Comprehensive Assessment for Breast Cancer Diagnosis Employing MATLAB Pre-trained Models with Machine and Deep Learning

MT13

ISPOR 2024 Presenter

Hao Mai Xuan[1,#], **Duong Cao Thi Thuy**[1], **Yankuba B. Manga**[1,2,*]

[1]*The Master Program in Smart Healthcare Management (SHM), International College of Sustainability Innovations, National Taipei University, New Taipei City, Taiwan*

[2]*The Bachelor Program in Smart Sustainable Development and Management (SSDM), International College of Sustainability Innovations, National Taipei University, New Taipei City, Taiwan*

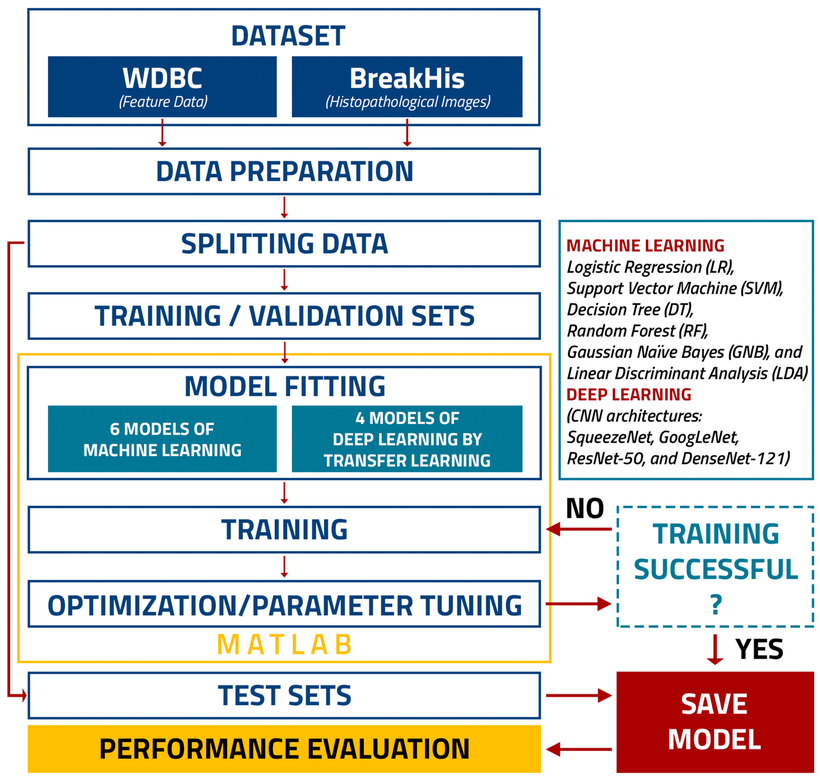
INTRODUCTION

Breast cancer is one of the global health concerns and a leading cause of death in the female population. It demands robust early diagnostic strategies for enhanced recovery rates and diminished mortality. This study aims to elevate breast cancer classification by integrating features and histopathological imaging data through advanced computational methods.

METHODS

A comparative analysis of six machine learning algorithms was conducted with the Wisconsin Diagnostic Breast Cancer (WDBC) dataset and four convolutional neural network (CNN) architectures were employed on the BreakHis dataset (Figure 1). Evaluation metrics such as the area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, and F1-score were used, with consistent settings facilitated by **MATLAB R2023b**.

Figure 1. Research flowchart



RESULTS

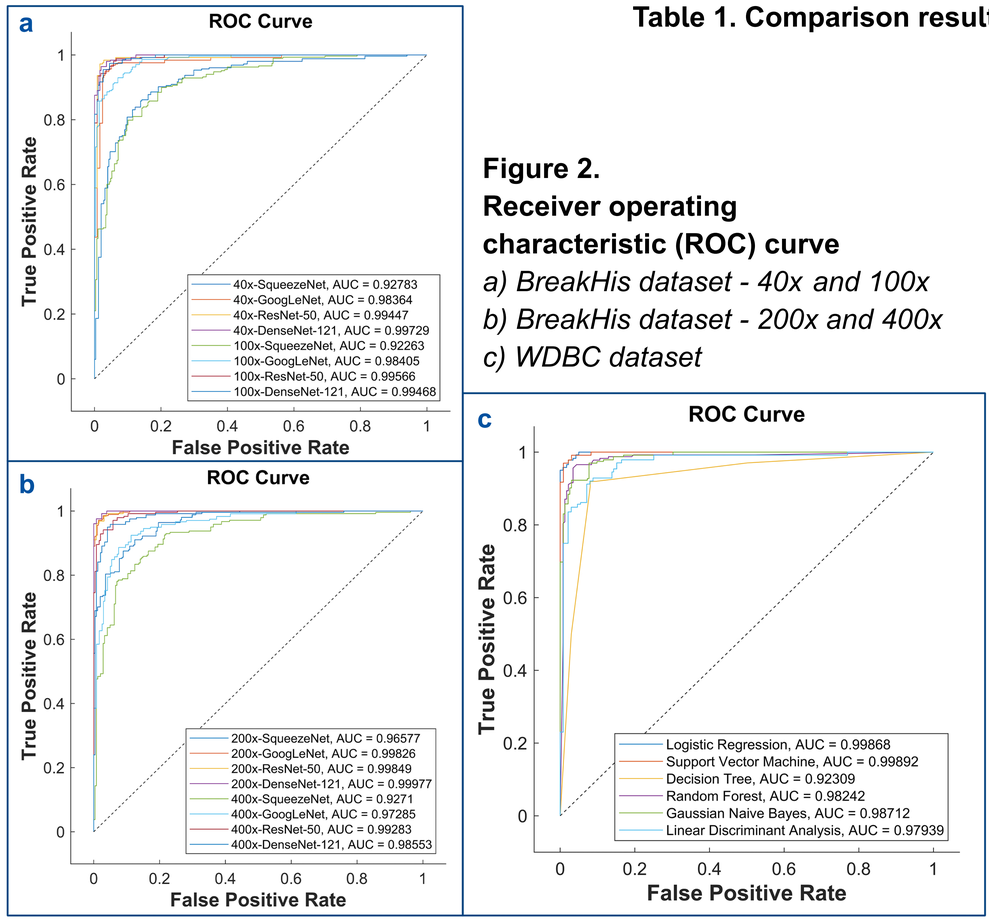


Figure 2.
Receiver operating characteristic (ROC) curve
a) BreakHis dataset - 40x and 100x
b) BreakHis dataset - 200x and 400x
c) WDBC dataset

Table 1. Comparison results of WDBC and BreakHis classification models

Models		Accuracy	Precision	Sensitivity	F1-score
LR		97.65%	0.955	0.985	0.970
SVM		98.24%	0.982	0.966	0.974
DT		92.35%	0.912	0.899	0.905
RF		95.88%	0.926	0.969	0.947
GNB		93.53%	0.915	0.900	0.908
LDA		94.12%	0.981	0.850	0.911
SqueezeNet	40×	88.94%	0.902	0.942	0.921
	100×	86.06%	0.908	0.889	0.898
	200×	91.04%	0.929	0.942	0.936
	400×	85.71%	0.922	0.862	0.891
GoogLeNet	40×	96.48%	0.971	0.978	0.975
	100×	93.27%	0.964	0.938	0.951
	200×	96.52%	1.000	0.950	0.974
	400×	91.76%	0.922	0.959	0.940
ResNet-50	40×	98.49%	0.979	1.000	0.989
	100×	96.63%	0.966	0.986	0.976
	200×	97.51%	0.979	0.986	0.982
	400×	95.05%	0.960	0.967	0.964
DenseNet-121	40×	97.49%	0.971	0.993	0.982
	100×	96.15%	0.972	0.972	0.972
	200×	98.51%	0.979	1.000	0.989
	400×	95.60%	0.967	0.967	0.967

The feature dataset demonstrates that the SVM outperforms other models in classification performance, as evidenced by an AUC of **0.999** and an accuracy rate of **98.24%**. In contrast, Logistic Regression and Random Forest show significant performance with accuracy rates of **97.64%** and **95.88%**, respectively (Figure 2). The histopathological image analysis also reveals that DenseNet-121 performs exceptionally well at 400x and 200x magnifications. Conversely, ResNet-50 performs suboptimally at lower resolutions (100x and 40x). SqueezeNet, GoogLeNet, ResNet-50, and DenseNet-121 achieve average accuracies of **87.94%**, **94.51%**, **96.92%**, and **96.94%**, respectively (Table 1).

CONCLUSIONS

The pre-trained models used in this research have shown outstanding performance compared to previous studies on different platforms. Emphasis is placed on the importance of diagnostic imaging in detecting breast cancer, highlighting the significance of these findings. The results validate the combination of machine learning algorithms and histopathological imaging, presenting a comprehensive and practical approach to classifying breast cancer.

Corresponding Author
Yankuba B. Manga, PhD.
Email: bydrmanga@gm.ntpu.edu.tw
Phone: +886-2-86741111 Ext. 68908