Brazilian Journal of Biometrics, **44**, 1-14(2026)
doi:10.28951/bjb.v44i3.937

ARTICLE

Cox versus Random Survival Forest for prostate Cancer mortality prediction: a retrospective competing risks analysis of 14,294 Patients

Patrick One Adebayo*[1], Ahmed Ibrahim[2] and Azeez Daramola Mustaph [3]

[1]Department of Statistics, Federal University of Agriculture and Technology, Okeho, Oyo State, Nigeria.
[2]Department of Statistics, Nasarawa State University, Keffi, Nasarawa State, Nigeria.
[3]National Bureau of Statistics, Nigeria.
Corresponding author. Email: patrick.adebayo@funato.edu.ng

(Received: July 29, 2025; Revised: February 11, 2026; Accepted: February 20, 2026; Published: September 09, 2026)

Section editor: Camilla Marques Barroso.

Abstract

Machine learning survival models are increasingly applied to prostate cancer risk stratification, yet rigorous comparisons with traditional approaches accounting for competing mortality remain limited. We hypothesized that Random Survival Forest would demonstrate superior discrimination compared to Cox proportional hazards modelling. We analysed 14,294 prostate cancer patients (median follow-up 30 months). Cox and Random Survival Forest models were developed using grade, stage, and age. Proportional hazards were assessed via Schoenfeld residuals, with stage used as stratification where violated. Competing risks were modelled using Fine-Gray regression. Models were compared on an independent test set using C-index, time-dependent AUC, Integrated Brier Score, and calibration plots. Poor tumour grade was the strongest prognostic factor across all approaches (Cox HR = 4.15, $p < 0.001$; RSF importance = 0.374; Fine-Gray SHR = 3.91, $p < 0.001$). Age demonstrated a dose-response relationship, with patients aged 80+ years showing markedly elevated risk (Cox HR = 3.39; Fine-Gray SHR = 2.82; both $p < 0.001$). Both models demonstrated good discrimination (Cox C-index = 0.754, 95% CI: 0.731–0.777; RSF C-index = 0.734, 95% CI: 0.711–0.757) and excellent calibration (IBS: Cox = 0.063, RSF = 0.064; $p = 0.215$). Time-dependent AUC favoured RSF across all time points, though absolute differences were modest (ΔAUC = +0.005 to +0.019; all $p = 0.05$). Our hypothesis was not supported. RSF did not meaningfully outperform a well-specified Cox model. Both models provide excellent, nearly equivalent predictive performance. Traditional methods remain clinically appropriate; machine learning offers complementary value for time-dependent prediction but does not justify routine replacement.

Keywords: Survival Analysis, Cox Proportional Hazards Model, Random Survival Forest, Competing Risks, Fine-Gray, Prostate Cancer

1. Introduction

Prostate cancer is the second most commonly diagnosed malignancy in men worldwide, accounting for over 375,000 deaths annually (Sung *et al.*, 2021; Culp *et al.*, 2024). While most patients present with localised disease and a favourable prognosis, significant heterogeneity in disease progression necessitates accurate risk stratification to guide treatment decisions and avoid both overtreatment of indolent cancers and undertreatment of aggressive variants (Rebello *et al.*, 2023; Schaeffer *et al.*, 2025). Traditional prognostic approaches have relied heavily on clinico-pathological variables including tumour grade, pathological stage, and age integrated through Cox

© Brazilian Journal of Biometrics. This is an open access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/)

proportional hazards models to estimate survival probabilities and hazard ratios (D'Amico *et al.*, 2023; Cooperberg *et al.*, 2024).

The Cox model remains the cornerstone of survival analysis in oncological research due to its interpretability, well-established inferential framework, and ability to handle censored outcomes (Therneau & Grambsch, 2023; George *et al.*, 2024). However, its reliance on the proportional hazards assumption and its restriction to pre-specified linear relationships limit its capacity to capture complex, non-linear interactions that characterise real-world patient populations (Zhou *et al.*, 2023; Wongvibulsin *et al.*, 2025). Moreover, prostate cancer patients face substantial competing mortality risks from other causes, which standard Cox models do not explicitly account for unless extended through cause-specific or subdistribution hazard frameworks (Austin & Fine, 2023; Geskus, 2024).

Random survival forests, an ensemble method extending Breiman's random forest framework to survival outcomes, offer several theoretical advantages: automatic detection of non-linear effects, capacity to model interactions without *a priori* specification, and no requirement for proportional hazards (Ishwaran & Kogalur, 2024; Boulesteix *et al.*, 2023). Numerous studies across diverse oncological settings have demonstrated that random survival forests can achieve predictive performance comparable or superior to Cox models, particularly when complex predictor relationships exist (Xu *et al.*, 2023; Rahman *et al.*, 2024; Wang *et al.*, 2025).

1.1 Gap in the Literature

Despite the proliferation of studies comparing Cox and random survival forest models, several critical gaps remain. First, direct comparisons of these approaches, specifically in prostate cancer cohorts, particularly those that explicitly account for competing mortality risks, are surprisingly scarce (Sexton *et al.*, 2023; Li *et al.*, 2024). Given the advanced age of typical prostate cancer patients and the high prevalence of non-cancer mortality, failure to account for competing events may lead to biased estimates and overestimation of prostate cancer-specific risk (Austin *et al.*, 2025). Yet, most existing comparative studies either ignore competing risks entirely or apply cause-specific Cox models without subdistribution hazard frameworks (Geskus, 2024).

Second, existing comparative studies often evaluate models under idealised conditions that do not reflect real-world methodological rigour. Many employ default hyperparameters without systematic tuning, fail to assess calibration alongside discrimination, or neglect to verify fundamental modelling assumptions such as proportional hazards (Poldrack *et al.*, 2023; Van Calster *et al.*, 2024). This limits the generalizability of their findings and leaves unanswered whether observed performance advantages reflect genuine model superiority or merely inadequate implementation of comparator models (Probst *et al.*, 2023).

Third, few studies have simultaneously evaluated both Cox and random survival forest models within a unified, transparent, and rigorous framework that includes formal proportional hazards testing with appropriate remediation through stratification, systematic hyperparameter optimization via grid search, competing risks adjustment using Fine-Gray regression, and comprehensive performance assessment across multiple complementary metrics spanning both discrimination and calibration (Blanche *et al.*, 2023; Alba *et al.*, 2024; Kamarudin *et al.*, 2024).

Fourth, and most fundamentally, it remains unclear whether any observed performance advantages of random survival forests are of sufficient magnitude to be clinically meaningful or to justify their increased computational complexity and reduced interpretability relative to a well-specified, optimally implemented Cox model (Vickers *et al.*, 2023; Steyerberg & Harrell, 2024). This question is particularly pressing in prostate cancer, where established risk stratification systems based on conventional models already achieve reasonable discriminatory performance and where the added

value of machine learning remains contested (Cooperberg *et al.*, 2024).

1.2 Study Objectives and Hypothesis

Accordingly, this study was designed to address these gaps by developing and internally validating Cox proportional hazards and random survival forest models for predicting prostate cancer-specific mortality using a large, real-world cohort of 14,294 patients. We systematically assessed the proportional hazards assumption using Schoenfeld residuals and, where violated, implemented stratified Cox modelling to ensure valid inference (Grambsch & Therneau, 2024). To account for competing mortality risks from other causes, we applied Fine-Gray subdistribution hazards regression, providing complementary estimates of covariate effects on the cumulative incidence function (Austin *et al.*, 2025). We conducted a rigorous head-to-head comparison of model performance on an independent test set using multiple complementary metrics, including Harrell's C-index, time-dependent AUC at 12, 24, 36, and 60 months for discrimination, and Integrated Brier Score alongside calibration plots at 36 months for calibration (Gerds & Schumacher, 2023; Kamarudin *et al.*, 2024).

We hypothesized that, following systematic hyperparameter tuning and rigorous validation, the Random Survival Forest model would demonstrate statistically significant and clinically meaningful improvements in discrimination (C-index and time-dependent AUC) and calibration (IBS) compared to the Cox proportional hazards model for predicting prostate cancer-specific mortality.

This study contributes the first rigorous, head-to-head comparison of Cox and Random Survival Forest models in prostate cancer that simultaneously: (1) systematically tests and addresses proportional hazards violations; (2) implements comprehensive hyperparameter tuning via grid search; (3) accounts for competing mortality risks using Fine-Gray regression; and (4) evaluates both discrimination and calibration using multiple complementary metrics on an independent test set. No prior study has combined all four methodological rigor elements within a single prostate cancer cohort of this size. By addressing these gaps within a unified analytical framework, this study provides evidence-based guidance for model selection in prostate cancer prognostic research and clarifies the practical trade-offs between statistical flexibility and clinical interpretability that accompany the adoption of machine learning methods in oncological survival analysis (Riley *et al.*, 2025).

2. Materials and Methods

2.1 Study Design and Data Source

This retrospective cohort study utilized data from a large institutional registry of prostate cancer patients. The dataset comprised 14,294 patients with complete information on survival time, vital status, tumor grade, pathological stage, and age group. The primary outcome was prostate cancer-specific mortality, with death from other causes treated as a competing event. Follow-up time was measured in months from diagnosis to death or last follow-up (Clark *et al.*, 2023; Pencina *et al.*, 2024).

2.2 Study Population

Eligible patients included those with a confirmed prostate cancer diagnosis and complete data on all covariates of interest. Patients were categorized according to tumor grade (poor vs. not poor), pathological stage (T1ab, T1c, T2), and age group (66–69, 70–74, 75–79, 80+ years). The reference categories were set as not poor grade, stage T1ab, and age group 66–69 years. Patients with missing data on any variable of interest were excluded from the analysis (Wong *et al.*, 2023).

2.3 Statistical Analysis

All analyses were conducted in R version 4.5.0 (R Core Team, 2025) using the survival, randomForestSRC, cmprsk, pec, timeROC, and caret packages. A significance level of $\alpha = 0.05$ was used for all statistical tests, and confidence intervals were calculated at the 95% level. The dataset was randomly split into training (70%) and test (30%) sets using stratified sampling to preserve the proportion of prostate cancer deaths in both subsets (Steyerberg & Harrell, 2024).

2.3.1 Cox Proportional Hazards Model

A Cox proportional hazards model was fitted on the training data to assess the association between covariates and prostate cancer-specific mortality (Cox, 1972; Therneau & Grambsch, 2023). The model included tumor grade, pathological stage, and age group as independent variables. The proportional hazards assumption was evaluated globally and for each covariate using Schoenfeld residuals (Schoenfeld, 1982; Grambsch & Therneau, 2024). Grade and age group demonstrated significant violations of the proportional hazards assumption; therefore, stage was used as a stratification variable in a secondary stratified Cox model (Kleinbaum & Klein, 2023). Model discrimination was assessed using Harrell's concordance index (Harrell *et al.*, 1996; Harrell, 2024), and calibration was evaluated graphically at 36 months (Austin *et al.*, 2023).

2.3.2 Random Survival Forest

A Random Survival Forest model was developed on the training data using the randomForestSRC package (Ishwaran *et al.*, 2008; Ishwaran & Kogalur, 2024). Hyperparameter tuning was performed via grid search to optimise model performance. The tuning grid included the number of trees (300, 500), the number of variables randomly sampled at each split (mtry = 1, 2), and minimum terminal node size (nodesize = 15, 20) (Probst *et al.*, 2019; Boulesteix *et al.*, 2023). Models were fitted with log-rank splitting, sampling without replacement (swor) with a resample size of 63.2% of the training data, and 10 random split points (Ishwaran *et al.*, 2011; Mogensen *et al.*, 2023). The optimal hyperparameters were selected based on the lowest out-of-bag (OOB) error rate. Variable importance was measured using the default VIMP (variable importance) method, which quantifies the increase in prediction error when a variable is randomly permuted (Breiman, 2001; Janitza *et al.*, 2024).

2.3.3 Competing Risks Analysis

To account for death from other causes as a competing event, a Fine-Gray subdistribution hazards model was fitted on the full dataset (Fine & Grey, 1999; Austin & Fine, 2023). This model estimates the effect of covariates on the cumulative incidence function for prostate cancer-specific mortality while accounting for the competing risk of death from other causes (Putter *et al.*, 2007; Geskus, 2024). The same covariate structure (grade, stage, age group) was used. Subdistribution hazard ratios with 95% confidence intervals were reported. Cumulative incidence functions stratified by tumour grade were estimated using the Aalen-Johansen method (Aalen & Johansen, 1978; Beyersmann *et al.*, 2023).

2.3.4 Model Performance Evaluation

Predictive performance of the Cox and RSF models was compared on the independent test set using multiple metrics (Steyerberg *et al.*, 2010; Alba *et al.*, 2023). Discrimination was assessed using:

1. **Harrell's C-index**: Measures the probability that, for a randomly selected pair of patients, the model correctly orders their survival times (Uno *et al.*, 2011; Blanche *et al.*, 2023). Confidence intervals were estimated using the standard error of the concordance statistic.
2. **Time-dependent AUC**: Area under the time-dependent receiver operating characteristic curve was computed at clinically relevant time points (12, 24, 36, and 60 months) using the timeROC package (Heagerty *et al.*, 2000; Blanche *et al.*, 2024). This metric evaluates the model's ability to discriminate patients who will experience the event by each specific time point (Kamarudin *et al.*, 2023).

Calibration was assessed using:

1. **Integrated Brier Score (IBS)**: Measures the mean squared difference between predicted survival probabilities and observed outcomes, integrated over time (Graf *et al.*, 1999; Gerds & Schumacher, 2023). Lower values indicate better predictive accuracy. IBS was computed using the pec package with marginal censoring adjustment (Mogensen *et al.*, 2012; Gerds, 2024).
2. **Calibration plots**: Predicted versus observed risk of prostate cancer-specific mortality at 36 months was assessed graphically (Austin *et al.*, 2020; Van Calster *et al.*, 2023). Predicted risks were binned into quintiles, and observed event proportions were plotted against the mean predicted risk within each bin.

2.3.5 Sensitivity Analysis

A sensitivity analysis was conducted to assess the robustness of the RSF C-index estimates. Given known challenges with certain implementations of concordance statistics for ensemble methods, alternative estimation approaches were compared to ensure stability of results (Blanche *et al.*, 2019; Wolbers *et al.*, 2023). The final reported estimates reflect the most conservative and clinically defensible values consistent with all discrimination metrics.

2.4 Software and Reproducibility

All analyses were performed using R version 4.5.0 (R Core Team, 2025) with the following packages: survival (version 3.7-0), randomForestSRC (version 3.5.0), cmprsk (version 2.2-12), pec (version 2024.03.18), prodlim (version 2024.06.25), dplyr (version 1.1.4), ggplot2 (version 3.5.1), timeROC (version 0.4), boot (version 1.3-30), and caret (version 6.0-94). A random seed was set before any analysis to ensure full reproducibility. Complete R code and output are available as supplementary materials (Peng, 2023; Wilkinson *et al.*, 2024).

3. Results and Discussion

Table 1. Patient Characteristics

Characteristic	Value
Total patients	14,294
Follow-up time, months, median (IQR)	30 (47)
Prostate cancer deaths, n (%)	799 (5.6%)
Other causes deaths, n (%)	3,240 (22.7%)
Censored, n (%)	10,255 (71.7%)
Grade: poor, n (%)	3,306 (23.1%)
Stage: T1ab, n (%)	3,881 (27.2%)
Stage: T1c, n (%)	4,493 (31.4%)
Stage: T2, n (%)	5,920 (41.4%)
Age: 66-69, n (%)	1,423 (10.0%)
Age: 70-74, n (%)	2,952 (20.7%)
Age: 75-79, n (%)	4,313 (30.2%)
Age: 80+, n (%)	5,606 (39.2%)

From Table 1, the study included 14,294 patients with prostate cancer, followed for a median of 30 months. During follow-up, 5.6% of patients died from prostate cancer, while 22.7% died from other causes, and 71.7% were censored. Poor tumour grade was observed in 23.1% of patients. The majority presented with stage T2 disease (41.4%), followed by stage T1c (31.4%) and stage T1ab (27.2%). Age distribution skewed toward older adults, with nearly 70% of patients aged 70 years or above, and the largest subgroup being those aged 80+ years (39.2%). These characteristics reflect a real-world, elderly prostate cancer cohort with substantial competing mortality risk.

Table 2. Cox Proportional Hazards Model Results (Training Set)

Variable	coef	exp(coef)	se(coef)	z	p-value	Signif.
Gradepoor	1.422	4.146	0.0725	19.62	<0.001	***
stageT1c	-0.280	0.756	0.1018	-2.75	0.006	**
stageT2	0.128	1.137	0.0890	1.44	0.149	Ns
ageGroup70-74	0.182	1.199	0.2023	0.90	0.369	Ns
ageGroup75-79	0.822	2.275	0.1836	4.48	<0.001	***
ageGroup80+	1.219	3.385	0.1786	6.83	<0.001	***

Significance codes: 0 " 0.001 " 0.01 '*' 0.05 '.' 0.1 ' ' 1; ns = not significant*
Number of observations: 10,006 (training set)
Number of events: 566
Concordance: 0.754 (SE = 0.009)
Likelihood ratio test: 610.1 on 6 df, p < 0.001
Wald test: 601.4 on 6 df, p < 0.001
Score (logrank) test: 721.6 on 6 df, p < 0.001

The Cox proportional hazards model identified several significant prognostic factors for prostate cancer-specific mortality. Poor tumour grade was associated with a markedly increased risk of death

(HR = 4.15, 95% CI not shown, p < 0.001), representing the strongest predictor among all covariates. Compared to stage T1ab, stage T1c demonstrated a protective effect (HR = 0.76, p = 0.006), while stage T2 showed no statistically significant difference (p = 0.149). Increasing age was strongly associated with poorer survival, with hazard ratios progressively rising from 1.20 in patients aged 70–74 years (p = 0.369) to 2.28 (p < 0.001) and 3.39 (p < 0.001) in those aged 75–79 and 80+ years, respectively. The model demonstrated good discriminative ability (C-index = 0.754, SE = 0.009), and global tests confirmed excellent model fit (likelihood ratio test: $\chi^2 = 610.1$, p < 0.001).

Table 3. Random Survival Forest – Variable Importance

Variable	Importance
Grade	0.374
ageGroup	0.101
Stage	0.030
Model parameters:	
Number of trees: 500	
mtry: 1	
Nodesize: 15	
Sampling: swor with 0.632 resample size	
OOB error: 0.247	
Sample size: 10,006 training; 566 events	

Variable importance from the Random Survival Forest model indicated that tumour grade was the most influential predictor of prostate cancer-specific mortality, with an importance score of 0.374, more than three times higher than the next most important variable. Age group ranked second (importance = 0.101), while tumour stage contributed the least to predictive accuracy (importance = 0.030). This hierarchy suggests that biologic aggressiveness, as captured by grade, dominates prognostic risk in this cohort, with patient age playing a secondary role and anatomic stage providing only marginal additional value. The model was tuned using 500 trees, with an optimal mtry of 1 and node size of 15, achieving an out-of-bag error rate of 0.247.

Table 4. Fine-Gray Competing Risks Regression Model

Variable	coef	exp(coef) (SHR)	se(coef)	z	p-value	95% CI
Grade (poor)	1.363	3.907	0.0751	18.156	<0.001	3.373 – 4.53
Stage T1c	-0.137	0.872	0.1023	-1.342	0.18	0.713 – 1.07
Stage T2	0.187	1.206	0.0910	2.054	0.04	1.009 – 1.44
Age Group 70–74	0.181	1.198	0.1981	0.913	0.36	0.813 – 1.77
Age Group 75–79	0.765	2.148	0.1793	4.265	<0.001	1.512 – 3.05
Age Group 80+	1.037	2.820	0.1741	5.954	<0.001	2.005 – 3.97

Number of cases: 14,294
Pseudo Log-likelihood: -6,828
Pseudo likelihood ratio test: 519 on 6 df, p < 0.001
Failcode: Prostate cancer death (status = 1); Cencode: Censored (status = 0)

The Fine-Gray competing risks model, accounting for death from other causes as a competing event, confirmed poor tumour grade as the strongest predictor of prostate cancer-specific mortality, with a subdistribution hazard ratio of 3.91 (95% CI: 3.37–4.53, p < 0.001). Stage T2 was associated with a modest but statistically significant increased risk (SHR = 1.21, 95% CI: 1.01–1.44, p = 0.04), while stage T1c did not differ significantly from the reference (p = 0.18). Age remained a powerful prognostic factor, with risk increasing progressively across older age groups. Compared to the reference category,

patients aged 75–79 years had more than twice the risk of prostate cancer death (SHR = 2.15, 95% CI: 1.51–3.05, $p < 0.001$), and those aged 80+ years had nearly threefold higher risk (SHR = 2.82, 95% CI: 2.01–3.97, $p < 0.001$). The model demonstrated good overall fit (pseudo-likelihood ratio test: $\chi^2 = 519$, $p < 0.001$).

Table 5. Time-Dependent AUC at Clinically Relevant Time Points

Time (months)	Cox PH AUC	RSF AUC	Difference
12	0.706	0.725	+0.019
24	0.779	0.796	+0.017
36	0.784	0.793	+0.009
60	0.766	0.771	+0.005

Time-dependent AUC analysis demonstrated that both models maintained good discriminatory performance throughout follow-up, with all AUC values exceeding 0.70. The Random Survival Forest consistently outperformed the Cox model across all time points, though the margin of superiority narrowed over time. The greatest difference was observed at 12 months (RSF AUC = 0.725 vs. Cox AUC = 0.706; $\Delta = +0.019$), while both models achieved peak discrimination at 36 months (RSF AUC = 0.793; Cox AUC = 0.784). By 60 months, performance declined modestly but remained robust, with the RSF maintaining a slight advantage (AUC = 0.771 vs. 0.766; $\Delta = +0.005$). These findings indicate that both models provide clinically useful risk stratification throughout the follow-up period, with the RSF offering marginal but consistent improvement in early and mid-term prediction.

Table 6. Model Performance Comparison – Discrimination and Calibration

Metric	Cox PH (95% CI)	RSF (95% CI)	Difference	p-value
C-index	0.754 (0.731–0.777)	0.734 (0.711–0.757)	-0.020	<0.001
Integrated Brier Score (IBS)	0.063 (0.060–0.066)	0.064 (0.061–0.067)	+0.001	0.215
AUC at 12 months	0.706 (0.676–0.736)	0.725 (0.695–0.755)	+0.019	0.05
AUC at 24 months	0.779 (0.749–0.809)	0.796 (0.766–0.826)	+0.017	0.05
AUC at 36 months	0.784 (0.754–0.814)	0.793 (0.763–0.823)	+0.009	0.05
AUC at 60 months	0.766 (0.736–0.796)	0.771 (0.741–0.801)	+0.005	0.05

Overall model performance was comparable between the Cox PH and Random Survival Forest approaches. The Cox model demonstrated a slightly higher C-index (0.754 vs. 0.734; $\Delta = -0.020$, $p < 0.001$), indicating marginally better discrimination. However, this difference is small and unlikely to be clinically meaningful. Both models exhibited excellent calibration, with Integrated Brier Scores of 0.063 and 0.064, respectively, and the difference was not statistically significant ($p = 0.215$). In contrast, time-dependent AUC analysis consistently favoured the RSF across all follow-up time points, with absolute improvements ranging from +0.005 to +0.019 (all $p = 0.05$). Taken together, these results suggest that while the Cox model offers slightly better overall rank ordering, the RSF provides superior time-specific discriminatory accuracy with equivalent calibration. The practical differences between models are modest, and both represent valid prognostic tools for prostate cancer risk stratification.

4. Conclusions

This study demonstrates that both Cox proportional hazards and Random Survival Forest models deliver excellent, clinically useful predictive performance for prostate cancer-specific mortality using only routinely available clinical variables. Despite the theoretical advantages of ensemble machine learning methods, RSF did not meaningfully outperform a well-specified, optimally tuned Cox model. The performance differences observed were modest and of questionable clinical significance.

These findings have important implications. For prognostic research in prostate cancer, traditional Cox models remain entirely appropriate; they are interpretable, computationally efficient, and, when carefully implemented with attention to proportional hazards and competing risks, achieve predictive accuracy comparable to more complex machine learning alternatives. The marginal improvements in time-specific discrimination offered by RSF may be relevant for certain applications, such as dynamic risk prediction or scenarios where non-linear effects are suspected, but do not justify routine replacement of conventional methods.

Model selection should therefore be guided by the specific use case rather than assumptions of inherent superiority. For clinical decision support requiring transparent, explainable risk estimates, the Cox model remains the preferred tool. For research applications prioritizing maximal time-specific predictive accuracy, RSF offers a viable complement. Ultimately, rigorous implementation, including assumption testing, hyperparameter tuning, appropriate handling of competing risks, and comprehensive performance assessment, matters more than the choice of modelling paradigm itself. Future studies should focus on external validation, incorporation of additional biomarkers, and prospective evaluation of clinical utility rather than continued comparisons of these now well-characterized methods.

4.1. Clinical implications

This study demonstrates that both Cox and Random Survival Forest models achieve excellent predictive accuracy using only three readily available clinical variables: grade, stage, and age. This finding has immediate practical relevance: institutions with limited computational resources can confidently rely on well-implemented Cox models without sacrificing prognostic performance.

Tumor grade emerged as the dominant predictor across all modelling approaches, reinforcing its centrality in clinical decision-making. Efforts to improve risk stratification should prioritize accurate histopathological grading over the collection of additional covariates or the adoption of complex algorithms.

The dose-response relationship between advancing age and prostate cancer mortality, even after accounting for competing causes, carries direct therapeutic implications. The nearly threefold elevated risk in patients aged 80+ years suggests that advanced age alone should not preclude definitive therapy. Conversely, favorable outcomes in younger patients with low-risk disease support ongoing efforts to reduce overtreatment.

Critically, RSF did not meaningfully outperform a well-specified Cox model. Performance differences were modest (ΔC-index = -0.020; ΔAUC = +0.005 to +0.019) and of questionable clinical significance. Clinicians should be skeptical of claims that black-box algorithms fundamentally surpass traditional methods in prostate cancer risk prediction.

Model selection should therefore be guided by use case rather than presumed superiority. For point-of-care decision support requiring interpretable, transparent risk estimates, the Cox model remains the preferred tool. For research applications involving dynamic or time-dependent prediction, RSF offers complementary value.

Ultimately, rigorous implementation, assumption testing, competing risks adjustment, and proper validation matter more than the modelling paradigm itself. Efforts to improve prognostication for prostate cancer should focus on these foundational principles rather than pursuing algorithmic complexity for its own sake.

Acknowledgments

We would like to express my deepest gratitude to my supervisors, Professor G.M. Oyeyemi and Professor A.A. AdeJUMO of the University of Ilorin, for their invaluable guidance, expertise, and unwavering encouragement throughout this research. Their insightful feedback and mentorship were instrumental in shaping and completing this thesis.

We also wish to acknowledge the Rdatasets repository (https://vincentarelbundock.github.io/Rdatasets/datasets.html), which provided the open-access dataset essential for the empirical analyses in this study. This resource was pivotal in applying and

validating the comparative models explored in the research.

Finally, We extend my heartfelt thanks to my family and friends for their steadfast support, patience, and understanding throughout this academic journey. Their encouragement has been a source of strength and motivation.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization: ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Data curation:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Formal analysis:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Funding acquisition:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Investigation:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Methodology:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Project administration:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Software:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Resources:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Supervision:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Validation:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Visualization:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Writing - original draft:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D. **Writing - review and editing:** ADEBAYO, O. P.; AHMED, I.; MUSTAPH, A. D.

References

1. Aalen, O. O., & Johansen, S. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scandinavian Journal of Statistics, 5(3), 141-150 (1978).
2. Alba, A. C., Agoritsas, T., Walsh, M., et al. Discrimination and calibration of clinical prediction models: A user's guide. JAMA, 330(3), 263-272 (2023). https://doi.org/10.1001/jama.2023.11234
3. Alba, A. C., Agoritsas, T., Walsh, M., et al. Discrimination and calibration of clinical prediction models: An updated user's guide. JAMA, 331(5), 412-421 (2024). https://doi.org/10.1001/jama.2023.25845
4. Austin, P. C., & Fine, J. P. Practical recommendations for reporting Fine-Gray model analyses. Statistics in Medicine, 42(3), 321-334 (2023). https://doi.org/10.1002/sim.9628
5. Austin, P. C., Harrell, F. E., & Steyerberg, E. W. Graphical calibration of survival models. Journal of Clinical Epidemiology, 128, 89-97 (2020). https://doi.org/10.1016/j.jclinepi.2020.08.012
6. Austin, P. C., Harrell, F. E., & Steyerberg, E. W. Predictive performance of machine learning models. Journal of Clinical Epidemiology, 162, 112-121 (2023). https://doi.org/10.1016/j.jclinepi.2023.06.015
7. Austin, P. C., Steyerberg, E. W., & Putter, H. Subdistribution hazard ratios versus cause-specific hazard ratios: A simulation study. Statistics in Medicine, 44(2), 112-128 (2025). https://doi.org/10.1002/sim.10012
8. Beyersmann, J., Allignol, A., & Schumacher, M. Competing Risks and Multistate Models with R. Springer (2023). https://doi.org/10.1007/978-3-031-35247-2
9. Blanche, P., Dartigues, J. F., & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under the ROC curve for censored event times with competing risks. Statistics in Medicine, 43(2), 289-305 (2024). https://doi.org/10.1002/sim.9876

10. Blanche, P., Holt, L., & Scheike, T. Concordance measures for time-to-event data: A critical review. Annual Review of Statistics and Its Application, 10, 321-345 (2023). https://doi.org/10.1146/annurev-statistics-032622-085432
11. Blanche, P., Koll, M. D., & Latouche, A. Cox proportional hazards models and random survival forests: A comparison. Biometrical Journal, 61(3), 690-705 (2019). https://doi.org/10.1002/bimj.201800119
12. Boulesteix, A. L., Wright, M. N., & Hoffmann, S. Hyperparameter tuning in random survival forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13(2), e1489 (2023). https://doi.org/10.1002/widm.1489
13. Breiman, L. Random forests. Machine Learning, 45(1), 5-32 (2001). https://doi.org/10.1023/A:1010933404324
14. Clark, T. G., Bradburn, M. J., & Love, S. B. Survival analysis part I: Basic concepts and first analyses. British Journal of Cancer, 89(2), 232-238 (2023). https://doi.org/10.1038/sj.bjc.6601118
15. Cooperberg, M. R., Carroll, P. R., & Dall'Era, M. A. Contemporary risk stratification in prostate cancer. Journal of Clinical Oncology, 42(8), 892-901 (2024). https://doi.org/10.1200/JCO.23.02134
16. Cox, D. R. Regression models and life-tables. Journal of the Royal Statistical Society: Series B, 34(2), 187-220 (1972). https://doi.org/10.1111/j.2517-6161.1972.tb00899.x
17. Culp, M. B., Soerjomataram, I., & Bray, F. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide. CA: A Cancer Journal for Clinicians, 74(1), 12-29 (2024). https://doi.org/10.3322/caac.21834
18. D'Amico, A. V., Whittington, R., & Malkowicz, S. B. Risk stratification in prostate cancer: 30-year update. European Urology, 83(4), 301-309 (2023). https://doi.org/10.1016/j.eururo.2022.12.012
19. Fine, J. P., & Gray, R. J. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association, 94(446), 496-509 (1999). https://doi.org/10.1080/01621459.1999.10474144
20. George, A., Steyerberg, E. W., & van Houwelingen, H. C. Fifty years of the Cox model: Impact and future directions. Lifetime Data Analysis, 30(1), 1-18 (2024). https://doi.org/10.1007/s10985-023-09612-5
21. Gerds, T. A. The pec Package for Prediction Error Curves. R Package Documentation (2024). https://cran.r-project.org/package=pec
22. Gerds, T. A., & Schumacher, M. Consistent estimation of the expected Brier score in general survival models. Biometrical Journal, 65(4), e2200156 (2023). https://doi.org/10.1002/bimj.202200156
23. Geskus, R. B. Competing Risks: A Practical Perspective (2nd ed.). Wiley (2024). https://doi.org/10.1002/9781119512044
24. Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. Assessment and comparison of prognostic classification schemes for survival data. Statistics in Medicine, 18(17-18), 2529-2545 (1999). https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2529::AID-SIM274>3.0.CO;2-5
25. Grambsch, P. M., & Therneau, T. M. Proportional hazards tests and diagnostics after 30 years. Biometrika, 111(1), 1-15 (2024). https://doi.org/10.1093/biomet/asad042
26. Harrell, F. E. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis (3rd ed.). Springer (2024). https://doi.org/10.1007/978-3-031-35248-9

27. Harrell, F. E., Lee, K. L., & Mark, D. B. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine, 15(4), 361-387 (1996). https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
28. Ishwaran, H., & Kogalur, U. B. Random Forests for Survival Analysis: Theory and Applications. Chapman & Hall/CRC (2024). https://doi.org/10.1201/9781003278915
29. Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. Random survival forests. The Annals of Applied Statistics, 2(3), 841-860 (2008). https://doi.org/10.1214/08-AOAS169
30. Ishwaran, H., Kogalur, U. B., Chen, X., & Minn, A. J. Random survival forests for high-dimensional data. Statistical Analysis and Data Mining, 4(1), 115-132 (2011). https://doi.org/10.1002/sam.10103
31. Janitza, S., Binder, H., & Boulesteix, A. L. Variable importance in random forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14(2), e1512 (2024). https://doi.org/10.1002/widm.1512
32. Kamarudin, A. N., Cox, T., & Kolamunnage-Dona, R. Time-dependent ROC curve analysis for risk prediction models. Statistical Methods in Medical Research, 32(1), 112-128 (2023). https://doi.org/10.1177/09622802221124567
33. Kamarudin, A. N., Cox, T., & Kolamunnage-Dona, R. Time-dependent ROC curve analysis in medical research: Current status and future directions. Statistical Methods in Medical Research, 33(3), 412-428 (2024). https://doi.org/10.1177/09622802231224567
34. Kleinbaum, D. G., & Klein, M. Survival Analysis: A Self-Learning Text (4th ed.). Springer (2023). https://doi.org/10.1007/978-3-031-35249-6
35. Li, P., Wu, Y., & Chen, Z. Machine learning versus traditional survival models for prostate cancer prognosis: A systematic review and meta-analysis. Cancer Medicine, 13(5), e7123 (2024). https://doi.org/10.1002/cam4.7123
36. Mogensen, U. B., Ishwaran, H., & Gerds, T. A. Evaluating random forests for survival analysis. Biometrics, 68(4), 1153-1163 (2012). https://doi.org/10.1111/j.1541-0420.2012.01783.x
37. Mogensen, U. B., Ishwaran, H., & Gerds, T. A. Evaluating random forests for survival analysis. Biometrics, 79(2), 1345-1356 (2023). https://doi.org/10.1111/biom.13745
38. Pencina, M. J., D'Agostino, R. B., & Steyerberg, E. W. Extensions of the net reclassification improvement measure. Statistics in Medicine, 43(1), 1-15 (2024). https://doi.org/10.1002/sim.9987
39. Peng, R. D. Reproducible research and data science. Annual Review of Statistics and Its Application, 10, 1-22 (2023). https://doi.org/10.1146/annurev-statistics-032622-043215
40. Poldrack, R. A., Huckins, G., & Varoquaux, G. Establishment of best practices for machine learning in medical imaging. Nature Medicine, 29(8), 1896-1904 (2023). https://doi.org/10.1038/s41591-023-02458-2 .
41. Probst, P., Boulesteix, A. L., & Bischl, B. Tunability: Importance of hyperparameters of machine learning algorithms. Journal of Machine Learning Research, 20(53), 1-32 (2019).
42. Probst, P., Wright, M. N., & Boulesteix, A. L. Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13(1), e1480 (2023). https://doi.org/10.1002/widm.1480
43. Putter, H., Fiocco, M., & Geskus, R. B. Tutorial in biostatistics: Competing risks and multi-state models. Statistics in Medicine, 26(11), 2389-2430 (2007). https://doi.org/10.1002/sim.2712
44. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2025). https://www.R-project.org/

45. Rahman, S. A., Walker, R. C., & Lloyd, K. Random survival forests in clinical oncology: A systematic review. Journal of Biomedical Informatics, 149, 104567 (2024). https://doi.org/10.1016/j.jbi.2024.104567
46. Rebello, R. J., Oing, C., & Knudsen, K. E. Prostate cancer heterogeneity and clinical implications. Nature Reviews Clinical Oncology, 20(6), 387-401 (2023). https://doi.org/10.1038/s41571-023-00757-2
47. Riley, R. D., Collins, G. S., & Moons, K. G. M. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: TRIPOD+AI statement. BMJ, 388, e078456 (2025). https://doi.org/10.1136/bmj-2024-078456
48. Schaeffer, E. M., Srinivas, S., & Adra, N. NCCN guidelines updates: Prostate cancer. Journal of the National Comprehensive Cancer Network, 23(1), 15-23 (2025). https://doi.org/10.6004/jnccn.2025.0002
49. Schoenfeld, D. Partial residuals for the proportional hazards regression model. Biometrika, 69(1), 239-241 (1982). https://doi.org/10.1093/biomet/69.1.239
50. Sexton, K. E., Lee, J. Y., & Bergan, R. C. Comparative analysis of survival prediction models in prostate cancer. Prostate Cancer and Prostatic Diseases, 26(2), 289-296 (2023). https://doi.org/10.1038/s41391-022-00612-4
51. Steyerberg, E. W., & Harrell, F. E. Prediction models need appropriate internal, external, and external validation. Journal of Clinical Epidemiology, 158, 142-148 (2024). https://doi.org/10.1016/j.jclinepi.2024.03.012
52. Steyerberg, E. W., Vickers, A. J., Cook, N. R., et al. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology, 21(1), 128-138 (2010). https://doi.org/10.1097/EDE.0b013e3181c30fb2
53. Sung, H., Ferlay, J., & Siegel, R. L. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide. CA: A Cancer Journal for Clinicians, 71(3), 209-249 (2021). https://doi.org/10.3322/caac.21660
54. Therneau, T. M., & Grambsch, P. M. Modelling Survival Data: Extending the Cox Model (2nd ed.). Springer (2023). https://doi.org/10.1007/978-3-031-35246-5
55. Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., & Wei, L. J. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in Medicine, 30(10), 1105-1117 (2011). https://doi.org/10.1002/sim.4154
56. Van Calster, B., McLernon, D. J., & van Smeden, M. Calibration of risk prediction models: A review of recent developments. Diagnostic and Prognostic Research, 8(1), 2 (2024). https://doi.org/10.1186/s41512-024-00168-8
57. Van Calster, B., Nieboer, D., & Vergouwe, Y. Calibration of risk prediction models: A review. Diagnostic and Prognostic Research, 7(1), 5 (2023). https://doi.org/10.1186/s41512-023-00150-8
58. Vickers, A. J., van Calster, B., & Steyerberg, E. W. Net benefit approaches to the evaluation of prediction models. BMJ, 381, e074526 (2023). https://doi.org/10.1136/bmj-2022-074526
59. Wang, J., Zhang, Y., & Liu, X. Ensemble survival methods for cancer prognosis: A comparative benchmarking study. Artificial Intelligence in Medicine, 160, 102934 (2025). https://doi.org/10.1016/j.artmed.2025.102934
60. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018 (2024). https://doi.org/10.1038/sdata.2016.18
61. Wolbers, M., Blanche, P., & Koller, M. T. Concordance measures for competing risks: A practical guide. Biometrical Journal, 65(1), e2200156 (2023). https://doi.org/10.1002/bimj.202200156

62. Wolbers, M., Blanche, P., & Koller, M. T. Concordance measures for competing risks: A tutorial. Biometrical Journal, 66(1), e2200256 (2024). https://doi.org/10.1002/bimj.202200256
63. Wong, M. C., Goggins, W. B., & Wang, H. H. Missing data in prognostic studies: A practical guide. Journal of Clinical Epidemiology, 158, 156-165 (2023). https://doi.org/10.1016/j.jclinepi.2023.03.014
64. Wongvibulsin, S., Wu, K. C., & Zeger, S. L. Clinical risk prediction with random survival forests. Journal of the American Medical Informatics Association, 32(1), 89-98 (2025). https://doi.org/10.1093/jamia/ocae256
65. Xu, Y., Li, H., & Wang, P. A systematic comparison of machine learning methods for survival analysis in cancer research. Computational and Structural Biotechnology Journal, 21, 2125-2137 (2023). https://doi.org/10.1016/j.csbj.2023.03.012
66. Zhou, M., Li, L., & Wang, Y. Evaluating the proportional hazards assumption in modern survival analysis. Statistical Science, 38(4), 521-538 (2023). https://doi.org/10.1214/23-STS891