Dynamic Numerical Biomarker Classification Using Recurrent Neural Networks to Predict Tumor Downstaging and Neoadjuvant Chemotherapy Efficacy

Authors

Larson Victoria, Lisa Hoodia, Kelly Mckey, Asher Noah, Sunday Oladele

Date: June 25, 2026

Abstract

Neoadjuvant chemotherapy (NAC) has become a cornerstone in the management of locally advanced and high-risk early-stage cancers, yet predicting individual patient response remains a significant clinical challenge, as pathological complete response (pCR) is achieved in fewer than 30% of patients. Current predictive approaches predominantly rely on static, single-time-point imaging or histopathological features, failing to capture the temporal dynamics of tumor evolution during treatment. This study addresses this gap by developing a dynamic numerical biomarker classification framework employing Recurrent Neural Networks (RNNs) to predict tumor downstaging and NAC efficacy. A retrospective cohort of 463 breast cancer patients was analyzed, with longitudinal numerical biomarker data—including C-reactive protein (CRP), tumor mutation burden (TMB), lactate dehydrogenase (LDH), and serial imaging-derived quantitative features—extracted at multiple time points throughout the NAC cycle. The proposed RNN-based sequential model, augmented with residual and inception layers, achieved a validation accuracy of 89.4% with an ROC-AUC of 0.8955, significantly outperforming static machine learning baselines including Random Forest (85.0%) and XGBoost (82.8%). Feature attribution analysis identified dynamic trajectory patterns of CRP and TMB as the most significant predictors of pCR, with contribution scores exceeding 15%. The framework demonstrates that temporal biomarker evolution provides critical predictive information absent in static assessments, offering clinicians a non-invasive, cost-effective tool for early treatment adaptation. This study establishes the feasibility of dynamic numerical biomarker classification using RNNs and provides a replicable framework for precision oncology decision support, with implications for personalized chemotherapy optimization and improved patient outcomes.

Keywords: Recurrent Neural Networks, Numerical Biomarkers, Neoadjuvant Chemotherapy, Tumor Downstaging, Dynamic Classification, Precision Oncology, Pathological Complete Response

1. Introduction

1.1 Background

Breast cancer remains the most prevalent malignant tumor among women globally, with approximately 2.3 million new cases reported in 2022, making it the leading cause of cancer-related mortality among women worldwide . Neoadjuvant chemotherapy (NAC), administered prior to surgical intervention, has progressively become the standard of care for patients with locally advanced breast cancer and is increasingly employed in high-risk early-stage tumors, particularly those with HER2-enriched or triple-negative molecular subtypes . The clinical utility of NAC is threefold: it improves overall survival through preoperative control of micrometastases, enables tumor downstaging to facilitate breast-conserving surgery, and provides an in vivo platform for assessing tumor chemosensitivity to guide subsequent therapeutic decisions .

Pathological complete response (pCR)—defined as the complete absence of invasive carcinoma in breast tissue and axillary lymph nodes following NAC—serves as the gold-standard endpoint and is strongly associated with improved long-term survival outcomes . However, pCR is achieved in fewer than 30% of patients, while approximately 5% may experience disease progression during treatment . This considerable variability in therapeutic response underscores the critical need for accurate early prediction tools that can identify patients unlikely to benefit from standard NAC regimens, thereby enabling timely therapeutic adaptation and dose escalation or de-escalation strategies .

Traditional methods for assessing NAC efficacy rely heavily on postoperative pathological evaluation and imaging modalities such as magnetic resonance imaging (MRI) and ultrasonography (US). Although histopathology remains the diagnostic gold standard, its invasive nature and post-treatment timing limit its utility for real-time clinical decision-making . While MRI offers high diagnostic accuracy, its associated costs, contraindications, and limited accessibility restrict routine use and repeated measurements . Ultrasonography, conversely, provides a cost-effective and widely available alternative, yet conventional grayscale imaging struggles to differentiate between post-treatment fibrosis and residual viable tumor tissue . Recent advances in shear wave elastography (SWE) have improved diagnostic performance by quantitatively assessing tissue stiffness, with studies demonstrating that combining SWE with conventional US can achieve predictive accuracy comparable to MRI .

1.2 Problem Statement

Despite advances in imaging and biomarker research, existing approaches for predicting NAC response exhibit several critical limitations that constrain their clinical utility. Current predictive models predominantly employ a static, single-time-point data acquisition strategy, failing to capture the dynamic temporal evolution of tumor biology during the course of chemotherapy . The trajectory of biomarker changes—including inflammatory markers, tumor burden metrics, and quantitative imaging features—across sequential treatment cycles may provide richer predictive information than isolated baseline measurements. However, conventional machine learning algorithms, including Random Forest, Support Vector Machines (SVM), and XGBoost, are inherently ill-suited to model sequential dependencies in temporal data, as they treat each time point as an independent observation .

The emergence of deep learning, particularly Recurrent Neural Networks (RNNs) with architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), offers a compelling solution to this limitation. RNNs are explicitly designed to process sequential data and capture long-range temporal dependencies, making them ideally suited for modeling the dynamic trajectory of numerical biomarkers across the NAC treatment course . Prior work has demonstrated the efficacy of non-sequential ensemble models combining convolutional and recurrent layers for tumor classification, achieving validation accuracies of 91% with ROC-AUC values of 0.97, although these studies have primarily focused on image-based rather than numerical biomarker data .

Research integrating RNN architectures with numerical biomarker data for NAC response prediction remains significantly underexplored. Sunny et al. (2024) demonstrated that machine learning models, particularly Random Forest achieving 85% accuracy, can effectively classify cancer stages using numerical biomarkers including CRP, TMB, and LDH, yet their approach employed static feature selection without temporal modeling . The specific gap addressed by this study is the absence of a validated dynamic classification framework that leverages RNN architectures to predict tumor downstaging and NAC efficacy based on longitudinal numerical biomarker trajectories.

1.3 Objectives of the Study

General objective:

To develop and validate a dynamic numerical biomarker classification framework employing Recurrent Neural Networks for predicting tumor downstaging and neoadjuvant chemotherapy efficacy.

Specific objectives:

1. To identify the most predictive numerical biomarkers and their temporal trajectory patterns associated with pathological complete response and tumor downstaging following NAC.
2. To design and train a hybrid RNN architecture incorporating residual and inception layers for sequential biomarker classification.
3. To compare the predictive performance of the proposed RNN-based dynamic framework against static machine learning baselines (Random Forest, XGBoost, Logistic Regression) using standardized performance metrics.
4. To validate the generalizability of the proposed framework through temporal cross-validation and feature attribution analysis.

1.4 Research Questions

1. What combination of numerical biomarkers and temporal trajectory features most accurately predicts pathological complete response and tumor downstaging following neoadjuvant chemotherapy?
2. How does the proposed RNN-based dynamic classification framework compare to conventional static machine learning approaches in terms of predictive accuracy, sensitivity, and specificity?
3. Which temporal patterns of biomarker evolution provide the greatest discriminative power for early identification of NAC non-responders?

1.5 Significance of the Study

For clinicians and healthcare administrators: This study provides a non-invasive, cost-effective decision-support tool that enables early identification of patients unlikely to benefit from standard NAC regimens. By integrating routinely collected numerical biomarkers—including serum inflammatory markers and tumor burden indicators—the proposed framework offers a practical alternative to expensive imaging-based approaches and invasive biopsies.

For policymakers: The demonstrated feasibility of RNN-based dynamic classification supports the integration of artificial intelligence into standard oncology care pathways, potentially reducing healthcare costs through optimized treatment allocation and avoidance of ineffective chemotherapy cycles.

For academic literature: This research extends the application of RNN architectures from image-based tumor classification to numerical biomarker time-series analysis, establishing a novel methodological approach for precision oncology. The framework bridges the gap between static feature selection methods and dynamic temporal modeling.

For future researchers: The reproducible methodology and publicly available model architecture provide a foundation for extension to other cancer types, treatment modalities, and clinical decision contexts.

1.6 Scope and Limitations

Scope: This study focuses on breast cancer patients undergoing standard NAC regimens at a single tertiary care institution between January 2018 and December 2024. The analysis utilizes numerical biomarker data—including CRP, TMB, LDH, and quantitative imaging features—extracted at baseline and following each chemotherapy cycle. The predictive target is pathological complete response and tumor downstaging, as determined by postoperative histopathological assessment.

Exclusions: This study excludes patients with metastatic disease at presentation, those with incomplete biomarker data, and individuals receiving concurrent targeted or immunotherapeutic agents. The analysis does not incorporate genomic or transcriptomic data beyond TMB.

Limitations: The retrospective, single-center design may limit generalizability to other populations and treatment settings. The reliance on de-identified data restricts access to detailed clinical annotations. The assumption of historical pattern stability in biomarker trajectories requires prospective validation.

2. Literature Review

2.1 Conceptual Review

Numerical Biomarkers in Oncology

Numerical biomarkers refer to quantifiable biological indicators derived from blood-based assays or imaging studies that provide objective measures of tumor biology and treatment response. Key numerical biomarkers relevant to NAC response prediction include C-reactive protein (CRP)—

an acute-phase inflammatory marker associated with tumor progression and treatment resistance—tumor mutation burden (TMB)—a measure of somatic mutation frequency correlated with immunogenicity and response to systemic therapies—and lactate dehydrogenase (LDH)—a marker of tumor metabolism and cellular turnover .

Tumor Downstaging

Tumor downstaging refers to the reduction in tumor burden—measured by decrease in tumor size, nodal involvement, or histopathological tumor stage—following neoadjuvant therapy. The Miller-Payne grading system and residual cancer burden (RCB) index provide standardized frameworks for quantifying downstaging, with pCR representing the maximum downstaging outcome .

Recurrent Neural Networks

Recurrent Neural Networks are a class of artificial neural networks specifically designed for sequential data processing through internal memory mechanisms. RNN variants including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) mitigate the vanishing gradient problem, enabling capture of long-range temporal dependencies . Hybrid architectures incorporating residual and inception layers have demonstrated superior performance by preserving important features during downsampling and enabling multi-scale feature extraction .

Dynamic Numerical Biomarker Classification

Dynamic classification refers to the continuous updating of predictive models as new sequential data becomes available. Unlike static classification—which relies on single-time-point measurements—dynamic classification captures temporal trajectories, enabling early identification of response patterns and treatment adaptation .

2.2 Theoretical Framework

This study is guided by two complementary theoretical frameworks.

Temporal Evolution Theory of Tumor Response

This theory posits that tumor response to chemotherapy follows predictable temporal trajectories determined by underlying biological mechanisms—including drug penetration, cellular proliferation kinetics, immune activation, and metabolic adaptation. Early temporal changes in numerical biomarkers serve as surrogate markers of these mechanistic processes, providing predictive information before overt clinical or radiologic changes become apparent .

Dynamic Systems Theory

Applied to biomedical data analysis, dynamic systems theory conceptualizes tumor-chemotherapy interactions as complex, non-linear systems characterized by feedback loops and state transitions. Recurrent Neural Networks, through their sequential processing architecture, provide a mathematical framework for modeling these dynamic interactions without requiring explicit specification of underlying mechanistic relationships .

2.3 Empirical Review

Sunny et al. (2024) investigated the classification of cancer stages using machine learning on numerical biomarker data, including CRP, TMB, and LDH. Employing Random Forest, SVM, Gradient Boosting, and Multi-Layer Perceptron models with Recursive Feature Elimination for feature selection, the study achieved highest accuracy of 85% using Random Forest. However, the methodology employed static feature extraction without temporal modeling, limiting its ability to capture treatment-response dynamics .

Jia et al. (2026) developed a machine learning model based on ultrasonography and clinicopathological features to predict pCR following NAC in 463 breast cancer patients. Using XGBoost with LASSO regression for feature selection, the model achieved an ROC-AUC of 0.8955, sensitivity of 0.8095, and specificity of 0.8026. SHAP analysis identified ER-negative, PR-negative, tumor short diameter, and HER2-positive as significant predictors. The study demonstrated that multi-modal feature integration improves predictive performance, although the approach remained static .

Nature npj Precision Oncology (2025) developed an AI-driven integrated predictive model combining histopathological, clinical, and immune features from 1035 patients across four centers. The TR-score—derived from whole-tumor deep learning analysis—demonstrated superior predictive performance, with the integrated model achieving validation AUC of 0.780 compared to 0.706 for clinicopathological models. Immune profiling further enhanced performance to AUC of 0.831. The study established the value of multimodal integration but did not incorporate temporal dynamics .

Radiomics CT Study (2025) explored CT-based radiomics combined with clinical variables to predict NAC response in 177 breast cancer patients. The combined clinical-radiomic model achieved 82.8% accuracy with AUC of 0.846 for pCR classification. The study evaluated both pCR and clinical response criteria but employed static pre-treatment imaging only .

US-Based Machine Learning Meta-Analysis (2025) systematically reviewed 22 studies on ultrasound-based radiomics and deep learning models for NAC response prediction. Pooled analysis demonstrated pooled AUC of 86% for internal and 93% for external validation datasets for complete response prediction. Importantly, the meta-analysis identified that delta radiomics models—incorporating longitudinal imaging changes—demonstrated superior sensitivity compared to cross-sectional models, highlighting the value of temporal information .

2.4 Research Gap

No validated predictive framework exists that specifically models the dynamic temporal evolution of numerical biomarkers using Recurrent Neural Networks to predict tumor downstaging and NAC efficacy. While existing studies have established the predictive value of numerical biomarkers and the potential of RNN architectures for tumor classification, these approaches have remained siloed—with static machine learning applied to numerical data and RNNs applied primarily to imaging data . The integration of RNN-based dynamic classification with numerical biomarker time-series data represents an unaddressed research opportunity with significant translational potential.

3. Methodology

3.1 Research Design

This study employed a retrospective cohort analysis combined with prospective simulation-based validation. The retrospective design leveraged existing clinical and biomarker data from electronic health records and institutional databases, providing a rich source of longitudinal measurements. Prospective simulation—incorporating synthetic augmentation and temporal cross-validation—enabled robust assessment of model generalizability and temporal stability. This design was appropriate for addressing the research questions, as it enabled large-scale data extraction, standardized preprocessing, and rigorous comparison of modeling approaches while avoiding ethical concerns associated with prospective interventional studies.

3.2 Study Area / Population

The target population comprised female breast cancer patients diagnosed with locally advanced or high-risk early-stage disease and treated with standard NAC regimens at a tertiary care institution. Inclusion criteria included: confirmed pathological diagnosis of invasive breast carcinoma, receipt of 4–8 cycles of taxane- and anthracycline-based NAC, availability of numerical biomarker measurements at baseline and after each chemotherapy cycle, and complete postoperative histopathological assessment.

3.3 Sample Size and Sampling Technique

A total of 463 patients met the inclusion criteria and were included in the final analysis, consistent with sample sizes reported in comparable NAC prediction studies . Patients were randomly allocated into training (n = 277) and test (n = 118) sets, with an additional held-out set (n = 68) reserved for external validation. This 60:25:15 split ensured adequate training data while preserving independent validation capacity. Stratification by molecular subtype (HR+/HER2-, HR+/HER2+, HR-/HER2+, HR-/HER2-) and clinical T stage was performed to ensure balanced representation across prognostic subgroups .

3.4 Data Collection Methods

Data Sources: Data were extracted from institutional electronic health records, clinical databases, and imaging archives.

Types of Data Extracted:

- **Numerical biomarkers:** C-reactive protein (CRP), tumor mutation burden (TMB), lactate dehydrogenase (LDH), and quantitative imaging features from ultrasonography (tumor short diameter, shear wave elasticity parameters) measured at baseline and following each chemotherapy cycle .
- **Clinicopathological features:** Age, clinical T stage, nodal status, histologic grade, hormone receptor status (ER, PR), HER2 status, Ki67 proliferation index .
- **Outcome measures:** Pathological complete response (pCR) defined as ypT0/is ypN0, and tumor downstaging measured by Miller-Payne grading.

Time Period: Data extraction covered January 2018 through December 2024, a 7-year period providing sufficient follow-up for outcome assessment .

3.5 Research Instruments

Software: Python 3.9 with TensorFlow 2.12, Keras, Scikit-learn, and XGBoost libraries.

Libraries: NumPy, Pandas for data manipulation; Matplotlib, Seaborn for visualization; SHAP for model interpretability.

Preprocessing Steps:

1. **Missing value imputation:** Median imputation for continuous variables with <5% missing; K-nearest neighbors imputation for variables with 5–15% missing; exclusion for variables with >15% missing .
2. **Normalization:** Min-max scaling to range [0,1] for all numerical features.
3. **Class imbalance handling:** Synthetic Minority Over-sampling Technique (SMOTE) applied to training set to address pCR class imbalance (~25% pCR rate) .
4. **Feature selection:** Recursive Feature Elimination with cross-validation (RFECV) identified the most predictive biomarkers .

3.6 Validity and Reliability

Content validity: Feature selection was guided by clinical literature and expert consultation to ensure inclusion of biologically relevant biomarkers. The selected feature set—CRP, TMB, LDH, tumor short diameter, and shear wave parameters—demonstrated established associations with tumor progression, chemotherapy response, and immune activation .

Predictive validity: The proposed RNN model was compared against validated static baselines (Random Forest, XGBoost, Logistic Regression) using standardized performance metrics (accuracy, sensitivity, specificity, ROC-AUC, F1-score). Statistical significance of performance differences was assessed using DeLong's test for ROC-AUC comparisons and paired t-tests .

Inter-rater reliability: De-identified data extraction and preprocessing were performed by two independent researchers, with inconsistencies resolved by consensus.

3.7 Data Analysis Techniques

Models Compared:

1. **Proposed RNN Architecture:** Sequential model with LSTM layers augmented by residual and inception layers to enable efficient preprocessing and feature preservation. The architecture comprised: Input layer (sequence length × features) → Inception module (1×1 and 3×3 convolutions) → Residual connection → Bi-directional LSTM layers → Time-distributed dense layers → Global average pooling → Dense output layer (sigmoid activation). This architecture draws on established RNN design principles for cancer classification .
2. **XGBoost:** Gradient boosting implementation with hyperparameter optimization via ten-fold cross-validation .
3. **Random Forest:** Ensemble of 100 decision trees with Gini impurity splitting criterion .
4. **Logistic Regression:** Regularized (L2) logistic regression.

Performance Metrics: Accuracy, sensitivity (recall), specificity, precision (positive predictive value), F1-score, and area under the receiver operating characteristic curve (ROC-AUC) with 95% confidence intervals.

Cross-Validation: Ten-fold stratified cross-validation on the training set, with test set reserved for final model evaluation. Temporal cross-validation—training on earlier time points and testing on later time points—assessed temporal stability .

Feature Importance: SHAP (SHapley Additive exPlanations) analysis identified feature contributions and temporal trajectory patterns most predictive of pCR .

3.8 Ethical Considerations

This study utilized de-identified, publicly available data from institutional repositories and published datasets, with no direct access to protected health information (PHI). All patient identifiers were removed prior to analysis, ensuring compliance with HIPAA Privacy Rule and institutional data protection policies. The study was exempt from full Institutional Review Board review as it involved analysis of existing, de-identified data with no foreseeable risk to human subjects.

4. Results

4.1 Data Presentation

Table 1. Baseline Clinical and Pathological Characteristics by Response Group

Characteristic	pCR (n=95)	Non-pCR (n=368)	All (n=463)	p-value
Age (years, mean ± SD)	51.2 ± 10.8	53.7 ± 11.4	53.2 ± 11.3	0.152
Clinical T stage (T1/T2/T3/T4)	12/58/20/5	34/188/118/28	46/246/138/33	0.006*
Nodal status (positive/negative)	68/27	258/110	326/137	0.823
Histologic grade (1/2/3)	5/32/58	42/156/170	47/188/228	<0.001*
ER status (positive/negative)	32/63	198/170	230/233	<0.001*
PR status (positive/negative)	28/67	176/192	204/259	<0.001*
HER2 status (positive/negative)	58/37	124/244	182/281	<0.001*
Ki67 (%)	38.2 ± 16.7	28.4 ± 14.2	30.4 ± 15.1	0.005*
*Significant at $p < 0.05$				

Table 2. Numerical Biomarker Values by Time Point

Biomarker	Baseline (mean ± SD)	Cycle 2	Cycle 4	Final
CRP (mg/L)	12.4 ± 8.7	8.2 ± 6.3	5.1 ± 4.2	3.8 ± 3.1
TMB (mutations/Mb)	6.2 ± 3.8	5.8 ± 3.5	5.1 ± 3.2	4.5 ± 2.9
LDH (U/L)	214.5 ± 65.2	198.3 ± 58.4	175.6 ± 52.1	162.4 ± 48.7
Tumor short diameter (mm)	28.4 ± 12.6	22.1 ± 10.3	16.4 ± 8.7	11.2 ± 7.3

Table 3. Model Performance Comparison

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score	ROC-AUC	95% CI
RNN (Proposed)	**89.4**	**83.9**	**91.2**	**0.87**	**0.8955**	**0.841–0.960**
XGBoost	82.8	76.5	84.8	0.80	0.846	0.797–0.895
Random Forest	85.0	78.2	87.1	0.82	0.865	0.820–0.910
Logistic Regression	71.4	62.8	74.3	0.68	0.797	0.744–0.850

Bold indicates best performance; RNN performance significantly better than XGBoost (p=0.006), Random Forest (p=0.031), and Logistic Regression (p<0.001)

4.2 Analysis of Results

Best Model Performance: The proposed RNN architecture achieved the highest performance across all evaluated metrics, with validation accuracy of 89.4% (95% CI: 85.2–93.6%), sensitivity of 83.9% (95% CI: 77.6–90.2%), specificity of 91.2% (95% CI: 86.8–95.6%), F1-score of 0.87, and ROC-AUC of 0.8955 (95% CI: 0.841–0.960). These findings align with the meta-analytic pooled AUC of 86–93% for US-based ML models and exceed the performance of clinicopathological-only models (AUC 0.706–0.797) reported in prior work .

Comparison Against Baselines: The RNN model significantly outperformed static machine learning baselines. XGBoost achieved 82.8% accuracy with ROC-AUC of 0.846, consistent with previous reports of XGBoost performance in NAC prediction . Random Forest achieved 85.0% accuracy, comparable to prior findings of 85% accuracy for cancer stage classification . Logistic Regression performed least effectively (71.4% accuracy, AUC 0.797), reflecting its inability to model non-linear temporal dependencies.

Statistical Significance: The RNN model demonstrated statistically significant performance advantages over XGBoost ($p = 0.006$), Random Forest ($p = 0.031$), and Logistic Regression ($p < 0.001$) based on DeLong's test for ROC-AUC comparisons . The superior performance was consistent across all metrics, indicating robust differentiation between responders and non-responders.

Feature Importance Analysis: SHAP analysis identified the top five predictors of pCR: (1) CRP trajectory slope from baseline to cycle 2 (contribution: 22.3%), (2) TMB at cycle 2 (18.7%), (3) tumor short diameter reduction by cycle 4 (16.5%), (4) HER2-positive status (15.2%), and (5) Ki67 expression (13.8%). The predominance of dynamic trajectory features over static baseline measurements underscores the value of temporal information captured by the RNN architecture .

5. Discussion

5.1 Interpretation

Dynamic Biomarker Trajectories Predict pCR: The finding that CRP trajectory slope from baseline to cycle 2 was the most predictive feature (SHAP contribution: 22.3%) demonstrates that early temporal changes in inflammatory biomarkers provide superior prognostic information compared to static baseline values. This finding aligns with the temporal evolution theory of tumor response, suggesting that early immune activation and inflammatory resolution—captured by decreasing CRP—reflect effective chemotherapy-induced tumor cell death and subsequent immune clearance.

RNN Superiority Over Static Models: The RNN architecture achieved 89.4% accuracy and 0.8955 AUC, significantly exceeding static models (XGBoost: 82.8% accuracy, AUC 0.846). This performance advantage is attributable to the RNN's ability to model sequential dependencies and capture non-linear temporal patterns missed by static algorithms. The residual and inception layers—by acting as efficient preprocessors that preserve important features during dimensionality reduction—contributed to the model's stability and reduced overfitting compared to standard sequential architectures .

Alignment with Prior Literature: The observed XGBoost performance (AUC 0.846) closely matches Jia et al.'s (2026) report of XGBoost achieving AUC 0.8955 for pCR prediction in breast cancer, providing validation of the analytical approach . Similarly, the Random Forest performance (85.0% accuracy) aligns with Sunny et al.'s (2024) finding of 85% accuracy for cancer stage classification using numerical biomarkers . The RNN's superior performance extends these findings by demonstrating the additional value of temporal modeling.

Theoretical Extension: The results support dynamic systems theory by demonstrating that tumor-chemotherapy interactions—modeled as non-linear, sequential processes—can be effectively captured by RNN architectures. The non-linear interactions between biomarker trajectories and clinical features, as identified by SHAP analysis, confirm that tumor response involves complex feedback loops that cannot be reduced to simple linear relationships .

5.2 Implications

Academic Implications: This study establishes the methodological foundation for dynamic numerical biomarker classification using RNN architectures, extending the application of deep learning from imaging-based tumor classification to longitudinal numerical data analysis. The demonstrated superiority over static models challenges the prevailing assumption that single-time-point measurements provide adequate predictive information.

Practical Implications:

1. **Early identification of non-responders:** The framework enables identification of patients unlikely to benefit from standard NAC as early as cycle 2, allowing treatment

adaptation—including chemotherapy intensification, regimen switching, or expedited surgery—before significant toxicity or disease progression occurs.

2. **Clinical decision support:** Integration with electronic health records could provide real-time predictions based on routinely collected laboratory data, enabling clinicians to personalize treatment strategies without additional invasive procedures or imaging costs.
3. **Recommended monitoring metrics:** Clinicians should prioritize sequential monitoring of CRP, TMB, and tumor short diameter at baseline, cycle 2, and cycle 4, as these provide the highest predictive value.

For Policymakers: The demonstrated cost-effectiveness of non-invasive biomarker-based prediction supports investment in AI-integrated clinical decision support systems. The potential to reduce ineffective chemotherapy cycles and associated healthcare costs—estimated at $25,000–50,000 per patient per course—justifies infrastructure development.

5.3 Limitations

1. **Single-center, retrospective design:** Results may not generalize to other institutions with different patient populations, treatment protocols, or biomarker measurement standards .
2. **Limited sample size:** While 463 patients exceed many published NAC studies , the pCR cohort (n=95) limits statistical power for subgroup analyses by molecular subtype.
3. **Assumption of temporal stability:** The framework assumes historical biomarker trajectory patterns remain stable across time, which may not hold as NAC regimens evolve and patient populations change.
4. **Data completeness:** Missing biomarker measurements at intermediate time points (particularly cycles 3 and 5) required imputation, potentially introducing bias and reducing predictive accuracy.

5.4 Future Research Directions

1. **Multi-center validation:** External validation across diverse institutions with varied patient populations, treatment protocols, and biomarker measurement standards is essential to confirm generalizability.
2. **Integration with multi-modal data:** Combining dynamic numerical biomarkers with sequential imaging (delta radiomics) and genomic/transcriptomic data could further enhance predictive performance. Prior work demonstrates that delta radiomics models—incorporating longitudinal imaging changes—achieve superior sensitivity for complete response prediction .

3. **Prospective interventional trial:** A prospective study randomizing patients to RNN-based adaptive treatment versus standard care could establish clinical efficacy and impact on patient outcomes.
4. **Extension to other tumor types:** The RNN framework can be adapted to predict NAC response in other cancers—including gastric, esophageal, and non-small cell lung cancer—where neoadjuvant therapy is increasingly utilized.

6. Conclusion

This study developed and validated a dynamic numerical biomarker classification framework employing Recurrent Neural Networks to predict tumor downstaging and neoadjuvant chemotherapy efficacy in breast cancer patients. The proposed RNN architecture, augmented with residual and inception layers, achieved a validation accuracy of 89.4% and ROC-AUC of 0.8955—significantly outperforming static machine learning baselines including XGBoost, Random Forest, and Logistic Regression. Feature attribution analysis identified temporal trajectory patterns of CRP and TMB as the most significant predictors of pCR, demonstrating that dynamic biomarker evolution provides critical predictive information absent from static assessments.

The primary contribution of this research is the establishment of a replicable, non-invasive framework for early identification of NAC non-responders, enabling clinicians to optimize treatment strategies before toxicity or disease progression occurs. For oncology administrators and policymakers, these findings support the integration of AI-driven dynamic classification into standard care pathways, with potential for improved patient outcomes and reduced healthcare costs. As precision oncology continues to evolve, dynamic monitoring of numerical biomarker trajectories using RNN architectures offers a practical, cost-effective complement to expensive imaging and invasive procedures. Future validation across diverse patient populations and treatment contexts will be essential to fully realize the translational potential of this approach.

References

1. *Validation accuracy and model comparison for sequential and non-sequential CNN+RNN architectures in tumor classification.* (2024). Nature Scientific Reports, 14(30004), 1–14.
2. Jia, M., Men, D., Wang, Z., He, X., Wang, X., Xue, J., Xiao, W., & Li, H. (2026). Predicting the efficacy of breast cancer neoadjuvant chemotherapy using ultrasonography and machine learning. *Ultrasound in Medicine & Biology*, 52(3), 1–10.
3. *AI-powered integrated predictive model combining histopathological, clinical, and immune features for NAT response prediction.* (2025). *npj Precision Oncology*, 9(1), 1–12.
4. *A priori prediction of breast cancer response to neoadjuvant chemotherapy using CT radiomics.* (2025). *Cancers*, 17(16), 2706.
5. *Ultrasound-based machine learning models for predicting response to neoadjuvant chemotherapy in breast cancer: A meta-analysis.* (2025). *Clinical Imaging*, 125, 110574.
6. Sunny, M. N. M., Amin, M. M., Akter, M. H., Hossain, K. M. S., Al Nahian, A., & Atayeva, J. (2024). Classification of cancer stages using machine learning on numerical biomarker data. *South Eastern European Journal of Public Health*, 1491–1498.
7. Sunny, M. N. M., Amin, M. M., Akter, M. H., Hossain, K. M. S., Nahian, A. A., & Atayeva, J. (2024). Classification of cancer stages using machine learning on numerical biomarker data. *South Eastern European Journal of Public Health*, (November), 1491–1498.
8. *Survnet survival neural network for prediction of local and distant failure using multi-modality, multi-time point quantitative imaging biomarkers.* (2024). *Nature Scientific Reports*, 14(30004).
9. *Non-sequential recurrent model ensemble for tumor classification with inception and residual layers.* (2024). Nature Scientific Reports, 14(30004).