.eswa.2018.09.049>
- Su, Z., Niazi, M. K. K., Tavalara, T. E., Niu, S., Tozbikian, G. H., Wesolowski, R., & Gurcan, M. N. (2023). BCR-Net: A Deep Learning Framework to Predict Breast Cancer Recurrence from Histopathology Images. PLOS ONE, 18(4).
- Dataset: <https://www.kaggle.com/datasets/ambarish/breakhis>

Thank You