Prognostic Imaging Biomarker Discovery in Survival Analysis for Idiopathic Pulmonary Fibrosis

An Zhao[1], Ahmed H. Shahin[1], Yukun Zhou[1], Eyjolfur Gudmundsson[1], Adam Szmul[1], Nesrin Mogulkoc[2], Frouke van Beek[3], Christopher J. Brereton[4], Hendrik W. van Es[5], Katarina Pontoppidan[4], Recep Savas[6], Omer Unat[2], Marcel Veltkamp[3,7], Mark G. Jones[4], Coline H.M. van Moorsel[3], David Barber[8], Joseph Jacob[1], and Daniel C. Alexander[1]

[1] Centre for Medical Image Computing, UCL, London, UK
an.zhao.19@ucl.ac.uk
[2] Department of Respiratory Medicine, Ege University Hospital, Izmir, Turkey
[3] Interstitial Lung Diseases Center of Excellence, Department of Pulmonology, St Antonius Hospital, Nieuwegein, Netherlands
[4] NIHR Southampton Biomedical Research Centre and Clinical and Experimental Sciences, University of Southampton, Southampton, UK
[5] Department of Radiology, St Antonius Hospital, Nieuwegein, Netherlands
[6] Department of Radiology, Ege University Hospital, Izmir, Turkey
[7] Division of Heart and Lungs, University Medical Center, Utrecht, The Netherlands
[8] Centre for Artificial Intelligence, UCL, London, UK

Abstract. Imaging biomarkers derived from medical images play an important role in diagnosis, prognosis, and therapy response assessment. Developing prognostic imaging biomarkers which can achieve reliable survival prediction is essential for prognostication across various diseases and imaging modalities. In this work, we propose a method for discovering patch-level imaging patterns which we then use to predict mortality risk and identify prognostic biomarkers. Specifically, a contrastive learning model is first trained on patches to learn patch representations, followed by a clustering method to group similar underlying imaging patterns. The entire medical image can be thus represented by a long sequence of patch representations and their cluster assignments. Then a memory-efficient clustering Vision Transformer is proposed to aggregate all the patches to predict mortality risk of patients and identify high-risk patterns. To demonstrate the effectiveness and generalizability of our model, we test the survival prediction performance of our method on two sets of patients with idiopathic pulmonary fibrosis (IPF), a chronic, progressive, and life-threatening interstitial pneumonia of unknown etiology. Moreover, by comparing the high-risk imaging patterns extracted by our model with existing imaging patterns utilised in clinical practice, we can identify a novel biomarker that may help clinicians improve risk stratification of IPF patients.

Keywords: Imaging biomarker discovery · Survival analysis · Contrastive learning · Clustering Vision Transformer · Idiopathic pulmonary fibrosis

1 Introduction

An imaging biomarker is defined as a characteristic derived from a medical image, that can be used as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention [12,2]. Clinicians often assess imaging biomarkers through visual assessment of medical images. Though some computer-based methods have been proposed for automated and quantitative measurement of imaging biomarkers [2], these methods often require expert labelling of potential biomarkers as training data. This is not only time-consuming, expensive and susceptible to inter-observer variability, but also restricts computer-based methods to evaluating limited numbers of predefined imaging biomarkers. Such strategies are insufficient to adequately mine the rich information contained within medical images [29].

There remains a long-term unmet need for quantitative and novel imaging biomarker development to inform the prediction of disease outcomes such as mortality. Many deep learning methods have been proposed recently for survival prediction using imaging data without the requirement of manual annotations [13,24]. One major limitation of these end-to-end deep learning survival models is that they often extract high-dimensional biomarkers which are difficult to interpret. Motivated by this, some researchers looking at histopathology images have tried to learn low-level prognostic imaging biomarkers that can underpin image-based survival models [1,30]. These methods often extract patch-level features and aggregate them to predict patient survival. Different strategies then associate the underlying patterns with high mortality risk. Accordingly, learning generalizable representations, aggregating sequences of patch information, and recognizing high-risk patterns become the main challenges for this task.

Recent success in self-supervised learning methods highlights the potential for learning discriminative and generalizable representations from unlabelled data. Contrastive learning, a dominant group of self-supervised learning methods, aims to group similar samples together while separating samples that differ [18]. Transformer neural networks have gained increasing interest from the medical imaging field because of their ability to capture global information [27]. The Vision Transformer (ViT) [8] accepts an image as a sequence of patch embeddings allowing for the fusion of patch information. However, medical images can often be split into thousands of patches, which results in a much longer sequence and a computationally expensive ViT when compared with natural images.

Motivated by challenges in prognostic biomarker discovery and survival analysis, we propose a framework that leverages plausible properties of contrastive learning and ViT, and validate its performance in a highly heterogeneous disease, idiopathic pulmonary fibrosis (IPF). IPF is a chronic lung disease of unknown cause, associated with a median survival of between 2.5 to 3.5 years [22]. The generally poor prognosis of IPF belies its highly heterogeneous disease progression between patients. A lack of reliable prognostic biomarkers hampers the ability to accurately predict IPF patient survival [29].

In our study, as shown in Fig. 1, we first learn patch representations via contrastive learning, on computed tomography (CT) imaging of IPF patients,

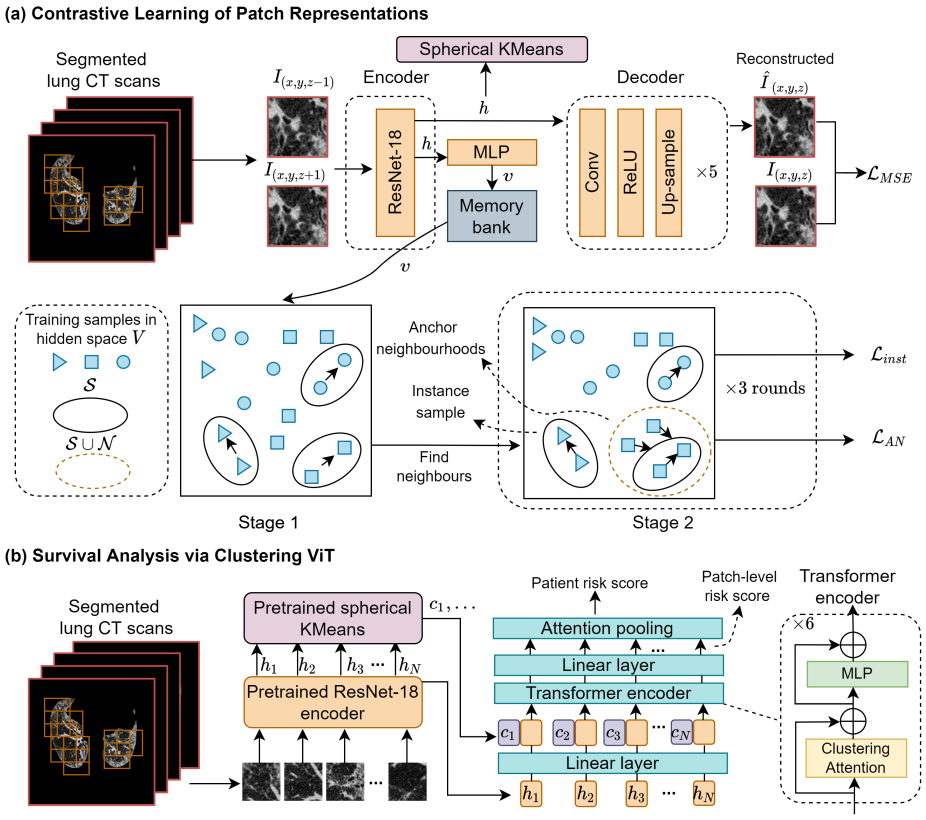

Fig. 1. An overview of the proposed model. (a) We first learn patch representations $\boldsymbol{h}$ by using a modified contrastive learning method [1]. All patch representations are clustered into K clusters by using spherical KMeans. (b) With trained models in (a), CT scans can be represented as a sequence of patch representations and their cluster assignments, which are then fed to a clustering ViT for survival prediction.

followed by a clustering method to extract underlying patterns in the patches. We then use an efficient clustering ViT to aggregate the extremely long sequence of patch representations for survival prediction and high-risk pattern recognition.

The contributions of this paper are as follows. First, we propose a framework tailored for prognostic imaging biomarker discovery and survival analysis based on large medical images. Experiments on two IPF datasets show that our model outperforms end-to-end ViT and deep 3D convolutional neural networks (CNN) in terms of mortality prediction performance and generalizability. Second, by comparing discovered biomarkers with existing visual biomarkers, we identify a novel biomarker that can improve the prediction performance of existing visual biomarkers which will aid patient risk stratification.

2 Method

2.1 Contrastive Learning of Patch Representations

We modify a contrastive learning method [1] to learn representations of lung tissue patches. For preparing the training data, we first segment the lung area by using a pre-trained U-Net [26,16]. Then we split the segmented lung area into patches. The network structure is a convolutional autoencoder as shown in Fig. 1. The contrastive learning process has two stages and is conducted in a divide-and-rule manner. The basic idea of contrastive learning is to make similar sample pairs (positive pairs) close to each other while dissimilar pairs (negative pairs) far apart. With the divide-and-rule principle, the method gradually expands the positive sample set by discovering class consistent neighbourhoods anchored to individual training samples within the original positive pair [17].

In the first stage, the model is optimized by minimizing a reconstruction loss and an instance-level contrastive loss jointly. Specifically, we denote patch as $\boldsymbol{I}_{(x,y,z)}$, where (x, y, z) is the location of the patch central voxel in original CT scans. The ResNet-18 [15] encoder maps input patches $\boldsymbol{I}_{(x,y,z-1)}, \boldsymbol{I}_{(x,y,z+1)}$ into the 512-dimensional latent representation $\boldsymbol{h}$, and the decoder reconstructs the patch $\boldsymbol{I}_{(x,y,z)}$ adjacent to them from $\boldsymbol{h}$. Mean squared error (MSE) loss $\mathcal{L}_{MSE}$ is used to measure the difference between patch $\boldsymbol{I}_{(x,y,z)}$ and the reconstructed patch. For contrastive learning, we consider patches 50% overlapped with patch i as its similar samples S_i and other patches in the training dataset as dissimilar samples. For calculating the similarity between samples, $\boldsymbol{h}$ is projected to a 128-dimensional variable $\boldsymbol{v}$ by a multi-layer perception (MLP) [5]. The instance-level contrastive loss is defined as $\mathcal{L}_{inst} = -\sum_{i \in B_{inst}} \log(\sum_{j \in S_i} p(j \mid i))$, where $p(j \mid i) = \frac{\exp(\boldsymbol{v}_j^\top \boldsymbol{v}_i / \tau)}{\sum_{k=1}^{N} \exp(\boldsymbol{v}_k^\top \boldsymbol{v}_i / \tau)}$, B_{inst} is the set of instance samples in a mini-batch.

In the second stage, the method discovers other similar patches (neighbour samples $\mathcal{N}_i$) for a given patch i based on relative entropy to expand the positive sample set. Lower entropy indicates a higher similarity between a patch i and its neighbourhoods, and these patches can be anchored together in the subsequent training process (anchored neighbourhoods, ANs). Higher entropy implies that the given patch is dissimilar with its neighbourhoods and so the pair should remain individuals (instance samples) rather than grouped together. AN-level contrastive loss is defined as $\mathcal{L}_{AN} = -\sum_{i \in B_{AN}} \log(\sum_{j \in S_i \cup \mathcal{N}_i} p(j \mid i))$, where B_{AN} is the set of ANs in a mini-batch. The total loss for two stages can be defined as $\mathcal{L} = \mathcal{L}_{inst} + \mathbb{1}(\text{stage} = 2) \cdot \mathcal{L}_{AN} + \mathcal{L}_{MSE}$.

The second stage can be performed over multiple rounds with an increasing number of ANs. This will progressively expand the local consistency to find the global class boundaries. A memory bank is used to keep track of the similarity matrix. $\boldsymbol{h}$ is used as patch representation after training the model. All patch representations are clustered into K clusters by using spherical KMeans [33]. These clusters are common patterns found in CT scans.

2.2 Survival Analysis via Clustering ViT

The objective of survival analysis is to estimate the expected duration before an event happens. The Cox proportional hazards model [6] is a widely used survival analysis method. Given the input data $\boldsymbol{x}$, the hazard function is modeled as $h(t|\boldsymbol{x}) = h_0(t)\exp(f(\boldsymbol{x}))$, where $h_0(t)$ is a base hazard function and $\exp(f(\boldsymbol{x}))$ is a relative risk function learned by models. Deep-learning based Cox methods [21,9] have been proposed to model more complicated nonlinear log-risk function $f(\boldsymbol{x})$. These methods optimize Cox log partial likelihood $L_{cox} = \prod_{i:E_i=1} \frac{\exp(\hat{f}(\boldsymbol{x}_i))}{\sum_{j\in\mathcal{R}(T_i)}\exp(\hat{f}(\boldsymbol{x_j}))}$, where T_i, E_i, $\boldsymbol{x_i}$, $\hat{f}(\boldsymbol{x}_i)$ are the event time (or censored), event indicator, baseline data, and estimated log-risk of patient i, and $\mathcal{R}(t)$ is the risk set of patients who are still alive at time t.

Inspired by [32], we propose a clustering ViT for survival analysis and high-risk pattern identification, that can handle long sequences (Fig. 1). For given CT scans, a sequence of patch representations and their cluster assignments are generated using the trained model in Sec. 2.1. We pad all input sequence to be a fixed length N. We first map patch representations to D dimensions with a linear layer, where D is the hidden size, and then pass the sequence into a Transformer encoder with L layers ($L = 6$). Each layer includes a clustering attention layer with 8 heads and a MLP. The original attention layer computes interaction between each pair of input patches, with a $O(N^2)$ complexity. The assumption of the clustering attention is that patches within the same cluster have similar attention maps. Queries within the same cluster can be represented by a prototype, which is the centroid of queries within this cluster. The clustering attention layer only calculates the attention maps between K prototypes and N keys, and then broadcasts it to queries within each cluster. This reduces complexity to $O(NK)$.

The sequential output of the L-layer ViT is defined as $\boldsymbol{x}_L \in \mathbb{R}^{B\times N\times D}$, where B is the batch size. After going through a linear layer g, $g(\boldsymbol{x}_L) \in \mathbb{R}^{B\times N\times 1}$ can be seen as sequences of patch-level risk scores. We then propose attention pooling, a variant of sequence pooling [14], to get the patient-level log-risk $\boldsymbol{r} = \mathrm{softmax}(g(\boldsymbol{x}_L)^\top)g(\boldsymbol{x}_L) \in \mathbb{R}^{B\times 1}$. This pooling method can assign different importance weights across patches with different risks. The network is trained based on average negative partial log-likelihood loss $\mathcal{L}_{neglog} = -\frac{1}{n_{E=1}}\sum_{i:E_i=1}(r_i - \log\sum_{j\in\mathcal{R}(T_i)}\exp(r_j))$, where r is log-risk estimated by the ViT, $n_{E=1}$ is the number of events observed. After training the ViT, we take the mean of risk scores of patches within a cluster as the cluster-level risk $R_k, k \in [1, K]$, which will help us to identify high-risk patterns.

2.3 Novel prognostic biomarker identification

High-risk imaging patterns discovered by our method may overlap the clinically-established patterns. With a CT dataset which includes annotated regions of lung tissue patterns predefined by clinicians (normal lung, ground-glass opacity, emphysema, and fibrosis), we propose an approach to identify novel prognostic

biomarkers by disentangling them from the established ones [9] (see Supplementary Fig. 1). The hypotheses are that: 1) The novel biomarker should not have a strong or moderate positive correlation (correlation coefficient > 0.3) with the extents of existing patterns [25]. 2) The novel biomarker should be significantly predictive of mortality (p-value < 0.05) independent of existing biomarkers when inputting both of them into the Cox model. 3) Centroids of representations of novel prognostic patterns should be relatively far from those of existing patterns.

3 Experiment

In this section, we first evaluate mortality risk prediction performance of the proposed model and compare it with CNN-based prediction models. A series of ablation studies are conducted to understand the contribution of each component. We also show representative patches of discovered high-risk patterns.

3.1 Datasets

For training the contrastive learning model, we use a dataset (Dataset 1) that contains 313 CT scans (186 death observed) of IPF patients from the Netherlands and Turkey. 1,547,467 patches are generated by using a 64×64 sliding window with a step size of 32 across the lung area. Dataset 1 is used for training and evaluating the ViT with 5-fold cross-validation. We randomly split the dataset into 5 folds. 1 fold is used as an internal test set, while the remaining 4 folds are randomly split into training and validation sets with a ratio of $4:1$. To evaluate the generalizability, we introduce an external test dataset (Dataset 2) from University Hospital Southampton, comprising 98 CT scans (48 death observed). For each split, we train the model on the training set in Dataset 1, choose the best model with the lowest loss on the validation set, and test the model on internal and external test sets. For novel prognostic biomarker identification, we use a subset of Dataset 1 and 2 with visual scores in step 1) and 2) of Sec. 2.3. In 253/313 (81%) CTs in Dataset 1, and all CTs in Dataset 2, fibrosis and emphysema extents have been visually scored by radiologists. We also use a publicly available interstitial lung disease dataset with annotated lung tissue patterns [7] for calculating centroids of existing patterns in step 3) of Sec. 2.3.

3.2 Implementation Details

For contrastive learning, we train the model with an Adam optimizer [20,23], a learning rate of 10^{-4}, and batch size of 128. We run the first stage for 1 round and the second stage for 3 rounds with $\tau = 0.05$, which takes about a week. Every round has 25 epochs. The number of clusters K is 64 for spherical KMeans. For clustering ViT, the sequence length N is 15,000 and the batch size is 6. We

[9] We use the extent of a high-risk pattern as a prognostic biomarker, obtained by calculating the percentage of this pattern within the whole lung.

Table 1. 5-fold cross-validation results on internal and external datasets compared with other survival prediction models. P-value shows significance of better performance of our proposed model ($^{\star}p < 0.05$,† $p < 0.01$,‡ $p < 0.001$).

Methods	# Pars	Internal Test Set		External Test Set	
		IPCW C-index	IBS	IPCW C-index	IBS
Ours	**21.96M**	**0.676±0.033**	**0.165±0.025**	**0.698±0.013**	**0.133±0.003**
K-M method	-	-	0.192±0.028	-	0.144±0.002†
3D ResNet-18	32.98M	0.639±0.032†	0.184±0.028	0.618±0.001‡	0.149±0.006*
3D ResNet-34	63.28M	0.657±0.048	0.180±0.032	0.631±0.032*	0.156±0.028

use Mixup [31] for data augmentation. The hidden size D is 256 and dropout rate is 0.1. We use a pre-trained model [3] for training our ViT. We adopt the Sharpness-Aware Minimization (SAM) algorithm [10] and use Adam as the base optimizer, with a learning rate of 2×10^{-5} and weight decay of 10^{-4}. The ViT is trained for 100 epochs (3 hours) [10]. We use an inverse-probability-of-censoring weighted version of C-index (IPCW C-index) [28] for assessing the discrimination of model, which quantifies the capability of discriminating patients with different survival times. For measuring the calibration (i.e. the capability of predicting true probabilities), we use the Integrated Brier Score (IBS) [11] which measures the difference between predicted probabilities and observed status by integrating Brier Score across the time span of test set. The paired t-test is used for testing statistical differences between the proposed method and other methods. More experiments of hyperparameters are provided in Supplementary Table 1.

3.3 Experiment Results

Comparison with Other Survival Prediction Models. We compare our method with end-to-end survival prediction models based on 3D ResNet-18 and 3D ResNet-34 [4]. 3D ResNets are trained with average negative partial log-likelihood loss [6]. Data splitting for each fold is the same as our model to ensure a fair comparison. Kaplan-Meier (K-M) method [19] is used as a baseline of the IBS. As shown in Table 1, our method achieves at least comparable or often significantly better performance in terms of discrimination and calibration with fewer parameters, especially in the external test dataset.

Ablation Study. To investigate the contribution of each component in the proposed method, we conduct ablation studies as shown in Table 2. The ablation study without contrastive learning in Sec. 2.1 is identical to train a regular ViT [8] end-to-end for survival prediction. We also remove attention pooling (using average pooling instead), Mixup [31] data augmentation and SAM algorithm [10],

[10] Our method is implemented by Pytorch 1.8. All models were trained on one NVIDIA RTX6000 GPU with 24GB memory. Code is available at https://github.com/anzhao920/PrognosticBiomarkerDiscovery

Table 2. 5-fold cross-validation results of ablation studies on internal and external datasets. P-values show the significance of better performance of the proposed model ($^\star p < 0.05, ^\dagger p < 0.01, ^\ddagger p < 0.001$).

Methods	Internal Test Set		External Test Set	
	IPCW C-index	IBS	IPCW C-index	IBS
Proposed model	**0.676±0.033**	**0.165±0.025**	0.698±0.013	0.133±0.003
w/o contrastive learning	0.633±0.042†	0.183±0.032*	0.561±0.017‡	0.163±0.011†
w/o attention pooling	0.659±0.025	0.177±0.034	**0.699±0.011**	**0.133±0.002**
w/o Mixup	0.660±0.038	0.177±0.034	0.686±0.023	0.139±0.006
w/o SAM	0.666±0.020	0.174±0.030	0.631±0.022†	0.154±0.010*

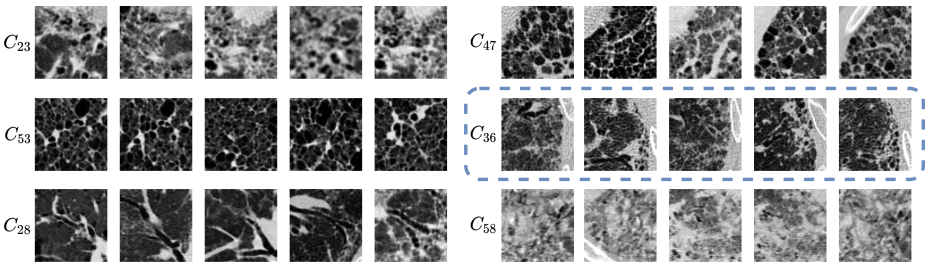

Fig. 2. Representative patches of top-6 high-risk clusters discovered by our model. Cluster 36 (C_{36}) is a novel prognostic pattern.

and compare their performance with the proposed method. Contrastive learning contributes the most to the performance. Using SAM optimizer significantly improves the generalizability, with 6.7% increase of IPCW C-index in the external test set. Mixup and attention pooling provide slightly better performance.

Biomarker Discovery. Based on cluster-level risk scores generated by the ViT, we show representative samples of clusters that have the 6 highest risk scores in Fig. 2. We can observe that these clusters not only focus on imaging patterns but also the location of the patterns. We also identify a novel pattern (C_{36} in Fig. 2) which is morphologically different from existing patterns. C_{36} is centered on the lateral border of the lung by the ribs, in a typical distribution for IPF-related fibrosis. The patches identify the commingling of high and low density extremes. The low density can be a combination of honeycombing, emphysema and traction bronchiectasis. The high density comprises primarily consolidation (not uncommonly representing a radiological pleuroparenchymal fibroelastosis pattern) and reticulation. The mortality prediction performance when using the novel biomarker (extent of C_{36}) and visual scores (fibrosis extent and emphysema extent) is better than using visual scores alone (Supplementary Table 2), with $\approx$ 1% increase of IPCW C-index in both test sets. This suggests the novel biomarker is likely to be complementary to existing biomarkers in predicting mortality risk.

4 Discussion

There are some limitations to this work. First, the homogeneity of some clusters is unsatisfactory (Supplementary Fig. 2a). The challenge is how to set the number of clusters to find a balance between having homogeneous clusters but ones that still have enough patches within the cluster to make them clinically useful. Second, the method of identifying novel biomarkers in Sec. 2.3 is intuitive and more research should be done. Third, this work mainly focuses on axial plane, and other planes need further exploration.

In this work, we propose a framework for prognostic imaging biomarker discovery and survival analysis. Experiments on two IPF datasets demonstrate that the proposed method performs better than its CNN counterparts in terms of discrimination and calibration. The novel biomarker discovered by our method provides additional prognostic information when compared to previously defined biomarkers used in IPF mortality prediction. This method can be potentially extended to broader applications for different diseases and image modalities.

Acknowledgements. AZ is supported by CSC-UCL Joint Research Scholarship. DCA is supported by UK EPSRC grants M020533, R006032, R014019, V034537, Wellcome Trust UNS113739. JJ is supported by Wellcome Trust Clinical Research Career Development Fellowship 209,553/Z/17/Z. DCA and JJ are supported by the NIHR UCLH Biomedical Research Centre, UK.

References

1. Abbet, C., Zlobec, I., Bozorgtabar, B., Thiran, J.P.: Divide-and-rule: self-supervised learning for survival analysis in colorectal cancer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 480–489. Springer (2020)
2. Abramson, R.G., Burton, K.R., John-Paul, J.Y., Scalzetti, E.M., Yankeelov, T.E., Rosenkrantz, A.B., Mendiratta-Lala, M., Bartholmai, B.J., Ganeshan, D., Lenchik, L., et al.: Methods and challenges in quantitative imaging biomarker development. Academic radiology **22**(1), 25–32 (2015)
3. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European conference on computer vision. pp. 213–229. Springer (2020)
4. Chen, S., Ma, K., Zheng, Y.: Med3d: Transfer learning for 3d medical image analysis. arXiv preprint arXiv:1904.00625 (2019)
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
6. Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) **34**(2), 187–202 (1972)
7. Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.A., Müller, H.: Building a reference multimedia database for interstitial lung diseases. Computerized medical imaging and graphics **36**(3), 227–238 (2012)

8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
9. Faraggi, D., Simon, R.: A neural network model for survival data. Statistics in medicine **14**(1), 73–82 (1995)
10. Foret, P., Kleiner, A., Mobahi, H., Neyshabur, B.: Sharpness-aware minimization for efficiently improving generalization. arXiv preprint arXiv:2010.01412 (2020)
11. Graf, E., Schmoor, C., Sauerbrei, W., Schumacher, M.: Assessment and comparison of prognostic classification schemes for survival data. Statistics in medicine **18**(17-18), 2529–2545 (1999)
12. Group, B.D.W., Atkinson Jr, A.J., Colburn, W.A., DeGruttola, V.G., DeMets, D.L., Downing, G.J., Hoth, D.F., Oates, J.A., Peck, C.C., Schooley, R.T., et al.: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clinical pharmacology & therapeutics **69**(3), 89–95 (2001)
13. Haarburger, C., Weitz, P., Rippel, O., Merhof, D.: Image-based survival prediction for lung cancer patients using cnns. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 1197–1201. IEEE (2019)
14. Hassani, A., Walton, S., Shah, N., Abuduweili, A., Li, J., Shi, H.: Escaping the big data paradigm with compact transformers. arXiv preprint arXiv:2104.05704 (2021)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
16. Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., Langs, G.: Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. European Radiology Experimental **4**(1), 1–13 (2020)
17. Huang, J., Dong, Q., Gong, S., Zhu, X.: Unsupervised deep learning by neighbourhood discovery. In: International Conference on Machine Learning. pp. 2849–2858. PMLR (2019)
18. Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F.: A survey on contrastive self-supervised learning. Technologies **9**(1), 2 (2021)
19. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. Journal of the American statistical association **53**(282), 457–481 (1958)
20. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
21. Kvamme, H., Borgan, Ø., Scheel, I.: Time-to-event prediction with neural networks and cox regression. arXiv preprint arXiv:1907.00825 (2019)
22. Ley, B., Collard, H.R., King Jr, T.E.: Clinical course and prediction of survival in idiopathic pulmonary fibrosis. American journal of respiratory and critical care medicine **183**(4), 431–440 (2011)
23. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
24. Pölsterl, S., Wolf, T.N., Wachinger, C.: Combining 3d image and tabular data via the dynamic affine feature map transform. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 688–698. Springer (2021)
25. Ratner, B.: The correlation coefficient: Its values range between+ 1/- 1, or do they? Journal of targeting, measurement and analysis for marketing **17**(2), 139–142 (2009)

26. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
27. Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S., Fu, H.: Transformers in medical imaging: A survey. arXiv preprint arXiv:2201.09873 (2022)
28. Uno, H., Cai, T., Pencina, M.J., D'Agostino, R.B., Wei, L.J.: On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in medicine **30**(10), 1105–1117 (2011)
29. Walsh, S.L., Humphries, S.M., Wells, A.U., Brown, K.K.: Imaging research in fibrotic lung disease; applying deep learning to unsolved problems. The Lancet Respiratory Medicine **8**(11), 1144–1153 (2020)
30. Wulczyn, E., Steiner, D.F., Moran, M., Plass, M., Reihs, R., Tan, F., Flament-Auvigne, I., Brown, T., Regitnig, P., Chen, P.H.C., et al.: Interpretable survival prediction for colorectal cancer using deep learning. NPJ digital medicine **4**(1), 1–13 (2021)
31. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
32. Zheng, M., Gao, P., Zhang, R., Li, K., Wang, X., Li, H., Dong, H.: End-to-end object detection with adaptive clustering transformer. arXiv preprint arXiv:2011.09315 (2020)
33. Zhong, S.: Efficient online spherical k-means clustering. In: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005. vol. 5, pp. 3180–3185. IEEE (2005)