Ind. J. Pharm. Edu. Res., 2026; IPC Issue:28-37.
https://www.ijper.org

Original Article

Artificial Intelligence Powered Digital Twin Framework for Personalized Breast Cancer Diagnosis and Treatment

Pallavi Gondale[1,*], Surendra Gattani[2,*]

[1]Department of Pharmacology, School of Pharmacy, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, INDIA.
[2]Department of Pharmaceutics, School of Pharmacy, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, INDIA.

ABSTRACT

Background/Objectives: Breast cancer remains a major global health challenge, with significant variability in patient outcomes necessitating personalized diagnostic and treatment approaches. The present study aimed to develop an artificial intelligence-powered digital twin framework for personalized breast cancer diagnosis and treatment simulation. **Materials and Methods:** A retrospective computational model was developed using the Wisconsin Breast Cancer Dataset comprising 569 samples and 30 radiomic features. The dataset was divided into training (80%) and testing (20%) sets, and five-fold cross-validation was performed. A gradient boosting algorithm was integrated within a digital twin architecture to enable patient-specific risk prediction and treatment pathway simulation. **Results:** The model achieved high diagnostic performance with an accuracy of 97.37%, sensitivity of 97.78%, specificity of 97.10%, precision of 0.976, F1-score of 0.975, and AUC-ROC of 0.9945. The framework enabled individualized risk stratification into low-, intermediate-, and high-risk categories, along with corresponding treatment pathway simulations. **Conclusion:** The proposed digital twin framework integrates predictive modeling with virtual patient simulation, enhancing interpretability and supporting precision oncology decision-making. This approach demonstrates strong potential for personalized breast cancer diagnosis, risk assessment, and clinical decision support.

Keywords: Artificial Intelligence, Breast Cancer, Digital Twin, Gradient Boosting, Precision Oncology.

Correspondence:
Ms. Pallavi Gondale
Department of Pharmacology, School of Pharmacy, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, INDIA.
Email: pallavigondale871@gmail.com
ORCID: 0009-0007-9671-5543

Prof. Surendra Gattani
Department of Pharmaceutics, School of Pharmacy, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, INDIA.
Email: sggattani@gmail.com
ORCID: 0000-0003-2179-0139

INTRODUCTION

Breast cancer continues to be a top cause of cancer-related deaths among women around the world. Breast cancer has high rates of occurrence and death globally, especially in low- and middle-income countries. Clinical results vary widely across people due to tumour heterogeneity and biological behaviour, despite breakthroughs in screening and treatment methods. Conventional breast cancer management relies on clinical criteria such tumour size, lymph node status, histological grading, and receptor expression. Although these variables provide important predictive information, they may not fully capture the complex link between tumour biology and patient-specific characteristics. As a result, therapy choices are often made using broad recommendations that might not be ideal for specific patients. This limitation has led to increased emphasis on precision oncology, which focuses on tailoring diagnosis and treatment according to individual patient profiles.[1-4]

Artificial intelligence is a powerful tool for analysing complex biomedical datasets and improving cancer diagnostic accuracy. Machine learning algorithms can analyse high-dimensional datasets, including imaging, histology, and clinical data.

Machine learning algorithms can find nonlinear relationships and reveal hidden patterns. Ensemble learning methods, like gradient boosting, provide better predictive performance because they combine several weak learners and lower prediction errors. These models have been used effectively in clinical decision-support systems, risk prediction, and breast cancer classification.[5-9] Digital twin technology represents an advanced application of artificial intelligence within the healthcare sector. A digital twin is a computational representation that mirrors a physical object by integrating dynamic data to simulate its behaviors and results.[10-14]

Digital twins make it possible to create patient-specific models in oncology that can assess treatment options, simulate the course of the disease, and aid in clinical decision-making. This method enhances individualised treatment planning and permits in silico experimentation.[11-16]

DOI: 10.5530/ijper.ipc004

Copyright Information :
Copyright Author (s) 2026 Distributed under Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia. [www.mstechnomedia.com]

Breast cancer is a heterogeneous illness, so the adoption of digital twin technology is quite promising. Digital twin frameworks can produce personalised risk assessments and predict possible treatment paths by combining machine learning with virtual patient modelling. Clinicians may be able to detect high-risk patients, optimise treatment plans, and reduce needless treatments with the use of such systems.[15,16]

Nonetheless, despite increased interest, there is still a scarcity of comprehensive frameworks that combine digital twin-based simulation and predictive modelling. Many existing studies focus either on classification algorithms or conceptual digital twin designs, with limited development of unified, clinically applicable systems.

Therefore, the present study aimed to develop an artificial intelligence-powered digital twin framework for personalized breast cancer diagnosis and treatment simulation using a gradient boosting machine learning approach. The proposed model integrates predictive analytics with virtual patient modeling to generate individualized risk stratification and treatment pathway simulation, thereby supporting precision oncology and clinical decision-making.[17-20]

MATERIALS AND METHODS

Study Design

The present research was structured as a retrospective computational model development study, aimed at establishing a digital twin framework utilizing artificial intelligence for personalized diagnosis and treatment simulation of breast cancer. The methodological workflow included data collection, preprocessing, feature engineering, model construction, internal validation, and digital twin-based simulation. The approach supports precision oncology decision-making by combining virtual patient modelling with machine learning-based prediction.[11,14]

The study focused on developing a gradient boosting-based predictive model embedded within a digital twin architecture capable of generating individualized risk stratification and treatment pathway simulation. The computational workflow followed standard machine learning practices including training-testing split, cross-validation, and performance evaluation using clinically relevant metrics.[5,17]

Dataset Description

The model was built on the Wisconsin Breast Cancer Dataset (Diagnostic). The collection includes 569 patient samples taken from digital pictures of breast lumps using a fine needle aspiration. Each sample is classified as benign or malignant, and 30 Numerical parameters that describe the morphology of cell nuclei.[18] The dataset characteristics used for digital twin model development are summarized in Table 1.

Table 1: Dataset Characteristics Used for Digital Twin Model Development.

Parameter	Description
Dataset	Wisconsin Breast Cancer Dataset
Total samples	569
Benign cases	357 (62.7%)
Malignant cases	212 (37.3%)
Number of features	30
Training set	455 (80%)
Testing set	114 (20%)
Validation method	5-fold cross-validation

The features were determined based on ten primary nuclear attributes, such as morphology, radius, perimeter, surface area, surface smoothness, compactness, concavity, concave regions, symmetry, and fractal dimension. Three calculations were made for each attribute: the mean value, standard deviation, and worst value. These characteristics indicate structural irregularities and nuclear morphology associated with malignant transformation.[6,9]

Data Preprocessing

The dataset was preprocessed to guarantee data consistency and quality before the model was developed. The absence of incomplete records in the dataset was verified by missing value analysis. Since every variable was numerical, machine learning analysis could be performed without the need for categorical encoding.

Standard normalisation was used for feature scaling in order to convert variables to zero mean and unit variance. In addition to ensuring that all features contributed equally, this step stopped high magnitude variables from controlling model training. Additionally, during gradient boosting optimisation, data normalisation enhances convergence.[7]

An 80:20 split was utilised to divide the dataset into training and testing sections. To eliminate sample bias and improve model generalisability, stratified sampling was utilised to ensure proportionate representation of benign and malignant cases in both groupings.

Digital Twin Framework Architecture

The proposed artificial intelligence-powered digital twin architecture consisted of four interconnected layers:Data assimilation layer, Feature processing layer, Predictive modeling layer. The overall architecture of the proposed AI-powered digital twin framework is illustrated in Figure 1.

Simulation and decision-support layer

The data assimilation layer receives patient-specific input features. The feature processing layer performs normalization

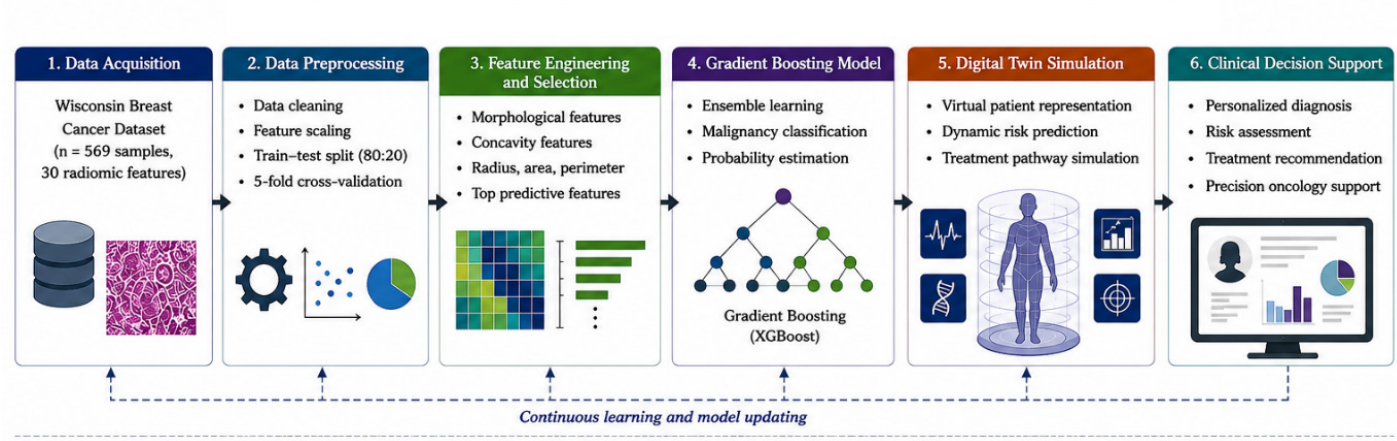


Figure 1: Architecture of the AI-powered digital twin framework for personalized breast cancer diagnosis and treatment simulation. The framework integrates data acquisition, preprocessing, feature engineering, gradient boosting-based prediction, digital twin simulation, and clinical decision-support modules to generate individualized risk assessment and treatment pathway recommendations.

and feature selection. The predictive modeling layer applies the gradient boosting classifier to generate malignancy predictions and probability scores. The simulation layer converts prediction outputs into a virtual patient model capable of risk stratification and treatment pathway simulation. This architecture enables dynamic representation of disease status and supports individualized decision-making in precision oncology.[10-20]

Feature Selection

All 30 features were initially included in model development. Feature importance ranking was performed using gradient boosting importance scores. Correlation-based filtering was applied to identify redundant variables and reduce multicollinearity.

Feature selection increases computing efficiency, decreases overfitting, and improves model interpretability. The most informative predictors contributing to malignancy classification were retained in the final model. These features correspond to morphological indicators of tumor aggressiveness.[17,21]

Machine Learning Model Development

Gradient boosting classifiers were chosen as the key prediction approach. Gradient boosting is an ensemble learning technique that minimises prediction error by gradually integrating weak learners. The model was trained using the training dataset (n=455). The technique creates decision trees iteratively, with each new tree removing residual flaws from the previous tree.[8]

Cross-validation was utilised to tune hyper parameter settings including training rate, count of estimators, and maximum tree depth. The algorithm's ability to capture complicated feature interactions and nonlinear correlations in structured biological datasets influenced its selection.[22,23]

Training and Validation

A five-fold cross-validation technique was employed to assess model stability while minimising overfitting. Five subsets were generated from the training dataset. The model was trained on four subsets, with the remaining subset utilised for validation. Average performance metrics were computed after this procedure was carried out five times. The independent testing dataset was not used for training; instead, it was held aside for the final model assessment. Cross-validation guarantees that performance is independent of a single data split and enhances reliability. Metrics for Performance Evaluation.[5,21]

Performance Evaluation Metrics

The following classification measures were used to evaluate model performance: accuracy, sensitivity, specificity, precision, F1-score, and Area Under Receiver Operational Characteristic Curve (AUC-ROC). Overall classification performance is measured by accuracy. While specificity assesses the accurate categorisation of benign instances, sensitivity assesses the accurate identification of malignant case. Positive predictive value is reflected in precision. The F1-score provides an accurate measure of recall and precision. AUC-ROC evaluates discriminatory power without respect for threshold.[9,19]

Digital Twin Simulation Workflow

Following prediction, the digital twin framework generated patient-specific outputs. The workflow included:

Patient data acquisition, Feature preprocessing, Gradient boosting prediction, Probability estimation, Risk stratification and Treatment pathway simulation.

Clinical decision-support output.

The digital twin translated prediction outputs into individualized risk categories.

Based on risk classification, the framework simulated conceptual treatment pathways including monitoring, diagnostic evaluation, surgical planning, and systemic therapy options. This virtual modelling supports precision oncology decision-making and individualized patient management.[11,12]

Digital twin framework architecture

The suggested artificial intelligence-powered digital twin system was designed to create a virtual depiction of breast cancer patients for personalised diagnosis and therapy simulation. The architecture integrates machine learning-based prediction with patient-specific modeling to support precision oncology decision-making. The digital twin operates as a computational surrogate that mirrors disease characteristics and produces individualized risk assessment and treatment pathway outputs.[10-20]

The framework consists of four primary layers: data assimilation layer, feature processing layer, predictive modeling layer, and simulation and decision-support layer. These components function sequentially to convert patient-specific input data into clinically meaningful outputs.

Data Assimilation Layer

The digital twin framework's input component is the data assimilation layer. Digitised fine needle aspiration images are used to generate patient-specific characteristics for this layer. The dataset includes the input variables include radius, texture, circumference, surface area, smoothness, compactness, concavity metric (concave points), symmetry, and fractal dimension. Each parameter consists of its mean, standard error, and worst values. Together, these metrics characterise the cellular morphology linked to malignant transformation. The input structure is standardised by the data assimilation layer, which also gets the dataset ready for further processing.[18,6]

Feature Processing Layer

Preprocessing, normalisation, and feature selection are carried out by the feature processing layer. To guarantee that each variable contributes equally during model training, features are converted into comparable ranges using standard scaling.This technique enhances model stability and avoids bias brought on by high-magnitude features.

Feature selection is performed using importance ranking and correlation filtering. Redundant and weakly predictive variables are minimized to enhance generalization and reduce overfitting. The processed feature set is then forwarded to the predictive modeling layer. This layer also improves interpretability by identifying clinically meaningful predictors contributing to malignancy classification.[7,17]

Predictive Modeling Layer

The predictive modeling layer represents the core computational engine of the digital twin framework. A gradient boosting classifier is implemented to generate malignancy predictions and probability scores. The algorithm sequentially combines multiple decision trees to minimize classification error and capture nonlinear relationships among features.

The predictive modeling layer produces three outputs:

- Binary classification (benign or malignant)
- Probability score
- Feature importance ranking

These outputs are subsequently utilized by the simulation layer to generate personalized virtual patient representations and risk stratification results.[7,5]

Simulation and Decision-Support Layer

The simulation layer transforms predicted outputs into a virtual patient model. Patients are classified into low, middle, and high risk categories based on their predicted chance.

The digital twin simulates treatment pathways according to risk classification. Low-risk patients are assigned to monitoring and routine screening. Intermediate-risk patients are recommended for additional diagnostic evaluation. High-risk patients are mapped to treatment planning pathways including biopsy, surgical consultation, and systemic therapy consideration.

The simulation layer generates individualized decision-support outputs. These outputs function as clinical guidance rather than prescriptive recommendations. The digital twin therefore serves as an assistive tool for precision oncology decision-making.[10-20]

Digital Twin Workflow

The overall digital twin workflow follows a sequential pipeline:

- Patient feature input.
- Data preprocessing.
- Feature selection.
- Gradient boosting prediction.
- Probability estimation.
- Risk stratification.
- Treatment pathway simulation.
- Clinical decision output.

This workflow enables transformation of static patient data into dynamic virtual patient models. The digital twin continuously

represents disease characteristics and supports individualized treatment planning.

Clinical Relevance of Digital Twin Architecture

The digital twin framework offers several advantages for breast cancer management. First, it enables personalized risk prediction using patient-specific data. Second, it supports treatment pathway simulation without direct clinical intervention. Third, it enhances decision-making through interpretable outputs. Fourth, the architecture is scalable and can integrate additional modalities including genomics, imaging, and longitudinal clinical data.

The proposed architecture therefore represents a practical implementation of digital twin technology for precision oncology applications and personalized breast cancer management.[10-20]

Ethical Approval

This study utilized a publicly available dataset (Wisconsin Breast Cancer Dataset) and did not involve human or animal subjects directly; therefore, ethical approval was not required.

RESULTS

Model Performance

The gradient boosting-based digital twin framework demonstrated strong performance in classifying breast tumors. An independent test dataset with 114 samples was used to assess the model. High diagnostic accuracy and efficient distinction between benign and malignant cases were demonstrated by the predictive model.

The model's overall classification accuracy was 97.37%. While the specificity was 97.10%, suggesting a dependable classification of benign cases, the sensitivity was 97.78%, indicating an exceptional capacity to identify malignant tumours. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which was 0.9945, showed near-perfect discriminative capacity. These findings demonstrate that the artificial intelligence-driven digital twin structure provides accurate diagnostic predictions and balanced categorisation performance. The performance metrics of the proposed AI-powered digital twin model are presented in Table 2. The Receiver Operating Characteristic (ROC) curve and digital twin-based risk stratification workflow are presented in Figure 2.

Cross-Validation Results

Five-fold cross-validation was applied on the training dataset to evaluate the model's stability and generalisation. The cross-validation results showed that all folds performed consistently. With little variance between folds, the average cross-validation accuracy was roughly 97.5%.

These findings indicate that the model generalized well and was not dependent on a single dataset partition. The consistency across folds confirms the robustness of the gradient boosting algorithm and supports its integration within the digital twin framework.

Confusion Matrix Analysis

To thoroughly assess categorisation performance, confusion matrix analysis was used. Most samples were correctly identified

Table 2: Performance Metrics of the AI-Powered Digital Twin Model.

Metric	Value
Accuracy	97.37%
Sensitivity	97.78%
Specificity	97.10%
Precision	0.976
F1 Score	0.975
AUC-ROC	0.9945

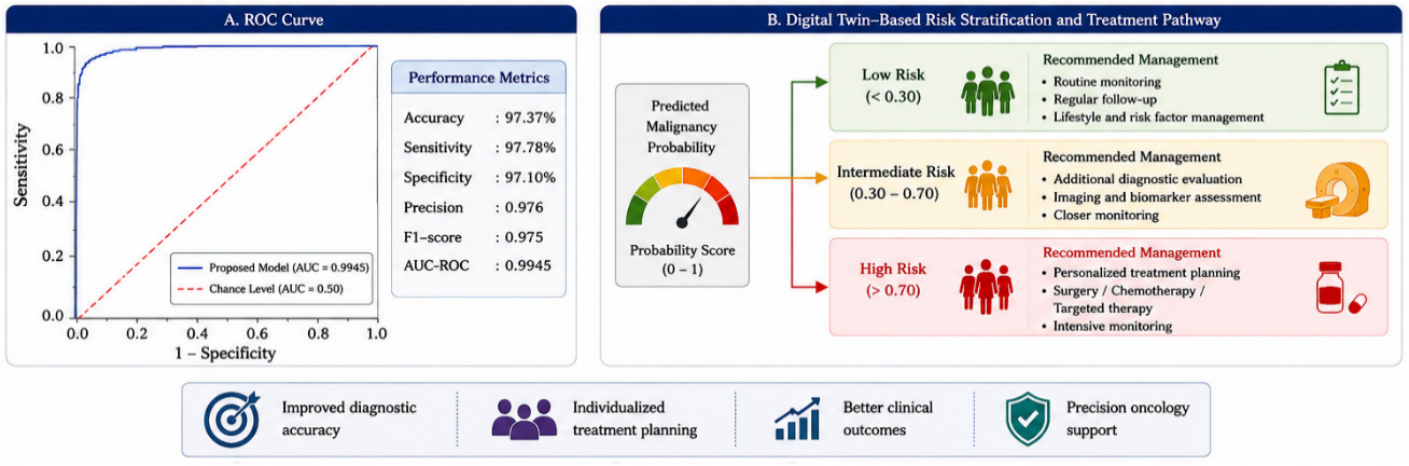


Figure 2: ROC curve performance and digital twin-based risk stratification framework. The ROC curve demonstrates the predictive performance of the gradient boosting model (AUC-ROC = 0.9945). The digital twin framework categorizes patients into low-, intermediate-, and high-risk groups and simulates corresponding treatment pathways to support precision oncology decision-making.

by the model, and just a few misclassifications were noted. The confusion matrix results are presented in Table 3.

The confusion matrix demonstrates high sensitivity and specificity. The low number of false negative cases is clinically significant because missed malignant diagnoses may delay treatment.

Analysis of Feature Importance

Gradient boosting importance scores were used in feature importance analysis to find factors influencing malignancy categorisation. The worst concave spots, worst area, worst perimeter, worst radius, and mean concavity were the most significant characteristics. The ranking of important predictive features is shown in Table 4.

Concavity-related and morphological features demonstrated the highest predictive value, indicating that structural irregularities of cell nuclei play a major role in distinguishing malignant from benign tumors.

Risk Stratification

The digital twin framework categorized patients into risk groups based on predicted probability scores. The risk categories included low risk, intermediate risk, and high risk.

Patients classified in the low-risk category demonstrated low probability of malignancy and were assigned to monitoring pathways. Intermediate-risk patients were recommended for additional diagnostic evaluation, whereas high-risk patients were mapped to treatment planning pathways. This risk stratification capability demonstrates that the digital twin model provides individualized patient assessment beyond binary classification.

Treatment Simulation Results

The digital twin framework simulated conceptual treatment pathways based on predicted risk categories. Low-risk patients were assigned to routine monitoring and follow-up screening. Intermediate-risk patients were recommended for diagnostic confirmation including imaging and biopsy. High-risk patients were mapped to treatment planning pathways including surgical consultation and systemic therapy consideration.

These simulation outputs provide individualized guidance for clinical decision-making and demonstrate integration of predictive modeling with treatment pathway simulation.

Clinical Decision Support Output

The digital twin generated patient-specific decision-support outputs including predicted diagnosis, probability score, risk category, and suggested clinical pathway. The proposed digital twin model demonstrated competitive accuracy and improved interpretability through integration of prediction and treatment simulation. Overall, the results indicate that the artificial intelligence-powered digital twin framework achieved high predictive accuracy and successfully generated individualized risk stratification and treatment simulation outputs for precision oncology applications.

DISCUSSION

The present study developed an artificial intelligence-powered digital twin framework for personalized breast cancer diagnosis and treatment simulation using a gradient boosting classifier. The proposed model demonstrated high diagnostic accuracy, strong discriminative capability, and effective risk stratification. Integration of machine learning prediction with virtual patient modeling represents an important advancement toward precision oncology-based clinical decision-support systems.[7-11]

With an AUC-ROC of 0.9945 and an overall accuracy of 97.37%, the model demonstrated exceptional classification performance. The model's high sensitivity (97.78%) demonstrates its capacity to accurately detect malignant tumours, which is essential for prompt diagnosis and treatment. In a similar vein, the specificity of 97.10% shows accurate benign case classification, which minimises needless diagnostic procedures. These results validate that gradient boosting methods may be successfully included into digital twin frameworks and are suitable for structured biomedical datasets.[7-9]

The proposed model's performance is on par with or superior to previously published machine learning models created with the Wisconsin Breast Cancer Dataset. Prior studies that employed logistic regression, support vector machines, and random forest classifiers found classification accuracy ranging from 93% to 98%.[30,37,45] However, most previous investigations focused only on prediction. In contrast, the present study integrates predictive modeling with digital twin-based simulation, thereby improving

Table 3: Confusion Matrix.

Type of tumor	Predicted Benign	Predicted Malignant
Actual Benign	60	1
Actual Malignant	2	51

Table 4: Feature Importance Ranking.

Rank	Feature	Importance
1	Worst Concave Points	0.42
2	Worst Area	0.18
3	Worst Perimeter	0.12
4	Worst Radius	0.09
5	Mean Concavity	0.07
6	Worst Texture	0.05
7	Mean Radius	0.03
8	Mean Area	0.02

clinical utility. The combination of prediction and simulation transforms a conventional classification model into a personalized decision-support system.

Digital twin technology provides several advantages compared with traditional artificial intelligence approaches. Conventional machine learning models generate static predictions, whereas digital twins create dynamic virtual patient representations. These virtual models enable simulation of disease progression, risk assessment, and treatment pathway evaluation. The digital twin framework developed in this study translates prediction outputs into individualized risk categories and simulated treatment pathways, improving interpretability and supporting precision oncology decision-making.[10-20]

The most significant predictors of malignancy, according to feature significance analysis, were nuclear morphological parameters and concavity-related traits. These findings are consistent with pathological studies that indicate uneven nuclear shape and structural distortion are associated with malignant transformation.[6,9] The agreement between model-derived predictors and biological understanding enhances interpretability and strengthens the reliability of the digital twin framework.

A crucial element of personalised medicine is risk classification. Three groups of patients were created: low-risk, intermediate-risk, and high-risk. using the suggested digital twin paradigm. This categorisation minimises needless actions for low-risk individuals while prioritising high-risk patients in need of immediate exam. The ability to stratify risk demonstrates that the digital twin extends beyond binary classification and supports individualized patient management.[9,19]

The treatment simulation component further improves clinical relevance. Based on predicted risk categories, the digital twin simulated conceptual treatment pathways including monitoring, diagnostic evaluation, and treatment planning. Although these simulations are not intended to replace clinical judgment, they provide decision-support guidance aligned with precision oncology principles. This capability may assist clinicians in evaluating management strategies and improving workflow efficiency.[24,25,31]

The digital twin architecture also demonstrates scalability and extensibility. While the present study utilized structured radiomic features, the framework can be expanded to incorporate multimodal data including genomic profiles, imaging data, histopathology findings, and longitudinal clinical records. Integration of multimodal data may improve predictive accuracy and enable dynamic updating of virtual patient models. Such capabilities are essential for real-world implementation of digital twins in oncology.[15,16,20,41-44]

Despite these advantages, there are a few drawbacks to take into account. As a benchmark dataset, the Wisconsin Breast Cancer Dataset might not accurately reflect clinical populations in the real world. To evaluate generalisability, independent datasets must be used for external validation.[34,38] Furthermore, the lack of genetic or longitudinal clinical data in the dataset restricts the ability to simulate the course of the disease over time. Before being implemented, the treatment simulation component must undergo prospective clinical validation.

Future research should focus primarily on integrating multimodal data sources, such as genetics, imaging, and electronic health records. To assess clinical value, prospective validation studies are required. Predictive performance and decision support may be further enhanced by real-time updating of digital twin models using patient-specific data. These advancements could make it possible to implement digital twin systems in standard healthcare procedures.[32-36,49]

Overall, the findings demonstrate that artificial intelligence-powered digital twin frameworks have strong potential in breast cancer precision oncology. Integration of gradient boosting prediction with virtual patient modeling enables personalized risk assessment and treatment simulation. This strategy is a significant advancement in the development of sophisticated clinical decision-support systems for the treatment of breast cancer.[25,48,49]

CLINICAL IMPLICATIONS

The proposed AI-powered digital twin paradigm has important implications for breast cancer screening and customised therapy planning. Precision oncology can be approached in a new way by combining virtual patient modelling with machine learning prediction. By creating individualised risk assessments and treatment simulations, the digital twin system can assist doctors in making well-informed decisions based on patient-specific characteristics.[10-16,25]

Increased diagnostic precision is one of this framework's main clinical consequences. The model's high sensitivity and specificity suggest that the digital twin can serve as a useful decision-support tool for the early identification of malignant tumours. Early detection of high-risk individuals is essential for managing breast cancer since prompt intervention increases survival rates. Therefore, the digital twin model may assist clinicians in prioritizing patients requiring urgent diagnostic evaluation.[9,19]

The risk stratification capability further enhances the clinical utility of the framework. Patients categorized as low risk may be managed with routine monitoring and follow-up, thereby reducing unnecessary invasive procedures. Intermediate-risk patients can be directed toward additional diagnostic evaluation including imaging and biopsy, whereas high-risk patients can be prioritized for treatment planning. This stratified approach aligns with personalized medicine principles and may improve healthcare resource utilization.[24,25]

The treatment simulation component provides an additional layer of clinical decision support. By mapping patients to conceptual treatment pathways, the framework assists clinicians in evaluating management strategies. Although the system does not replace clinical judgment, it offers evidence-based guidance that may improve workflow efficiency and decision-making.[9,19] This capability may be particularly useful in multidisciplinary tumor board discussions where treatment planning requires integration of multiple clinical parameters.

The digital twin concept may also be useful in precision pharmacology. Breast cancer is frequently treated with systemic therapies such as chemotherapy, hormone therapy, and targeted therapy.

Using patient-specific data, the digital twin architecture can be expanded to model drug response and toxicity risk. This kind of simulation capabilities could help minimise side effects and promote customised therapy selection. Pharmacogenomic data integration could improve treatment customisation even more.[24,32,33]

Another important implication is the use of digital twins in clinical education and training. Virtual patient models enable clinicians and trainees to simulate treatment scenarios and evaluate outcomes without risk to patients. This approach may improve understanding of disease progression and treatment response. Digital twin platforms may therefore serve as educational tools in oncology training programs.[34,39,40]

Additionally, the platform facilitates integration with healthcare systems and scalability. The architecture can be extended to include genomic data, imaging data, and electronic health records. Real-time updating of digital twin models would enable continuous monitoring of disease progression and treatment response. Such integration could transform digital twins into dynamic clinical decision-support systems for oncology practice.[20,41-44]

Overall, the artificial intelligence-powered digital twin framework demonstrates substantial clinical potential in breast cancer management. The ability to combine diagnostic prediction, risk stratification, and treatment simulation supports precision oncology and individualized patient care. To evaluate practical efficacy and integration into standard healthcare workflows, more clinical validation research is needed.[45-49]

LIMITATIONS

Although the proposed digital twin framework works well, there are a few disadvantages to consider. The work started with the Wisconsin Breast Cancer Dataset, a common dataset used in machine learning research. This dataset may not fully capture the complexity and diversity of real clinical populations, even if it was carefully chosen. Independent datasets must be used for external validation in order to confirm the model's generalisability.[18,6]

Second, the dataset contains only morphological features derived from fine needle aspiration images. Important clinical variables such as patient demographics, tumor staging, receptor status, genomic profiles, and treatment outcomes were not available. The absence of multimodal data limits the ability of the digital twin to simulate comprehensive disease behavior and treatment response.[41-44]

Third, the digital twin treatment simulation is conceptual and not intended as direct clinical recommendations. The simulated pathways are based on risk stratification rather than real treatment outcomes. To validate the simulation component and assess its clinical utility, prospective clinical investigations are necessary.[24,34]

The study's retrospective design is another drawback. Without a prospective assessment, the model was trained and validated using data that already existed.

Real-world deployment would require prospective validation and calibration using clinical datasets from multiple institutions. Additionally, model performance may vary across different populations and healthcare settings.[34,38]

The current framework also does not incorporate longitudinal data. A true digital twin should dynamically update using serial patient information, including follow-up imaging, laboratory values, and treatment response. The absence of longitudinal data limits the ability to simulate disease progression over time.[20,41]

Finally, interpretability and clinical integration remain challenges. Although feature importance analysis provides partial interpretability, further development of explainable artificial intelligence techniques is necessary. Integration with hospital information systems and clinical workflows will also require additional validation and regulatory considerations.[21,33,34]

FUTURE SCOPE

Future studies should concentrate on adding multimodal data sources to the suggested digital twin paradigm. Combining genetic profiles, radiological imaging, histopathology pictures, and clinical factors may increase tumour heterogeneity representation and predictive accuracy. Multimodal digital twin systems can capture complex disease characteristics and support more precise personalized treatment planning in breast cancer management.[11,41-44]

Prospective clinical validation represents an essential next step for translation into real-world practice. Evaluation of the digital twin framework in clinical settings would help determine its impact on diagnostic accuracy, treatment decision-making, and patient outcomes. Multicenter validation studies involving diverse patient populations may improve model generalizability, reliability, and clinical applicability.[24,34,38]

Real-time updating of digital twin models is another important future direction. Continuous integration of patient-specific data, including treatment response, follow-up imaging, and laboratory findings, would enable dynamic modeling of disease progression. Adaptive digital twins could support longitudinal monitoring, early detection of disease progression, and optimization of therapeutic strategies over time.[10-20]

Integration with precision pharmacology also represents a promising area for future research. Advanced digital twin models may simulate drug response, toxicity risk, and optimal dosing strategies based on patient-specific characteristics. Incorporation of pharmacogenomic data could further personalize therapy selection and minimize adverse drug reactions, thereby improving treatment efficacy and patient safety.[24,32,33]

Scalable clinical implementation may also be made possible by the deployment of digital twin systems on cloud-based platforms. Integration with electronic health records and hospital information systems would make normal clinical use easier. Adoption in healthcare settings will depend on clinician-interpretable dashboards, explainable AI outputs, and user-friendly interfaces.[21,33,34]

Overall, continued research, clinical validation, and technological advancements will enable translation of artificial intelligence-powered digital twin systems from computational prototypes to real-world precision oncology applications. Future advancements in digital twins could drastically change how breast cancer is diagnosed, tracked, and treated.[15,20,45-50]

CONCLUSION

The current work developed an artificial intelligence-powered digital twin architecture for personalised breast cancer diagnosis and therapy simulation using a gradient boosting machine learning technique. The proposed model demonstrated outstanding diagnostic performance with an accuracy of 97.37%, sensitivity of 97.78%, specificity of 97.10%, and AUC-ROC of 0.9945. These findings demonstrate that the model accurately distinguishes between benign and malignant breast tumours and support its application in precision oncology.

Integration of machine learning prediction with digital twin architecture enables transformation of patient-specific data into virtual representations capable of risk stratification and treatment pathway simulation. Unlike conventional classification models, the proposed framework generates individualized outputs that support personalized decision-making. The ability to categorize patients into risk groups and simulate management pathways highlights the clinical relevance of the digital twin approach.

The framework also demonstrates scalability and adaptability. Although the present study utilized structured radiomic features, the architecture can be extended to incorporate multimodal data including imaging, genomics, histopathology, and clinical records. Such integration may improve predictive accuracy and enable dynamic updating of virtual patient models. The digital twin approach therefore represents a promising foundation for future precision oncology systems.

Despite these benefits, prior to practical deployment, prospective clinical testing and external validation are required. The treatment simulation component remains conceptual and requires validation using real-world clinical outcomes. Future research should focus on multimodal data integration, explainable artificial intelligence development, and clinical effectiveness evaluation.

All things considered, the digital twin framework driven by artificial intelligence shows great promise for enhancing breast cancer diagnosis and assisting with individualised treatment planning. An essential step toward intelligent clinical decision-support systems and precision oncology applications is the integration of gradient boosting prediction with virtual patient modelling.

ACKNOWLEDGEMENT

The author sincerely thanks Dr. Surendra Gattani for his invaluable advice, unwavering support, and perceptive recommendations during the course of this research project. Additionally, the author thanks School of Pharmacy, Swami Ramanand Teerth Marathwada. University, Nanded for offering academic assistance and a favourable research setting so that this work could be finished. The institutional resources and support offered throughout the writing of this work are also appreciated by the author.

ABBREVIATIONS

AI: Artificial Intelligence; **AUC:** Area Under the Curve; **AUC-ROC:** Area Under the Receiver Operating Characteristic Curve; **DT:** Digital Twin; **GBM:** Gradient Boosting Machine; **KNN:** K-Nearest Neighbours; **ROC:** Receiver Operating Characteristic; **SVM:** Support Vector Machine; **WHO:** World Health Organization; **NCCN:** National Comprehensive Cancer Network.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTION

Pallavi Gondale contributed to concept development, digital twin model design, data analysis, and manuscript preparation. Dr. Surendra Gattani provided supervision, critical review, scientific guidance, and final approval of the manuscript.

SUMMARY

An AI-powered digital twin framework for breast cancer diagnosis and treatment simulation was developed.

• The gradient boosting model achieved 97.37% accuracy and an AUC-ROC of 0.9945.

• The framework enabled individualized risk stratification into low-, intermediate-, and high-risk groups.

• Patient-specific treatment pathway simulation supported precision oncology decision-making.

• The digital twin architecture demonstrated potential as a future clinical decision-support system.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide. CA Cancer J Clin. 2021;71(3):209-49.
2. World Health Organization. Breast cancer fact sheet [Internet]. Geneva: WHO; 2023.
3. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, *et al*. Breast cancer. Nat Rev Dis Primers. 2019;5:66.
4. Bray F, Laversanne M, Weiderpass E, Soerjomataram I. The global cancer burden 2020. Lancet Oncol. 2021;22(11):1493-502.
5. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, prospects. Science. 2015;349:255-60.
6. Erickson BJ, Korfiatis P, Akkus Z, *et al*. Machine learning for medical imaging. Radiographics. 2017;37(2):505-15.
7. Friedman JH. Greedy function approximation: gradient boosting machine. Ann Stat. 2001;29(5):1189-232.
8. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proc ACM SIGKDD Int Conf Knowl Discov Data Min. 2016:785-94.
9. McKinney SM, Sieniek M, Godbole V, *et al*. International evaluation of AI for breast cancer screening. Nature. 2020;577:89-94.
10. Laubenbacher R, Sluka JP, Glazier JA. Using digital twins in medicine. Nat Comput Sci. 2024;4(3):184-91.
11. Corral-Acero J, Margara F, Marciniak M, *et al*. Digital twin for healthcare. NPJ Digit Med. 2020;3:143.
12. Bruynseels K, Santoni de Sio F, van den Hoven J. Digital twins in health care. Nat Biotechnol. 2018;36:1139-47.
13. Giansanti D. Digital twins in oncology: from predictive modelling to personalised treatment strategies. Crit Rev Oncol Hematol. 2026;105171.
14. Viceconti M, Hunter P, Hose R. Big data, digital twins and *in silico* medicine. J R Soc Interface. 2016;13:20151089.
15. Olawade DB, Wada OZ, David-Olawade AC, *et al*. Digital twins in oncology: a narrative review. Front Digit Health. 2025;7:1583466.
16. Erickson BJ, Bartholmai B. Digital twins in medicine. Radiology. 2021;299:1-2.
17. Parmar C, Grossmann P, Bussink J, *et al*. Machine learning methods for radiomics. Sci Rep. 2015;5:13087.
18. Wolberg WH, Street WN, Mangasarian OL. Wisconsin Diagnostic Breast Cancer (WDBC) Dataset [Internet]. Irvine (CA): University of California, Irvine Machine Learning Repository; 1995 [cited 2026 Jun 23]. Available from: https://archive.ics.uci.edu
19. Hosny A, Parmar C, Quackenbush J, *et al*. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18:500-10.
20. Jiang Y, Yin S, Kaynak O. Data-driven monitoring for digital twins. IEEE Trans Ind Inform. 2021;17:1-11.
21. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765-74.
22. Murphy KP. Machine Learning: A Probabilistic Perspective. Cambridge: MIT Press; 2012.
23. Bishop CM. Pattern Recognition and Machine Learning. New York: Springer; 2006.
24. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Breast Cancer [Internet]. Version 3.2024. Plymouth Meeting (PA): NCCN; 2024 [cited 2026 Jun 23]. Available from: https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf
25. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.
26. Esteva A, Kuprel B, Novoa RA, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115-18.
27. Rajpurkar P, Irvin J, Ball RL, *et al*. Deep learning for chest radiograph diagnosis. PLoS Med. 2018;15(11):e1002686.
28. Ching T, Himmelstein DS, Beaulieu-Jones BK, *et al*. Opportunities and obstacles for deep learning in biology. J R Soc Interface. 2018;15(141):20170387.
29. Bibault JE, Giraud P, Burgun A. Big data and machine learning in radiation oncology. Lancet Oncol. 2016;17(12):e559-e66.
30. Kourou K, Exarchos TP, Exarchos KP, *et al*. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015;13:8-17.
31. Komorowski M, Celi LA, Badawi O, et al. The artificial intelligence clinician learns optimal treatment strategies. Nat Med. 2018;24:1716-20.
32. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317-18.
33. Char DS, Shah NH, Magnus D. Implementing machine learning in health care. N Engl J Med. 2018;378:981-83.
34. Sendak MP, D'Arcy J, Kashyap S, *et al*. A path for translation of machine learning into healthcare. NPJ Digit Med. 2020;3:17.
35. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436-44.
36. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge: MIT Press; 2016.
37. Breiman L. Random forests. Mach Learn. 2001;45:5-32.
38. Obermeyer Z, Emanuel EJ. Predicting the future-big data, machine learning. N Engl J Med. 2016;375:1216-19.
39. Litjens G, Kooi T, Bejnordi BE, *et al*. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60-88.
40. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221-48.
41. Wu J, Gensheimer MF, Dong X, et al. Radiomics in precision oncology. Radiology. 2020;295:328-38.
42. Lambin P, Rios-Velazquez E, Leijenaar R, *et al*. Radiomics: extracting more information from medical images. Eur J Cancer. 2012;48:441-46.
43. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures. Radiology. 2016;278:563-77.
44. van Griethuysen JJM, Fedorov A, Parmar C, *et al*. Computational radiomics system. Cancer Res. 2017;77:e104-e107.
45. Rodriguez-Ruiz A, Lang K, Gubern-Merida A, *et al*. Stand-alone AI for breast cancer detection. Nat Med. 2019;25:116-24.
46. Doi K. Computer-aided diagnosis in medical imaging. IEEE Trans Med Imaging. 2007;26(6):774-81.
47. Wang S, Summers RM. Machine learning and radiology. Med Image Anal. 2012;16(5):933-51.
48. Bi WL, Hosny A, Schabath MB, *et al*. Artificial intelligence in cancer imaging. CA Cancer J Clin. 2019;69:127-57.
49. Samei E, Badano A, Chakraborty DP, *et al*. Virtual clinical trials in medical imaging. J Med Imaging. 2020;7(4):042805.
50. Hinton G, Deng L, Yu D, *et al*. Deep neural networks for acoustic modeling. IEEE Signal Process Mag. 2012;29(6):82-97.

Cite this article: Gondale P, Gattani S. Artificial Intelligence Powered Digital Twin Framework for Personalized Breast Cancer Diagnosis and Treatment. Indian J of Pharmaceutical Education and Research. 2026; IPC Issue:28-37.