Department of Computer Science and Engineering
Bangladesh University of Business and Technology (BUBT)

CSE 478: Literature Review Records

Student's Id and Name	**Name:** Fabia Zaman Ekah and **ID:** 19202103225
Project Title	Deep Learning in Healthcare: Breast Cancer Detection and classification using Image Processing and CNN
Supervisor Name & Designation	**Name:** Khan Md. Hasib & **Designation:** Assistant Professor, Department of CSE, BUBT
Course Teacher's Name & Designation	**Name:** Khan Md. Hasib & **Designation:** Assistant Professor, Department of CSE, BUBT

Aspects	Paper # 5 (Title)
Title / Question (What is problem statement?)	A Hybrid Dependable Deep Feature Extraction and Ensemble-Based Machine Learning Approach for Breast Cancer Detection
Objectives / Goal (What is looking for?)	The main aim of the work was to to learn and extract hidden patterns from complex breast cancer images, while ML algorithms contribute interpretability and generalization capabilities.
Methodology / Theory (How to find the solution?)	The work was divided into some phases. • Dataset Acquisition: It encompasses capturing and acquiring a histopathological image-based (HPI) dataset in this model. • Preprocessing: To prepare the dataset for DL model training, we apply effective preprocessing techniques such as resizing images, sharpening, normalizing pixel values, and label encoding. • Dataset Split: We evaluate the performance of the models using the k-fold cross-validation technique. • Deep Feature Extraction: We use a pre-trained TL model to extract deep features from breast cancer images, ResNet50V2. • Machine Learning Models: We apply various ensemble-based ML models, such as DT, RF, ET, AdB, HGBC, GBC, XGB, and LGB classifiers, to analyze the deep features obtained. • Performance Evaluation: We use various performance metrics to analyze the experiments to evaluate how well the suggested methodology works.
Software Tools (What program/software is used for design, coding and simulation?)	Implementation work was carried out at Intel(R) Core (TM) i7 CPU M60 @ 2.80 GHz in Jupyter.
Test / Experiment How to test and characterize the design/prototype?	For the experimental work, the datasets were divided into the ratio of 80% and 20%. 80% of the datasets were used to train classification algorithms, and the remaining 20% used as test data.
Simulation/Test Data (What parameters are determined?)	Datasets was collected from - Kaggle.
Result / Conclusion (What was the final result?)	The results indicate that the DT algorithm achieved an accuracy of 78%, a precision of 77.39%, a recall of 79.47%, and an F1-score of 77.42%. Additionally, the DT algorithm exhibited an MAE of 22, MSE of 22, and RMSE of 46.9. On the other hand, both the RF and ET algorithm demonstrated higher accuracy than the DT, achieving an accuracy of 94.5%. The RF algorithm exhibited a precision of 94.45%, recall of 93.63%, and F1-score of 94.01%, while the ET algorithm achieved a precision of 94.18%, recall of 93.92%, and F1-score of 94.05%. Furthermore, the RF and ET algorithms had an MAE of 5.5, MSE of 5.5, and RMSE of 23.45. The AdB, HGB, GBC, XGB, and LGB algorithms also showcased competitive performance. AdB achieved an accuracy of 90.5%, while HGB, GBC, XGB, and LGB achieved accuracies of 94.5%, 94%, 94.5%, and 95%, respectively.

Obstacles/Challenges (List the methodological obstacles if authors mentioned in the article)	• Firstly, our study focuses specifically on Invasive Ductal Carcinoma (IDC), which may restrict the generalizability of our approach to other breast cancer subtypes. Further research is needed to assess its performance across a broader range of cancer types. • Secondly, the performance of our model is impacted by the quality and variety of the training dataset. Although we utilized a comprehensive dataset, the availability of more extensive and diverse datasets could further enhance the model's generalization capabilities. • Lastly, while our model demonstrates high accuracy, there are still areas for improvement, particularly in scenarios with challenging image conditions, such as noise and artifacts. Further optimization and refinement of the model could enhance performance in real-world clinical settings.
Terminology (List the common basic words frequently used in this research field)	Deep learning, Machine learning, ResNet50V2,Invasive Ductal Carcinoma (IDC), HPI.
Review Judgment (Briefly compare the objectives and results of all the articles you reviewed)	• Anjum et al. proposed an ML approach using histopathological images of breast tissues for detecting malignant cells. They combined HOG and Canny Edge detection techniques to extract features and utilized PCA for dimensionality reduction. SVM achieved 94 • Singh and Kumar focused on histopathological image analysis for breast cancer detection using a cubic Support Vector Machine (SVM), utilizing histopathology-based features
Review Outcome (Make a decision how to use/refer the obtained knowledge to prepare a separate and new methodology for your own research project)	For my own research I would use the DL-based pred-trained ResNet50V2 TL model in extracting complex patterns and representations.