British Journal of Nursing Studies
ISSN: 2755-080X
DOI: 10.32996/bjns
Journal Homepage: www.al-kindipublisher.com/index.php/bjns

BJNS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Early Detection of Breast Cancer Using Machine Learning: A Tool for Enhanced Clinical Decision Support

Mustafizur Rahaman[1], Ekramul Hasan[2], Dipta Paul[3]✉, Md Al Amin[4], Md Tuhin Mia[5]
[12] *College of Technology & Engineering*, *Westcliff University,* Irvine, California, USA
[3] *Department of Electrical and Electronic Engineering, American International University-Bangladesh (AIUB),* Dhaka, Bangladesh
[45] *School of Business*, *International American University,* Los Angeles, CA, USA
Corresponding Author: Dipta Paul, **E-mail**: diptapal4@gmail.com

| ABSTRACT
Breast cancer arises when there is an abnormal increase in breast tissue, resulting in the creation of lumps or irregular cell layers. This cancer ranks as the second most common among women worldwide, trailing only melanoma, and primarily impacts those over 50 years old, although it can manifest at any age. Timely identification and robust preventive measures are essential for minimizing health risks associated with cancer. Clinical trials in cancer prevention are persistently investigating innovative approaches for early diagnosis and treatment. This research utilizes machine learning methods to categorize breast cancer tumours as benign or malignant, facilitating prompt clinical decision-making. The dataset utilized for this study was obtained from Kaggle and underwent preprocessing and exploratory data analysis, incorporating correlation matrices and Principal Component Analysis (PCA) for dimensionality reduction and data visualization. Four supervised machine learning algorithms were assessed: Decision Tree, Logistic Regression, Bagging, and Random Forest. The evaluation of the models was conducted using metrics such as accuracy, precision, recall, F1-score, confusion matrix, and Area Under the Curve (AUC). The Random Forest classifier demonstrated the highest accuracy at 98%, showcasing its exceptional ability to manage the provided dataset effectively. Furthermore, the Bagging algorithm exhibited the highest AUC value at 99%, underscoring its effectiveness in differentiating between benign and malignant cases. The results indicate that ensemble techniques, especially Random Forest and Bagging, serve as valuable instruments for breast cancer classification, potentially aiding clinicians in achieving early and precise diagnoses.

| KEYWORDS
Breast Cancer Classification, Machine Learning, Tumour Diagnosis, Random Forest, Bagging Algorithm, PCA, Cancer Detection.

| ARTICLE INFORMATION
ACCEPTED: 01 June 2025 **PUBLISHED:** 12 June 2025 **DOI:** 10.32996/bjns.2025.5.1.6

1. Introduction

Breast cancer counts as a primary contributor to cancer-related fatalities among women worldwide[1] [2] [3]. Timely and precise diagnosis is essential, as early detection greatly enhances treatment outcomes[4]. Conventional diagnostic techniques, including mammography and biopsy, present certain limitations, such as the occurrence of false positives and negatives. Recent advancements in artificial intelligence (AI) and machine learning (ML) present significant opportunities to improve diagnostic accuracy and efficiency [5] [6]. These advanced systems possess the ability to scrutinize extensive datasets, discern intricate patterns, and generate data-informed predictions that can enhance clinical decision-making. For example, AI models that have been trained using labeled mammography and histopathological data have shown diagnostic accuracies that are on par with or even surpass those of seasoned radiologists [7] [8]. Moreover, the incorporation of ensemble learning methods and hybrid models has significantly enhanced classification accuracy by mitigating overfitting and improving model generalization. Deep learning architectures, especially convolutional neural networks (CNNs), have demonstrated considerable potential in segmenting tumor regions and extracting distinguishing features from medical images, all while minimizing the requirement for extensive manual annotation. These capabilities position AI as an essential tool in diagnostic workflows, providing valuable

insights while also playing a significant role in mitigating the worldwide shortage of skilled radiologists. As a result, AI-driven diagnostics are steadily moving from research environments to practical clinical settings, fundamentally altering the field of cancer detection and management. The implementation of AI in clinical environments has demonstrated encouraging outcomes. Large-scale trials conducted in Europe and North America have shown enhanced cancer detection rates through AI-assisted mammogram reading, effectively lowering false negatives while maintaining false positives at the same level [9]. The recent developments significantly improve early detection and contribute to reducing the burden on radiologists, who are encountering heightened demands as screening programs expand[10].

Algorithms in machine learning possess the capability to scrutinize intricate medical data, revealing patterns that may suggest the presence of breast cancer. For example, investigations have shown that the combination of convolutional neural networks (CNNs) with explainable AI (XAI) methods enhances the interpretability of diagnostic models, thereby promoting increased trust among clinicians [11] [12] [13]. Additionally, models based on transformers that handle multi-modal and longitudinal imaging data have demonstrated enhanced effectiveness in risk assessment and classification tasks [14]. In recent years, the incorporation of artificial intelligence (AI) and machine learning (ML) into healthcare has transformed numerous facets of patient care and other industries. In the realm of healthcare[15], diagnostic tools powered by AI are increasingly aiding clinicians in the identification of diseases, including the analysis of radiology images for early tumor detection. Additionally, these tools facilitate the development of personalized treatment plans through the use of predictive analytics. In addition to healthcare and in other sectors, the use of AI in sectors such as transportation, manufacturing[16] [17], and finance has enhanced operational efficiency[18] [19] [20] [21], bolstered safety, and fostered innovation through the automation of intricate decision-making processes [22] .

The use of AI in breast cancer screening has been confirmed in practical, real-world environments [23]. A comprehensive investigation conducted in Germany revealed that the implementation of AI-assisted screening led to a 17.6% increase in cancer detection rates, all while maintaining stable false-positive rates. In a similar vein, the UK's National Health Service has launched the largest trial globally to assess the effectiveness of AI in diagnosing breast cancer, with the goal of optimizing workflows and alleviating the workload of radiologists[24].

The research employs machine learning algorithms to categorize breast cancer stages, differentiating between benign and malignant cases. Utilizing tools like WEKA and methods such as Principal Component Analysis (PCA), our objective is to pinpoint the most precise and effective algorithms for early detection[25]. This methodology focuses on minimizing false negatives to guarantee prompt treatment interventions. Through this investigation, researchers aim to enhance the creation of dependable, AI-powered diagnostic tools that can be incorporated into clinical practice, ultimately leading to better patient outcomes.

2. Methodology

2.1 Explanation of the data

The data was sourced from Kaggle. The dataset comprises 683 entries and includes 10 attributes, such as :

- Clump Thickness: 1 - 10
- Uniformity of Cell Size: 1 - 10
- Uniformity of Cell Shape: 1 - 10
- Marginal Adhesion: 1 - 10
- Single Epithelial Cell Size: 1 - 10
- Bare Nuclei: 1 - 10
- Bland Chromatin: 1 - 10
- Normal Nucleoli: 1 - 10
- Mitoses: 1 - 10
- Class: (2 for benign, 4 for malignant)

Figure 1. Attribue names

2.2. Description of the algorithm

2.2.1. Logistic regression

This model is commonly employed in statistics to simulate the likelihood of a specific class or event occurring, such as the probability of a team achieving success or a patient being in optimal health, among other scenarios. This could be broadened to encompass various other scenarios, including the identification of an image depicting a creature like a dog, cat, lion, or others. It was expected that a single prominent object would be identified in the image, with the importance of each object varying from zero to one. The log-odds of the logistic model for the value labeled "1" represent a linear combination of one or more relationships among the predictions; both parameters can either be a binary classification variable or any real number. The logistic function, as indicated by the title, converts file to likelihood; the mean diameter of the value designated as "1" may vary between 0 and 1, hence the labeling. The standard unit of measurement for the logarithmic scale is the logit, derived from the logistic unit ; thus, it is relatively distinct. The key feature of the logistic regression model is that enhancing one of the individual variables related to computer resources consistently influences the probability of the specific outcome, with each predictor variable assigned its own parameter; for binary predictors, this extends to the hazard ratio. Alternative models, like the probit model, may serve as substitutes for the likelihood of stagnation. A binary logistic regression model with two levels indicates the dependent variable. In cases where the output consists of more than two results, multinomial logistic regression is employed for modeling purposes. When the various categories are arranged, ordinal logistic regression is employed.

2.2.2. Random Forest

A call tree and possibly a fabric classification possess hyperparameters that resemble those found in a random forest. The unpredictability of the model is heightened by cultivating these plants within a Random Forest framework. When tearing apart a node, it examines the simplest attribute from a random assortment of options instead of focusing on the most significant one. Functioning of the Random Forest algorithm. The subsequent phases can aid in comprehending the operational mechanics of the Random Forest algorithm. The initial phase involves selecting random samples from a specific dataset.
A single set of steps. Subsequently, the algorithmic software can generate the wire tree for each sample. Subsequently, each call tree will undergo the effects of the prediction. Actions are implemented for every anticipated result in the third step. In the final step, choose the predicted result that garnered the highest number of votes as the outcome. When utilizing the Random Forest method for regression scenarios, the mean square error is employed to determine how your data diverges from each node.

$$MSE = 1ni = 1N(xi - yi)2$$

2.2.3 Bagging

The Bootstrapping integration is a machine learning technique that merges a mechanism designed to enhance the legality and reliability of algorithms used in processing and retrieval, alongside what is perceived to be mere content. Through the addition, variability is minimized and the risk of overdosing is mitigated. Bulk predictions are often generated through the use of bootstrap predictive models, allowing for the incorporation of various predictive models. The subdivision or retrospective rule was applied to every random subset, and a new forecasting measurement tool predicts outcomes based on the viewers of each foundation in the context of hindsight. This resource could serve as a straightforward guideline for all duplicates of the bootstrapping measurement apparatus utilized in a primary training session conducted. Once the issue of segregation has been handled, the lowest student forecast grading system includes a bully vote in mass or by evaluating any open division possibilities. X is a guessable record, fbag is the bagged forecast, and f1(X), f2(X). fb(X) are forecasts from users of each basis. It's going to be connected to Calculation.

$$fbag = \ f1(X) + f2(X) + \ldots . + fb(X)$$

The aggregation strategy employed in bagging effectively reduces the variation of a personal learning algorithm essentially, averaging diminishes variability; however, it does not consistently improve the performance of an individual base learner. Bagging proves to be highly advantageous when dealing with volatile, multi-variable trainees, as the accuracy of predictions can fluctuate significantly due to slight changes in coaching input. This encompasses the call tree and K nearest algorithms. On the contrary, sacking results in a smaller increase in expected outcomes for systems characterized by high unit stability or bias due to reduced fluctuation.

2.2.4 Decision tree

In a decision tree, the terminal nodes signify outcomes or class labels, whereas the internal nodes denote decision points. Each internal node is analyzed for one or more designations, leading to the formation of one or more connections or branches. The government has assigned a decision value to this connection. The connections discussed are characterized by both exclusivity and inclusivity. This suggests that by simply clicking on one of the links, you will be protected from all possible scenarios. Decision trees are the most effective instruments for evaluating multiple options. This framework provides a valuable structure for delineating alternatives and assessing the possible results associated with each choice. Every node in a binary decision tree represents a comparison to be executed or an alternative to be selected. Two edges are associated with each node, both entering and exiting. The result "yes" or "true" is indicated by one edge, while the result "no" or "false" is indicated by the opposing edge. The coins are marked with the letters A, B, C, and D. Among them, three coins possess equal weight, while one coin is distinctly lighter. Identify the coin that has a greater weight. The model for predicting this issue is illustrated in Figure *. The bodyweights of A + B and C + D are evaluated and analyzed at the root node. The response is affirmative, and the left subsidiary holds true if A + B exceeds C + D. In addition to that, A + D presents a greater challenge as it requires the application of the correct branching techniques. The node on the left branch analyzes the training processes of a and b concerning their weights. The greater load coin is selected if the answer to this question is affirmative. If you reply with "no," it designates b as the currency with the significantly higher value. The identical approach is utilized for c and d when the primary node's result is "no."

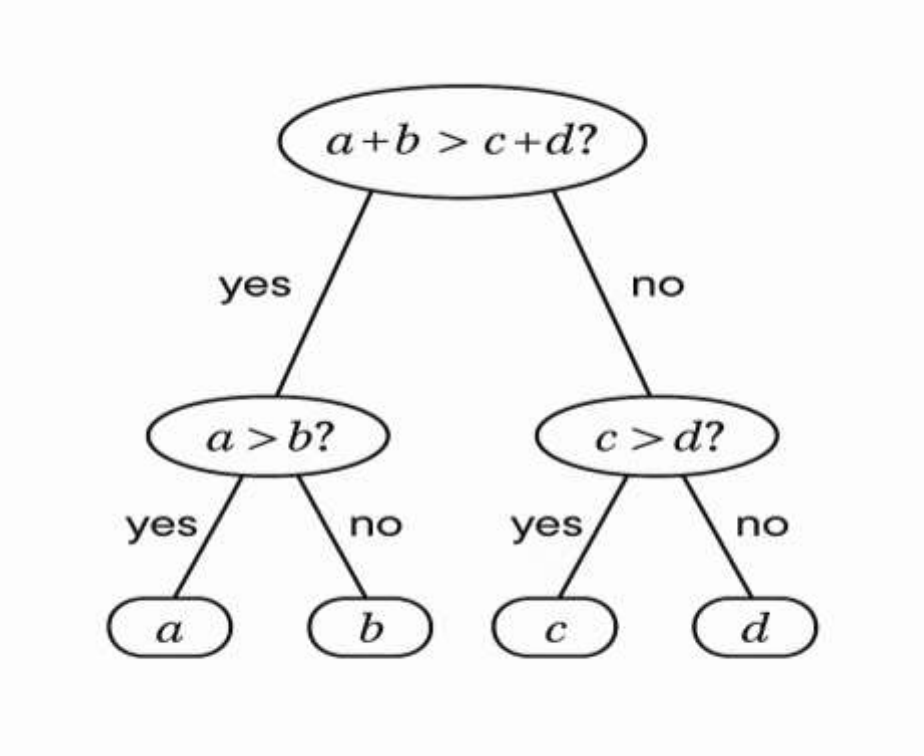

Figure 2. Decision tree figure

An uncomplicated decision tree relevant to the organization is illustrated in the example in Figure *. The subsequent matters were noted :

The four possible outcomes are illustrated by four-leaf nodes. Each leaf node represents the weight of a coin. A conclusive result or a leaf node necessitates two pairwise comparisons. Every weighing procedure is associated with a specific stage. From the roots to each leaf, numerous decision points exist. Every tree follows a set of principles, extending from its roots to its leaves. In the context of a decision-making process, the criterion can be articulated as follows : "if A + B > C + D and A > B, then the outcome is favorable."

2.3. Data Implementation

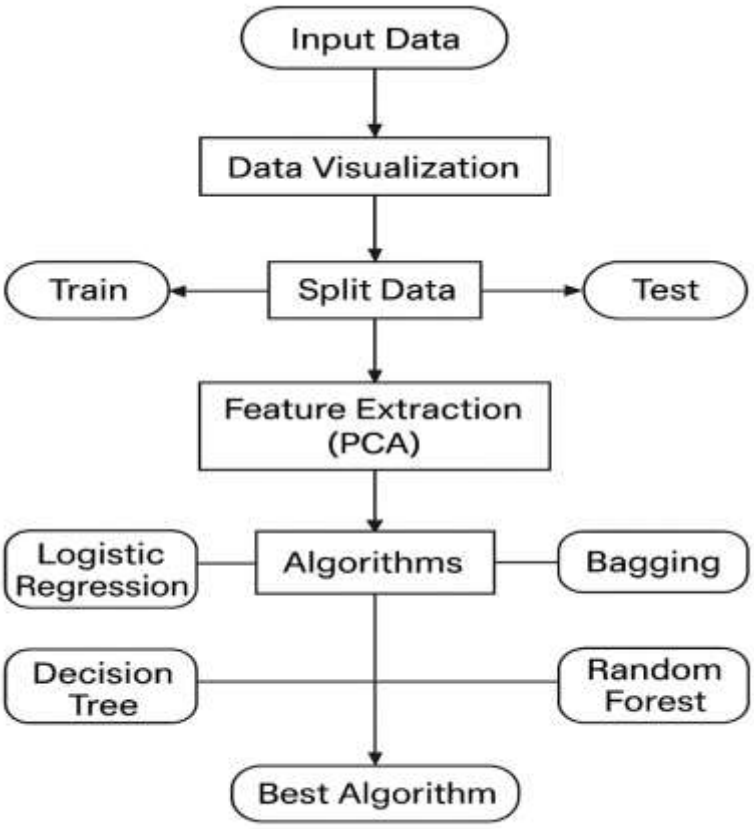

Figure 3. Data implementation

2.3.1. Input data

A system file containing information that will be used as input by a system or software is referred to as input data. the original document Computational engineering represents a specialized area within engineering technology that employs machines to analyze processes and structures that can be quantified.

2.3.2. Split data

The practice of segmenting data into distinct categories is commonly utilized in machine learning. In the process of model training, it is common practice to divide the data into distinct training and testing sets. Select the hyperparameters of the model and evaluate the prediction error or accuracy of the model. The training datasets will not provide sufficient data for the model to effectively learn the translation from source to destination when the datasets are divided into training and testing sets. Furthermore, the testing method would not provide sufficient data to accurately assess the model's performance.

2.3.3. Algorithm

Four algorithms are utilized, and they are :

1. Logistic Regression. (LR)
2. Decision Tree
3. Ensemble Method : Bagging
4. Random Forest

2.3.4. Categorization Result

Following the application of these four algorithms, researchers have identified the optimal outcome.

2.3.5. Optimal Algorithm

The algorithm that demonstrates the highest performance among the four applied is referred to as the optimal algorithm.

2.4. Data visualization

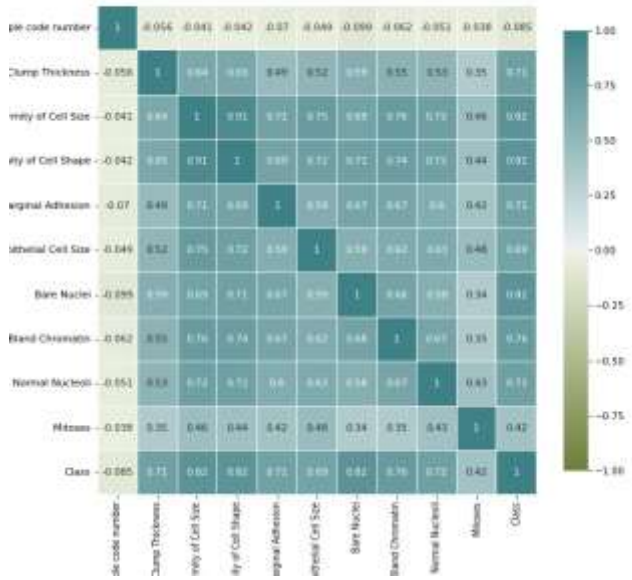

Figure 4. Corelation matrix

2.4.1. Correlation matrix

Compute and sum the component products for each position of the two columns in relation to the first to obtain the cross-correlation of the two matrices. This can be utilized to ascertain the offset required for two matrices of related values to align, though there are specific limitations to consider. A correlation matrix provides a concise summary of a dataset. A correlation matrix provides a straightforward summary of the relationships among all the variables in a dataset. Regression diagnostics employ a correlation matrix.

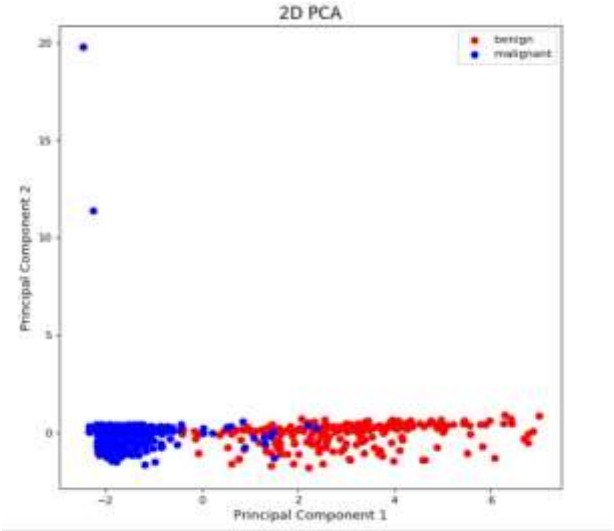

Figure 5. Principle component analysis

2.4.2. PCA

There is information available online regarding PCA, with numerous high-quality articles. However, many of these delve into excessive technical detail, while most of us prefer a straightforward understanding of its workings.

Unsupervised machine learning algorithms, including Principal Component Analysis, are commonly utilized across diverse applications, particularly in exploratory contexts.

1. Examination of data
2. Reduction of information, 3. Minimization of dimensions
3. Preparation of data, among other tasks.
4. Result Analysis

3. Experiment result analysis

Table 1. Evaluated Result For all Algorithm

Classification	class	Precision	Recall	F1-score
Logistic regression	2	0.65	1.00	0.79
	4	0.00	0.00	0.00
Random Forest	2	1.00	0.97	0.98
	4	0.95	1.00	0.97
Bagging	2	0.98	0.97	0.97
	4	0.95	0.96	0.95
Decision tree	2	0.97	0.95	0.96
	4	0.91	0.94	0.93

In Table 1, researchers assessed the outcomes for the algorithm. In this section, researchers will present four distinct types of machine learning algorithms, which include. The methods include decision trees, logistic regression, bagging, and random forest. In class 2, it indicates a benign condition, whereas class 4 signifies malignancy. In this context, researchers analyze the precision, F1 score, and recall.

Table 2. Result of confusion matrixes

Classification	Accuracy	AUC	Label	Predictive Negative (%)	Predictive Positive (%)
Logistic regression	64%	54%	Actual Negative	133	0
			Actual Positive	72	0
Random Forest	98%	98%	Actual Negative	129	4
			Actual Positive	0	72
Bagging	96%	99%	Actual Negative	129	4
			Actual Positive	3	69

Decision tree	94%	94%	Actual Negative Actual Positive	126 4	7 68

In table 2, the evaluation of confusion matrices was conducted to determine the accuracy and AUC value. The accuracy rates are as follows : decision trees at 94%, logistic regression at 64%, bagging at 96%, and random forest at 98%. The results indicate that the random forest model provides the highest accuracy. The AUC rates are as follows: decision trees at 94%, logistic regression at 54%, bagging at 99%, and random forest at 98%. Upon comparing the results of these algorithms, it is evident that bagging yields the highest AUC value at 99%. The bagging algorithm demonstrates superior performance in this table. This table presents the findings from the confusion matrix analysis. Researchers am designating it as Actual Negative and Actual Positive. The logistic regression algorithm demonstrates the highest true positive rate at 133%, while maintaining a false negative rate of 0%.

4. Conclusion

Reducing health risks and improving preventative strategies may lead to the prevention of cancer. Numerous elements can affect our likelihood of developing breast cancer over the span of our lives. Maintaining our health can potentially reduce the risk of developing breast cancer. However, certain factors, such as aging or a family history of the disease, remain outside our influence. Resrachers employed machine learning techniques in that study to differentiate between benign and malignant tumors. If the illness is mild, the doctor can initiate treatment promptly. The research aims to determine the most effective method by evaluating accuracy, precision, recall, and the confusion matrix. Data visualization employs principal component analysis (PCA) and correlation matrices. The bagging technique demonstrates the highest AUC value at 99% and the datasets indicate that Random Forest achieves the highest accuracy at 98%. Researchers employed four algorithms to predict breast cancer in this study, and researchers aim to incorporate additional methods in the future for enhanced identification. A range of techniques can be utilized to attain optimal and improved outcomes. Utilizing deep learning or artificial neural networks in upcoming research will yield optimal results.

Funding: This research received no external funding. This is a self-funded study.
Conflicts of Interest: The authors declare no conflict of interest.
ORCID ID: 0009-0001-2693-8449

References

[1] S. Hussain *et al.*, "Breast cancer risk prediction using machine learning: a systematic review," *Front Oncol*, vol. 14, p. 1343627, Mar. 2024, doi: 10.3389/FONC.2024.1343627/BIBTEX.
[2] A. K. Das, S. K. Biswas, A. Mandal, A. Bhattacharya, and S. Sanyal, "Machine Learning based Intelligent System for Breast Cancer Prediction (MLISBCP)," *Expert Syst Appl*, vol. 242, p. 122673, May 2024, doi: 10.1016/J.ESWA.2023.122673.
[3] A. K. Das, S. K. Biswas, A. Bhattacharya, and E. Alam, "Introduction to Breast Cancer and Awareness," *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, pp. 227–232, Mar. 2021, doi: 10.1109/ICACCS51430.2021.9441686.
[4] M. G. Hanna *et al.*, "Abstract PD6-03: Clinical-grade detection of breast cancer in biopsies and excisions using machine learning," *Cancer Res*, vol. 81, no. 4_Supplement, pp. PD6-03, Feb. 2021, doi: 10.1158/1538-7445.SABCS20-PD6-03.
[5] N. Meti *et al.*, "Machine Learning Frameworks to Predict Neoadjuvant Chemotherapy Response in Breast Cancer Using Clinical and Pathological Features," *JCO Clin Cancer Inform*, no. 5, pp. 66–80, Dec. 2021, doi: 10.1200/CCI.20.00078/ASSET/30A04FED-D845-4FAA-B583-FEF90D6B8160/ASSETS/IMAGES/LARGE/CCI.20.00078TA2.JPG.
[6] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, "A deep learning mammography-based model for improved breast cancer risk prediction," *Radiology*, vol. 292, no. 1, pp. 60–66, May 2019, doi: 10.1148/RADIOL.2019182716/ASSET/IMAGES/LARGE/RADIOL.2019182716.FIG5B.JPEG.
[7] R. K. Sachdeva and P. Bathla, "A Machine Learning-Based Framework for Diagnosis of Breast Cancer," *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJSI.301221*, vol. 10, no. 1, pp. 1–11, Jan. 1AD, doi: 10.4018/IJSI.301221.
[8] H. Sinha and M. Shah, "Early Prediction and Classification of Breast Cancer Survival Based on Machine Learning Models," *2025 IEEE 15th Annual Computing and Communication Workshop and Conference, CCWC 2025*, pp. 1185–1193, 2025, doi: 10.1109/CCWC62904.2025.10903817.
[9] B. Arianna Bunnell, P. Sadowski, and C. John Shepherd Peter Washington, "EARLY BREAST CANCER DIAGNOSIS VIA BREAST ULTRASOUND AND DEEP LEARNING," 2023.
[10] T. Umamaheswari and Y. Murali Mohan Babu, "Literature Review on Breast Cancer Diagnosis using 3D Images: Methods and Performance Analysis," *Proceedings - 2nd International Conference on Smart Electronics and Communication, ICOSEC 2021*, pp. 1728–1734, 2021, doi: 10.1109/ICOSEC51865.2021.9591861.

[11]O. K. Pal, D. Paul, E. Hasan, M. Mohammad, M. A. H. Bhuiyan, and F. Ahammed, "Advanced Convolutional Neural Network Model to Identify Melanoma Skin Cancer," *Proceedings of IEEE InC4 2023 - 2023 IEEE International Conference on Contemporary Computing and Communications*, 2023, doi: 10.1109/INC457730.2023.10262990.

[12]E. Hasan *et al.*, "CANCER DRUG SENSITIVITY THROUGH GENOMIC DATA: INTEGRATING INSIGHTS FOR PERSONALIZED MEDICINE IN THE USA HEALTHCARE SYSTEM," *The American Journal of Medical Sciences and Pharmaceutical Research*, vol. 6, no. 12, pp. 36–53, Dec. 2024, doi: 10.37547/TAJMSPR/VOLUME06ISSUE12-06.

[13]M. Al Amin *et al.*, "Enhancing Patient Outcomes with AI: Early Detection of Esophageal Cancer in the USA," *Journal of Medical and Health Studies*, vol. 6, no. 1, pp. 08–27, Jan. 2025, doi: 10.32996/JMHS.2024.6.1.2.

[14]G. Wadhwa and A. Kaur, "Various Image Modalities Used in Computer-Aided Diagnosis System for Detection of Breast Cancer Using Machine Learning Techniques: A Systematic Review," pp. 281–292, 2022, doi: 10.1007/978-981-16-1249-7_27.

[15]D. Paul, S. M. Aliuzzaman, M. F. Khan, M. T. Shakil, MD. M. Ali, and A. Rabbi, "An Innovative Embedded Ventilator for Accessible and Intelligent Respiratory Support," *Journal of Medical and Health Studies*, vol. 6, no. 1, pp. 99–108, Feb. 2025, doi: 10.32996/JGCS.2025.6.1.14X.

[16]A. K. Nuhel, D. Paul, E. Hasan, A. Azad, F. F. Rafi, and P. H. Roy, "A Microcontroller based Automated Waste Recycling Management System for SMEs," *2023 International Conference on Artificial Intelligence and Smart Communication, AISC 2023*, pp. 78–82, 2023, doi: 10.1109/AISC56616.2023.10085182.

[17]A. K. Nuhel, M. Al Amin, D. Paul, D. Bhatia, R. Paul, and M. M. Sazid, "Model Predictive Control (MPC) and Proportional Integral Derivative Control (PID) for Autonomous Lane Keeping Maneuvers: A Comparative Study of Their Efficacy and Stability," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 537, pp. 107–121, 2024, doi: 10.1007/978-3-031-48891-7_9.

[18]D. Paul *et al.*, "AI-Enhanced Multifunctional Smart Assistive Stick for Enhanced Mobility and Safety of the Visually Impaired," *Journal of Computer Science and Technology Studies*, vol. 7, no. 1, pp. 283–301, Mar. 2025, doi: 10.32996/JCSTS.2025.7.1.21X.

[19]A. T. Z. Prince *et al.*, "Design And Implementation of a Smart Wireless Parking System," *Journal of Computer Science and Technology Studies*, vol. 7, no. 1, pp. 113–121, Feb. 2025, doi: 10.32996/JCSTS.2025.7.1.9.

[20]A. Tamim *et al.*, "Design & Development of a Road Safety Device," *International Journal of Advanced Engineering Research and Science*, vol. 11, no. 12, pp. 2456–1908, Dec. 2024, doi: 10.22161/ijaers.1112.8.

[21]D. Paul, "An Autonomous Firefighting Robot for Industrial Safety Applications", doi: 10.13140/RG.2.2.19197.27361.

[22]A. K. Nuhel, M. M. Sazid, D. Paul, E. Hasan, P. H. Roy, and F. P. Sinojiya, "A PV-Powered Microcontroller-Based Agricultural Robot Utilizing GSM Technology for Crop Harvesting and Plant Watering," *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2023*, 2023, doi: 10.1109/SCEECS57921.2023.10062995.

[23]K. S. Tipatet, K. Hanna, L. Davison-Gates, M. Kerst, and A. Downes, "Subtype-Specific Detection in Stage Ia Breast Cancer: Integrating Raman Spectroscopy, Machine Learning, and Liquid Biopsy for Personalised Diagnostics," *J Biophotonics*, vol. 18, no. 1, p. e202400427, Jan. 2025, doi: 10.1002/JBIO.202400427.

[24]P. Oza, P. Sharma, S. Patel, and P. Kumar, "Deep convolutional neural networks for computer-aided breast cancer diagnostic: a survey," *Neural Comput Appl*, vol. 34, no. 3, pp. 1815–1836, Feb. 2022, doi: 10.1007/S00521-021-06804-Y/METRICS.

[25]P. Patil, M. Sharma, R. Rewatkar, and B. Fulkar, "Detecting Breast Cancer: A Comparative Study of Various Machine Learning Models," *2024 Parul International Conference on Engineering and Technology, PICET 2024*, 2024, doi: 10.1109/PICET60765.2024.10716141.