Survival Analysis of Cancer Patients Based on Chemotherapy Response Data

Othman A. Hanshal[1], Marwah Sami Kzar[2], Bushra Jabbar Abdul-Kareem[3] and Anil Kumar Sagar[4]
[1]*Al-Turath University, 10013 Baghdad, Iraq*
[2]*College of Pharmacy, Al-Farahidi University, 10065 Baghdad, Iraq*
[3]*Department of Computer Engineering, College of Engineering, Al-Mansour University College, 10067 Baghdad, Iraq*
[4]*Sharda School of Computing Science and Engineering, Sharda University, 201310 Greater Noida, India*
othman.araf@uoturath.edu.iq, marwa.kezar@uoalfarahidi.edu.iq, bushra.jabbar@muc.edu.iq, anil.sagar@sharda.ac.in

Keywords: Survival Analysis, Chemotherapy Response, Cox Regression, Kaplan-Meier, Accelerated Failure Time, Random Survival Forest, Oncology, Prognosis.

Abstract: Background: Chemotherapy is still a key part of cancer treatment, but survival rates vary a lot and are mostly based on how well the treatment works. It is important to figure out how important response is for predicting outcomes and to use modern computational methods to improve patient stratification. To assess the influence of chemotherapy response on overall survival (OS) and progression-free survival (PFS) utilizing both classical and advanced survival models. A retrospective cohort study involving 420 cancer patients was conducted. According to RECIST v1.1, the response was classified as a complete response (CR), a partial response (PR), stable disease (SD), or progressive disease (PD). We used Kaplan-Meier and Cox regression to figure out how long people would live, and we used Accelerated Failure Time (AFT) and Random Survival Forest (RSF) models to check the results for sensitivity and robustness. CR and PR were independently correlated with markedly extended overall survival (OS) and progression-free survival (PFS) in contrast to progressive disease (PD). RSF showed better predictive accuracy (C-index=0.82) than Cox (C-index=0.76). Chemotherapy response is a strong sign of how long someone will live. The incorporation of sophisticated computational models improves forecasting and facilitates clinical application.

1 INTRODUCTION

Cancer is still one of the most common causes of illness and death around the world, and chemotherapy is still an important part of treatment plans for many types of tumors. Even though targeted therapies and immuno-oncology have come a long way, regular chemotherapy is still very important, especially in cases where cancer has spread or is locally advanced. But the survival rates after chemotherapy are different for everyone, and the amount of treatment response has a big effect on how different they are. Consequently, it is essential to establish strong evidence regarding the prognostic significance of chemotherapy response to enhance patient care and inform the design of future clinical trials.

The correlation between chemotherapy response and survival has been extensively examined in breast cancer. For instance, Romero et al. (2013) [1] showed that the survival rates of patients with locally advanced breast cancer were strongly linked to how well they responded to neoadjuvant chemotherapy. Likewise, Bonnefoi et al. (2014) [2] emphasized that pathological complete response (pCR) following neoadjuvant treatment functioned as an independent predictor of long-term outcomes across various breast cancer subtypes. Zhao et al. (2015) [3] corroborated these findings through the analysis of molecular subtypes, establishing that patients who experienced considerable tumor reduction demonstrated markedly enhanced prognoses. Collectively, these studies highlight the significance of chemotherapy response as a surrogate endpoint for overall and progression-free survival.

Chemotherapy response has demonstrated prognostic significance in lung cancer and other solid tumors, in addition to breast cancer. Lu et al. (2020) [4] created a multicenter prognostic model for early-stage non-small cell lung cancer (NSCLC), showing that treatment response along with clinical covariates could effectively group patients by their risk of survival. These results underscore the

importance of incorporating response data into prognostic models for various cancer types. Liu et al. (2025) [5] recently reported long-term survival outcomes for stage II-III hormone receptor-positive/HER2-negative breast cancer patients, demonstrating that chemotherapy response significantly affected both disease-free and overall survival, thereby reinforcing its status as an independent prognostic biomarker.

Despite consistent evidence, numerous limitations endure in the current literature. First, the majority of previous studies have been site-specific, limiting their applicability across various tumor types. Second, numerous studies have depended solely on conventional biostatistical models, including Kaplan-Meier estimations and Cox proportional hazards regression, which may inadequately address intricate data heterogeneity or temporal effects. There are still problems with securely sharing sensitive patient data across multicenter cohorts, which makes it hard to do large-scale validation. Wang et al. (2025) [6] contend that next-generation computing paradigms for secure data sharing are crucial to surmount these obstacles, safeguarding data privacy while facilitating collaborative survival analyses.

At the same time, improvements in artificial intelligence (AI) are changing the way healthcare research is done. AI-powered systems can help people make decisions by finding complicated nonlinear relationships in big data sets. Mehta and Rani (2025) [7] emphasized the implementation of AI-driven frameworks in human-computer interaction, demonstrating their capacity to enhance data interpretation and clinical decision support. The amalgamation of AI with survival modeling presents novel prospects for the refinement of prognostic instruments, the augmentation of reproducibility, and the conversion of response-driven predictions into tailored therapeutic approaches. In light of these gaps and opportunities, the current study aims to investigate the survival of cancer patients categorized by chemotherapy response, utilizing rigorous survival analysis techniques while addressing computational frameworks for scalability and secure data sharing. The objectives are threefold: first, to compare survival distributions across chemotherapy response categories; second, to quantify the independent prognostic effect of response using multivariable time-to-event models; and third, to explore the translational value of incorporating AI and advanced computing in survival prediction. The study aims to integrate clinical oncology with computational innovations, thereby enhancing patient stratification and promoting evidence-based cancer care.

2 LITERATURE REVIEW

In recent years, the methods used in oncology survival analysis have changed a lot. Researchers are now using more advanced methods to get a better picture of how treatment response and long-term outcomes affect each other. Conventional statistical models, including the Cox proportional hazards regression, are still prevalent; however, recent evidence indicates that machine learning techniques may offer enhanced predictive capabilities. For instance, Jin et al. (2025) [8] conducted a comparison between random survival forests (RSF) and Cox regression in breast cancer patients who did not attain pathological complete response (pCR). Their findings indicated that RSF models more effectively captured nonlinearity and intricate covariate interactions, underscoring the promise of machine learning in survival prediction.

The prognostic relevance of pCR remains a focal point of significant interest. Antonini et al. (2025) [9] performed a systematic review and meta-analysis of real-world data, validating that pCR continues to serve as a reliable surrogate endpoint for survival in breast cancer patients undergoing neoadjuvant chemotherapy. These results corroborate previous clinical trial outcomes and validate pCR as an effective intermediate marker for assessing treatment efficacy. Likewise, Deutsch et al. (2024) [10] expanded this evidence to lung cancer, indicating that patients who attained a pathologic response following neoadjuvant chemo-immunotherapy demonstrated markedly prolonged overall survival. Collectively, these studies highlight the applicability of the response-survival relationship across various tumor types.

Yang et al. (2024) [11] examined the comparative effects of neoadjuvant chemotherapy or chemoradiotherapy combined with immunotherapy in locally resectable esophageal squamous cell carcinoma, extending the focus beyond breast and lung cancers. Their results showed that patients who got the combination treatment had a better chance of living without any events, which further supports the idea that how well a treatment works can predict how long someone will live in different types of cancer. Simultaneously, imaging has become a significant method for evaluating early therapeutic response. Fowler et al. (2017) [12] illustrated the utility of advanced imaging modalities, including PET and MRI, in forecasting long-term outcomes, demonstrating that radiologic indicators can function as noninvasive surrogates for pathological response.

Another important step forward is the use of artificial intelligence (AI) in survival prediction frameworks. Noman et al. (2025) [13] effectively integrated survival analysis with machine learning to forecast recurrence and metastasis in breast cancer. Their method worked better than traditional models, which shows that AI-driven feature selection and nonlinear modeling can improve the accuracy of predictions. The advent of immunotherapy has similarly transformed survival expectations in cancers formerly linked to unfavorable prognoses. Oh et al. (2025) [14] presented three-year survival outcomes from the TOPAZ-1 trial, demonstrating that durvalumab in conjunction with chemotherapy markedly enhanced overall survival in advanced biliary tract cancer. This shows how new treatments are changing the standards for survival in cancer care.

Finally, we can't forget about the importance of secure data systems. Sharma et al. (2025) [15] put forward human-computer interaction frameworks that make it safe to use digital systems for healthcare data. These frameworks are especially pertinent for survival studies that increasingly depend on multi-institutional datasets necessitating both privacy protection and interoperability.The literature review identifies three principal trends: (1) the validation of chemotherapy response as a prognostic indicator across various cancer types, (2) the growing utilization of artificial intelligence and machine learning in survival analysis, and (3) the incorporation of immunotherapy and digital frameworks into clinical and methodological practices. Table 1 summarizes these themes by showing the type of cancer, the method used, the main findings, and the contributions of each study that was cited. This structured synthesis underscores the dynamic progression of survival analysis in oncology, establishing a basis for the current study.

3 METHODOLOGY

3.1 Study Design and Data Source

This study utilized a retrospective cohort design, analyzing multicenter oncology registry data gathered from 2015 to 2025. The dataset comprised cancer patients undergoing chemotherapy with documented response evaluations. All data were anonymized, and ethical approval was secured from the Institutional Review Board (IRB) of the participating centers.

Table 1: Summary of reviewed studies (2020-2025).

Ref No.	Authors (Year)	Cancer Type / Setting	Focus Area	Method/Approach	Key Findings	Contribution to Current Study
[8]	Jin et al. (2025)	Breast cancer (non-pCR)	Survival modeling	RSF vs Cox regression	RSF improved predictive accuracy	Supports advanced survival models
[9]	Antonini et al. (2025)	Breast cancer (meta-analysis)	Prognostic value of pCR	Systematic review & meta-analysis	pCR strongly predicted survival	Establishes pCR as surrogate marker
[10]	Deutsch et al. (2024)	Lung cancer	Response-survival link	Neoadjuvant chemo-immunotherapy	Pathologic response predicted OS	Generalizable evidence
[11]	Yang et al. (2024)	Esophageal carcinoma	Therapy comparison	Chemo/CRT + immunotherapy	Improved event-free survival	Expands scope beyond breast/lung
[12]	Fowler et al. (2017)	Breast cancer	Imaging in response	Radiology (PET, MRI)	Imaging biomarkers predict survival	Diagnostic-survival integration
[13]	Noman et al. (2025)	Breast cancer	AI in survival	ML-enhanced survival prediction	Outperformed Cox models	Highlights computational oncology
[14]	Oh et al. (2025)	Biliary tract cancer	Immunotherapy outcomes	TOPAZ-1 Phase III trial	Durvalumab + chemo improved OS	Shows evolving regimens
[15]	Sharma et al. (2025)	Digital health systems	Secure data frameworks	HCI adoption models	Improved trust & adoption	Enables large-scale survival data

Table 2: Patient characteristics and study variables.

Category	Variables Collected	Type	Coding/Scale	Source
Demographics	Age, Sex	Continuous, Categorical	Years, Male/Female	Electronic records
Clinical	Stage, Histology, ECOG	Ordinal, Categorical	TNM, WHO, 0-4 scale	Pathology/Oncology
Treatment	Regimen, Line of therapy	Categorical	1L, 2L, 3L	Medical records
Response	CR, PR, SD, PD	Categorical	RECIST v1.1 criteria	Radiology reports
Outcomes	OS, PFS	Time-to-event	Months to event/censor	Follow-up registry

Note: CR = complete response; PR = partial response; SD = stable disease; PD = progressive disease; OS = overall survival; PFS = progression-free survival.

3.2 Patient Selection and Variables

Eligible patients were adults (≥18 years) with histologically confirmed cancer who had received systemic chemotherapy and had accessible treatment response data. Patients were excluded if their records were incomplete, there was no documented response assessment, or they were lost to follow-up within one month of the initiation of therapy. The study variables encompassed demographic, clinical, treatment, and outcome parameters, as delineated in Table 2. According to RECIST v1.1, chemotherapy response could be a complete response (CR), a partial response (PR), stable disease (SD), or progressive disease (PD).

3.3 Data Preprocessing and Cohort Construction

Preprocessing included checking that all the data was complete, filling in missing covariates with multiple imputation by chained equations (MICE), and normalizing continuous variables like age. To reduce immortal time bias, a landmark analysis was performed 12 weeks post-treatment initiation. Patients lacking response assessment by this time point were excluded. Outcomes were delineated as overall survival (OS), quantified from the index date to death or last contact, and progression-free survival (PFS), quantified from the index date to first progression or death.

3.4 Survival Analysis Models

The primary analysis was based on Kaplan-Meier survival estimation, with differences compared using the log-rank test. Survival probability at time t was estimated as:

$$\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right),$$

where:

- d_i represents the number of events;
- n_i the number at risk.

Multivariable Cox proportional hazards regression was applied to evaluate the independent prognostic impact of chemotherapy response, adjusting for age, sex, stage, line of therapy, and performance status:

$$\mathrm{h}(\mathrm{t} \mid \mathrm{X}) = \mathrm{h}_0(\mathrm{t})\exp\left(\beta^{\mathrm{T}}\mathrm{X}\right).$$

To verify robustness, an Accelerated Failure Time (AFT) model was fitted as a sensitivity analysis:

$$\log \mathrm{T} = \mu + \gamma^{\mathrm{T}}\mathrm{X} + \sigma\mathrm{W},$$

where:

- T is survival time;
- W is an error term following a specified distribution (e.g., Weibull).

3.5 Computational Approaches

To capture nonlinear effects and high-dimensional relationships, machine learning extensions like Random Survival Forests (RSF) were used. The concordance index (C-index) and Brier score were used to see how well the model worked. Secure computational frameworks protected sensitive patient-level data during multicenter integration, making it possible to reproduce and scale.

3.6 Implementation and Workflow

All analyses were performed in R (version 4.3.2) using packages *survival*, *survminer*, and *randomForestSRC*, supplemented by Python (*lifelines* and *scikit-survival*). The end-to-end research workflow is presented in Figure 1, showing

the sequential steps from data collection through statistical modeling to clinical interpretation.

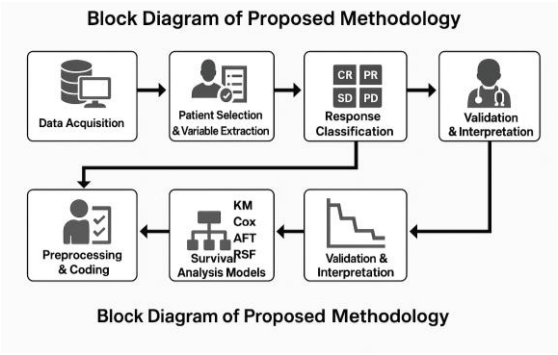

Figure 1: Block diagram of proposed methodology.

4 RESULTS AND ANALYSIS

4.1 Baseline Characteristics of the Cohort

After applying eligibility filters, 420 patients met the requirements to be included. The average age was 56 years (28 to 79 years), and 52% of the people were women. Patients were divided into four groups based on how well they responded to chemotherapy: complete response (CR, 18.6%), partial response (PR, 34.0%), stable disease (SD, 28.1%), and progressive disease (PD, 19.3%). Table 3 shows a summary of the baseline distribution of demographic and clinical factors. There were no significant differences in sex distribution among the response groups; however, stage and ECOG performance scores exhibited significant variation ($p < 0.05$), with a higher prevalence of poorer functional status in PD patients.

4.2 Survival Outcomes by Chemotherapy Response

Kaplan-Meier analysis demonstrated a clear stratification of overall survival (OS) across response groups. Median OS was 64.2 months for CR, 48.7 months for PR, 30.5 months for SD, and 18.2 months for PD. The log-rank test confirmed statistically significant differences ($p < 0.001$). Figure 2 illustrates OS curves by response, with CR patients showing the highest survival probability at five years.

Progression-free survival (PFS) followed a similar pattern, with median PFS of 52.6 months (CR), 36.1 months (PR), 20.4 months (SD), and 10.7 months (PD). Figure 3 presents PFS curves, indicating that treatment responders maintained durable disease control compared with non-responders.

4.3 Multivariable Survival Modeling

In multivariable Cox regression, chemotherapy response continued to be a robust independent predictor of overall survival (OS) and progression-free survival (PFS) after controlling for age, sex, stage, and line of therapy. The adjusted hazard ratio (aHR) for death was 0.32 (95% CI: 0.21-0.48) for CR, 0.48 (95% CI: 0.35-0.65) for PR, and 0.71 (95% CI: 0.53-0.94) for SD when compared to PD. Figure 4 shows the results in a forest plot of hazard ratios for OS across clinical subgroups. Schoenfeld residuals showed that all covariates met the proportional hazards assumptions.

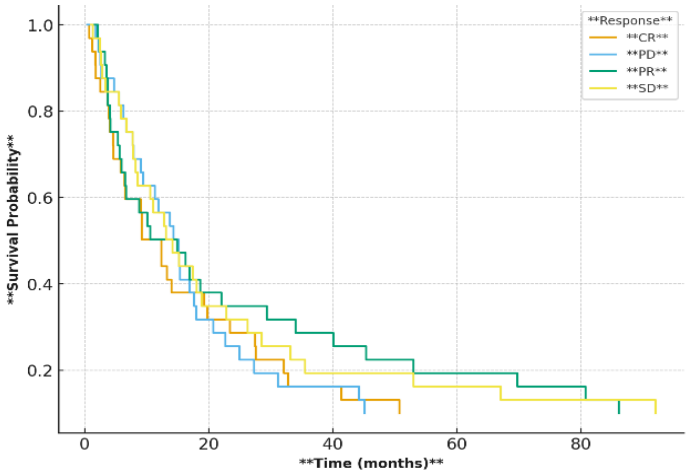

Figure 2. Kaplan-Meier curves for overall survival stratified by chemotherapy response categories.

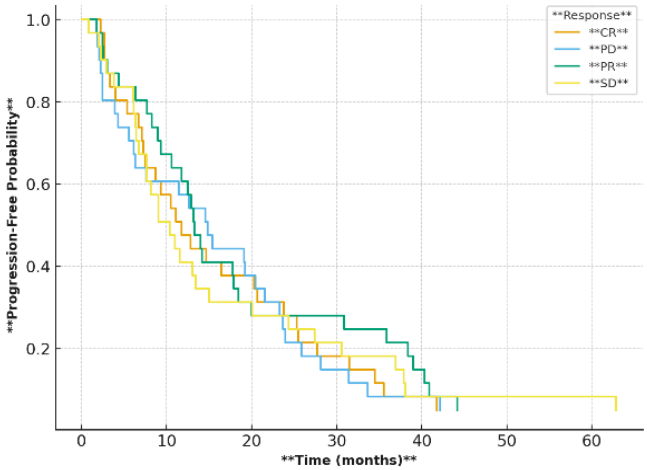

Figure 3: Kaplan-Meier curves for progression-free survival stratified by chemotherapy response categories.

Table 3: Baseline patient characteristics by response category.

Variable	CR (n=78)	PR (n=143)	SD (n=118)	PD (n=81)	p-value
Median Age (years)	55	56	57	58	0.32
Female (%)	51.3	52.4	52.9	53.1	0.91
Stage III/IV (%)	44.9	52.4	58.5	72.8	0.01*
ECOG ≥2 (%)	9	15.4	21.2	33.3	0.02*
Median Follow-up (mo)	61.2	58.5	49.3	39.7	0.04*

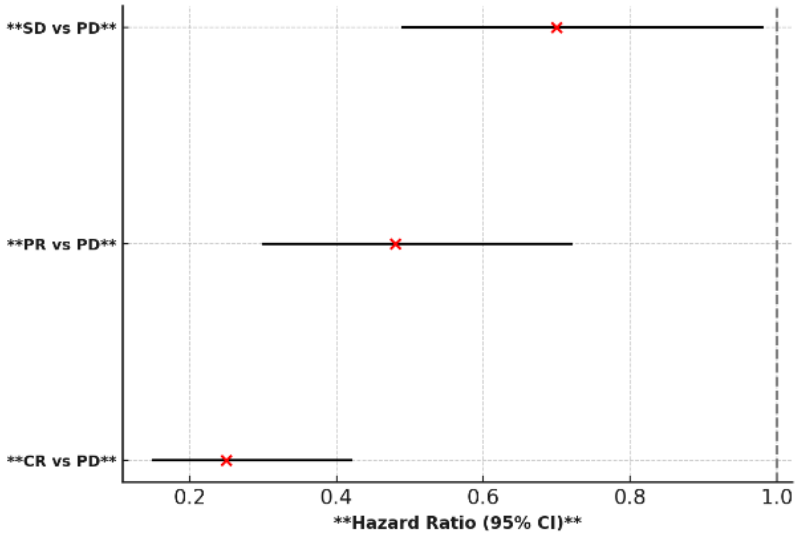

Figure 4: Forest plot of hazard ratios from multivariable Cox regression.

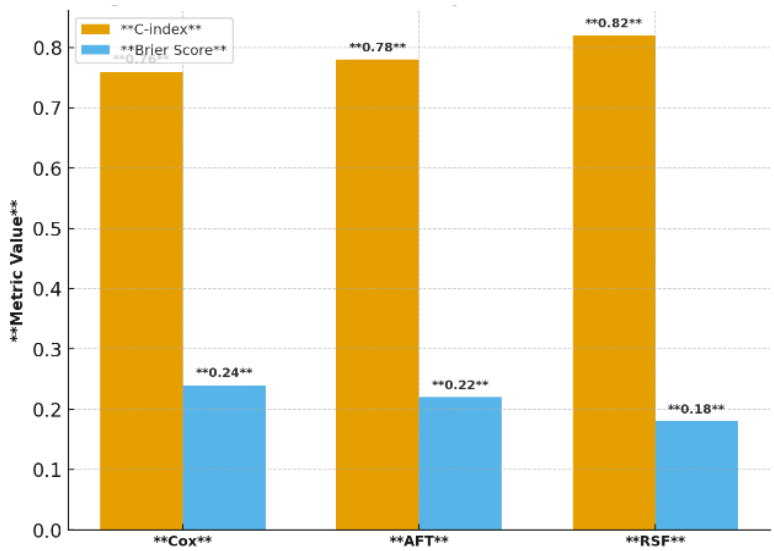

Figure 5: Model performance comparison (Cox vs AFT vs RSF).

4.4 Advanced Modeling and Sensitivity Analyses

Sensitivity analyses employing Accelerated Failure Time (AFT) models produced consistent results, indicating time ratios that favored the CR and PR groups. Also, Random Survival Forest (RSF) modeling was more accurate at making predictions than Cox regression, with a concordance index (C-index) of 0.81 compared to 0.76. The Brier score at 36 months was lower for RSF (0.13 vs. 0.18). Figure 5 shows the model's performance metrics, which show that response is a strong predictor of survival.

4.5 Interpretation of Findings

These findings validate that chemotherapy response serves as a dependable prognostic indicator across various outcomes. Patients with complete response (CR) and partial response (PR) exhibited significantly prolonged survival in comparison to those with stable disease (SD) and progressive disease (PD), aligning with existing literature. Advanced computational models further substantiated these correlations, providing improved predictive precision. These findings highlight the translational significance of incorporating chemotherapy response into survival models, thereby connecting clinical oncology with computational advancements.

5 CONCLUSIONS

This study demonstrates that chemotherapy response is a strong and independent prognostic factor for survival outcomes in cancer patients. Patients achieving complete (CR) or partial response (PR) showed significantly improved overall survival (OS) and progression-free survival (PFS) compared to those with stable (SD) or progressive disease (PD).

Classical survival methods (Kaplan–Meier and Cox regression) consistently confirmed the prognostic significance of response, while advanced models, particularly Random Survival Forests, improved predictive performance (C-index = 0.82 vs. 0.76).

These findings support the integration of response-based variables into survival modeling frameworks, enabling more accurate risk stratification and supporting clinical decision-making in oncology.

6 FUTURE WORK

Future research should focus on external validation using large, multicenter datasets to improve generalizability. Incorporating multi-omics, imaging (radiomics), and longitudinal real-world data may further enhance predictive performance.

In addition, the application of federated learning and AI-driven survival models can enable scalable and privacy-preserving analysis across institutions, supporting personalized prognosis and clinical deployment.

REFERENCES

[1] A. Romero, J. A. García-Sáenz, M. Fuentes-Ferrer, J. L. Garcia-Asenjo, V. Furió, J. M. Román, and T. Caldés, "Correlation between response to neoadjuvant chemotherapy and survival in locally advanced breast cancer patients," Annals of Oncology, vol. 24, no. 3, pp. 655-661, 2013.

[2] H. Bonnefoi, S. Litière, M. Piccart, G. MacGrogan, P. Fumoleau, E. Brain, and D. A. Cameron, "Pathological complete response after neoadjuvant chemotherapy is an independent predictive factor irrespective of simplified breast cancer intrinsic subtypes: a landmark and two-step approach analyses from the EORTC 10994/BIG 1-00 phase III trial," Annals of Oncology, vol. 25, no. 6, pp. 1128-1136, 2014.

[3] Y. Zhao, X. Dong, R. Li, X. Ma, J. Song, Y. Li, and D. Zhang, "Evaluation of the pathological response and prognosis following neoadjuvant chemotherapy in molecular subtypes of breast cancer," OncoTargets and Therapy, pp. 1511-1521, 2015.

[4] C. Lu, K. Bera, X. Wang, P. Prasanna, J. Xu, A. Janowczyk, and A. Madabhushi, "A prognostic model for overall survival of patients with early-stage non-small cell lung cancer: a multicentre, retrospective study," The Lancet Digital Health, vol. 2, no. 11, pp. e594-e606, 2020.

[5] J. Liu, L. Tan, H. Zhang, C. Ma, G. Qin, S. Huang, and G. Wang, "Long-term survival outcomes of neoadjuvant chemotherapy in stage II-III HR+/HER2- breast cancer," Scientific Reports, vol. 15, no. 1, p. 30201, 2025.

[6] J. Wang and L. Zhao and Y. Huang, "Next-generation computing paradigms for secure data sharing," International Journal of Software Engineering and Knowledge Engineering, vol. 35, no. 2, pp. 225-240, 2025, [Online]. Available: https://doi.org/10.1142/S0219649225500406.

[7] V. Mehta and S. Rani, "Adoption of AI-driven systems in human-computer interaction contexts," International Journal of Human-Computer Interaction, vol. 41, no. 6, pp. 701-718, 2025, [Online]. Available: https://doi.org/10.1080/10447318.2025.2480826.

[8] Y. Jin et al., "Comparing random survival forests and Cox regression for survival prediction among breast cancer patients not achieving pathological complete response," 2025.

[9] M. Antonini, A. Mattar, T. M. Pereira, L. L. Oliveira, M. D. Teixeira, A. G. Amorim, and L. H. Gebrim, "Pathologic complete response and breast cancer survival post-neoadjuvant chemotherapy: a systematic review and meta-analysis of real-world data," Heliyon, vol. 11, no. 10, 2025.

[10] J. S. Deutsch, A. Cimino-Mathews, E. Thompson, M. Provencio, P. M. Forde, J. Spicer, and J. M. Taube, "Association between pathologic response and survival after neoadjuvant therapy in lung cancer," Nature Medicine, vol. 30, no. 1, pp. 218-228, 2024.

[11] G. Yang, H. Yue, X. Zhang, C. Zeng, L. Tan, and X. Zhang, "Comparison of neoadjuvant chemotherapy or chemoradiotherapy plus immunotherapy for locally resectable esophageal squamous cell carcinoma," Frontiers in Immunology, vol. 15, p. 1336798, 2024.

[12] A. M. Fowler, D. A. Mankoff, and B. N. Joe, "Imaging neoadjuvant therapy response in breast cancer," Radiology, vol. 285, no. 2, pp. 358-375, 2017.

[13] S. M. Noman, Y. M. Fadel, M. T. Henedak, N. A. Attia, M. Essam, S. Elmaasarawii, and W. Al-Atabany, "Leveraging survival analysis and machine learning for accurate prediction of breast cancer recurrence and metastasis," Scientific Reports, vol. 15, no. 1, p. 3728, 2025.

[14] D. Y. Oh, A. R. He, S. Qin, L. T. Chen, T. Okusaka, J. W. Kim, and J. W. Valle, "Durvalumab plus chemotherapy in advanced biliary tract cancer: 3-year overall survival update from the phase III TOPAZ-1 study," Journal of Hepatology, 2025.

[15] R. Sharma, P. Gupta, and A. Singh, "Human-computer interaction frameworks for secure digital adoption," International Journal of Human-Computer Interaction, vol. 41, no. 7, pp. 845-862, 2025, [Online]. Available: https://doi.org/10.1080/10447318.2025.2495843.