**Methods**

**Study Design and Data Preparation**

This study utilised survival analysis to investigate the association between clinical and demographic factors and patient outcomes. The dataset included Censoring occurred for patients who did not experience the event during the follow-up period or were lost to follow-up. X patients (x%) had a 30 day follow up and X patients (%) were followed up at the one-year period.

The data was split into a training and test set, where 80% of the dataset was using to train the models, select features and hyperparameters of the models. The other 20% was used as a hold-out test set to evaluate the performance of each of the trained models.

**Handling Censored Data**

Censored data, a hallmark of survival analysis, arises when the outcome of interest is not observed for all study participants. In this study, both right-censoring and administrative censoring were present. To ensure robust analysis, we utilised techniques designed to handle censored observations appropriately, maintaining the integrity of survival estimates and hazard ratios.

**Data Preprocessing and Feature Engineering**

**Interval Censoring**

136 participants had the outcome of interest, death, (15.54% of those died within study) in between the follow up periods. For these participants, the latest follow up time was assigned as the time of death within the study.

**Missing Data**

Features with more than 20% missing values were removed from the dataset. Of those remaining, they were imputed with the median value.

**Feature Engineering**

Features such as albumin, haemoglobin and platelet count were transformed by including a nonlinear penalty to values that were outside of normal range. The following transformation was used where:

Where is a chosen scaling constant and the upper and lower ranges for each of these features are given in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Sex | Lower Range | Upper Range |
| Albumin | M/F | 33g/L | 48g/L |
| Haemoglobin | M  F | 130g/L  120g/L | 170g/L  150g/L |
| Platelet count | M/F | 150\*109/L | 450\*109/L |

Furthermore, features were standardised to improve training stability and convergence.

**Statistical Modelling Approaches**

**Cox Proportional Hazards Model**

The Cox proportional hazards model was employed to assess the relationship between covariates and the hazard of the event. The model assumes a proportional hazards structure, expressed as:

where is the hazard at time given covariates, is the baseline hazard, and represents the coefficients for covariates.

**Penalised Cox Proportional Hazards Models**

To address potential multicollinearity and improve predictive accuracy, penalised Cox proportional hazards models, including ridge, lasso, and elastic net, were applied (https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780132307). These models incorporate regularisation terms in the partial likelihood function:

Where represents the penalty term, such as for ridge, for lasso, or a combination of both for elastic net. Hyperparameters, such as the penalty coefficient, were tuned using cross-validation.

**Survival Tree**

To explore non-linear relationships and potential interactions between covariates, we constructed survival trees (ST). This approach partitions the data recursively based on covariate values to create homogenous groups with respect to survival. The survival tree was built using recursive partitioning with log-rank splitting criteria (https://europepmc.org/article/med/4042086).

**Random Survival Forest**

Random survival forests (RSFs) were used to capture complex interactions and non-linear effects. This ensemble learning method aggregates multiple survival trees to improve prediction accuracy. Each tree in the forest was grown using bootstrapped samples, with random feature selection at each split. The cumulative hazard function was estimated for each observation by averaging across the forest. Variable importance was assessed using permutation-based measures. (https://projecteuclid.org/journals/annals-of-applied-statistics/volume-2/issue-3/Random-survival-forests/10.1214/08-AOAS169.full)

**Gradient Boosted Survival Tree**

Gradient boosted survival trees (GBSTs) were applied to further enhance predictive performance by sequentially building trees that minimise the loss function related to survival outcomes. The approach used the negative log partial likelihood as the loss function, iteratively improving model fit. Hyperparameters, including the learning rate, maximum tree depth, and number of boosting iterations, were optimised through grid search with cross-validation.

**Model Evaluation**

Model performance was assessed using multiple metrics to ensure a comprehensive evaluation of the models:

**Concordance Index (C-index)**: The C-index, as described by Harrell et al. (<https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780030207>) was used to measure the discriminative ability of the models. This metric reflects the proportion of correctly ordered pairs of observations based on predicted survival times.  Harrell et al. C-index is known to be biased upwards if the amount of censoring in the test data is high. As such, the C-index was also calculated using Uno et al’s C-index (<https://onlinelibrary.wiley.com/doi/10.1002/sim.4154>) which is more robust in such situations.

**Cumulative Dynamic AUC**: The cumulative dynamic area under the curve (AUC) was used to assess time-dependent discrimination. This metric evaluates the ability of a model to correctly rank survival probabilities over time and is particularly useful in survival settings where predictive accuracy evolves as time progresses (Blanche et al., 2013) (https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.5958). The cumulative dynamic AUC was chosen for its robustness in quantifying model performance in the presence of censored data and enables the evaluation of the models across different time horizons.

**Integrated Brier Score (IBS)**: The IBS was employed to quantify the prediction error across the follow-up period and is often used to assess calibration. It evaluates the deviation of predicted survival probabilities from the observed outcomes, integrating over time ([https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0258%2819990915/30%2918%3A17/18<2529%3A%3AAID-SIM274>3.0.CO%3B2-5?ref=https://githubhelp.com](https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0258%2819990915/30%2918%3A17/18%3c2529%3A%3AAID-SIM274%3e3.0.CO%3B2-5?ref=https://githubhelp.com)).

**Software**

All analyses were conducted using Python 3.12. All code can be viewed at https://github.com/harrisonnguyen/tavi/