Robust and *fair* time-to-event framework for predicting cancer-associated *Venous Thromboembolism (VTE)* using routinely-collected clinical and panel-sequencing data

Intae Moon[1,3], Hyewon Jeong[1], Alexander Gusev[2,3], and Marzyeh Ghassemi[1]

[1]Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Division of Population Sciences, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA
[3]The Broad Institute of MIT and Harvard, Cambridge, MA, USA

Intae Moon
PhD candidate at MIT
Contact info

Motivation

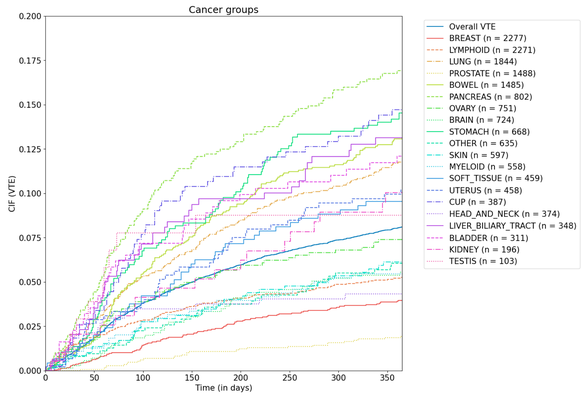
- **Venous Thromboembolism (VTE)** is a frequent, yet **fatal complication in patients with active cancer**, especially while they are receiving chemotherapy.
- Accurate stratification of the VTE risk among patients with cancer may allow clinicians to improve clinical outcome while minimizing side effects due to overtreatment.
- A major challenge with accurately identifying patients at high risk for cancer-associated VTE lies in the **heterogeneity of the VTE risk across diverse patient subpopulations**.
- **Our goal** is to **address the heterogeneity in cancers** and **improve the prediction accuracy** of cancer-associated VTE **across diverse patient groups** defined by cancer types and demographics.

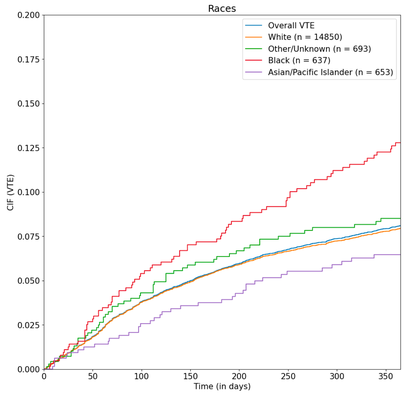
Patient Analysis Cohort

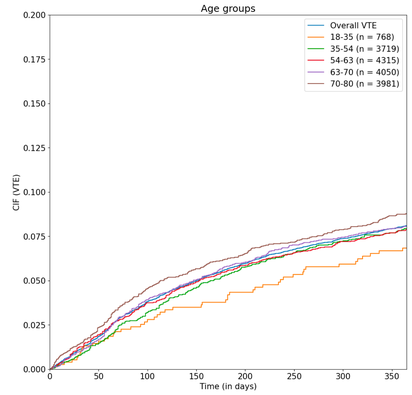
- **16,833 ambulatory patients with cancer** aged 18-80, who were treated and followed up at Dana-Farber Cancer Institute (DFCI) since June 1, 2015.
- None of these patients had an acute VTE episode in the six months leading up to their treatment.

I. Heterogeneity of cancer-associated VTE incidence across diverse patient subgroups

- We utilized Aalen–Johansen estimator to estimate Cumulative Incidence Function (CIF) for VTE event for each group while considering all-cause mortality as a competing event.
- "Time zero" for each patient is the date they began their first treatment regimen.
- We considered various subgroup including **cancer groups**, **ethnicities**, **age groups**, and **biological sexes.**
- We observed **highly heterogeneous VTE incidence** across the considered patient subgroups.



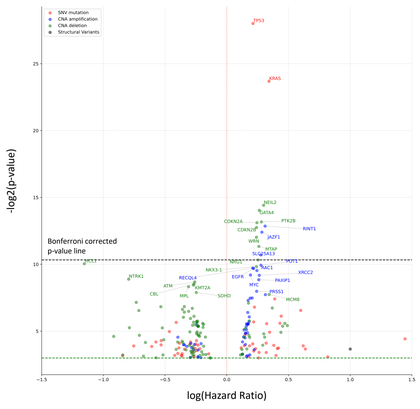




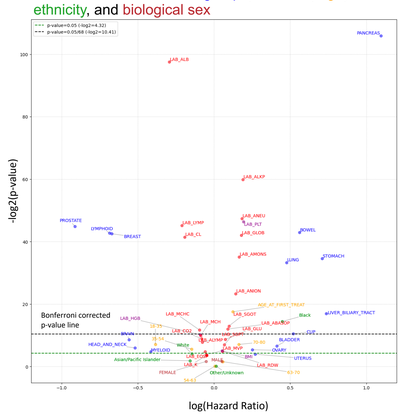
II. Association of panel sequencing and clinical features with VTE

- We investigated **panel sequencing somatic variant features (left)**, **clinical features (middle)** including cancer groups, lab tests, age, ethnicity, and biological sex, and finally **first treatment regimens (right)**.
- To determine Hazard Ratio and p-value of each feature.
  - For panel sequencing somatic variant features and clinical features, we iteratively ran **univariate cause-specific Cox Proportional Hazard regression analysis.**
  - For treatment regimens, we ran **multi-variable analysis adjusting for cancer groups and features for Khorana scores** (clinical baseline for VTE).

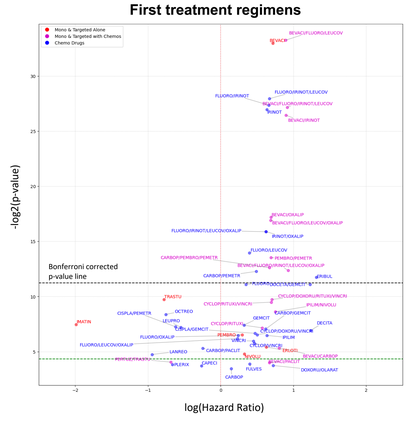
Panel sequencing somatic variant features



Clinical features: cancer groups, lab tests, age, ethnicity, and biological sex



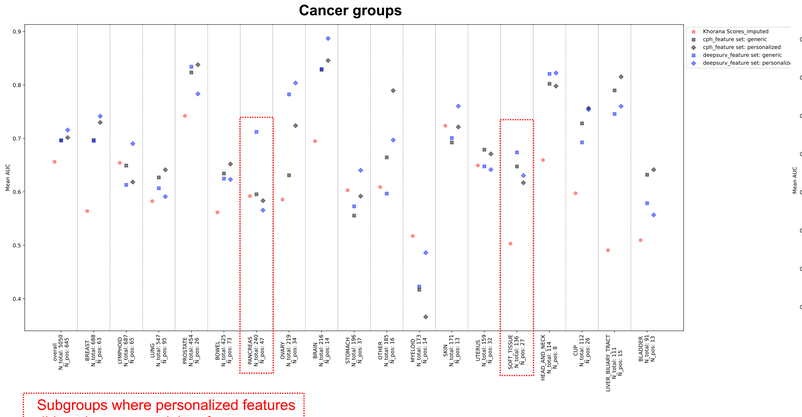
First treatment regimens



III. Prediction of time-to-cancer associated VTE

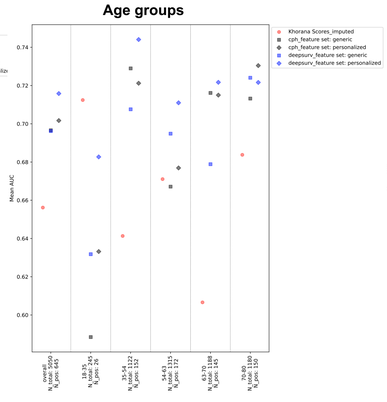
- We utilized Cox Proportional Hazard model and DeepSurv (Katzman et al., 2018).
- We also plotted performance of **Khorana score**, the most widely utilized risk stratification tool for VTE (Khorana et al. 2008).
- We considered two feature sets: **generic** (clinical and treatment features without cancer groups, age, ethnicity, and sex) and **personalized** (all clinical and treatment features).
- Overall, a configuration with more features (i.e., personalized feature set) provides a better performance; but **no single model configuration was universally beneficial to all groups we considered**.

Cancer groups

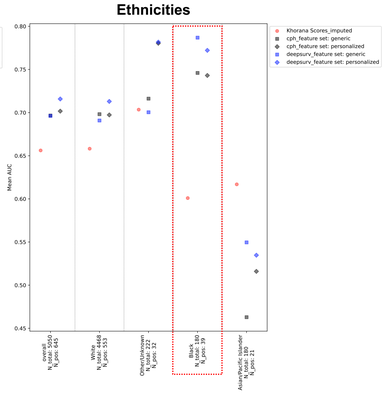


Subgroups where personalized features did not improve model performance.

Age groups



Ethnicities



Ongoing and future work

- Utilize advanced optimization techniques like **Distributionally Robust Optimization (DRO)** to encourage fairness in model predictions.
- Propose a framework that effectively combines models based on their subgroup performances, aiming to **mitigate the harms of any individual model configuration**.
- Outperform current state-of-the-arts for predicting VTE risks as well as **serve a more diverse group of cancer patients**.