Alzheimer’s Disease Progression and Survival Analysis

Leah Zhang , Sarah Baalbaki

Introduction:

Alzheimer's disease (AD) is a progressive brain disease characterized by memory loss and cognitive decline, with no known treatment, and medications can offer only temporary relief. Approximately 1 in 9 individuals over 65 has Alzheimer's, and 1 in 3 seniors dies due to Alzheimer's or a related dementia. In order to study Alzheimer’s, different types of data can be used: genomic data, demographic data, MRI imaging data, cognitive testing data, cerebrospinal fluid (CSF) data. Previous research concerning survival analysis has been performed and proven by Bewick et al [4], where potential models to estimate survival curves are proposed. Moreover, survival analysis in Alzheimer’s was also studied, as seen from Sharma et al [2] in their paper Time-to-event prediction using survival analysis methods for Alzheimer's disease progression. This research focuses on a deep learning based survival analysis for Alzheimer’s. Successfully predicting survival time is important because it will help in aspects such as personalized medicine, extracting important hidden features, helping in improving medical equipment assignment (knowing which patients need more attention), and potentially implementing into other areas of research. Our project focuses on a survival analysis of Alzheimer's disease progression to gain a deeper understanding of Alzheimer's disease and aid in the development of effective treatments using techniques designed for survival analysis (time-to-event data) to explore the factors influencing the disease's course, aiming to provide insights for future treatment strategies. Our research builds upon the foundation laid by Sharma et al [2], expanding and improving it in several key aspects. Firstly, we have enriched their study by integrating a cerebrospinal fluid dataset with the Uniform Data Set. Furthermore, by utilizing a variety of feature extraction methods and examining a wider array of machine learning models, such as Random Survival Forest and DeepSurv, we achieved significantly improved model performance (while they reported a best performance c-index of 0.78, our approach yielded a c-index of 0.957). Through our investigations, we establish that the Cox Mixture Model is a leading-edge machine learning method for handling time-to-event data. DeepSurv also emerges as a promising model, delivering comparably high accuracy. Notably, DeepSurv offers a significant advantage in terms of computational efficiency over the Cox Mixture Model, though it slightly lags in accuracy. This efficiency could offer substantial benefits in future applications.

Data:

We created our dataset from the data provided by the NACC (National Alzheimer's Coordinating Center). We combined different datasets, which are: UDS (Uniform Data Set) that contains features for demographics, medications, health history, physical parameters, parkinson's disease ratings, and a neuropsychiatric questionnaire, a Neuropathology dataset, a genetics dataset including APOE (apolipoprotein E), FTLD (Frontotemporal lobar degeneration) and Lewy body dementia (LBD), and finally a CSF Biomarker dataset. We requested access to the data on the NACC website, and combined our data based on patient ID’s, and had a total of 18,688. We removed columns with one unique value across all entries, and filtered out columns that have more than 70% missing data. We chose this cutoff since we have many missing values as we are combining multiple data sources and not all patients necessarily have all information documented or exams done. We then imputed the missing values based on the mean of each feature, since NaN values are not accepted among most of the machine learning models.

Since we have many data sources, we performed feature selection to keep only the relevant features. We started out with 1,958 features, and performed both mutual information feature selection and feature selection based on F-1 score. We did those by running 10-fold cross validation and obtaining the

common features. The resulting top features were very similar, so we used the information from the mutual information score, and made sure to choose a number of features such that we retain 95% of our data, and ensure no significant information loss, such that we are left with 563 features. A plot of top 100 features returned by mutual information is included in the Appendix.
For our different data tasks, we restructured and reformatted the data differently:

- AD stage prediction: we kept multiple entries corresponding to the same patient in the dataset, where different entries signify different visits that have updated AD stages or updated results for other factors (neuropsychiatric data or CSF data). We also made sure to map the AD level to integers instead of floats, which represents our outcome (label) in this case, resulting in the corresponding labels:

AD Level	Encoded AD Level (Label)	Significance
0.0	0	No impairment
0.5	1	Questionable impairment
1.0	2	Mild impairment
2.0	3	Moderate impairment
3.0	4	Severe impairment

Table 1: AD levels

- Survival Analysis for Disease Progression Prediction: we merged the multiple entries for a single patient together, based on their NACCID, and added a new column to the dataset where we calculated the ‘cdr_difference’ (by subtracting ‘last_cdr’ and ‘first_cdr’) and added that as an outcome to track/predict disease progression, which we then binarized to indicate disease progression occurrence as a 1, and no progression as a 0, which resembles our “event” variable in the time to event analysis. We used the variable of NACCDAYS which specifies the difference between the initial and final visit as the duration of progression, which represents the “time” in our time to event analysis.
- Survival Analysis for Survival Prediction: We used the NACCDIED variable as an indicator of the patients' vital status. A value of 0 denotes that the patient is alive, while a value of 1 signifies that the patient has passed away.

Methods:

To predict the AD stage, we use a random forest model, with 10-fold cross validation. The random forest model is an ensemble learning method that will make multiple decision trees and output the mode of the classes, which in our case is the AD stage. We use 10-fold cross validation and since we have multiple entries for some patients, we make sure to keep all entries for the same patient either in training, or test, or validation, when we split up the dataset, in order to avoid data leak and unreliable results. We evaluate our model based on accuracy that assesses the percent of correct predictions made out of the total predictions. We also evaluate the precision, which measures the proportion of true positive predictions among all positive predictions, recall, which assesses the fraction of true positives among all positives, and F1 score which is the harmonic mean of precision and recall.
To track the disease progression in our survival analysis, we employ a diverse set of models designed to capture and interpret the complexities of time-to-event data, and use the following models:

- Random Survival Forest (Baseline)
- Cox Proportional Hazards Mixture Model
- DeepSurv

We use the Random Survival Forest as our baseline model, which leverages an extension of the random forest model to handle the survival data, where our outcome is measuring the time until a certain event

occurs. This model provides insights into the progression of AD. The Cox Proportional Hazards Mixture Model, on the other hand, is a robust statistical approach designed specifically for survival analysis. It focuses on modeling the hazard function, which represents the probability that an individual under observation at a time t experiences an event at that time. In our context, we aim for the event to signify the progression of the Alzheimer's Disease (AD) stage or death. Utilizing the hazard function, the Cox Mixture Model generates predictions for risk scores (described below) based on the provided dataset.
We also use the DeepSurv model, which is a deep learning model specifically for survival analysis. It has a neural network architecture, and so it does really well at predicting the time until a specific event of interest occurs, which is the AD stage progression, or death. As a result of using a deep learning model, it is able to learn patterns and capture complex relationships in the data. The neural network also focuses on modeling the hazard function stated above, which further improves the capacity to detect differences in the progression of diseases over time.
For all these models, we use already existing packages that are able to perform the survival analysis given data in a specific format, which we have as two columns in the dataframe: ‘time’, and ‘event’ , where time corresponds to the time between events and event is a binary variable of ‘1’ indicating event occurrence (disease progression or death) and ‘0’ indicating otherwise (no event occurrence).
When evaluating, for each data point, the models calculate the risk score, which is the predicted survival probability (a value between 0 and 1) for each individual in the dataset- a measure of the likelihood that an individual will survive beyond a certain time point. Since we use RSF, the ensemble prediction is simply the average resulting risk score across all trees in the forest. When testing our models on unseen data, unlike the model that provides survival probabilities over time, the Cox Proportional-Hazards Model, which estimates hazard ratios. The Cox model and deepsurv assess this based on the input features associated with each observation, and it is calculated as the exponentiated sum of the products of their feature values and the corresponding coefficients in the model. It represents the relative hazard for the individual compared to the baseline hazard, where a higher risk score indicates a higher estimated risk of experiencing the event of interest. The prediction from DeepSurv works similarly with the Cox Proportional-Hazards Model.
To evaluate these survival models, we use the concordance index (C-index) and integrated brier score (IBS). C-index is a metric for survival analysis that assesses how well the model's predicted risk scores (or hazard scores for Cox model) rank the survival times, and so a higher C-index indicates between model performance since it means the model’s predicted scores rank the survival times more accurately. IBS is a single metric that summarizes the accuracy of a probabilistic forecast for time-to-event data by averaging the squared differences between observed outcomes and predicted probabilities over time. A higher C-index indicates improved model performance, while a lower IBS signifies enhanced predictive accuracy of the model.

Results:

For the task of AD level prediction using random forest, we perform the AD classification and obtain an accuracy of 94% for our predictions, with a standard deviation of 0.03, as the results from training our model with 10-fold cross validation.
For the survival analysis prediction, we run the three models (using already created packages) and as input we specify the ‘time’ and ‘event’ columns and get the risk scores and calculate the C-index as mentioned above.
We get the following results for each model with best performance:

	RSF (Baseline)	DeepSurv	Cox Mixture Model
C-index	0.910	0.933	0.957
IBS	0.0302	0.0433	0.0196

Table 2: Best Accuracies for Different Models

These results show that the RSF, which is our baseline, performs well, with a C-index of 91.0%. The other two models perform even better than the baseline, with DeepSurv attaining a C-index of 93.3% and Cox Mixture Model attaining a C-index of 95.7%. We can see that the models perform at predicting the risk or hazard scores, which reflect the likelihood for survival at a certain time. It's noteworthy that DeepSurv's performance is marginally inferior to the Cox Mixture Model, yet it operates significantly faster. While the Cox Mixture Model requires 15 minutes to process the test dataset, DeepSurv accomplishes this in approximately 1 minute. This efficiency positions DeepSurv as a promising model for real-time clinical data prediction.
For the Cox mixture model, we plot the hazard scores as a function of survival time, separated for both groups where AD progression occurs and AD progression doesn't occur. We can see that for the group with AD disease progression, their scores are relatively higher than those for the non-AD progression group, for the same survival time. We can also see the scores decreasing as survival time progresses, so we can say that this trend in is consistent with the nature of survival analysis, since as individuals survive for longer durations, the hazard scores tend to decrease, indicating a diminishing risk of experiencing the event, and that over time, the likelihood of the event (AD progression) occurring decreases for individuals who have not experienced it. Also, this graph provides a nice visual insight into the model's discriminative ability between the AD and non-AD progressed groups, where the separation of hazard curves between the groups reinforces the model's ability to distinguish between different risk profiles within the studied population.

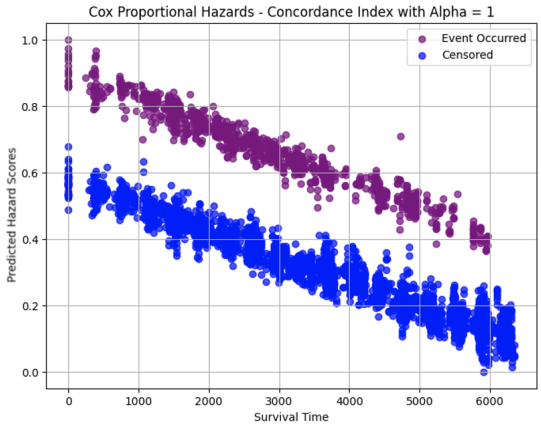

Figure 1: Predicted Hazard Score as a function of Survival Time

We also plot the Kaplan Meier curve which is a graphical representation used in survival analysis to depict the estimated probability of survival over time. We plot it for the different AD groups (0-5) in order to compare the survival outcomes between different AD groups.

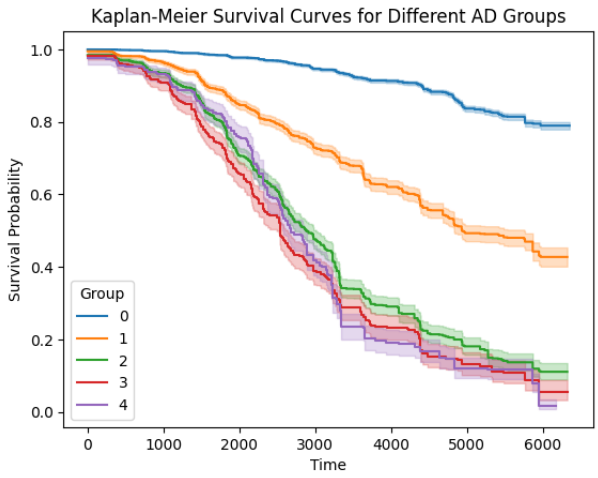

Figure 2: Kaplan Meier Curves for Different AD Stages

We can see that for the group 0, which has no impairment, we get a relatively slowly decreasing curve for the survival of this group, which depicts a relatively normal survival curve where the probability of survival decreases as time increases due to normal death trends. For group 1, with questionable impairment, we see a more intensely decreasing curve, where the decrease might signify patients getting an Alzheimer's diagnosis after having questionable impairment, leading to more cases of death across the same time. For groups 3,4, and 5, with mild impairment, moderate impairment, and severe impairment, we see a very steep curve drop reflecting a sharp decrease in the survival probability of the groups over time, reflecting that Alzheimer's diagnosis does affect the diagnosed groups' survival times, as they have lower survival probability at the same time as other more normal groups that have no Alzheimer's diagnosis. Also, we can see that the first two groups have a steadier decrease in survival probability over time, but the diagnosed groups have a very sharp decrease over time, meaning that not many make it that far.

Conclusion and Future Work:

In this project, we worked with Alzheimer's data where we were able to both predict the AD stage, and perform a survival analysis to assess the time to disease progression and patient death. We saw that our models performed well at both predicting the AD stage and performing the survival analysis. From this, we can conclude that our data did indeed give insights into helping diagnose AD stage and also in performing the survival analysis. The important features that contributed were seen to be from CSF data, as well as neurological examinations indicating different tasks and also some demographic data.

We can also see that from our survival analysis we were able to predict if AD stage progression happens with more than 90% accuracy, and in the relative time frame it happens, which can help doctors in selecting the next course of action for their patients. As no known treatments exist so far, this might help in simply prescribing some medications that offer temporary relief, how to proceed with the patients and to know how much care they should allocate to them based on the speed of their deterioration and the stage they are at, and maybe even help select candidates for possible future clinical trials.

For future work, we can hypothesize that adding MRI images and scans can also help in diagnosis and might also be a possible next step for such a project, to achieve even better and more accurate results. So, we can further their research by incorporating MRI imaging data also, using a multi-modal approach that will hopefully improve our findings.

References:

[1] Breijyeh, Z., & Karaman, R. (2020). Comprehensive Review on Alzheimer's Disease: Causes and Treatment. Molecules (Basel, Switzerland), 25(24), 5789. https://doi.org/10.3390/molecules25245789

[2] Sharma R, Anand H, Badr Y, Qiu RG. Time-to-event prediction using survival analysis methods for Alzheimer's disease progression. Alzheimer's Dement. 2021;7:e12229. https://doi.org/10.1002/trc2.12229

[3] Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Medical Research Methodology, 18(1), 24.

[4] Bewick V, Cheek L, Ball J. Statistics review 12: survival analysis. Crit Care. 2004 Oct;8(5):389-94. doi: 10.1186/cc2955. Epub 2004 Sep 6. PMID: 15469602; PMCID: PMC1065034.

Appendix:

Figure 1: Survival Analysis of Cox Proportional Hazard measured by c-index. Predicted Risk Scores as a function of Survival Time, with different alpha values:

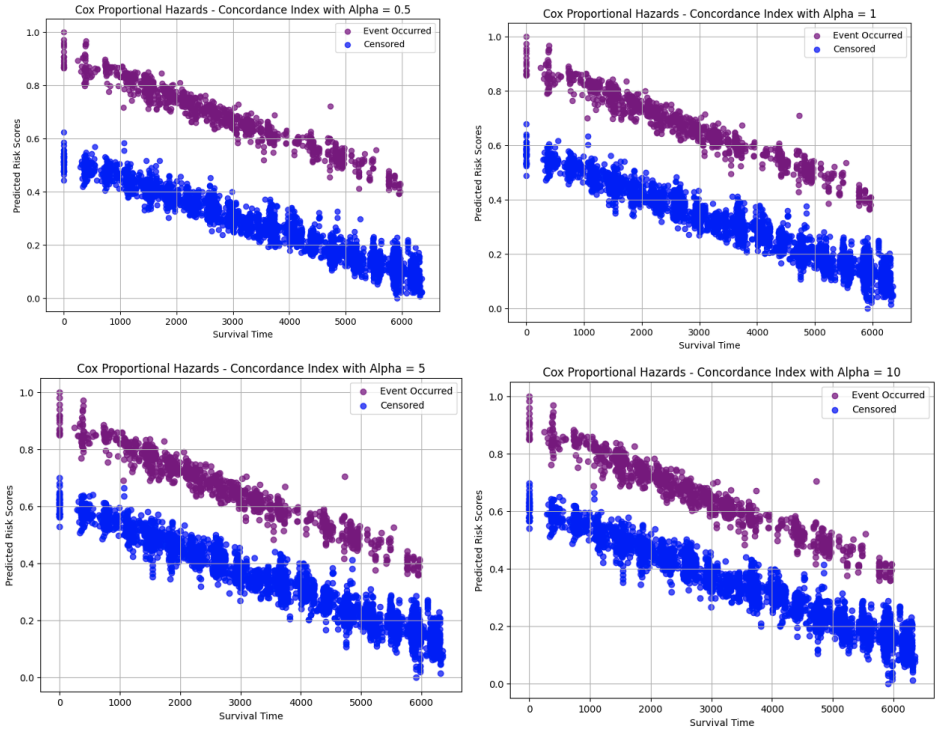

Figure 2: Progression Analysis of Cox Proportional Hazard measured by c-index. Predicted Risk Scores as a function of Survival Time, with different alpha values:

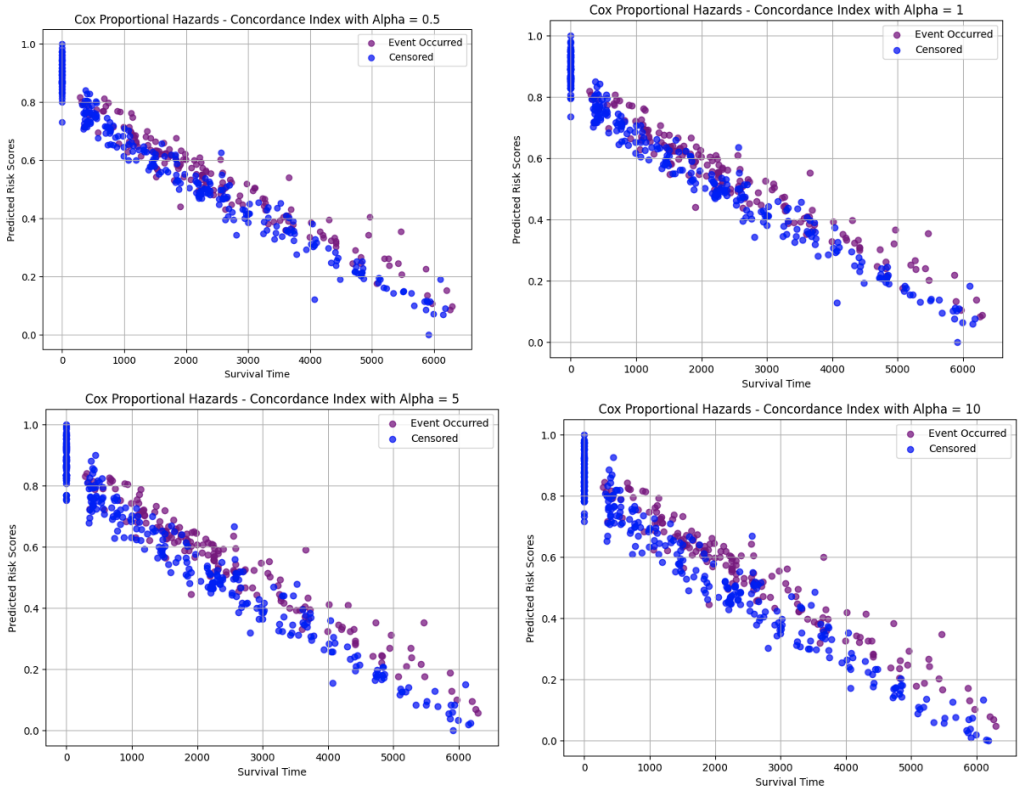

Table 1: Models run on both Survival dataset and Progression dataset

	Parameter1	Parameter2	Parameter3	Parameter4
Cox	Alpha=0.5	Alpha=1	Alpha=5	Alpha=10
DeepSurv	net1	net2	net3	net4
Random Survival Forest	N_estimator =100	N_estimator =300	N_estimator =500	N_estimator =1000

Table 2: Survival results measured by c-index and integrated brier score (IBS)

	Architecture 1	Architecture 2	Architecture 3	Architecture 4
Cox	C-index:0.987 IBS: 0.0080	C-index: 0.987 IBS:0.0079	C-index: 0.987 IBS:0.0079	C-index: 0.987 IBS: 0.0079
DeepSurv	C-index:0.985 IBS: 0.0120	C-index:0.986 IBS: 0.0121	C-index:0.986 IBS: 0.0131	C-index:0.984 IBS: 0.0148
Random Survival Forest	C-index:0.981 IBS: 0.0122	C-index:0.982 IBS: 0.0121	C-index:0.981 IBS: 0.0121	C-index:0.982 IBS: 0.0120

Table 3: Progression results measured by c-index and integrated brier score (IBS). The model in bold showed the best performance overall.

	Architecture 1	Architecture 2	Architecture 3	Architecture 4
Cox	C-index:0.955 IBS: 0.0212	C-index: 0.956 IBS:0.0203	**C-index: 0.957** **IBS:0.0196**	C-index: 0.956 IBS: 0.0201
DeepSurv	C-index:0.930 IBS: 0.0476	C-index:0.930 IBS: 0.0444	C-index:0.930 IBS: 0.0469	**C-index:0.933** **IBS: 0.0433**
Random Survival Forest	C-index:0.909 IBS:0.0301	C-index:0.909 IBS: 0.0302	C-index:0.909 IBS: 0.0305	**C-index:0.910** **IBS: 0.0302**

The architecture used in DeepSurv:

```
net1 = torch.nn.Sequential(
    torch.nn.Linear(survive_X_train.shape[1], 32), nn.Linear(survive_X_train.shape[1], 32),
    torch.nn.ReLU(),
    torch.nn.Linear(32, 32),
    torch.nn.ReLU(),
    torch.nn.Linear(32, 1)
)
net2 = torch.nn.Sequential(
    torch.nn.Linear(survive_X_train.shape[1], 32),
    torch.nn.ReLU(),
    torch.nn.Linear(32, 64),
    torch.nn.ReLU(),
    torch.nn.Linear(64, 64),
    torch.nn.ReLU(),
    torch.nn.Linear(64, 1)
)
net3 = torch.nn.Sequential(
    torch.nn.Linear(survive_X_train.shape[1], 64),
    torch.nn.ReLU(),
    torch.nn.Linear(64, 64),
    torch.nn.ReLU(),
    torch.nn.Linear(64, 64),
    torch.nn.ReLU(),
    torch.nn.Linear(64, 1)
)
net4 = torch.nn.Sequential(
    torch.nn.Linear(survive_X_train.shape[1], 128),
    torch.nn.ReLU(),
    torch.nn.Linear(128, 128),
    torch.nn.ReLU(),
    torch.nn.Linear(128, 64),
    torch.nn.ReLU(),
    torch.nn.Linear(64, 1)
)
```

Figure 4: Survival Analysis Concordance Index Plots for Predicted Risk Scores as a function of Survival Time, with different number of estimators, for Random Survival Forest:

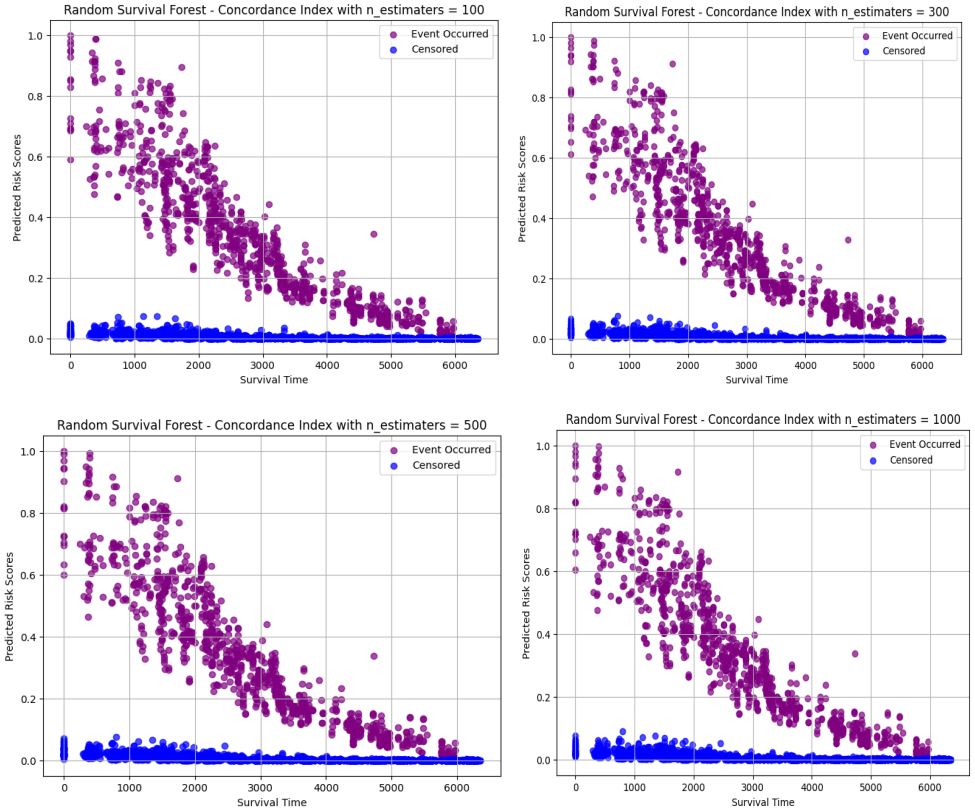

Figure 5: Survival Analysis Concordance Index Plots for Predicted Risk Scores as a function of Survival Time, with different number of estimators, for DeepSurv:

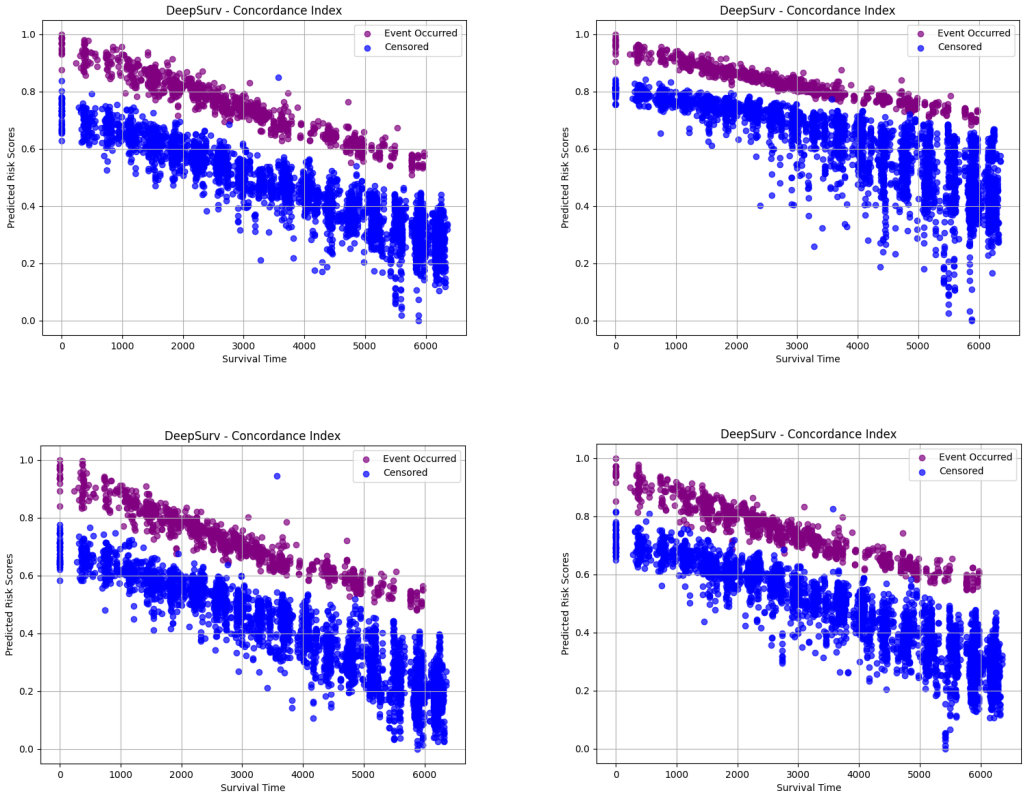

Figure 6: Progression Analysis Concordance Index Plots for Predicted Risk Scores as a function of Survival Time, with different number of estimators, for DeepSurv:

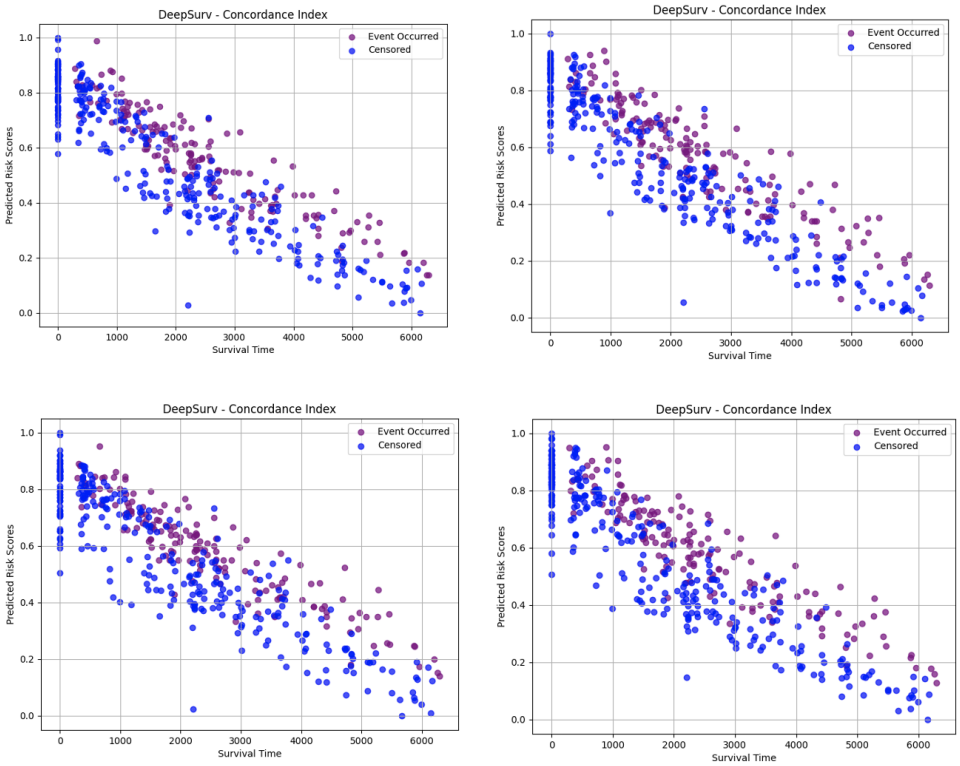

Figure 7: Progression Analysis Concordance Index Plots for Predicted Risk Scores as a function of Survival Time, with different number of estimators, for Random Survival Forest:

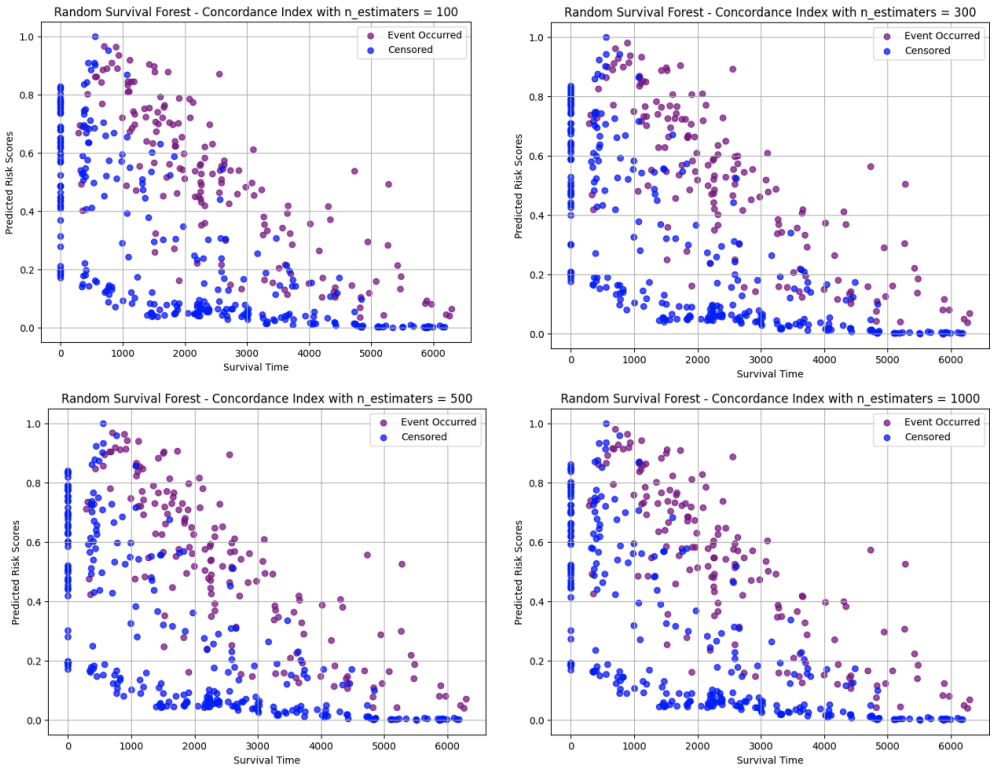

Figure 8: Top 100 Features obtained from Mutual Information feature Selection Method