MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 4611-4613

Web-based Interface for Breast Cancer Prediction for classification of benign and malignant using Machine Learning.

Name: Ms Rasika Patil
M.Tech Data Science Scholar,
Department of Computer Engineering,
Vidyalankar Institute of Technology wadala
rasika.patil@gmail.com

Guide Name: Dr. Umesh Kulkarni
Department of Computer Engineering,
Vidyalankar Institute of Technology wadala
umesh.kulkarni@vit.edu.in

ABSTRACT
This paper emphasizes directions for future research and various gaps in research for a better understanding, accuracy, reliability, and clinical implementation of breast cancer diagnostic systems. This web-based UI will allow doctors to upload patient data to get instant predictions based on selected algorithms. A final AI pipeline, from preprocessing of the data to the deployment of the model, will be shown by this system, so that it will target to help to find early diagnosis of breast cancer.
Worldwide, among cancer-related deaths, it has become one of the prime causes in females, contriving early and precise diagnosis is a serious challenge in healthcare. Orthodox diagnostic techniques depend deliberately on expert interpretation and are arduous. To deal with such restraints, data–centric task and machine learning–based approaches are increasingly adopted to support clinical decision-making. Due to its usefulness and clinical applicability, the Wisconsin Breast Cancer Dataset has become a broadly used benchmark for measuring computational models for tumour classification. This review critically analyses the present research on the diagnosis of breast cancer, leveraging this dataset, employing statistical methods, techniques for classical machine learning, accumulated models, and a deep learning approach. The survey contrast report performance speech analyses, systematic practice, conversation challenges, constraint and future research methods in checkup artificial intelligence.

Keywords: *Machine Learning, Classification, Benign, Malignant, Breast Cancer*

1. INTRODUCTION:

Breast cells develop a malevolent tumour which is known as breast cancer, and it remains serious people health issue over the worldwide. Global cancer stats state that cancer of breast is the major reason of cancer related death in women. Initial stage cancer's precise diagnosis plays serious role in operative treatment planning and increasing survival rates of patient. A long-established methods such as, analysis through histopathology, biopsy with the help of ultrasound, X-rays and other clinical exams which require high level interpretation from oncology experts which consume a lot of time. Detection in early stage causes human errors while following above methods. To overcome such problems or limitations CAD systems are used simultaneously with techniques of machine learning is observed widely. Large amount of medical data is analysed through machine learning techniques, numerous complex patterns are identified, and it helps experts in making right and precise decisions. This study paper highlights present research related to breast cancer detection using ML techniques, methodologies comparison and performance results. It targets to systematically study the research efforts that will be use this information set and evaluated the usefulness of multi computational methods.

2. OVERVIEW:

2.1 LITERATURE SURVEY:

Breast cancer is one of the common and life ending diseases among women worldwide. The patient survival rate can be raised if cancer is diagnosed early so treatment complications are reduced "Traditional diagnostic methods such as mammography, ultrasound, and biopsy" [1]. depend largely on the experience of experts, which is sometimes prone to false conclusions. To overcome this limitation, Machine and Deep Learning (ML & DL) systems are being used extensively to automate breast cancer diagnosis and classification. Some early surveys focused on traditional machine learning algorithms using structured datasets. Various classification methods, such as SVM i.e Support Vector Machine, MLP Multilayer perceptron, logistic regression, Closest Nearest Neighbour, were compared using the "Wisconsin Diagnostic Breast Cancer dataset. In practice, MLP was found to achieve approximately 99% classification accuracy, which clearly shows the usefulness of machine learning systems in cancer diagnosis".
Similarly, Zhou et al analyzed various ML models with statistical feature selection methods and reported an AdaBoost–Logistic Regression–based model achieving 99.12% accuracy, which is higher than other classification methods [3].
Due to the exponential increase in medical image processing data , deep learning methods such as CNN are becoming more important. Arshad Adinni Histopathological pattern comparison exercise of pre-trained CNN architectures like VGG-16, DenseNet-121, ResNet152V2 and MobileNetV2. According to their findings, DenseNet-121 achieved accuracy as high as 99%, which proved the superiority of deep learning in pattern-based cancer diagnosis [4].
Aliquiel's amendments are not limited to classification alone, but extend to localization and segmentation of tumors. Anas Adini proposed an improved YOLOv5 model aggregated with Mask R-CNN for the analysis of mammogram patterns. Or the hybrid method significantly reduced false positive and false negative results and achieved a Matthew Correlation Coefficient (MCC) of 92.02%, which shows greater reliability for direct clinical use [5].
Explainable Artificial Intelligence (XAI) system is used to overcome the problem. Khater et al developed ML models using SHAP and permutation importance methods to explain the impact of important features on breast cancer classification. This method provides higher accuracy and greater transparency, making the AI system more reliable for medical experts [6].
Alternative imaging methods are still in practice. Al Husaini Adiini Detailed review of breast cancer research using thermography and neural networks Regards Banana. Or AI-assisted thermographic analysis can achieve high accuracy in practice; However, small datasets, lack of standardization and lack of clinical acceptability still exist [7].
Going beyond the paradigm, the subtypes of breast cancer have been identified using molecular modification, genomic and multi-omics data. Zhang proposed a multi-omics model based on Logistic Regression based on ELMO to collect RNA-seq and DNA methylation data, which achieved greater accuracy by using the reduction feature. Similarly, an autoencoder-based deep learning model shows better correlation with cognitive function than conventional PAM50 classification [8].
To improve ease of understanding and visual presentation, Wang developed a machine learning model using colored Petri nets to model breast tumor growth and metastasis. While the help of Logistic Regression maintains the accuracy, the hybrid method provides a visual assessment of the progress [9].

2.2 SUMMARY AND RESEARCH GAP

From the literature, it is apparent that deep learning models surpass traditional machine learning techniques in breast cancer detection, particularly in checkup image analysis. However, challenges such as express dataset handiness, lack of explainability, class imbalance, and insufficient real-time clinical validation stay unsolved. Therefore, there is a strong research scope for development accurate, explainable, and computationally efficient model that integrates multi-modal data for reliable breast cancer diagnosis.

2.3 EXISTING MODEL:

Several researchers have explored machine learning techniques for breast cancer detection:
Support Vector Machines (SVM) have shown high accuracy in classifying tumours using structured datasets.
Decision Trees and Random Forests provide interpretable models with competitive performance.
K-Nearest Neighbour (KNN) is simple but sensitive to noise.
Artificial Neural Networks (ANN) and deep learning models achieve high accuracy, especially with image-based datasets.
Convolutional Neural Networks (CNN) are widely used for mammogram and histopathology image classification.
Studies commonly use publicly available datasets such as the Wisconsin Breast Cancer Dataset (WBCD) and mammography image databases

2.4 PROPOSED MODEL:

Breast cancer occurs when breast tissues cells are growing uncontrollably. While it affects women as well as males can also diagnosed this disease. As per (ACS) American Cancer Society, from 1 women out of 8 women is in danger and they would be diagnosed with cancer of breast in their lifespan [1]. So, for prior detection, we are broadly classifying it into:
Bbenign tumours – non-cancerous growths
Malignant tumours – cancerous and potentially life-threatening

2.5 TOOLS & TECHNOLOGIES CATEGORY

This section expands the tools and technologies, hardware, and software requirements for implementing the proposed breast cancer prediction system.
2.5.1 SOFTWARE REQUIREMENTS
Tools Programming:
python 3. 10+ ML libraries such as NumPy, pandas, scikit-learn
Web Framework:

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 4611-4613

HTML and CSS deployment: it is used to design the user interface for the software login, registration page, dashboard, patient data forms, analysis page, and visit history page.

2.5.2 HARDWARE REQUIREMENTS

Operating System	Windows 11 Sun Valley
Ram	16,0 Giga bytes
Processor	13th Generation Intel , i7 Processor
System type	64-based processor and x64- bit operating system.

2.5.3 DATASET SOURCE

UCI Breast Cancer Wisconsin (diagnostic). "Numeric data having 30 features from fine-needle aspirate (FNA) images" [2]. Model Design Overview for Numeric model (prediction)

Dataset: UCI BREAST CANCER

Models: KNN, RANDOM FOREST, XGBOOST, LOGISTIC REGRESSION, DECISION TREE

Metrics: Accuracy, Precision, Recall, Auc, F1-Score

Explainability: Shap Values

Database Used: Mysql, It Is Used as a relational database management system to store, Edit, delete, User Authentication Details from the login page, Registration Page, patient information, Patient Data Forms, analysis results, and the Visit History page. It Will Make Sure that data Persistence and Efficient Query Management.

3. OBJECTIVES

1. Collect and preprocess the numeric data of breast cancer datasets.
2. Develop machine learning models for tabular data-based cancer prediction.
3. Implement explainability using Shap for numeric data.
4. Develop a web-based UI for Prediction, visualization, and to display the result.
5. Calculate the system's performance metrics and make the final report.

3.1 RISK ELEMENTS IN BREAST CANCER:

Genetic mutations: (BRCA1, BRCA2) [2]

Personal Health: If anyone of your breast is diagnosed with cancer automatically it increases the chances of another breast will diagnose too.

Gender: Cancer of breast is generally seen more in females as compared to male in ratio of 100 is to 1.

Age: At the age of 55 2 out of 3 females with breast cancer are diagnosed.

Race: Fair skinned women are majorly prone to cancer of breast as compared to darker skinned ones.

Genetic & Family History : If any of the family member is diagnosed with cancer of breast, than you too may increase great risk of suffering with cancer. Especially this possibility increases if any of them is suffered this cancer at prior age of 55.

3.2 Importance of Prior Detection :

Early detection significantly enhances the treatment success and reduces mortality. This has motivated the integration of automated and intelligent diagnostic tools in healthcare systems

3.2 TRADITIONAL METHODS FOR BREAST CANCER DETECTION:

Traditional diagnostic techniques include:

"Fine Needle Aspiration (FNA) and Core Needle Biopsy" (AME GROUPS, n.d.) : Invasive diagnostic methods used for histopathological examination. They are accurate but time-consuming and require expert analysis.

Mammography: "Screening mammography is a specific type of breast imaging that uses low-dose X-rays to detect cancer early – before women experience symptoms – when it is most treatable". (Clevel and Clinic Abudhabi, n.d.)

Ultrasound imaging: "The sound waves are produced by a handheld probe called a transducer, which detects the waves as they bounce off your breast tissue and create echoes. The transducer, which is connected to a computer, turns the echoes into a picture". (www.bcrf.org, n.d.)

Magnetic Resonance Imaging : "Breast MRI (magnetic resonance imaging) uses radio waves and strong magnets to make detailed pictures of the inside of the breast". (Marco Tulio Ribeiro, Sameer Singh,Carlos Guestrin, 9 Aug 2016) Although effective, these techniques depended heavily on expert radiologists and pathologists, leading to variability in diagnosis and increased workload. Machine learning in breast cancer diagnosis, the machine learning algorithms enable the automated analysis of medical data to detect whether the tumours are benign or malignant.

3.4 ML-BASED SYSTEMS TYPICALLY FOLLOW THESE STEPS:

1. Data collection
2. Preprocessing
3. Feature extraction
4. Classification
5. Performance evaluation

3.5 COMPARATIVE ANALYSIS OF ML ALGORITHMS

Machine learning techniques	Type of dataset	Pros	Cons
SVM	Tabular data used	It has high frequency and effctivi in high dimension	limited to kernel selection
KNN	tabular data used	implementation is simple	Requires high computations
DECISION TREE	Tabular data used	Interpreetationis easy	Couses overfitting
RANDOM FOREST ALGORITHM	Tabular data used	It is robust and having high accuracy	Executes in less time
ANN	Tabular dataset used	Handles comlex data pattern	requires large no of data
CNN	Image data used	It is having better feature extraction	Requires high cost fofr computation

4. PERFORMANCE EVALUATION METRICS:

"Basically, the Sensitivity is mainly important in medical diagnosis field to minimize false negatives. Some of the common features used in breast cancer detection studies are Accuracy, Precision, Recall (Sensitivity), Specificity, F1-SCORE, ROC-AUC" [3]

5. CHALLENGES AND RESEARCH GAPS:

1. Availability is limited for the medical data which is labelled.
2. Data imbalance issues
3. Lack of model interpretability
4. Generalization across different populations
5. Integration into real clinical environments

Mainly, most of the studies focused on accuracy while neglecting explainability and real-world deployment.

6. FUTURE SCOPE

Future research should focus on integrating structural datasets with imagination and genomic data to better diagnostic accuracy. Explainable AI (XAI) techniques are required for increasing transparency and trust to checkup AI system. [9] Also, real-time clinical deployment requires the large, data privacy, diverse, multi-centric datasets and standardize evaluation protocol to guarantee dependability and fairness. [10]

6. REFERENCES

[1] vol. 6, no. 2.
[2] "www.ettaatlantic.com," [Online]. Available: https://www.ettaatlantic.com/breast-cancer-what-you-need-to-know/.
[3] "Applying Deep Learning for Breast Cancer Detection in Radiology," [Online]. Available: Mahoro E, Akhloufi MA. Applying Deep Learning for Breast Cancer Detection in Radiology. Curr Oncol. 2022 Nov 16;29(11):8767-8793. doi: 10.3390/curroncol29110690. PMID: 36421343; PMCID: PMC9689782..
[4] "www.nationalbreastcancer.org," [Online]. Available: https://www.nationalbreastcancer.org/breast-cancer-risk-factors/.
[5] "https://www.radiologyinfo.org/en," [Online]. Available: https://www.radiologyinfo.org/en/info/mammo.
[6] "www.cancer.org," [Online]. Available: https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/breast-mri-scans.html.
[7] Vishwas V. Patange,Sanjay L. Nalbalwar, Jagadish B. Jadhav, Suchitra Shankar Hyam, "Deep Hybrid Learning for Accurate Lung Nodule Classification in CT Scans," International Journal of Computer Applications (0975 – 8887), vol. 187, p. 5, 2025.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 4611-4613

[8] Marco Tulio Ribeiro, Sameer Singh,Carlos Guestrin, ""Why Should I Trust You?" Explaining the Predictions of Any Classifier," arXiv:1602.04938v3 [cs.LG], vol. 1602.04938, no. v3, p. 10, 9 Aug 2016.
[9] Andre Esteva1,3*, Alexandre Robicquet1,3, Bharath Ramsundar1, Volodymyr Kuleshov1, Mark DePristo2,Katherine Chou2, Claire Cui2, Greg Corrado2, Sebastian Thrun1 and Jeff Dean2, "A guide to deep learning in healthcare," Nature Medicine.
[10] Abhishek Umre, Vaibhav Patel,Anurag Shrivastav, "A Systematic Review On Autism Spectrum Disorder Detection Using Machine Learning Techniques," International Journal of Creative Research Thoughts (IJCRT), vol. 13, no. 2, p. 6, 2025.
[11] Gani Esen, Marat Nurtas, „ "Reinforcement Learning for Breast Cancer Diagnosis with Visual Interpretability, Bazargul Matkerim," vol. 272, p. 8, 2025.
[12] A. F. M. Agarap, "On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset," arXiv:1711.07831v4 [cs.LG] 7 Feb 2019, 2019.
[13] Sheng Zhou1 , Chujiao Hu2 ,Shanshan Wei1 ,Xiaofan Yan2, "Breast Cancer Prediction Based on Multiple Machine Learning Algorithms," Technology in Cancer Research & treatment, vol. 23, no. 28.
[14] WISHAL ARSHAD 1 , TEHREEM MASOOD 2 , TARIQ MAHMOOD 3,4, ARFAN JAFFAR2 , FATEN S. ALAMRI 5 , SAEED ALI OMER BAHAJ 6 , AND AMJAD R. KHAN3, "Cancer Unveiled: A Deep Dive Into Breast Tumor Detection Using Cutting-Edge Deep Learning Models," ieee access .
[15] MUHAMMAD ANAS 1 , IHTISHAM UL HAQ 2 , GHASSAN HUSNAIN 3 , AND SYED ALI FARAZ JAFFERY , "Advancing Breast Cancer Detection: Enhancing YOLOv5 Network for Accurate Classification in Mammogram Images," IEEE ACCESS , 2024.
[16] TAREK KHATER1 , ABIR HUSSAIN 1,2, (Senior Member, IEEE), RIYAD BENDARDAF 3,4 , IMAN M. TALAAT3,5, HISSAM TAWFIK1 , (Senior Member, IEEE), SAM ANSARI 1 , AND SOLIMAN MAHMOUD (Senior Member, IEEE) , "An Explainable Artificial Intelligence Model for the Classification of Breast Cancer," IEEE ACCESS , 2023.
[17] MOHAMMED ABDULLA SALIM AL HUSAINI1 , MOHAMED HADI HABAEBI 1 , (Senior Member, IEEE), SHIHAB A. HAMEED2 , (Senior Member, IEEE), MD. RAFIQUL ISLAM 1 , (Senior Member, IEEE), "A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks," IEEE ACCESS, 2020.
[18] RICHARD LUPAT 1,2, RASHINDRIE PERERA 1,3, (Member, IEEE), SHERENE LOI 1,2, AND JASON LI 1,2, (Member, IEEE), "Moanna: Multi-Omics Autoencoder-Based Neural Network Algorithm for Predicting Breast Cancer Subtypes," IEEE ACCESS, 2023.
[19] Xuyue Wang, Wangyang Yu*, Zeyuan Ding, Xiaojun Zhai, and Sangeet Saha, "Modeling and Analyzing of Breast Tumor Deterioration Process with Petri Nets and Logistic Regression," COMPLEX SYSTEM MODELING AND SIMULATION , vol. 2, p. 8, 2022.
[20] M. Swain, "The GH Falcon," 2025. [Online]. Available: https://ghfalcon.com/20247/news/breast-cancer-awareness-month-spreading-hope-saving-lives/.
[21] [Online]. Available: https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data.
[22] K. Raja Guru , Sai Sree, "Comparison of DenseNet201 and ResNet50 for Breast Cancer Detection using Deep Learning," UtilitasMathematica, Vols. Volume122,, no. ISSN0315-3681, p. 7, 2025.
[23] "Clevel and Clinic Abudhabi," [Online]. Available: https://www.clevelandclinicabudhabi.ae/en/health-hub/health-resource/diagnostics-and-testing/mammography.
[24] "AME GROUPS," [Online]. Available: ccts.amegroups.com.
[25] [Online]. Available: https://www.bcrf.org/about-breast-cancer/breast-ultrasound/.