A Transductive Approach to Survival Ranking for Cancer Risk Stratification

Ethar Alzaid, Muhammad Dawood, Fayyaz Minhas

Tissue Image Analytics (TIA) Centre, Department of Computer Science, University of Warwick, Coventry, UK

{ether.alzaid, muhammad.dawood, fayyaz.minhas}@warwick.ac.uk

ABSTRACT

How can we stratify patients into subgroups based on their expected survival in a purely data-driven manner? Identifying cancer patients at higher risk is crucial in planning personalized treatment to improve patient survival outcomes. The main challenge with existing approaches is the underlying complexity of handling censoring in the survival data and manually setting a precise threshold to stratify patients into risk groups. In this paper, a Transductive Survival Ranking (TSR) model for patient risk stratification is proposed. The model handles samples in pairs to make use of instances with censored survival information. It incorporates unlabeled test samples in the training process to maximize the margin between their predicted survival scores resulting in automatic patient stratification into subgroups without the need for any additional post-processing or manual threshold selection. The model was evaluated on several datasets with varying sets of covariates, and all stratification were significant ($p << 0.05$) with high concordance indices of up to 0.78 in Disease Specific Survival and 0.75 in Overall Survival.

1. INTRODUCTION

Cancer is a disease where normal cells transform into tumor cells due to changes in the patient's genetic factors [1]. The underlying genetic makeup of cancer can determine patient trajectories which can be quite different even for the same cancer type across patients [2]. Handling such differences in genetic makeup across patients can be achieved by precision or personalized medicine which is the current practice of planning a patient's treatment journey based on a number of factors such as tumor variability in genes and tumor environment [3]. Determining a patient's risk group can play a major role in selecting an intervention or treatment as well as identifying the underlying determinants of such differences.

Numerous methods exist to aid personalized treatment planning where risk stratification is the main step prior to making any treatment decisions. Many models rely on the classic Cox Proportional Hazard (CoxPH) model [4]. CoxPH assesses the effect of a set of covariates on patient survival. Cox-net and DeepSurv [5,6] adapted CoxPH and trained a neural network model to minimize the partial log-likelihood loss function. Support Vector Machines (SVMs) have also been adapted in [7] to handle censored data by training the SVM for both ranking and regression purposes with squared hinge loss function and optimized with the truncated Newton optimization algorithm. Similarly, Gradient Boosting Machines (GBMs) have also been repurposed as survival prediction models by nonparametric concordance index learning [8]. A variation of Random Forest (RF) assesses the relation between covariates changing over time (longitudinal data) and baseline data for dynamic survival prediction [9]. Yang et al. evaluated the performance of different survival prediction models such as GBM, SVM, RF and Logistic Regression for prognosis predictability of renal clear cell cancer and found that GBM's predictions produces optimal cutpoint [10].

All the survival prediction approaches discussed above generate a real-valued score for a test patient and do not inherently stratify a patient into high risk or low risk groups. Such stratification is

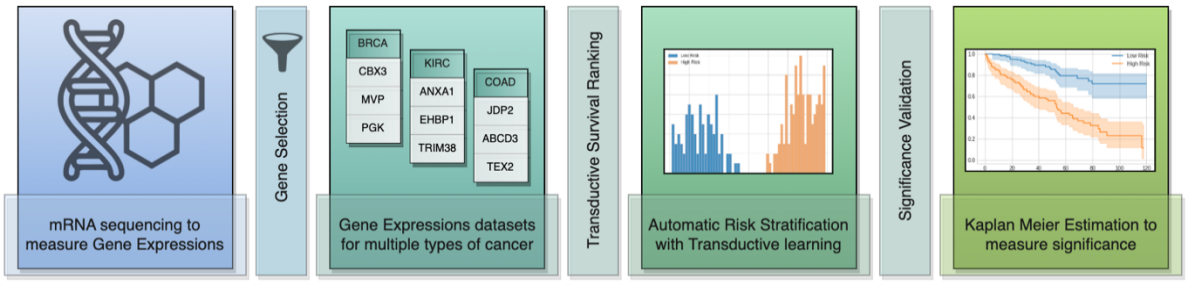

Figure 1: Concept diagram of the proposed work. The input to the model is a set of mRNA gene expressions which are selected based on their prognostic value. Transductive learning is applied to produce a risk score used for automatic risk stratification. Significance is measured and KM curves are plotted.

achieved by manually selecting an optimal cutoff point so that patients below the selected threshold are stratified into one group and the remaining ones into the other group. Methods for estimating the optimal threshold varies in terms of complexity and can be as simple as using a fixed threshold such as the median [11] or based on the area under the receiver operating characteristic (ROC) curve to select the threshold that produces the highest sensitivity and specificity [12] or as complicated as evaluating multiple methods [13]. "cutpointr" is a package that evaluates several thresholding techniques and selects one based on the highest Youden's index. Another package with the same purpose is OptimalCutpoints [14] which is designed specifically for diagnostic tests and takes into account the prevalence of the disease and cost of various diagnostic decision before selecting a cutpoint.

In this work, we have developed a transductive approach for automatically selecting an optimal point as part of the fitting of the survival prediction model without any manual post-processing or threshold selection. Transductive learning was first introduced by Vapnik in [15] as a method for incorporating unlabeled data in the training process. It works by learning a large margin hyperplane using labeled data and forcing the decision boundary to pass through a low data density region for classification problems. This method has proven to improve generalization in the case of a small set of labeled samples compared to larger sets of available unlabeled samples [16]. In this work, we have incorporated transductive learning into a survival prediction model to allow automatic patient stratification. Figure 1 demonstrates the main steps of the proposed model. The major contributions of the proposed model are as follows:

1. We propose an approach for automatic subgroup discovery based on transductive learning which automatically selects an optimal cutoff point for a given dataset.
2. We show that the proposed approach can work for different datasets using gene expression signatures of cancer patients.

2. METHODS AND MATERIALS

2.1 Datasets

In this work we have utilized gene expression data from the Cancer Genome Atlas (TCGA) [17]. It provides clinicopathologic, molecular profiles, and survival outcomes for 33 cancer types. In this study, five types of cancer from TCGA were selected to train and evaluate the proposed model. We restricted the survival prediction analysis to the set of genes known to be associated in a favorable or unfavorable patient prognosis as indicated by The Human Protein Atlas [18]. Table 1 presents details about the selected datasets. The m-RNA gene expressions were log-transformed and z-score normalized across all samples of the same dataset. For survival duration, 10 years censoring was applied as it is more likely that a sample with high survival duration is lost to follow up [19].

Cancer Type	Samples	Covariates	Events (OS)	Events (DSS)
Breast Carcinoma (BRCA)	1082	554	14%	8%
Colorectal Adenocarcinoma (COAD)	592	594	20%	12%
Lung Adenocarcinoma (LUAD)	566	644	33%	20%
Liver Hepatocelluar Carcinoma (LHIC)	372	2856	35%	21%
Kidney Renal Clear Cell Carcinoma (KIRC)	512	5926	33%	21%

Table 1 : Datasets selected from TCGA to evaluate the method, with the number of samples in each set, the number of selected genes (covariates), and the percentage of events taken place in Overall survival and Disease Specific Survival endpoints.

2.2 Transductive Survival Ranking (TSR)

For survival prediction, we consider a training set $R = \{(\boldsymbol{x_i}, t_i, \delta_i), i = 1 \dots N\}$ consisting of N patients each with a vector $\boldsymbol{x_i} \in \boldsymbol{R^d}$ of d-dimensional covariates, survival time t_i and an event indicator variable δ_i which is set to 1 to indicate if the event for that individual has taken place at time t_i or not ($\delta_i = 0$). We want to develop a predictor that produces survival prediction scores $f(\boldsymbol{x}; \boldsymbol{w})$ which can be used to rank test patients based on their expected survival time as well as automatically stratify them into low or high risk groups. Here, $\boldsymbol{w}$ represent the learnable parameters or weights of the model. The ranking of patients can be implemented using a ranking loss function during model training [20]. To achieve automatic stratification of test patients, we would like to enforce a constraint that the prediction scores $f(\boldsymbol{x}; \boldsymbol{w})$ from the model for one subgroup of test patients to be greater than zero and less than zero for the other. However, the subgroups need to be selected in an automatic manner such that $f(\boldsymbol{x}; \boldsymbol{w}) = 0$ acts as the stratification boundary. For this purpose, we use transductive learning which implies incorporating samples from the test set $R' = \{\boldsymbol{x_k}, k = 1 \dots M\}$ in the model training process. It is important to note that these test samples have information about the covariates only without any knowledge of their true survival times or event indicator variable status. The aim here is to force the curve $f(\boldsymbol{x}; \boldsymbol{w}) = 0$ to pass through a region of low data density while minimizing ranking errors over the training dataset. It can be achieved by formulating an optimization problem for penalizing predictions that contradict the ranking dictated by the actual survival times in the training dataset in addition to penalizing scores close to zero for samples in the test data. We use a ranking based formulation in which we first select pairs of samples in the training set $P(R) = \{(i, j) | t_i > t_j, \delta_j = 1, i, j = 1 \dots N\}$ whose survival times can be compared with each other and then use a ranking loss function to minimize the expected number of mis-ranking errors, i.e., when the prediction from the model for two samples does not align with the expected order of their known survival times (for further details, the interested reader is referred to our previous work [20]). For risk stratification based on transductive learning, we add an additional loss function $l\big(f(\boldsymbol{x_k}; \boldsymbol{w})\big) = exp(-3\, f(\boldsymbol{x_k}; \boldsymbol{w})^2)$ which penalizes prediction scores close to zero to force the stratification boundary to pass through an area of low data density [16]. We use a simple linear model with a bipolar sigmoid (hyperbolic tangent) activation $f(\boldsymbol{x}; \boldsymbol{w}) = tanh(\boldsymbol{w}^T \boldsymbol{x})$ which constrains the model output to the range $f(\boldsymbol{x}; \boldsymbol{w}) \in [-1, +1]$. The model is trained, and its weights are obtained by solving the following optimization formula:

$$Q(\boldsymbol{w}; R, R') = \underbrace{\lambda_w \|\boldsymbol{w}\|_2}_{Regularization} + \underbrace{\frac{1}{|P(R)|} \sum_{(i,j) \in P(R)} max\left(0, 1 - \left(f(\boldsymbol{x_i}; \boldsymbol{w}) - f(\boldsymbol{x_j}; \boldsymbol{w})\right)\right)}_{Ranking\ Loss} + \underbrace{\lambda_u \sum_{k \in R'} exp(-3\, f(\boldsymbol{x_k}; \boldsymbol{w})^2)}_{Transductive\ loss}$$

Here, λ_w is a hyperparameter that controls the strength of the L_2 regularization and the average ranking loss for all comparable pairs in $P(R)$ and λ_u controls the loss over the prediction scores of the samples in the test set. The proposed model has three major components: Survival ranking loss term, Regularization term and a Transductive loss term. Average survival ranking loss for all comparable pairs $P(R)$ is calculated using a hinge loss and averaged across all such pairs in the training data. L_2 (Ridge) regularization is performed to ensure that small changes in the input covariates do not have large effects on the prediction score so that the resulting predictor can generalize to test samples. This is a crucial step as typically the number of genes in the datasets is considerably large and minor changes can occur. Transductive learning was applied using symmetric sigmoid loss [16]. This term is calculated by taking the exponent of the square of the predicted scores of all test examples. The expected effect is that the model would produce little to no prediction scores equal to zero which is equivalent to the survival function passing through a region of low data density in the test data. Patient stratification into subgroups with the value of zero as a threshold is applied automatically with this approach without additional post-processing. The model was implemented in Python using PyTorch [21] library to construct the neural network.

2.3 Performance Evaluation

To ensure fair evaluation for the method, out of sample bootstrap analysis was conducted for each dataset where sampling with replacement is done in each bootstrap iteration to select the training set and all examples not included in the training set are used for as the test set. A maximum of 1000 iterations was used. For each bootstrap iteration, we calculate the concordance index (c-index) of the model as well as the p-value of the log-rank test over the prediction scores examples stratified into the two subgroups by the model. Concordance index (c-index) is a measure of how well a model is in producing survival prediction scores [22] by comparing a pair of samples' actual survival against the predicted score, the pair is concordant if the sample with lower survival score has a lower prediction. A c-index of 1 is perfect concordance, while 0.5 corresponds to a random model. The scores of the test samples in each bootstrap run are automatically split around zero as a constant threshold into low versus high risk groups. To establish statistical significance of the difference between the two groups identified by the model in terms of their survival, a p-value is determined using the log-rank test [23]. A combined p-value is calculated by taking the median of the p-values across all bootstrap runs and multiplying it by 2 [24].

3. RESULTS AND DISCUSSION

3.1 Concordance and Automatic Subgroup Discovery

Table 2 shows the average c-index, and the combined p-value for each of the five datasets for Overall Survival (OS) and Disease Specific Survival (DSS) endpoints. BRCA and KIRC shown highest mean c-index of 0.75 and 0.72 respectively for OS and 0.78 and 0.76 for DSS. Note that these two types have the smallest and largest set of covariates which proves the effectiveness of the regularization in maintaining the performance. The reported combined p-value shows significant difference (p <<0.05) in risk groups for all cancer types indicating that the proposed model is able to automatically discover subgroups in a completely data driven manner.

The distribution of test prediction scores for a single run of the proposed model is shown in figure 2 (a), and it can be clearly noted that the TSR is pushing the scores away from zero creating a separation margin to facilitate stratification. This effect can be seen on all datasets regardless of the number of covariates in each one. The survival curves for each of the subgroups identified by the model were generated using a Kaplan Meier estimator [25]. The estimation supports the stratification as it can be seen in figure 2 (b) that the survival of the low risk group is significantly higher. In contrast, figure 2 (c) shows that there is no clear point of separation when no transduction was used in the training.

Cancer Type	OS		DSS	
	Mean c-index	Combined p-value	Mean c-index	Combined p-value
Breast Carcinoma (BRCA)	0.75 (0.03)	6×10^{-8}	0.78 (0.03)	1×10^{-5}
Kidney Renal Clear Cell Carcinoma (KIRC)	0.72 (0.02)	6×10^{-8}	0.76 (0.03)	4×10^{-9}
Colorectal Adenocarcinoma (COAD)	0.69 (0.02)	2×10^{-4}	0.70 (0.04)	1×10^{-3}
Liver Hepatocelluar Carcinoma (LHIC)	0.67 (0.03)	3×10^{-3}	0.70 (0.04)	1×10^{-2}
Lung Adenocarcinoma (LUAD)	0.65 (0.02)	1×10^{-3}	0.65 (0.03)	2×10^{-2}

Table 2: Average Concordance index for all bootstrap runs, Standard Deviation, and the combined p-value for Overall Survival and Disease Specific Survival endpoints.

3.4 Robustness Analysis

For all bootstrap runs, the box plot in figure 3 shows the distribution of c-indices and the distribution of p-values are shown in figure 4. The low values of the standard deviation of the c-indices and the low p-values across bootstrap runs for datasets show robustness and consistency of the model in automatically discovering subgroup of patients with different survival patterns. It is notable that the model produces higher c-indices for DSS as these events are more specific, i.e., death due to the target disease which is directly linked to the changes in prognostic genes, whereas OS events can be death due to any cause which may not be directly associated with the covariates.

3.5 Analysis of Differences in Gene Expression Across Subgroups

It is expected for significant prognostic genes to be expressed differently in the two risk groups identified by the model. To visualize such differences, prognostic genes were clustered using Hierarchical clustering which takes every single gene as its own cluster and begin merging clusters based on the average distance between pairs of genes within different clusters [26]. As a result, genes with the same expression level across all patients will be grouped in one cluster. Due to the space limitation, only BRCA is shown in the cluster map in figure 5. In the map we can see the clear vertical separation as an effect of significant differences in expression values between risk groups. GSTK1 is among the genes that were expressed differently in the two groups and it was previously proved to be predictive of survival in one of BRCA subtypes [27]. In addition, SBDS gene has a negative correlation with survival for breast cancer patients and was highly expressed in higher risk patients [28]. Similarly, SLC16A2 was proved to be linked to poor prognosis (which puts the patient at higher risk) for BRCA and multiple cancer types when highly expressed [29].

3.6 Ablation and Comparison with Existing Methods

To ensure that the proposed model is non-inferior in terms of its predictive quality based on the concordance index, the same set of data and experiments were carried out on different models. The performance of the model was compared to CoxPH and the Survival Ranking (SR) model which has the same baseline as TSR formulation but with no transduction. The comparison was performed in terms of mean c-index and combined p-value and is shown in Table 3. The stratification threshold for both CoxPH and SR was calculated as the median of prediction scores and the p-value was calculated accordingly. The performance of TSR is comparable to the other two models which means that the automation of the stratification by transductive learning works effectively and did not interfere with predictions' quality. On the other hand, both CoxPH and SR required additional post-processing to achieve stratification where the significance of the p-value relies on the selected thresholding method.

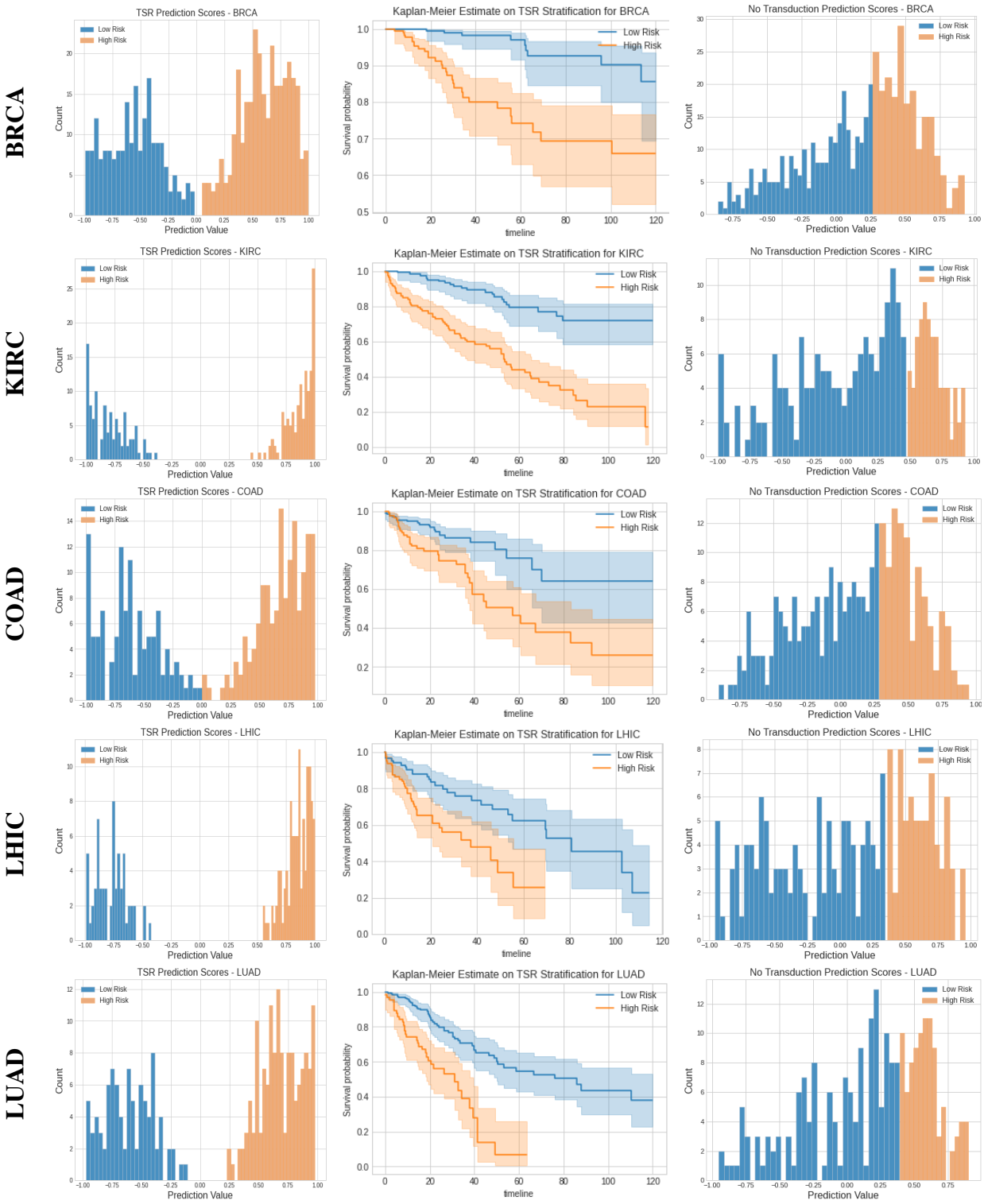

Figure 2: (a) TSR prediction scores across the datasets (b) Kaplan Meier estimation curves for stratified patients for all datasets (c) prediction scores with no transduction across datasets.

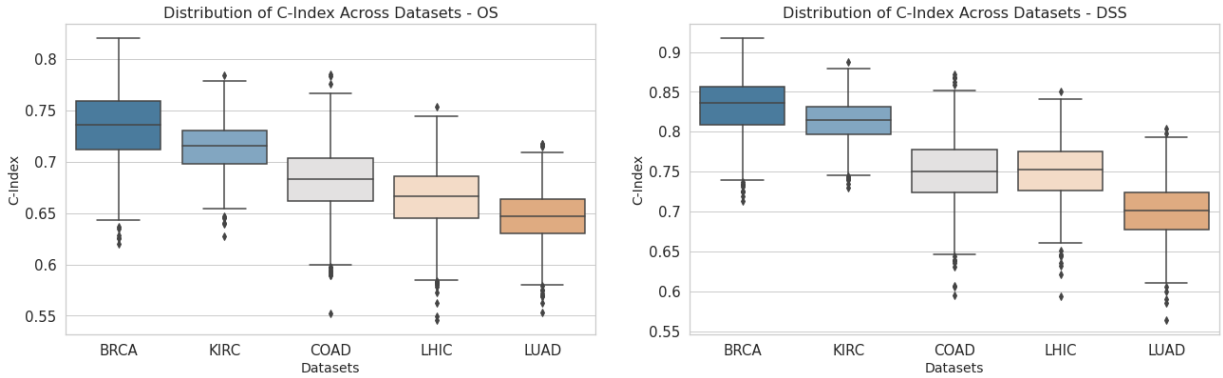

Figure 3: Box plot of the distribution of concordance indices across bootstrap runs for different datasets (left is for Overall Survival (OS), and right is Disease Specific Survival (DSS))

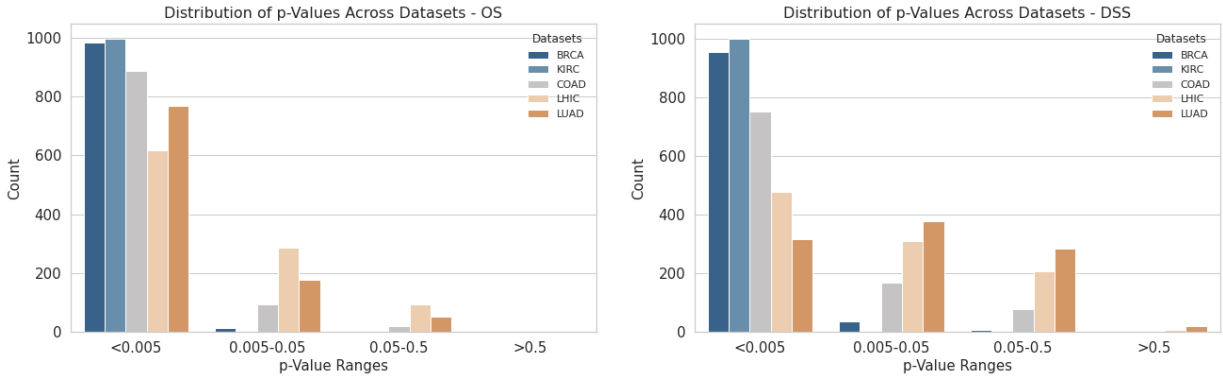

Figure 4: Histogram of p-values across all bootstrap runs for different datasets (left is for Overall Survival (OS), and right is Disease Specific Survival (DSS))

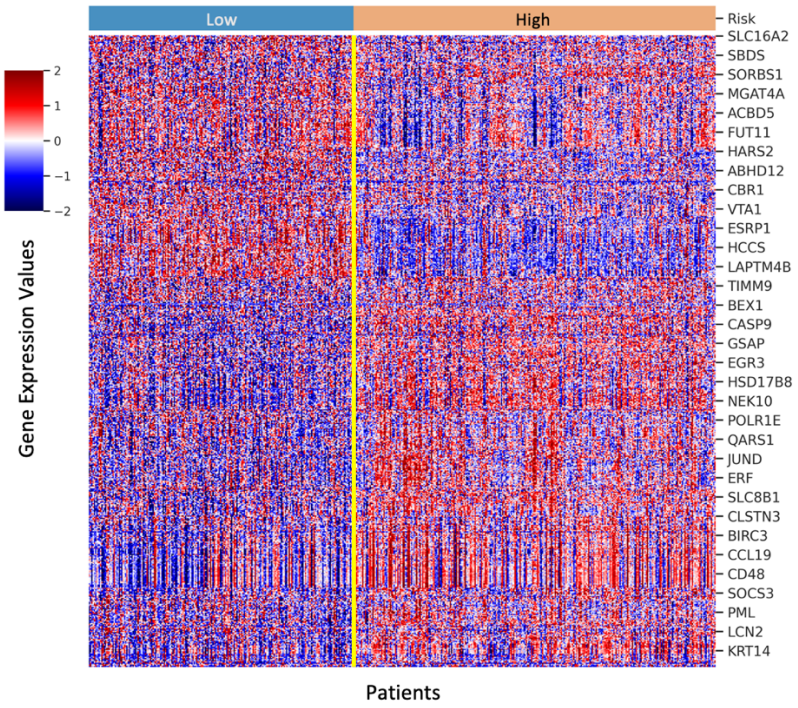

Figure 5: Clustered heatmap of gene expression values for all prognostic genes across BRCA patients

Model	BRCA		KIRC		COAD		LHIC		LUAD	
	Mean c-index	Combined p-value	Mean c-index	Combined p-value	Mean c-index	Combined p-value	Mean c-index	Combined p-value	Mean c-index	Combined p-value
CoxPH	0.75 (0.03)	8×10^{-7}	0.69 (0.02)	1×10^{-5}	0.69 (0.03)	0.0025	0.63 (0.03)	7×10^{-2}	0.63 (0.03)	2×10^{-2}
SR	0.75 (0.03)	2×10^{-6}	0.71 (0.02)	6×10^{-5}	0.69 (0.03)	0.0036	0.65 (0.03)	9×10^{-2}	0.64 (0.03)	4×10^{-2}
TSR	0.75 (0.02)	6×10^{-8}	0.72 (0.02)	6×10^{-8}	0.69 (0.02)	0.0002	0.67 (0.03)	3×10^{-3}	0.65 (0.02)	1×10^{-3}

Table 3: Comparison between the performance of CoxPH, Survival Ranking model, and the proposed TSR in terms of average c-index, standard deviation, and the combined p-value for Overall Survival endpoint.

3.7 Code Availability

The code is available in the following github repository: https://github.com/EtharZaid/TSR.git. The interested reader may use the code replicate the experiments and check other experiment results that are not included in this paper due to page limitations.

4. CONCLUSIONS

In this work, a Transductive learning approach to survival ranking was proposed to facilitate automatic stratification of cancer patients into risk groups without post-processing. It was evaluated on five cancer datasets with various sets of prognostic gene expressions. The model proved its ability to significantly stratify patients based on their predicted survival score using test set predictions and the proposed transductive loss. The impact of the added loss made the model produce little to no prediction scores that are equal to zero and creating the optimal cutpoint automatically. Personalized treatment decision support models can benefit from this reliable stratification method as identifying patients at risk is crucial in setting a direction for the treatment plan. In the future, the loss term may be incorporated with other survival prediction models to explore its stratification ability with other loss terms.

5. ACKNOWLEDGMENTS

EAZ is supported by the Saudi Cultural Bureau in London, UK. MD acknowledges PhD funding support from GlaxoSmithKline (GSK) outside of this research. FM is supported by a grant from EPSRC EP/W02909X/1. We also acknowledge support from the Tissue Image Analysis (TIA) centre at the University of Warwick.

REFERENCES

[1] WHO, "Cancer." https://www.who.int/news-room/fact-sheets/detail/cancer (accessed Sep. 19, 2023).

[2] Cancer Research UK, "What is personalised medicine?" https://www.cancerresearchuk.org/about-cancer/treatment/personalised-medicine (accessed Sep. 19, 2023).

[3] A. Hoeben, E. A. J. Joosten, and M. H. J. van den Beuken-van Everdingen, "Personalized Medicine: Recent Progress in Cancer Therapy," *Cancers (Basel)*, vol. 13, no. 2, p. 242, Jan. 2021, doi: 10.3390/cancers13020242.

[4] D. R. Cox, "Regression Models and Life-Tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972, doi: 10.1111/j.2517-6161.1972.tb00899.x.

[5] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network," *BMC Medical Research Methodology*, vol. 18, no. 1, p. 24, Feb. 2018, doi: 10.1186/s12874-018-0482-1.

[6] T. Ching, X. Zhu, and L. X. Garmire, "Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data," *PLOS Computational Biology*, vol. 14, no. 4, p. e1006076, Apr. 2018, doi: 10.1371/journal.pcbi.1006076.

[7] S. Pölsterl, N. Navab, and A. Katouzian, "Fast Training of Support Vector Machines for Survival Analysis," in *Machine Learning and Knowledge Discovery in Databases*, A. Appice, P. P. Rodrigues, V. Santos Costa, J. Gama, A. Jorge, and C. Soares, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 243–259. doi: 10.1007/978-3-319-23525-7_15.

[8] Y. Chen, Z. Jia, D. Mercola, and X. Xie, "A Gradient Boosting Algorithm for Survival Analysis via Direct Optimization of Concordance Index," *Computational and Mathematical Methods in Medicine*, vol. 2013, p. e873595, Nov. 2013, doi: 10.1155/2013/873595.

[9] K. L. Pickett, K. Suresh, K. R. Campbell, S. Davis, and E. Juarez-Colunga, "Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker," *BMC Medical Research Methodology*, vol. 21, no. 1, p. 216, Oct. 2021, doi: 10.1186/s12874-021-01375-x.

[10] W. Yang *et al*., "Machine learning to improve prognosis prediction of metastatic clear-cell renal cell carcinoma treated with cytoreductive nephrectomy and systemic therapy," *Bosn J of Basic Med Sci*, Nov. 2022, doi: 10.17305/bjbms.2022.8047.

[11] D. Iacobucci, S. S. Posavac, F. R. Kardes, M. J. Schneider, and D. L. Popovich, "Toward a more nuanced understanding of the statistical properties of a median split," *Journal of Consumer Psychology*, vol. 25, no. 4, pp. 652–665, 2015, doi: 10.1016/j.jcps.2014.12.002.

[12] C.-H. Yang, S.-H. Moi, F. Ou-Yang, L.-Y. Chuang, M.-F. Hou, and Y.-D. Lin, "Identifying Risk Stratification Associated With a Cancer for Overall Survival by Deep Learning-Based CoxPH," *IEEE Access*, vol. 7, pp. 67708–67717, 2019, doi: 10.1109/ACCESS.2019.2916586.

[13] C. Thiele and G. Hirschfeld, "cutpointr: Improved Estimation and Validation of Optimal Cutpoints in R," *Journal of Statistical Software*, vol. 98, pp. 1–27, Jun. 2021, doi: 10.18637/jss.v098.i11.

[14] M. López-Ratón, M. X. Rodríguez-Álvarez, C. Cadarso-Suárez, and F. Gude-Sampedro, "OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests," *Journal of Statistical Software*, vol. 61, pp. 1–36, Nov. 2014, doi: 10.18637/jss.v061.i08.

[15] V. N. Vapnik, "The Vicinal Risk Minimization Principle and the SVMs," in *The Nature of Statistical Learning Theory*, V. N. Vapnik, Ed., in Statistics for Engineering and Information Science. New York, NY: Springer, 2000, pp. 267–290. doi: 10.1007/978-1-4757-3264-1_9.

[16] R. Collobert, F. Sinz, J. Weston, and L. Bottou, "Large Scale Transductive SVMs," *Journal of Machine Learning Research*, vol. 7, no. 62, pp. 1687–1712, 2006.

[17] J. Liu *et al*., "An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics," *Cell*, vol. 173, no. 2, pp. 400-416.e11, Apr. 2018, doi: 10.1016/j.cell.2018.02.052.

[18] M. Uhlen *et al*., "A genome-wide transcriptomic analysis of protein-coding genes in human blood cells," *Science*, vol. 366, no. 6472, p. eaax9198, Dec. 2019, doi: 10.1126/science.aax9198.

[19] C. R. Lesko, J. K. Edwards, S. R. Cole, R. D. Moore, and B. Lau, "When to Censor?," *American Journal of Epidemiology*, vol. 187, no. 3, pp. 623–632, Mar. 2018, doi: 10.1093/aje/kwx281.

[20] F. Minhas, M. S. Toss, N. ul Wahab, E. Rakha, and N. M. Rajpoot, "L1-regularized neural ranking for risk stratification and its application to prediction of time to distant metastasis in luminal node negative chemotherapy na\"ive breast cancer patients." arXiv, Aug. 23, 2021. doi: 10.48550/arXiv.2108.10365.

[21] A. Paszke *et al*., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[22] F. E. Harrell Jr, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati, "Evaluating the Yield of Medical Tests," *JAMA*, vol. 247, no. 18, pp. 2543–2546, May 1982, doi: 10.1001/jama.1982.03320430047030.

[23] J. M. Bland and D. G. Altman, "The logrank test," *BMJ*, vol. 328, no. 7447, p. 1073, May 2004, doi: 10.1136/bmj.328.7447.1073.

[24] C. J. DiCiccio, T. J. DiCiccio, and J. P. Romano, "Exact tests via multiple data splitting," *Statistics & Probability Letters*, vol. 166, p. 108865, Nov. 2020, doi: 10.1016/j.spl.2020.108865.

[25] E. L. Kaplan and P. Meier, "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 457–481, Jun. 1958, doi: 10.1080/01621459.1958.10501452.

[26] F. Nielsen, "Hierarchical Clustering," 2016, pp. 195–211. doi: 10.1007/978-3-319-21903-5_8.

[27] S. Luthra, U. Chandran, B. Diergaarde, M. Becich, A. V. Lee, and C. A. Neumann, "Expression of reactive species related genes is associated with patient survival in luminal B breast cancer," *Free Radical Biology and Medicine*, vol. 120, pp. 170–180, May 2018, doi: 10.1016/j.freeradbiomed.2018.03.011.

[28] J. Lee *et al*., "Shwachman-Bodian-Diamond syndrome protein desensitizes breast cancer cells to apoptosis in stiff matrices by repressing the caspase 8-mediated pathway," *Animal Cells and Systems*, vol. 23, no. 6, pp. 414–421, Nov. 2019, doi: 10.1080/19768354.2019.1666030.

[29] J. Li *et al*., "A pan-cancer analysis revealed the role of the SLC16 family in cancer," *Channels (Austin)*, vol. 15, no. 1, pp. 528–540, doi: 10.1080/19336950.2021.1965422.