Assignment 4: Survival Analysis

Deadline: Friday, July 11th 2025

Part 4: Short-Answer Questions (Upload to Blackboard as a .PDF)

Answer the following based on your work in **Part 3**:

1. **Please provide the link to your public GitHub repository.**

https://github.com/n01751968/assignment4_survival_analysis

2. **Based on your work in Part 3, please fill out the following table:**

Method	Strengths	Limitations	Key Use Cases	Your Observations
Kaplan-Meier	Simple to interpret and visualize. Non-parametric and works well for comparing groups.	Cannot handle covariates. Assumes homogeneous groups.	Comparing survival between two or more groups (e.g., Male vs Female).	The KM plot showed no significant survival difference between males and females (p = 0.67).
Cox Proportional Hazards	Can model multiple covariates. Offers interpretability through hazard ratios.	Assumes proportional hazards. Sensitive to multicollinearity.	Evaluating the influence of clinical features like age or treatment.	Dropping low variance covariates like Stage helped improve model stability.
Random Survival Forests	Handles complex interactions and nonlinear relationships. No strict assumptions.	Less interpretable. Requires more computational resources.	Predictive modeling in complex clinical data.	RSF identified important variables like Age and ECOG PS, and showed better predictive performance.

3. **What are the primary differences between Kaplan-Meier (KM) analysis, Cox regression, and Random Survival Forests?**

Kaplan-Meier is a non-parametric method used to estimate survival probabilities over time and compare groups. It does not include covariates. Cox regression is a semi-parametric model that incorporates covariates to assess their effect on survival. It assumes proportional hazards. Random Survival Forests are fully non-parametric ensemble models that use decision trees to capture nonlinear relationships and interactions without making distributional assumptions.

4. **What assumptions are made by Cox Proportional Hazards regression? How can these be evaluated?**

Cox regression assumes proportional hazards, meaning the hazard ratios between groups remain constant over time. This assumption can be evaluated using Schoenfeld residuals or log-minus-log plots. If the assumption is violated, the model may produce biased estimates.

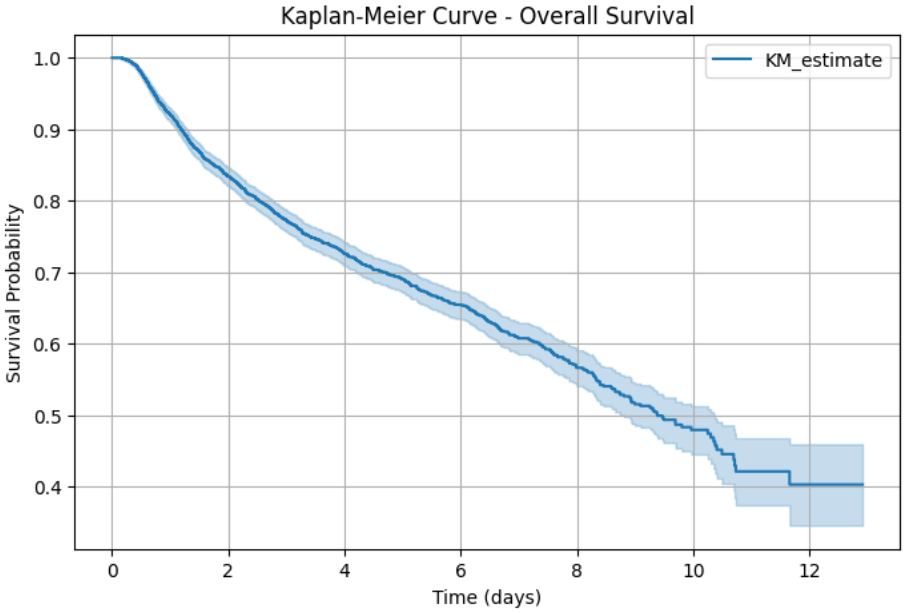
5. **Which method provided the best balance between interpretability and predictive performance?**

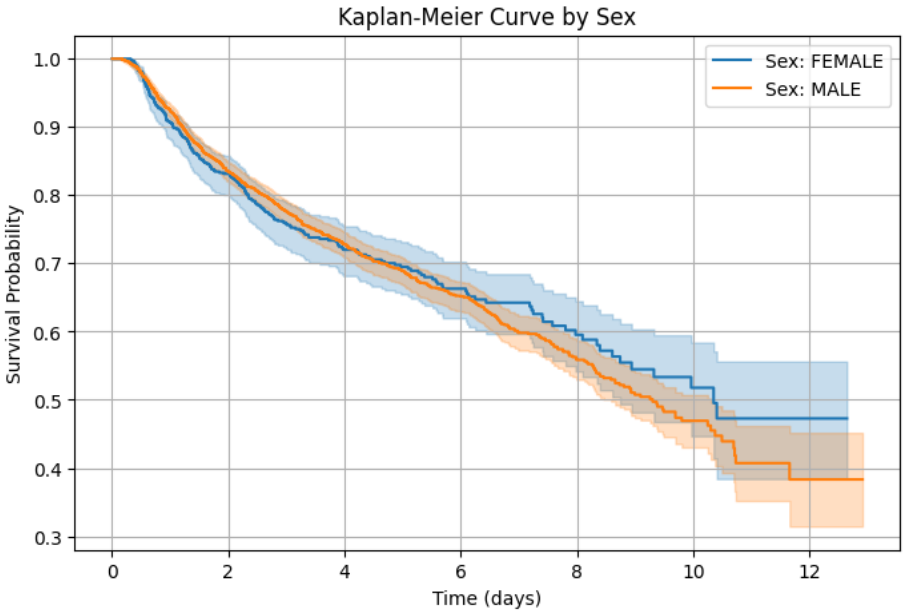
Cox regression offered the best balance. It allowed interpretation through hazard ratios while still incorporating covariates. Random Survival Forests had better predictive performance but were less interpretable. Kaplan-Meier was simple to use but limited in flexibility.

6. **Identify any features that consistently demonstrate predictive power across different methods and highlight their potential clinical significance.**

Age and ECOG PS consistently showed predictive power across Cox regression and RSF. These features may reflect a patient's baseline health and ability to tolerate treatment, which are critical factors in predicting survival outcomes in cancer patients.

Deadline: Friday, July 11th 2025





Deadline: Friday, July 11th 2025