Deep Integrative Analysis for Survival Prediction

Chenglong Huang, Albert Zhang and Guanghua Xiao
Colleyville Heritage High School, Colleyville, TX, 76034, USA
Highland Park High School, Dallas, TX, 75205, USA
Department of Clinical Science, The University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA

Survival prediction is very important in medical treatment. However, recent leading research is challenged by two factors: 1) the datasets usually come with multi-modality; and 2) sample sizes are relatively small. To solve the above challenges, we developed a deep survival learning model to predict patients' survival outcomes by integrating multi-view data. The proposed network contains two sub-networks, one view-specific and one common sub-network. We designated one CNN-based and one FCN-based sub-network to efficiently handle pathological images and molecular profiles, respectively. Our model first explicitly maximizes the correlation among the views and then transfers feature hierarchies from view commonality and specifically fine-tunes on the survival prediction task. We evaluate our method on real lung and brain tumor data sets to demonstrate the effectiveness of the proposed model using data with multiple modalities across different tumor types.

Keywords: Survival Prediction, Integrative Analysis, Deep Learning

1. Introduction

Survival analysis aims at modeling the time that will elapse from the present to the occurrence of a certain event of interest (e.g. biological death). The prognostic models generated by survival analysis can be used to explore interactions between prognostic factors in certain diseases, and also predict how a new patient will behave in the context of known data. In survival analysis, the Cox proportional hazards model[1] and parametric survival distributions[2] have long been used as important fundamental techniques. Clinicians and researchers usually apply these models to test for significant risk factors affecting survival. In order to handle the high-dimensional data, dimension reduction and penalized regression have been proposed in the Cox model.[3–7] However, the Cox model and its extensions are still built based on the assumption that a patient's risk is a linear combination of covariates. The parametric censored regression approaches[2,8] are highly dependent on the choice of the distribution. In fact, there are too many complex interactions that can affect the event (death) in various ways, and thus a more comprehensive survival model is needed to better fit data in real-world applications. To formulate the survival problem without any additional hypothesis, Li et al. modeled the prediction problem as standard multi-task learning using an additional indicator matrix.[9] However, the number of tasks corresponds to the maximum follow-up time of all the instances. In fact, recent cancer datasets are collecting patient electronic health records (EHR) with a very long follow-up time. Another limitation for existing survival models is that they mainly focus on one view and cannot efficiently handle multi-modalities data. Since more comprehensive multi-source data are available to health-care research, a powerful survival

analysis that can learn from those multi-view data is required.

One good way to learn highly complex survival functions is by using recent neural network techniques.[10–12] Katzman *et al.* proposed a deep fully-connected network (DeepSurv) to represent the nonlinear risk function.[10] They demonstrated that DeepSurv outperformed the standard linear Cox proportional hazard model. However, DeepSurv is still too simple to handle real cancer data. First, real datasets contain complex imaging and genomic data from different views. Although using multiple pieces of information can provide complementary characterizations of tumors at different levels, the view discrepancy and heterogeneity will bring challenges for survival prediction. Second, compared to computer vision applications, survival prediction problems only provide a very small training set due to the cost of multiple comprehensive data collections. To integrate multiple modalities and eliminate view variations, a good solution is to learn a joint embedding space in which different modalities can be compared directly. Such an embedding space will benefit the survival analysis since recent studies have suggested that common representation from different modalities provides important information for prognosis.[13,14] For example, molecular profiling data and pathological images actually share representations to describe the same event in tumor growth, which is very important for diagnosis. Stromal tissue has been verified to have a surprising role in predicting the overall survival of breast cancer patients.[13] The proportion of stromal cells correlated with the overexpression of genes, including FBLN1, FBLN2, COL6A2 and COL6A3, that encode extracellular matrix proteins .[14]

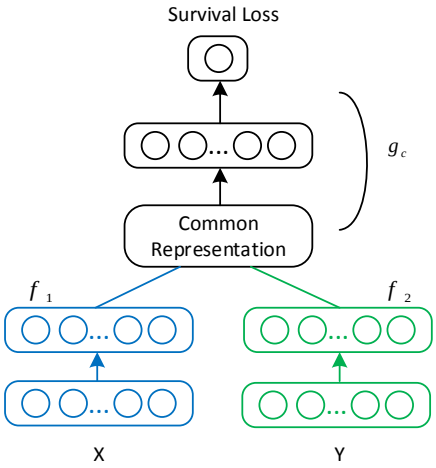

Fig. 1. An overview of the proposed model.

In order to take the advantage of both histopathological information and molecular profiles from imaging-genetics data, we developed an integrative pipeline as shown in Fig.1. It consists of two sub-networks, view-specific sub-network f_1, f_2 and common sub-network g_c. The view-specific sub-networks reduce the discrepancy between the view and the commonality of all views. The common sub-network is shared by all views and can extract a view-invariant representation for survival prediction. One advantage of the proposed architecture is that it

has good generality, since the network can handle any kind of data sources with well-designed view-specific sub-networks. Motivated by recent deep learning methods,[15,16] we proposed Convolutional Neural Networks (CNNs) and Fully Connected Networks (FCNs) to learn deep representations from pathological images and molecular profiling data, respectively.

To handle multi-modalities data, we integrate outputs of two networks into a common space where the two modalities have maximal correlation. The primary motivation of using such a model is to eliminate the view variations and find the maximum correlated representation from the views of pathological images and molecular data. Although the commonality of two views reduces the view-discrepancy, it still cannot guarantee that the common space is directly associated with survival outcomes. To address this issue, the proposed model transfers feature hierarchies from such common spaces and specifically fine-tunes on the survival regression task. This will not only help to avoid over-fitting, but also accelerates the model training. Moreover, it has the ability to discover important markers that cannot be found by previous deep correlational learning methods, which will benefit the survival prediction. The contribution of this paper can be summarized as: 1) We proposed a deep learning approach which can model very complex view distributions and learn good estimators for predicting patients' survival outcomes with insufficient training samples. 2) Our model used CNNs to represent much more abstract features from pathological images for survival prediction. Traditional survival models usually adopted hand-crafted imaging features. 3) Extensive experiments on TCGA-LUSC and GBM demonstrate that the proposed model can achieve better predictions across different tumor types.

2. Related Work

In this section, we give a brief survey on recent survival analysis methods with basic notations and then briefly review recent deep multi-modal embeddings.

2.1. *Survival Analysis*

Survival analysis aims to analyze the expected duration of time until events happen. It covers many topics as the event can be defined very broadly such as failure in mechanical systems and death in biological organisms. Survival analysis tries to find the answer of questions like: how does the proportion of a population survive past a certain time (e.g. 5 years)? what rate will they die or fail? Given a set of N patients, $\{x_i\}, i = 1 \dots N$, each patient has the label (t_i, δ_i) indicating the survival status where t_i is the observed time, δ_i is the indicator: 1 is for a uncensored instance (death event happens during the study), and 0 is for a censored instance (death not observed). If and only if $t_i = \min(O_i, C_i)$ can be observed during the study, the dataset is said to be right-censored.[17]

In Survival Analysis, the survival function $S(t|\mathbf{x}) = Pr(O \geq t|\mathbf{x})$ is used to identify the probability of being still alive at time t where $\mathbf{x} = (x_1, ...x_p)^\intercal$ is the covariates of dimension p. The hazard function is defined as

$$h(t|\mathbf{x}) = \lim_{\triangle t \to 0} \frac{Pr(t \leq O \leq t + \triangle t | O \geq t; \mathbf{x})}{\triangle t}, \tag{1}$$

which assesses the instantaneous rate of failure at time t. In the modeling methods, Cox proportional hazard model[1] is among the most popular one. The hazard function for the Cox proportional hazard model has the form

$$h(t|\mathbf{x}_i) = h_0(t)\exp(\beta^\intercal \mathbf{x}) \tag{2}$$

where $\beta = (\beta_1, ..., \beta_p)^\intercal$ is a vector of regression parameters, and $h_0(t)$ is the baseline hazard. We can define $f(\mathbf{x}) = \beta^\intercal \mathbf{x}$ as a risk function. This gives the hazard rate at time t for the patient i with covariate vector $\mathbf{x}_i$.

A major challenge is that the number of features p is much larger than the number of patients n. To handle high-dimensional data, many feature selection methods have been adapted to the Cox regression setting for censored survival data.[3–7,18] Another type of hazard model is estimated by logistic regression such that the probability of surviving beyond t is $Pr(O \geq t|x) = (1 + \exp[x^\top \beta(t) + th])^{-1}$ with a threshold th.[19,20] Instead of defining the hazard function, one recent work transforms the original survival analysis problem into a multi-task learning problem by decomposing the regression component into related classification tasks; the new objective function can be solved by popular ADMM based optimization.[9] It is a good way to learn highly complex survival functions by using the advanced neural networks techniques.[10,12] We can get the risk score through neural networks and now denote the risk for the patient i as $\mathbf{o}_i$. Deepsurv[10] is the earlier attempt to learn a nonlinear risk function by replacing the linear part $\beta^T x$ in $f(x)$ with a nonlinear deep fully connected network.

One very simple way for data fusion is to create a concatenated feature vector comprising of all features selected individually from each modality.[21] However, a powerful feature selection is required to search for those important biomarkers from the original features, and each modality is processed individually without considering their inter-connections. The inherent challenge in combining data streams for survival analysis is that individual data sources are very heterogeneous due to the heterogeneity of tumors. However, recent studies have shown that different views actually share common representations to describe tumor morphology, which is very important for diagnosis.[14] A key challenge for survival analysis is how to eliminate view-discrepancies and learn such common representations.

2.2. *Deep multi-modal embeddings*

Recent deep multi-modal embeddings[22–26] provide a very good solution to the above challenge. They have been successfully applied in computer vision applications such as image-text matching[23,26] and image reconstruction utilizing multiple auto-encoders.[24,25]

In finding a correlated meta-space for data fusion, recent DNN-based multi-view methods provide very complex representation learning using deep neural networks (DNNs) that maximizes signals which are common to data from multiple modalities. They can learn much more comprehensive representation and more easily process large amounts of training data. However, these methods belong to unsupervised feature learning, which is incapable of survival analysis since it cannot guarantee that the integrated feature space is highly associated with patients' survival outcome. In addition, recent cancer datasets cannot provide multi-modalities data with sufficient patient samples, while deep multi-modal embeddings need large amounts of data.

3. Methodology

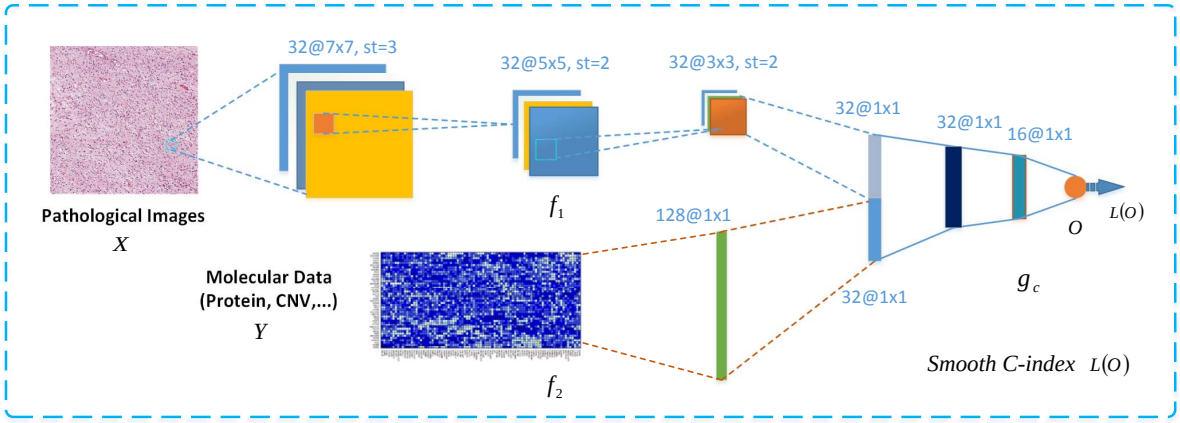

Fig. 2. The architecture of our framework. 'st' is short for 'stride'.

Figure 2 shows the pipeline of the proposed framework. f_1, f_2 is the view-specific sub-network and g_c is the common sub-network. The proposed model uses Convolutional Neural Networks (CNNs) as one image-view sub-network f_1 and Fully Connected Neural Networks (FCNs) as another view-specific sub-network f_2 to learn deep representations from pathological images and molecular profiling data, respectively. More details about the sub-network f_1 can be seen in 1. It consists of 3 convolutional layers, 1 max-pooling layer and 1 fully-connected layer. In each convolutional layer, we employ ReLU as the nonlinear activation function.

The sub-network f_2 has two fully connected layers equipped with ReLU activation function, with 128 and 32 neurons, respectively.

Table 1. The architecture of CNNs

Layer	Filter size, stride, number
Conv (ReLU)	7×7, 3, 32
Conv (ReLU)	5×5, 2, 32
Conv (ReLU) Max-pool	3×3, 2, 32 2×2
FC	32

3.1. *Deep Correlational Learning*

Denote $\mathbf{x}_i, \mathbf{y}_i$ from two views as i-th sample, its representation passing through the corresponding view sub-network is denoted as $f_1(\mathbf{x}_i; \mathbf{w_x})$ and $f_2(\mathbf{y}_i; \mathbf{w_y})$ respectively. $\mathbf{w_x}, \mathbf{w_y}$ represent all

parameters of the two sub-networks. The outputs of the two branches will be connected to a correlation layer to form the common representation.

In the correlational layer, deep correlational learning tries to find pairs of projections that maximize the correlation of two outputs from each network $f_1(\mathbf{x}_i; \mathbf{w_x}), f_2(\mathbf{y}_i; \mathbf{w_y})$. If $\mathbf{w_x}, \mathbf{w_y}$ represent all parameters of two networks, then the commonality is enforced by maximizing the correlation between two views as

$$L = corr(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{i=1}^{m}(f_1(\mathbf{x}_i) - \overline{f_1(\mathbf{X})})(f_2(\mathbf{y}_i) - \overline{f_2(\mathbf{Y})})}{\sqrt{\sum_{i=1}^{m}(f_1(\mathbf{x}_i) - \overline{f_1(\mathbf{X})})^2 \sum_{i=1}^{m}(f_2(\mathbf{y}_i) - \overline{f_2(\mathbf{Y})})^2}}, \tag{3}$$

where networks' parameters $\mathbf{w_x}, \mathbf{w_y}$ are omitted in the loss function (3). We can maximize the correlation loss function to generate the shared representation indicating the most correlated features from two modalities. Although different views of patients' data are very heterogeneous, there still share some common information for survival prediction. Correlational learning provides a very good way to find such common representation using the correlation function (3). However, it belongs to unsupervised learning and thus this procedure has a risk of losing the discriminant markers for predicting patients' survival outcomes.

3.2. *Survival prediction with smooth C-index loss function*

Denote $\mathbf{O} = [o_1, ..., o_N]^\top$ as the outputs of common sub-network $\mathbf{g}_c$, i.e., $o_i = \mathbf{g}_c(\mathbf{z}_i)$. The final model will be fine-tuned on the survival prediction task using the knowledge from the deep correlational learning. This will give the proposed model the ability to discover important markers that are ignored by the correlational model, and learn the best representation for survival prediction. Different from the use of negative log partial likelihood as survival loss in recent deep survival learning,[11] we propose to minimize the smoothed empirical risk function[27] which is from the concordance index (C-index) estimator and differentiable with respect to the predictor o_i.

During the past few decades, the C-index, a general discrimination measure for the evaluation of prediction models, has gained enormous popularity in biomedical research. The concordance index (C-index) quantifies the ranking quality of rankings and is calculated as

$$c = P(o_i > o_j | T_i < T_j) \tag{4}$$

where T_i, T_j and o_i, o_j are the event times and the predicted risk values. The C-index measures whether large values of o are associated with short survival times T and vice versa. Uno et al. proposed a modified C-index estimation as follows:[28]

$$C_{uno} = \frac{\sum_{i,k} \delta_i (G_m(T_i))^{-2} I(T_i < T_k) I(o_i > o_k)}{\sum_{i,k} \delta_i (G_m(T_i))^{-2} I(T_i < T_k)}. \tag{5}$$

where $G_m(t)$ denotes the Kaplan-Meier estimator of the unconditional survival function of Censored time (C_{cens}) estimated from the learning data. However, the Uno estimator is unfeasible because it is not differentiable to o_i. To solve this problem, the indicator function $I(o_i > o_k)$ is approximated by the sigmoid function:

$$L(\mathbf{o}) = \sum_{i,k} w_{i,k} \frac{1}{1 + \exp(\frac{o_k - o_i}{\sigma})}, \tag{6}$$

where o_i is the output of the i-th patient. We implement the smoothed C-index function (6) as the survival loss function in our method. The weights w_{ik} are defined as

$$w_{i,k} = \frac{\delta_i (G_m(T_i))^{-2} I(T_i < T_k)}{\sum_{i,k} \delta_i (G_m(T_i))^{-2} I(T_i < T_k)}. \tag{7}$$

where $I(T_i < T_k)$ is an indication function that indicates whether T_i is larger than T_k or not. It is easy to check the smoothed empirical risk is differentiable with respect to the predictor o_i. The derivative is given by

$$\frac{\partial L}{\partial o_i} = -\sum_{k} w_{i,k} \frac{\exp(\frac{o_k - o_i}{\sigma})}{\sigma(1 + \exp(\frac{o_k - o_i}{\sigma}))} \tag{8}$$

Compared with recent deep survival models,[10,29] which can only handle one specific view of data, our model can achieve more complex architecture for the integration of multi-modalities data, which can be used for practical applications on more challenging datasets.

4. Experiments

4.1. *Dataset Description*

TCGA (The Cancer Genome Atlas) data cohort[30] is a very large dataset which contains both high resolution whole slide pathological images and molecular profiling data. In TCGA-cohort, we focused on glioblastoma multiforme (GBM) and lung squamous cell carcinoma (LUSC). For each cancer type, we adopted a core sample set from UT MD Anderson Cancer Center[31] in which each sample has information for the overall survival time, pathological images, and molecular data related to gene expression. For model evaluation, 80% of patients were randomly selected for training and the remaining 20% were used for testing.

- **TCGA-LUSC**: Lung squamous cell carcinoma (LUSC) is one major type in Non-Small-Cell Lung Carcinoma (NSCLC). 106 patients with pathological images and protein expression (reverse-phase protein array, 174 proteins) are collected in our experiments.
- **TCGA-GBM**: Glioma is a type of brain cancer, and it is the most common malignant brain tumor. 126 patients are selected from the core set with images and CNV data (Copy number variation, 106 dimension).

4.2. *Comparison approaches*

We compare our model with four state-of-the-art survival approaches and three baseline deep survival models. The four survival methods include LASSO-Cox,[18] Parametric censored regression models with components with Weibull, Logistic distribution,[2] and Boosting concordance index (BoostCI).[27] Those above methods need hand-crafted features as inputs. To calculate imaging hand-crafted features, we used CellProfiler[32] to analyze pathological images in comparison survival models. CellProfiler is widely used as a state-of-the-art medical image feature

extracting and quantitative analysis tool. Motivated by the pipeline,[33] a total of 1,795 quantitative features were calculated from each image tile.

The three baseline deep survival models are as follows:

- **CNN-Surv**: Deep convolutional survival model;[29] we use the same architecture as the sub-network f_1 .
- **FCN-Surv**: FCN sub-network f_2 followed by negative log partial likelihood loss.[10]
- **DeepCorr+DeepSurv**: The shared representation learned by deep correlational learning is directly fed to another DeepSurv model.

To make fair comparisons, the architectures of different deep survival models are kept the same as the corresponding parts in the proposed method.

4.3. *Results and Discussion*

To evaluate the performances in survival prediction, we take the concordance index (CI) as our evaluation metric.

Table 2. Performance comparison of the proposed methods and other existing related methods

Data	Model	LUSC	GBM
Images	LASSO-Cox[18]	0.3411	0.5775
	BoostCI[27]	0.5088	0.5565
	Weibull[2]	0.4261	0.4787
	Logistic[2]	0.4217	0.4921
	CNN-Surv[29]	0.5797	0.5154
Protein/CNV	LASSO-Cox[18]	0.6231	0.4920
	BoostCI[27]	0.5714	0.4676
	Weibull[2]	0.4851	0.5659
	Logistic[2]	0.3915	0.4218
	FCN-Surv[10]	0.5462	0.5221
Integration	DeepCorr+DeepSurv	0.5622	0.5900
	Proposed	**0.6638**	**0.6045**

Results in Table 2 presents the C-index values by various survival methods on TCGA-LUSC and TCGA-GBM. It can be seen that the integration of both modalities in the proposed model achieves the best performance, for both lung and brain cancer. That is because the proposed method can remove view discrepancy as well as learn the survival-related common representations from both modalities. The difference in the DeepCorr+DeepSurv from ours is that those two models are trained separately. Performance shows that the common representation by maximizing the correlation in an unsupervised manner still has the risk of discarding markers that are highly associated with survival outcomes. In fact, the proposed model used a similar smoothed C-index as the survival loss function compared with BoostCI,[27] but the proposed method outperforms BoostCI in evaluation. This demonstrates that the proposed method can efficiently learn deep representation from two modalities and achieve better predictions.

From the results, we can see that it is not easy to find a general model that can successfully estimate patients' survival outcomes across different tumor types using only one specific view, either images or molecule data. The reason might be the heterogeneous of different tumor types and the original data in each view might contain variations or noise and thus affect the estimation of survival models. Because the proposed model can effectively integrate two views, it can achieve good prediction performance across different tumor types.

5. Conclusion

In this paper, we proposed a deep survival model to efficiently integrate multi-modalities from lung and brain tumor patients. Eliminating the view discrepancy between imaging data and molecular profiling data, deep correlational learning provides a good solution to maximize the correlation of two views and find the common embedding space. However, deep correlational learning belongs to an unsupervised learning which cannot ensure the common representation from correlational layer is suitable for survival prediction. To overcome this issue, the proposed model fine-tunes the whole network using smooth C-index loss after transferring knowledge from the embedding space. Experiments have demonstrated the proposed method can discover important markers that might be ignored by correlational learning. Our model can find non-linear relationships between factors and prognosis; it achieved quite promising performance with improvements. In the future, we will extend the proposed framework to directly process original whole slide images (WSIs).

References

1. D. R. Cox, *Journal of the Royal Statistical Society. Series B (Methodological)* , 187 (1972).
2. J. D. Kalbfleisch and R. L. Prentice, *The statistical analysis of failure time data* (John Wiley & Sons, 2011).
3. E. Bair and R. Tibshirani, *PLoS Biol* **2**, p. E108 (2004).
4. E. Bair, T. Hastie, D. Paul and R. Tibshirani, *Journal of the American Statistical Association* **101** (2006).
5. H. C. van Houwelingen, T. Bruinsma, A. A. Hart, L. J. van't Veer and L. F. Wessels, *Statistics in medicine* **25**, 3201 (2006).
6. M. Y. Park and T. Hastie, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 659 (2007).
7. H. M. Bøvelstad, S. Nygård, H. L. Størvold, M. Aldrin, Ø. Borgan, A. Frigessi and O. C. Lingjærde, *Bioinformatics* **23**, 2080 (2007).
8. Y. Li, K. S. Xu and C. K. Reddy, Regularized parametric regression for high-dimensional survival analysis, in *In Proceedings of SIAM International Conference on Data Mining. SIAM*, 2016.
9. Y. Li, J. Wang, J. Ye and C. K. Reddy, A multi-task learning formulation for survival analysis, in *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
10. J. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang and Y. Kluger, *arXiv preprint arXiv:1606.00931* (2016).
11. J. Yao, X. Zhu, F. Zhu and J. Huang, Deep correlational learning for survival prediction from multi-modality data, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017.
12. X. Zhu, J. Yao, F. Zhu and J. Huang, Wsisa: Making survival prediction from whole slide

histopathological images, in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
13. A. H. Beck, A. R. Sangoi, S. Leung, R. J. Marinelli, T. O. Nielsen, M. J. van de Vijver, R. B. West, M. van de Rijn and D. Koller, *Science translational medicine* **3**, 108ra113 (2011).
14. Y. Yuan, H. Failmezger, O. M. Rueda, H. R. Ali, S. Gräf, S.-F. Chin, R. F. Schwarz, C. Curtis, M. J. Dunning, H. Bardwell *et al.*, *Science translational medicine* **4**, 157ra143 (2012).
15. A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, in *Advances in neural information processing systems*, 2012.
16. K. Chatfield, K. Simonyan, A. Vedaldi and A. Zisserman, Return of the devil in the details: Delving deep into convolutional nets, in *British Machine Vision Conference*, 2014.
17. C. K. Reddy and Y. Li, A review of clinical prediction models, in *Healthcare Data Analytics*, (Chapman and Hall/CRC, 2015) pp. 343–378.
18. R. Tibshirani *et al.*, *Statistics in medicine* **16**, 385 (1997).
19. H.-c. Lin, V. Baracos, R. Greiner and J. Y. Chun-nam, Learning patient-specific cancer survival distributions as a sequence of dependent regressors, in *Advances in Neural Information Processing Systems*, 2011.
20. X. Song and C.-Y. Wang, *Statistics in medicine* **32** (2013).
21. X. Zhu, J. Yao, X. Luo, G. Xiao, Y. Xie, A. Gazdar and J. Huang, Lung cancer survival prediction from pathological images and genetic data - an integration study, in *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016.
22. G. Andrew, R. Arora, J. A. Bilmes and K. Livescu, Deep canonical correlation analysis., in *ICML*, 2013.
23. F. Yan and K. Mikolajczyk, Deep correlation for matching images and text, in *CVPR*, June 2015.
24. W. Wang, R. Arora, K. Livescu and J. Bilmes, On deep multi-view representation learning, in *Proc. of the 32st Int. Conf. Machine Learning (ICML 2015)*, 2015.
25. S. Chandar, M. M. Khapra, H. Larochelle and B. Ravindran, *Neural computation* (2016).
26. L. Wang, Y. Li and S. Lazebnik, *arXiv preprint arXiv:1511.06078* (2015).
27. A. Mayr and M. Schmid, Boosting the concordance index for survival data–a unified framework to derive and evaluate biomarker combinations (1) (Public Library of Science, 2014) p. e84483.
28. H. Uno, T. Cai, M. J. Pencina, R. B. D'Agostino and L. Wei, *Statistics in medicine* **30**, 1105 (2011).
29. X. Zhu, J. Yao and J. Huang, Deep convolutional neural network for survival analysis with pathological images, in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016.
30. C. Kandoth, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski *et al.*, *Nature* **502**, 333 (2013).
31. Y. Yuan, E. M. Van Allen, L. Omberg, N. Wagle, A. Amin-Mansour, A. Sokolov, L. A. Byers, Y. Xu, K. R. Hess, L. Diao *et al.*, *Nature biotechnology* **32**, 644 (2014).
32. A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat *et al.*, *Genome biology* **7**, p. R100 (2006).
33. J. Yao, S. Wang, X. Zhu and J. Huang, Imaging biomarker discovery for lung cancer survival prediction, in *MICCAI 2016, Part II*, eds. S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal and W. Wells, LNCS, Vol. 9901 (Springer, Heidelberg, 2016).