# CTD^2 BeatAML DREAM Challenge: Predicting ex-vivo Drug Sensitivity and Survival Analysis

Ruibo Zhang, Daniel Nolte, and Saugato Rahman Dhruba

Department of Electrical and Computer Engineering, Texas Tech University, Lubbock, Texas, USA

### Abstract

Screening for cancer drugs is often done simultaneously for multiple drugs, and the correlations between drugs often hold a wealth of information. We consider the problem of predicting multidrug sensitivity in cancer cell lines and propose a predictive approach to better exploit drug response correlations. Our model utilizes multitask bagged linear regression with post covariance alignment of the prediction values to predict the area under the curve (AUC) values of individuals inhibitor responses from genomic features. Our approach ranked 4th among 27 teams when applied in CTD^2 BeatAML DREAM Challenge. The results show that the incorporation of drug responses covariance structures raises the prediction correlations, and proper stacking method lowers the prediction errors, illustrating the effectiveness of our approach in multi-task regression problem with significant task correlations.

### Introduction

In the era of precision medicine, AML patients have few therapeutic options, with “7 + 3” induction chemotherapy having been the standard for decades (Bertoli et al. 2017). While several agents targeting the myeloid marker CD33 or alterations in FLT3 or IDH2 have demonstrated efficacy in patients (Wei and Tiong 2017), responses are uncertain in some populations (Castaigne et al. 2012) and relapse remains prevalent (Stone et al. 2017). These drugs highlight both the promise of targeted therapies in AML and the urgent need for additional treatment options that are tailored to more refined patient subpopulations in order to achieve durable responses.

The BeatAML initiative was launched as a comprehensive study of the relationship between molecular alterations and ex-vivo drug sensitivity in patients with AML. One of the primary goals of this multi-center study was to develop a discovery cohort that could yield new drug target hypotheses and predictive biomarkers of therapeutic response. Patient samples were subjected to whole-exome sequencing (WES), transcriptomic sequencing (RNA-seq), and ex-vivo functional drug sensitivity screens \href{<www.vizome.org>}{(Tyner et al. 2018)}. This rich resource should enable the discovery of molecular correlates of drug response and putative patient populations most likely to respond to targeted agents. Indeed, analysis of these data has already revealed numerous correlations of drug sensitivity or resistance with a variety of mutational subsets of disease as well as numerous gene expression signatures that correlated with drug sensitivity/resistance (Tyner et al. 2018).

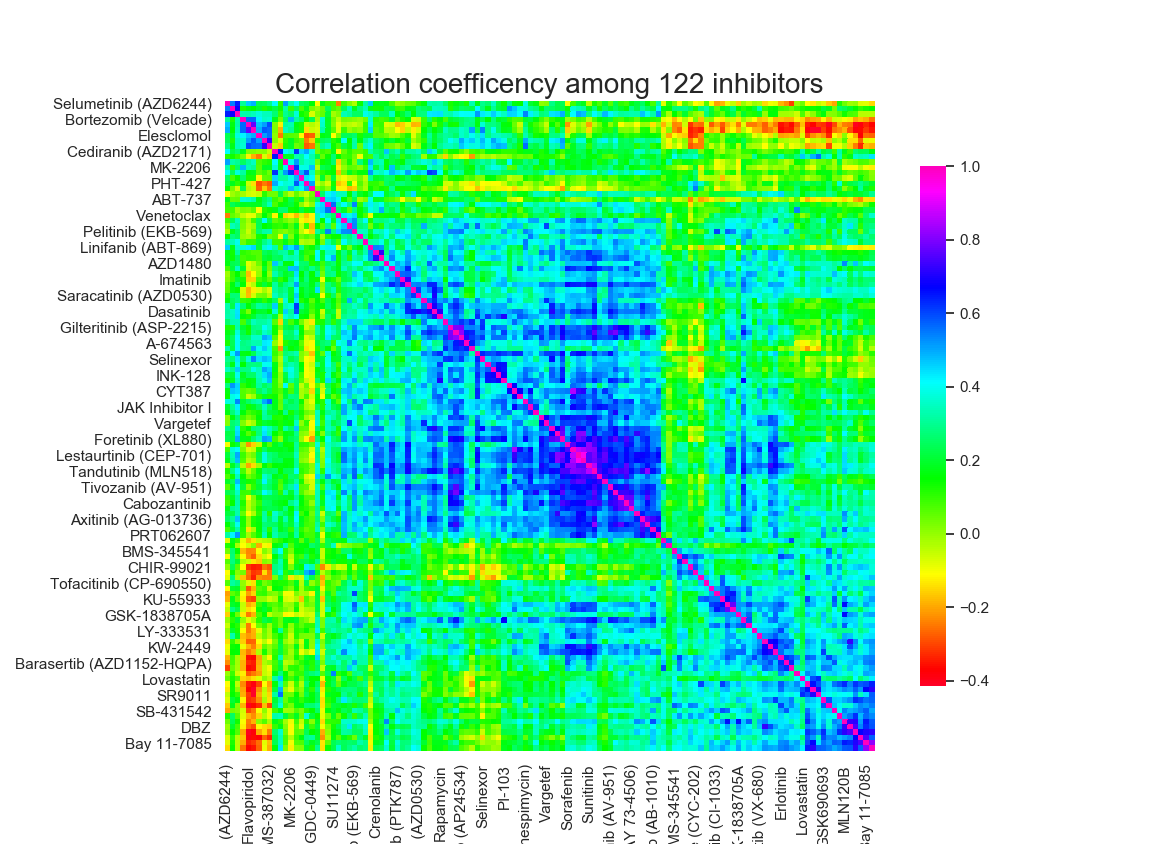
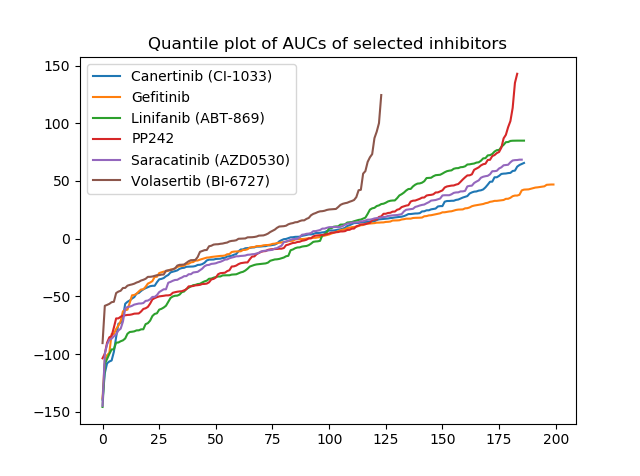
Genetic characterization has long been used in assisting the screening of chemotherapy agents and targeted drugs. The recent large-scale data collection projects, including NCI-60 database, CCLE, GDSC, and CTD2 (*name and refs*), measured the drug sensitivity of genetic characterized cancer cell lines, providing a source database for response prediction tasks. Various of approaches have been proposed to address this problem. The representative methods are, linear regressions, such as elastic net in CCLE [ref](https://www.nature.com/articles/nature11003), elastic net associated with auto-encoder feature extraction [ref](https://pubmed.ncbi.nlm.nih.gov/29133589/?dopt=Abstract); random forest, [ref](Wan%20Q,%20Pal%20R%20(2014)%20An%20Ensemble%20Based%20Top%20Performing%20Approach%20for%20NCI-DREAM%20Drug%20Sensitivity%20Prediction%20Challenge.%20PLoS%20ONE%209(6):%20e101183.%20doi:10.1371/journal.pone.0101183), [ref](https://www.nature.com/articles/s41598-017-11665-4), [ref](Rahman,%20Raziur,%20John%20Otridge,%20and%20Ranadip%20Pal.%20%22IntegratedMRF:%20random%20forest-based%20framework%20for%20integrating%20prediction%20from%20different%20data%20types.%22%20*Bioinformatics*%2033.9%20(2017):%201407-1410.), ; SVM [ref](https://bmccancer.biomedcentral.com/articles/10.1186/s12885-015-1492-6), ; kernel-based methods 引用tdb; and methods based on similarities and nearest neighbors, [ref?]

Just as there is genetic correlation between cell lines and similarity in chemical structure between drugs, there is also a correlation between different drugs responses, which stem from similarities between cell lines and drugs. (ref?). The covariate structure of responses between inhibitors provides a wealth of information that can be used to improve prediction accuracy however often ignored in modeling. Many approaches are proposed for this purpose. Based on random forest, Haider et.al. incorporated the difference between the two empirical copulas of the training data and node samples into the cost criterion. [ref](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4684346/) .

selecting the most promising therapy is often more important than accurately predicting the sensitivity to all potential drugs [ref](https://academic.oup.com/bioinformatics/article/34/16/2808/4924716)

In sub challenge 1, the AUC values of tissue specimens after the application of 122 inhibitors are given for a training set where the goal is to predict the AUC values for the target Leaderboard set using relevant genomic and/or clinical features. In Figure 1, the AUC values are plotted against their ranking within each of the selected inhibitors after subtracting the mean. From Figure 1, we observe significant nonlinearity in the AUC distributions for all the selected inhibitors, specifically in the tails, inferring that the highly responsive and highly resistant specimens exhibit extreme AUC values. This inspired us to use the normalized ranking of the AUCs instead of the actual values to better utilize a linear model, which have been shown to perform well in anticancer drug sensitivity prediction [1].

Cancer drug design usually follows the logic of pathway control and relationships between chemical groups and biochemical reactions. Therefore, similar inhibitors are assumed to cause similar responses in similar cell specimens. This can be represented by the covariance structures of the AUC values for different inhibitors. The covariance heatmap for the AUC values of 122 inhibitors are shown in Figure 2. Inhibitors are first clustered according to their dissimilarity measures (i.e., 1 – correlation) and shown following the clustering order . As shown in Figure 2, the correlations between inhibitor-pairs show a clear pattern i.e., the clustered inhibitors have high concordance of up to 0.8 while the non-clustered ones have poor concordance demonstrated by values as low as -0.3. It is, therefore, reasonable to assume that the predicted AUC values for the inhibitors should preserve this covariance structure. To pass the original covariance structure to the predictions from a predictive model, we adopted a covariance alignment method, CORAL, which is well accepted in the cross-domain learning [2] . To elaborate, we first remove the covariance structure from our initial model predictions for all inhibitors through “whitening” (i.e., dividing by the square root of the covariance matrix), and then perform “re-coloring” of the whitened prediction values (i.e., multiply by the square root of the covariance matrix of the training set). From another aspect, we are in effect minimizing the higher order prediction residuals (covariance, in the order of 4), besides the mean square error (MSE) in linear regression, which is in the order of 2. To utilize the multitask intrinsic qualities of our task, we use bagged linear regression as the base model. Bagging compensates the instability of simple linear regression and provides improved prediction accuracy.



**Figure 1. (a) Distribution of the AUC values of 6 inhibitors shown in order. (b) Correlation coefficient among 122 inhibitors.**

### 2 Results

#### 2.1 Results on CTD^2 Beat AML DREAM Sub-Challenge 1 (SC1)

In this drug sensitivity prediction challenge, a total of 122 drugs are applied on [xxx] cell lines (leader board phase and xx for validation phase). The Area-under-the-Curve (AUC) values are given as the responses, and contain NaNs since not all possible combinations between drug and cell lines are tested. The RNA expression, DNA mutations, and clinical data are provided for all cell lines in the dataset. The data are provided pre-publication by [xxxx]

Our approach contains a combination of pre-processing and post-prediction processing, which includes covariance alignment and a values matching map the rankings/quantiles back to actual AUC values. As described before, the rankings of each cell lines are used as the training data to obtain better linearity, and the post-prediction correlation alignment is adopted to minimize the higher order errors. In our approach, a bagged linear regression estimator (n estimators = 5000) is used as our predictive model. To avoid overfitting, we only use a subset of total features (<115 each time) for training each of the 5000 regressors.

The prediction performance is evaluated by 5-fold cross validation. The average Spearman correlation coefficient, Pearson correlation coefficient of all 122 drugs are used as the performance metrics. The 5-fold average performances are evaluated with comparison to commonly used multi-task regression models Ridge regression, random forest, KBMTL, and KRL. *All the parameters are determined by actually using the same data in cross-validation. Best lmbd=0.15*.

to be done..

**Figure 2. Prediction correlations comparisons for different models on CTD^2 BeatAML challenge dataset.**

#### 2.2 Ablation tests on CTD^2 Beat AML DREAM Sub-Challenge 1

We performed ablation tests for our approach to illustrate the effects of each components in the approach. As shown in Table 1, 9 models are evaluated with 5 fold cross validation using both training and leaderboard data. As with the criteria used in the competition, Pearson correlation and Spearman correlation are used here as criteria. First, Ridge regression and bagged linear regression on actual AUC values are shown as the baseline. AUC rankings are used instead of AUC values in training models #3 and #4. Further, post prediction covariance alignment is shown as model #5 – 8. *λ* = 1 is used for CORAL regulation term factor for all models. As a comparison, Bagged LR with Ridge feature selection but not postprocessing is given as #9.

**Table 1. Summary of model performances.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Training values | Features kept after 1st feature selection | the 2nd feature selection | Post processing | Regression model | Model parameters | Spearman correlation | Pearson correlation |
| #1 | values | 30k | - | - | Ridge | \alpha=1*α*=1 | 0.322344 | 0.318868 |
| #2 | values | 30k | - | - | Bagged LR | Max feature portion=0.005 | 0.323710 | 0.312633 |
| #3 | ranks | 30k | - | - | Ridge | \alpha=1*α*=1 | 0.327609 | 0.314593 |
| #4 | ranks | 30k | - | - | Bagged LR | Max feature portion =0.005 | 0.328202 | 0.314058 |
| #5 | ranks | 30k | - | CORAL | Ridge | \alpha=1*α*=1 | 0.336019 | 0.323188 |
| #6 | ranks | 30k | Ridge | CORAL | Ridge | \alpha=1*α*=1 | 0.330447 | 0.320249 |
| #7 | ranks | 30k | - | CORAL | Bagged LR | Max feature portion =0.005 | 0.333266 | 0.323085 |
| #8 | ranks | 30k | Ridge | CORAL | Bagged LR | Max feature portion =0.005 | 0.341710 | 0.326697 |
| #9 | ranks | 30k | Ridge | - | Bagged LR | Max feature portion =0.005 | 0.333703 | 0.317601 |

###

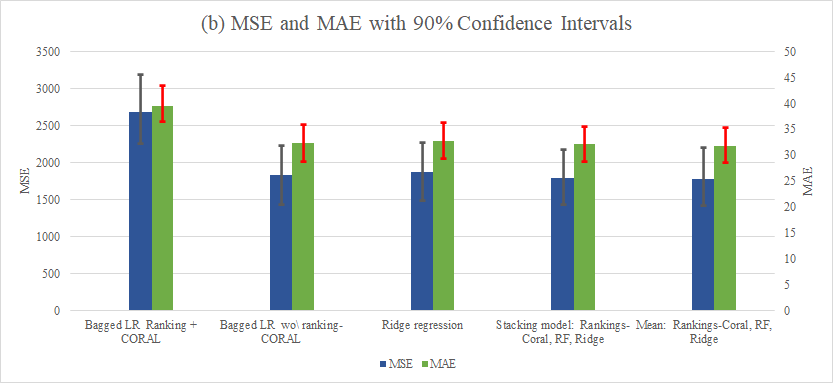
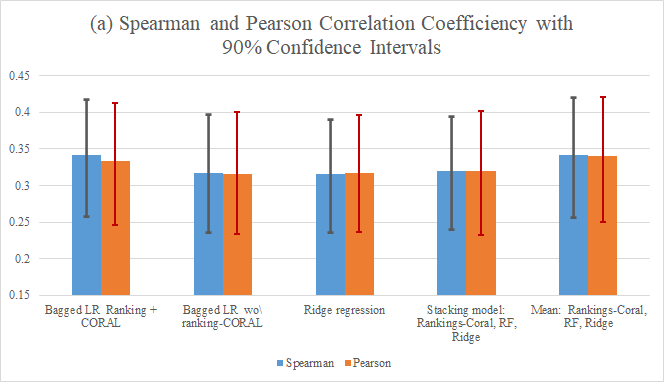
#### 2.3 Ensemble Model and Robustness Analysis

Our approach focuses on the correlation between prediction and ground truth, omitting the errors of the predicted AUC. To improve the performance in terms of errors, and a value matching step is taken to transform the predicted rankings to AUC values (described in Methods). To further reduce the errors, we ensemble the bagged linear regression model with multiple non-linear models. Random forest, KNN regressor, and Gaussian Regressor, which are able to capture the actual values of y and apply to multi-task regression, are considered. The models are stacked with a top linear-regression or simply take the mean of model outputs as the prediction. The Spearman correlation coefficient, Pearson correlation coefficient, mean square error (MSE), and mean absolute error (MAE) of single models and ensemble models are shown in Table 2. By simply taking the mean of prediction, the errors for predictions are further reduced.

Table 2. Performance of single and ensembled models. "LR" denotes for linear regression.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Model description | Stacking method | Spearman | Pearson | MSE | MAE |
| #1 | Rankings + Bagged LR + CORAL + Back mapping | - | 0.344792 | 0.302907 | 1888.021 | 32.58503 |
| #2 | Bagged LR | - | 0.323421 | 0.31494 | 1824.378 | 32.36096 |
| #3 | Ridge Regression | - | 0.322344 | 0.318868 | 1869.757 | 32.73517 |
| #4 | Random Forest | - | 0.335788 | **0.341463** | 1860.911 | 32.69214 |
| #5 | Gaussian Process Regressor | - | 0.304717 | 0.301775 | 1964.07 | 33.44859 |
| #6 | KNN Regressor | - | 0.245159 | 0.276327 | 1968.771 | 33.88596 |
| #7 | #1 + #2 + #3 | Mean | 0.331298 | 0.320397 | 1809.298 | 32.12484 |
| #8 | #1 + #2 + #4 | Stacking | 0.334471 | 0.317776 | 1794.032 | 32.08207 |
| #9 | #1 + #3 + #4 | Mean | **0.349332** | 0.338449 | 1781.876 | **31.88853** |
| #10 | #1 + #3 + #4 | Stacking | 0.336582 | 0.32884 | **1774.413** | 31.98663 |
| #11 | #1 + #3 + #4 + #5 | Stacking | 0.29684 | 0.292823 | 1827.081 | 32.48612 |
| #12 | #1 + #3 + #4 + #6 | Stacking | 0.331903 | 0.323139 | 1788.084 | 32.06569 |

The prediction confidence interval is calculated by non-parametric bootstrapping the test sets in 5-fold cross validation (See Methods). We compared these models: 1) Ranking + Bagged linear regression + CORAL (without back mapping); 2) Bagged linear regression (same to #2 in Table 2); 3) Ridge regression (same to #3 in Table 2); 4) Model #1 ensembled with ridge regression and RF, stacked with a linear regression as the top layer (same to #10 in Table 2); 5) Mean of model #1, ridge regression and RF (same to #9 in Table 2).



**Figure 3. Average Bootstrap correlation/error and 90% confidence intervals for 6 models.**

### 3 Discussion and Conclusion

In this paper we propose a new approach for multi-inhibitor responses prediction task based on the intuition of the distribution and correlations of AUC values. This approach include a preprocessing of data that mapping the AUC values to the rankings, which transform the data to a ranking space. A post-prediction covariance alignment is used to reduce the higher order errors between predictions and ground truth. The result on CTD^2 BeatAML data shows that our approach outperforms common linear and non-linear models, and state-of-art models such as KBMTL, KRL. Ablation tests show that the combination of ranking and CORAL contribute to the prediction performance.

To predict the actual AUC values instead of rankings, the predictions are mapped back to value space non-parametrically. Ensemble this model with multi-task models that captures the actual values can reduce the error of our model. Though ideally stacking single models with a top-layer linear regression possess better performance since the square error is minimized, our result shows that taking the mean of single models, i.e. all 3 models share the same weight, provides a better performance in general.

### 4. Methods

#### 4.1 SC1 preprocessing

Both training and leaderboard sets are preprocessed in the way described below. In SC1, RNA sequence, DNA sequence, and clinical data (numerical/categorical) are provided as input for modeling. However, we only used RNA sequence i.e. gene expression as predictors. First, the expression values are normalized by dividing the Euclidean norm of each specimen, followed by a feature selection process. Next, the expression values are standardized to achieve zero mean and unity variance for each gene (feature) before modeling. Instead of using the actual values directly, we used the normalized AUC rankings of the specimens as the target after subtracting the mean normalized ranking for each inhibitor. NaN values in AUC are replaced by 0 before model fitting. In the given RNA-seq data, expression values for 63,677 genes are available, while the amount is too high for modeling. First, the dataset is filtered to keep the 30,000 genes with largest average Spearman correlations to the 122 drugs. Then, a ridge regression model is fitted and the features with the highest weights are kept.

#### 4.2 Post-prediction correlation alignment and back mapping

**Correlation alignment**

To reduce the covariance discrepancy between training and predicted AUC values, we used the method of covariance alignment (CORAL) described by Sun et al [3]. For efficiency, stability, and to avoid matrix singularity, we simply add a regularization term to the covariance matrix. First, predictions are whitened i.e., divided by the square root of its covariance, and then re-colored i.e., multiplied by the square root of training data covariance. This can be described as:

[*fuction*]

P*{match}=P\_0 ⋅ C\_p^{(-1/2)} ⋅ C*{tr}^{(1/2)}*P****m****a****t****c\*\*h*=*P*0⋅*C\*\*p*(−1/2)⋅*C****t****r*(1/2)

where C\_p=Cov(P\_0 )+\lambda I*C\*\*p*​=*C****o****v*(*P*0​)+*λ\*\*I* is the regularized prediction covariance, C\_{tr}*C****t****r*​ is the regularized training AUC covariance, P\_0*P*0​ is the pristine prediction.

**Back mapping**

As described before, in our approach, the training responses is the rankings of cell lines normalized between 0 and 1. To predict the AUC values, another back mapping process is introduced to transform predicted rankings to real values. Due to the non-linearity of the actual AUC distribution, the responses distribution is hard to capture, therefore we mapped the ranking to values using a non-parametric way. The real AUC are preserved for the post-prediction processing step, and the predicted values are found by looking up the quantile of the actual values.

#### 4.3 Confidence interval

The confidence intervals are calculated with the method described below. In a 5-fold cross validation, each splitted test set is sampled with replacement for 400 times, generating 400 test sets. 2000 performance values (correlation coefficient, error) are obtained. For the distribution is skewed from normal distribution, the CIs are determined non-parametrically, i.e. find the upper 5% and bottom 5% quantile for the 90% CI. Figure 4 shows the distribution of the Spearman correlation coefficient for our approach on the CTD^2 BeatAML Challenge dataset.

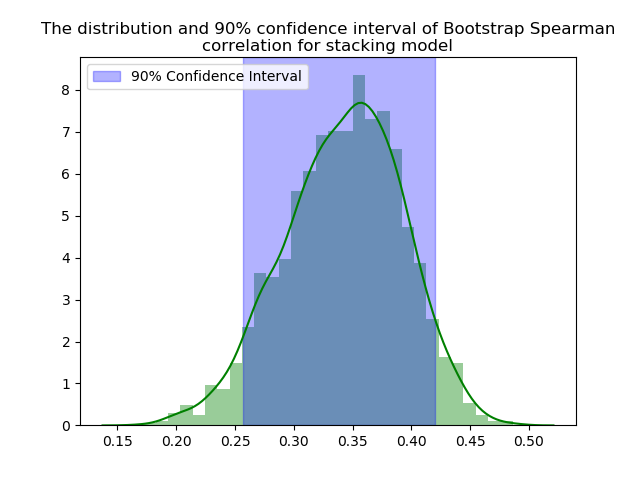


Figure 4.

## References

1. Barretina, Jordi, et al. "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity." Nature 483.7391 (2012): 603-607.
2. Sun B., Saenko K. (2016) Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In: Hua G., Jégou H. (eds) Computer Vision – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science, vol 9915. Springer, Cham
3. Sun, Baochen, Jiashi Feng, and Kate Saenko. "Return of frustratingly easy domain adaptation." Thirtieth AAAI Conference on Artificial Intelligence. 2016.
4. A. Hyvarinen and E. Oja, Independent Component Analysis: Algorithms and Applications, Neural Networks, 13(4-5), 2000, pp. 411-430
5. Pölsterl, S., Navab, N., and Katouzian, A., “Fast Training of Support Vector Machines for Survival Analysis”, Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Porto, Portugal, Lecture Notes in Computer Science, vol. 9285, pp. 243-259 (2015)

[<https://pubmed.ncbi.nlm.nih.gov/29133589/?dopt=Abstract>]: [<https://www.nature.com/articles/nature11003>]: