



HAL
open science

Learned Deep Radiomics for Survival Analysis with Attention

Ludivine Morvan, Cristina Nanni, Anne-Victoire Michaud, Bastien Jamet, Clément Bailly, Caroline Bodet-Milin, Stephane Chauvie, Cyrille Touzeau, Philippe Moreau, Elena Zamagni, et al.

► **To cite this version:**

Ludivine Morvan, Cristina Nanni, Anne-Victoire Michaud, Bastien Jamet, Clément Bailly, et al.. Learned Deep Radiomics for Survival Analysis with Attention. Predictive Intelligence in Medicine (PRIME) 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, pp.35-45, 2020, 10.1007/978-3-030-59354-4_4. hal-03266299

HAL Id: hal-03266299

<https://hal.science/hal-03266299v1>

Submitted on 21 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learned deep radiomics for survival analysis with attention^{*}

Ludivine Morvan^{1,2}, Cristina nanni³, Anne-Victoire Michaud⁴, Bastien Jamet⁴,
Clément Bailly^{2,4}, Caroline Bodet-Milin^{2,4}, Stephane Chauvie⁵, Cyrille
Touzeau⁶, Philippe Moreau⁶, Elena Zamagni⁷, Françoise Kraeber-Bodéré^{2,4,8},
Thomas Carlier^{2,4}, and Diana Mateus¹

¹ Ecole Centrale de Nantes, Laboratoire des Sciences Numériques de Nantes (LS2N),
CNRS UMR 6004, Nantes, France

² CRCINA, INSERM, CNRS, University of Angers, University of Nantes, Nantes,
France

³ Nuclear Medicine, AOU Policlinico S. Orsola-Malpighi, Bologna, Italy

⁴ University Hospital of Nantes, Nuclear Medicine Department, Nantes, France

⁵ Medical Physics Division, Santa Croce e Carle Hospital, Cuneo, Italy

⁶ University Hospital of Nantes, Haematology Department, Nantes, France and
CRCINA

⁷ Seragnoli Institute of Hematology, Bologna University School of Medicine, Italy

⁸ Nuclear Medicine, ICO Cancer Center, Saint-Herblain, France

ludivine.morvan@ls2n.com

Abstract. In the context of multiple myeloma, patient diagnosis and treatment planning involve the medical analysis of full-body Positron Emission Tomography (PET) images. There has been a growing interest in linking quantitative measurements extracted from PET images (radiomics) with statistical methods for survival analysis. Following very recent advances, we propose an end-to-end deep learning model that learns relevant features and predicts survival given the image of a lesion. We show the importance of dealing with the variable scale of the lesions, and propose to this end an attention strategy deployed both on the spatial and channels dimensions, which improves the model performance and interpretability. We show results for the progression-free survival prediction of multiple myeloma (MM) patients on a clinical dataset coming from two prospective studies. We also discuss the difficulties of adapting deep learning for survival analysis given the complexity of the task, the small lesion sizes, and PET low SNR (signal to noise ratio).

Keywords: Survival Analysis · Multiple Myeloma · PET imaging · Convolutional NN · Attention models · Spatial Pyramidal Pooling

1 Introduction

Survival analysis aims to quantitatively link patient data to disease progression over time. It enables identifying bio-markers useful for splitting patient popula-

^{*} This work has been partially funded by the SIRIC ILIAD and the MILCOM Connect Talent region pays de la loire and nantes metropole.

tion into risk (e.g. low and high) subgroups. It is also used for training models that, given the data of a new patient, are able to predict the time to the next event (e.g. relapse, progression, death, etc.). Traditional survival methods rely on Cox proportional hazards regression and Kaplan Meier curves [13].

The goal of this paper is to adapt deep learning methods to survival analysis for prognosis in multiple myeloma (MM) patients using PET images. The use of quantitative image analysis for survival analysis based on PET images is a relatively recent field, commonly addressed with handcrafted (radiomics) features [5]. There are several challenges in adapting CNNs to instead learn discriminative features for survival. Firstly, PET images have low resolution, especially when considering the size of MM lesions. Secondly, prospective datasets that control patient treatment and follow-up are needed but often limited to a low number of patients. Finally, survival data often suffers from censorship (missing data) related to patients presenting no event during the study.

In this paper, we revisit recent deep learning methods for survival analysis, and propose a method that handles the challenges described previously. We manage the small-sized lesions and their scale variability of lesions with SPP [8, 15] or with an attention model [21]. At the same time, the low number of data samples imposes architectural and learning choices such as the reduction of the number of neurons in the fully connected layers, leaky ReLU, instance normalization, learning rate decay, and kernel regularisation to avoid overfitting. Finally, a specific survival loss function (negative partial log-likelihood) enables us to deal with missing data in the form of censorship.

This work investigates for the first time the use of learned deep radiomics for survival prognosis in the context of MM using PET images. To the best of our knowledge, we are the first to combine channel and spatial attention in the context of survival analysis using images as input. Our work improves prediction results over state-of-the-art methods for image-based multiple myeloma survival prediction.

2 Related Work

Early adaptations of Deep Learning (DL) to survival tasks extract deep features with a pre-trained Neural Network and feed them to prediction models such as Lasso Cox [14] or Random Survival Forest (RSF) [19]. The survival problem has also been simplified to the *classification* of different risk groups (e.g., low, middle, high risk) [2], or to the *regression* of the time-to-event [19]. However, such formulations do not natively handle *censored data*.

Risk-predicting methods adapt the learning loss to take into account the censorship. Faraggi and Simon [6] adapted the linear Cox Proportional Hazards (CPH) model with a more flexible feedforward network. Katzman et al. revisits (in the so called DeepSurv approach) the Faraggi-Simon’s loss in the context of deeper networks [12]. Some other variants stick to the CPH linear model, $\mathbf{w}^\top \mathbf{x}$, but use the network either for non-linear dimensionality reduction of the input

\mathbf{x} [4], or for predicting the weights \mathbf{w} [17]. Recent alternatives include combining a regression and a ranking loss [4] or using discrete survival model [7].

Regarding the data type, Xinliang *et al.* [22] were the first to adapt DeepSurv to Convolutional Layers (DeepConvSurv) and dealt with images data as input. Zhu *et al.*, extend in [1] and subsequent papers the DeepConvSurv to handle very large whole slide histopathological images. We are instead interested in the problems raised by the low-resolution and SNR of PET images. Both [2] and [15] propose 3D CNN models for survival analysis from PET images. Amyar *et al.* [2] target radio-chemotherapy response in esophageal cancer by simplifying the survival problem to a classification without censorship and using an input layer of relative large size ($100 \times 100 \times 100$) compared to our MM lesions. To predict treatment response in colorectal cancer, Li et al [15] modify the DeepConvSurv model with an additional spatial pyramid pooling (SPP) layer [8] in order to deal with small lesions of multiple scales. Both PET and CT data are considered as inputs, showing that multi-modality improves the performance. Although our study is performed on PET images only, we borrow from the idea of using SPPs to handle multiple-scales of even smaller lesions for the MM survival analysis. In addition, we propose a novel strategy to handle the variability of small lesions, by means of attention models. Attention has been successfully used in a large variety of medical applications including reconstruction [10], segmentation, detection [20] or classification [9]. Kaji et al. [11] use LSTM to predict daily sepsis from clinical data while Liu et al.[16] focus on spatial attention. Instead, our work integrates a CBAM model, which also includes channel attention to determine the most predictive "learned radiomics" filters.

Moreover, our work is the first to adapt CNNs for MM survival from PET lesion images only.

3 Method

The input to our model is a data-set of N samples, $\{\mathbf{x}_i, t_i, \delta_i\}_{i=1}^N$, each consisting of an image \mathbf{x}_i associated to a target time-to-event t_i and a binary censorship δ_i . $\delta_i = 0$ means that the event of interest did not happen during the studied period, while $\delta_i = 1$ means that the event did occur. In the particular case of study, our images come from manually identified lesions in full-body PET images and time is given in days. We consider both 2D and 3D images. The proposed method adapts a CNN to predict progression free survival (PFS), i.e. the time to (or risk of) the next disease progression, from images and in presence of censored data. The output of the method is the scalar risk value. An overview of the method is presented in Fig1. In the following, we describe the loss function for survival (sec. 3.1) and the deep learning model including the SPP and attention strategies (sec. 3.2).

3.1 From Cox survival model to loss function

A common approach to survival analysis with CNNs has been to derive a loss function from the Cox proportional hazard model. Cox assumes that each vari-

able independently affects a baseline hazard $h_0(t)$ (a measure of risk at time t), multiplied by a constant factor independent of time. The risk for patient i is then modelled as $h(\mathbf{x}_i, t) = h_0(t)e^{h_\beta(\mathbf{x}_i)} = h_0(t)e^{\beta^\top \mathbf{x}_i}$ where $\mathbf{x}_i \in \mathbf{R}^m$ and $\beta \in \mathbf{R}^m$ are respectively, the vector of input variables and their associated coefficients. The partial likelihood for one coefficient is the product of probabilities over time, where for each event-time t_i the probability is computed as the ratio between the risk of patient i and the cumulative risk of all individuals still at risk at time t_i :

$$L_{\text{cox}}(\beta) = \prod_{\{i|\delta_i=1\}} \frac{e^{h_\beta(\mathbf{x}_i)}}{\sum_{\{j|t_j \geq t_i\}} e^{h_\beta(\mathbf{x}_j)}}, \quad (1)$$

where the product is done over the defined (uncensored) time events.

DeepSurv Following [6, 12], the linear model $h_\beta(\mathbf{x})$ is replaced by the output of a neural network $h_\theta(\mathbf{x})$ parameterised by weights θ while the same partial likelihood is kept. Computing the negative log likelihood from Eq. 1 leads to the following loss to optimise the parameters θ :

$$l_{\text{cox}}(\theta) = - \sum_{\{i|\delta_i=1\}} [h_\theta(\mathbf{x}_i) - \log \sum_{\{j|t_j \geq t_i\}} e^{h_\theta(\mathbf{x}_j)}] \quad (2)$$

This loss function pushes the network to predict risks that explain the order of events in the dataset. Note that the risk of the current patient depends on all patients at risk at the event time. We use the loss in Eq.3.1 to optimize the parameters of the CNN architecture described next.

3.2 CNN model for survival analysis

The core of the risk prediction model $h(\mathbf{x}_i, \theta)$ is a CNN learning radiomics features and whose architecture was inspired from [15]. We consider both a 2D and 3D version of the model and two additional (optional) blocks: i) a Spatial Pyramidal block to deal with small lesions of multiple-sizes [8] ii) a spatial and channel attention block [21] to localise respectively the lesion and the filters, and enhance model interpretability.

Radiomics Feature Learning block This block is a standard CNN transforming the 3-D image of a lesion \mathbf{x}_i into C feature maps each of size $H \times W \times L$. It is composed of three convolutional blocks. Each block has one convolution with leaky RELU activations (+ Maxpooling when SPP is not used).

In practice, the input layer is set to $36 \times 36 (\times 36)$, according to the distribution of MM lesions' scale in our database, where the bounding polygons surrounding the lesions were shown to be between $3 \times 3 (\times 3)$ and $32 \times 40 (\times 53)$ pixels size.

The kernel size of the convolutions are $3 \times 3 (\times 3)$, $5 \times 5 (\times 5)$ and $3 \times 3 (\times 3)$ with padding and a stride at 1. The difference with model [15] is in the added layers as Leaky RELU, dropout or Instance normalisation.

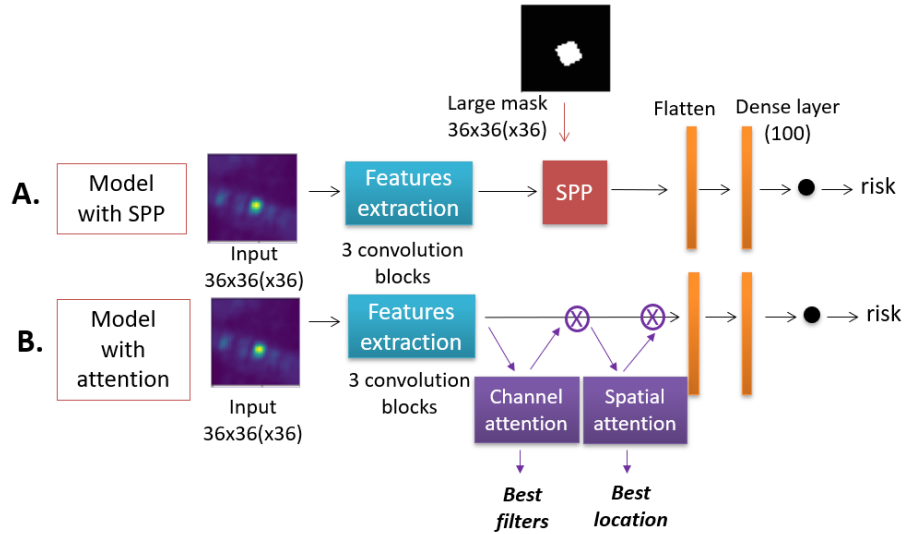


Fig. 1. Models with SPP (A) or Attention part (B)

Spatial Pyramid Pooling (SPP) was designed to handle multi-scale images efficiently. Multiple max pooling layers at different scales are applied to the features from the last convolutional layer to later flattened and concatenated (see Fig 2). Different to classical pooling, the size of the resultant feature maps is fixed, and thus the output feature vector dimension remains constant irrespective of the input image size.

The SPP method was used in [15] to handle PET/CT rectal-cancer lesions of different sizes. A bounding box delimiting the tumour was given as an auxiliary input to the SPP. Moreover, each feature map was pooled to three 2-D matrices of size $8 \times 8 \times 8$, $4 \times 4 \times 4$ and $2 \times 2 \times 2$ (in the case of 3D) and 8×8 , 4×4 and 2×2 (in the case of 2D). In this work, we also use the tumour mask. However, for our MM application, many of the lesions are small (the smallest is less than $3 \times 3(\times 3)$), and thus a choice has to be made (on our 2D model) between either keeping the 8×8 as in [15] dominated by background information, or retaining the limited amount of information from the smaller 4×4 and 2×2 maps. An alternative is to use a spatial attention model.

CBAM Attention Model An attention mechanism learns to assign weights to the extracted features according to their relevance for a given task, with the side effect of improving the task performance. The mechanism is flexible in the sense the weights adapt to the input image. After the feature-extraction, we add a Convolutional Block Attention Module (CBAM) to sequentially infer attention maps for both the feature channels and spatial axes [21]. Channel attention is computed squeezing spatial information from the convolutional layers with max

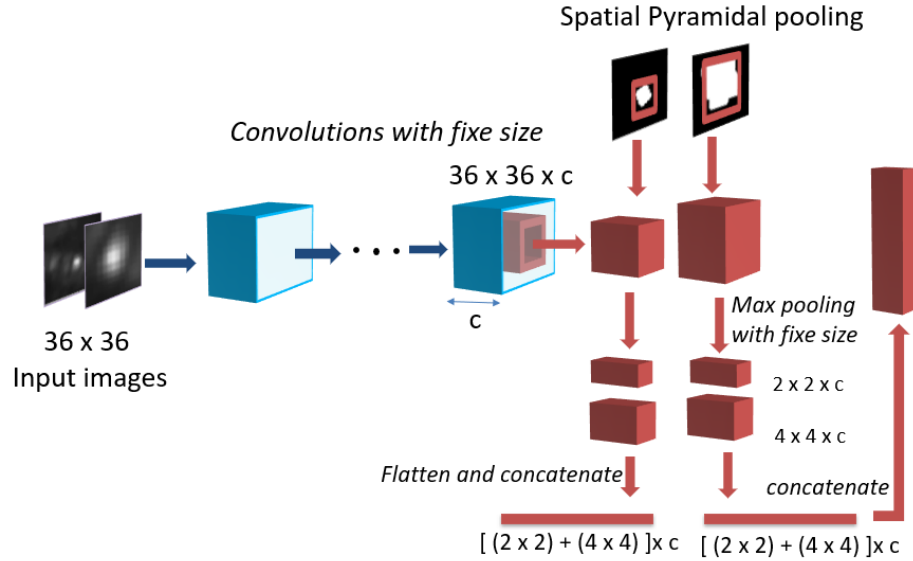


Fig. 2. SPP block for a 2D model.

and average pooling operations. The pooled values are passed through a shared MultiLayer Perceptron network to predict the attention weights. The MLP is trained along with the survival prediction layers following a parallel branch from the output back to the convolutional layers (See Fig. 3).

Spatial attention weights are calculated at each location of the image. They are computed applying max pooling and average pooling to squeeze the channels. They are then applied to the feature maps by an element-wise multiplication. In this way, CBAM focuses spatial attention within the lesion and not around. For this reason, it is an interesting alternative to the SPP. Spatial attention maps can also, in some cases, show the most important part within the lesion. Channel attention provides information about the most informative filters and thus, give interpretability to the model.

Survival prediction output The prediction layer consists of 3 fully connected layers, and an output layer with a single neuron predicting the risk $h(\mathbf{x}_i, \theta)$.

Optimisation details An ADAM optimizer was used. Hyperparameter optimisation was done on the batch size (10, 32, 64), the learning rate (1e-03, 1e-04, 1e-05, 1e-06), the learning rate decay (1e-06, 1e-07, 1e-08, 1e-09) and the number of epochs (until 150). Dropout was added after the first fully connected layer and also on the input but was not retained. Batch and instance normalisation were enforced. The Leaky RELU activation is set to 0.1. Leaky RELU, instance normalisation, kernel regularisation and dropout are used to avoid overfitting.

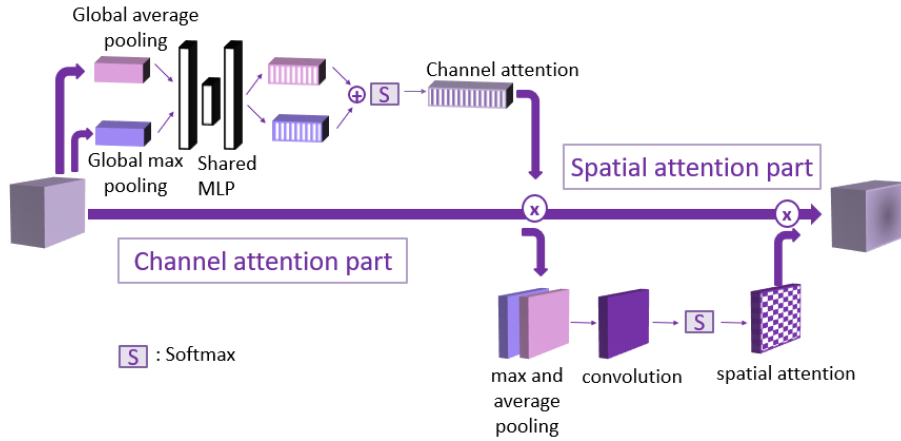


Fig. 3. Attention block for a 2D model [21].

4 Experimental Validation

Experimental set-up We evaluate our survival model on a dataset composed of two prospective multi-centric MM studies with respectively 87 and 65 patients [18, 3]. The 3D baseline PET-images and survival time (until 7 years) are available for each patient. The most intense lesion for each patient was selected and a global polygon (no segmentation step involved) around the lesion was drawn by a nuclear physician. According to the version, 2D or 3D information was used for both the PET image and its corresponding mask. For the 2D case, we extract the 2D transverse, coronal, and sagittal sections from the middle of the bounding polygon, and use them individually. For 3D we use a single cube encompassing the bounding box. Finally, we use data augmentation (rotation, translation, zoom, flip) to obtain 30 different images per input image. This leads to a total of 13850 images in 2D and 4560 in 3D, split patient-wise into 4 sets for cross-validation.

Regarding the input layer size, the median lesion size is $116 \times 116 \times 12$, and only 4 of the 155 patients have a box bigger than $36 \times 36 \times 36$. Therefore, we define the input size to $36 \times 36 \times 36$. This choice avoids deteriorating the smallest lesions with the interpolation step or having too much background information (in the case of spatial attention). For the SPP and attention methods, we extract a $36 \times 36 (\times 36)$ cube around the center of polygon. For the baseline CNNs for comparison (without SPP or attention), we take the parallelepiped englobing the bounding box of the lesion, and we resize it with cubic interpolation to a size of $36 \times 36 (\times 36)$.

Evaluation Metrics To evaluate the model’s performance, we rely on the C-index measuring if the predicted risk respects the events order.

$$\text{Cindex} = \frac{\sum_{ij} \mathbf{1}_{t_j < t_i} \cdot \mathbf{1}_{h(\mathbf{x}_j, \theta) > h(\mathbf{x}_i, \theta)} \cdot \delta_j}{\sum_{ij} \mathbf{1}_{t_j < t_i} \cdot \delta_j} \quad (3)$$

Table 1. Comparison vs baseline models. Average c-index over 4-fold cross validation.

Model	C-index
Lasso-Cox	0.494 (± 0.0025)
RSF	0.529 (± 0.036)
VIMP+RSF	0.583 (± 0.017)
CNN 2D	0.6100 (± 0.041)
Our model (3D+Attention)	0.6414 (± 0.023)

4.1 Experiment 1. CNNs vs classical methods for survival analysis

Herein, we compare a CNN-based model against three baselines: Lasso-Cox, and Random Survival Forest (RSF) with and without Variable importance selection (VIMP). Input to the baseline models were 19 handcrafted radiomics. VIMP feature selection was done on 100 runs of the initial RSF model trained on all the variables. Both models were optimized to their best performance. The results of a 4-fold cross-validation are shown in Table 1. We report the average and standard deviation across the 4 folds. Our simplest CNN model, adapted from DeepConvSurv [1] and trained on 2D input data, shows improvements over the baselines.

4.2 Experiment 2. Evaluation of deep learning methods

Here, we perform a comparative study based on the 4-fold cross-validation results of different configurations of the method. We consider 2D and 3D versions to explore the performance variability for two extremes of the model capacity: with the 2D models we search for a compact model capable of predicting meaningful risks; with the 3D models we look for a high complexity model we can still train with the available data. The results are reported in Table 2. The reported values in the table are taken at the best validation performance.

Among the 2D models, the SPP seems to most effectively learn discriminative features for survival probably due to the concentration of the information in the compact pooling layers, it also corresponds to the most compact model. Attention adds relatively few additional parameters, but does not improve the results any further. Actually, when looking at the learning curves we notice instead an acceleration of the overfitting. Conversely, the 3D model with attention can take full advantage of the extra model parameters, effectively guiding the convergence and significantly improving the results over the baseline. The difference might come from the spatial constraints enforced by the 3D data. Despite performing similar to 2D+SPP, the cost of a larger complexity is worthwhile when interpretability of the model is sought.

5 Discussion and conclusions

In this paper, we address survival analysis of MM patients from PET image lesions. We present a deep learning approach that adapts CNNs to learn relevant radiomics features from the data in an end-to-end fashion. We develop two

Table 2. Comparison of different deep models. The result is for each model the average C-index over 4-cross validation.

Model	Validation c-index	Train c-index	number of parameters
2D simple	0.6100 (± 0.041)	0.7814 (± 0.066)	554,003
2D+SPP	0.6396 (± 0.061)	0.7671 (± 0.045)	163,603
2D+Attention	0.5059 (± 0.017)	0.7526 (± 0.201)	555,197
2D + SPP +Attention	0.4869 (± 0.017)	0.7398 (± 0.101)	163,783
3D simple	0.5712 (± 0.028)	0.5525 (± 0.048)	4,800,557
3D+Attention	0.6414 (± 0.023)	0.6384 (± 0.044)	4,802,339

strategies to deal with the small and variable size of the input lesions, SPP and Attention, and explore their effect on model capacity and performance. When looking for compact models, SPP seems to be more appropriate. If more complexity is affordable, then the 3D + CBAM attention model is preferred: for a similar performance, it removes the need for segmentation masks, it can be used to verify if the model has correctly focused on a lesion, and provides additional hints to the importance of the learned features. Such double attention mechanism is new in the context of deep survival analysis with images.

In future work, we will further analyse the clinical relevance of our method, study the attention matrix and the incidence of considering a multi-modal PET/CT approach.

References

1. Zhu, Xinliang and Yao, Jiawen and Zhu, Feiyun, Huang, J.: WSISA: Making Survival Prediction from Whole Slide Histopathological Images. In: *cvpr*, pp. 970–975 (2017)
2. Amyar, A., Ruan, S., Gardin, I., Chatelain, C., Decazes, P., Modzelewski, R.: 3-d rpet-net: Development of a 3-d pet imaging convolutional neural network for radiomics analysis and outcome prediction. *IEEE Transactions on Radiation and Plasma Medical Sciences* **3**(2), 225–221 (2019)
3. Carlier, T., Bailly, C., Leforestier, R., Touzeau, C., Moreau, P., Bodere, F., Bodet-Milin, C.: Prognostic added value of pet textural features at diagnosis in multiple myeloma. *Journal of Nuclear Medicine* **58**(supplement 1), 111 (2017)
4. Ching, T., Zhu, X., Garmire, L.X.: Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Computational Biology* **14**(4), 1–18 (2018). <https://doi.org/10.1371/journal.pcbi.1006076>
5. Cook, G.J.R., Siddique, M., Taylor, B.P., Yip, C., Chicklore, S., Goh, V.: Radiomics in PET: principles and applications. *Clinical and Translational Imaging* **2**(3), 269–276 (2014). <https://doi.org/10.1007/s40336-014-0064-0>. URL <https://doi.org/10.1007/s40336-014-0064-0>
6. Faraggi, D., Simon, R.: A neural network model for survival data. *Statistics in Medicine* **14**(1), 73–82 (1995). <https://doi.org/10.1002/sim.4780140108>
7. Gensheimer, M.F., Narasimhan, B.: A scalable discrete-time survival model for neural networks. *PeerJ* **2019**(1), 1–17 (2019). <https://doi.org/10.7717/peerj.6257>

8. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (eds.) *Computer Vision – ECCV 2014*, pp. 346–361. Springer International Publishing, Cham (2014)
9. Herent, P., Schmauch, B., Jehanno, P., Dehaene, O., Saillard, C., Balleyguier, C., Arfi-Rouche, J., Jégou, S.: Detection and characterization of mri breast lesions using deep learning. *Diagnostic and Interventional Imaging* **100**(4), 219 – 225 (2019). <https://doi.org/https://doi.org/10.1016/j.diii.2019.02.008>
10. Huang, Q., Yang, D., Wu, P., Qu, H., Yi, J., Metaxas, D.: Mri reconstruction via cascaded channel-wise attention network. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1622–1626 (2019). <https://doi.org/https://doi.org/110.1109/ISBI.2019.8759423>
11. Kaji DA Zech JR, K.J.e.a.: An attention based deep learning model of clinical events in the intensive care unit. *PLoS One* **14**(2:e0211057) (2019). <https://doi.org/https://doi.org/10.1371/journal.pone.0211057>
12. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Yuval: Deepsurv: Personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Med Res Methodol* **18**(24) (2018)
13. Kleinbaum, D.G., Klein, M.: *Survival Analysis: A Self-Learning Text*. Springer Science and Business Media, LLC (2012)
14. Lao, J., Chen, Y., Li, Z.C., Li, Q., Zhang, J., Liu, J., Zhai, G.: A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Scientific Reports* **7**(1) (2017). <https://doi.org/10.1038/s41598-017-10649-8>
15. Li, H., Boimel, P., Janopaul-Naylor, J., Zhong, H., Xiao, Y., Ben-Josef, E., Fan, Y.: Deep convolutional neural networks for imaging data based survival analysis of rectal cancer. In: *isbi*, vol. 2019-April, pp. 846–849 (2019). <https://doi.org/10.1109/ISBI.2019.8759301>
16. Liu, Z., Sun, Q., Bai, H., Liang, C., Chen, Y., Li, Z.: 3d deep attention network for survival prediction from magnetic resonance images in glioblastoma. In: *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 1381–1384 (2019)
17. Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D.A., Barnholtz-Sloan, J.S., Velázquez Vega, J.E., Brat, D.J., Cooper, L.A.D.: Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences* **115**(13), E2970–E2979 (2018). <https://doi.org/10.1073/pnas.1717139115>. URL <https://www.pnas.org/content/115/13/E2970>
18. Nanni, C., Versari, A., Chauvie, S., et al.: Interpretation criteria for fdg pet/ct in multiple myeloma (impetus): final results. impetus (italian myeloma criteria for pet use). *European Journal of Nuclear Medicine and Molecular Imaging* **45**, 712–719 (2018). <https://doi.org/https://doi.org/10.1007/s00259-017-3909-8>
19. Shboul, Z.A., Alam, M., Vidyaratne, L., Pei, L., Elbakary, M.I., Iftekharuddin, K.M.: Feature-guided deep radiomics for glioblastoma patient survival prediction. *Frontiers in Neuroscience* **13**, 966 (2019). <https://doi.org/10.3389/fnins.2019.00966>. URL <https://www.frontiersin.org/article/10.3389/fnins.2019.00966>
20. Tong, Q., Li, C., Si, W., Liao, X., Tong, Y., Yuan, Z., Heng, P.A.: Rianet: Recurrent interleaved attention network for cardiac mri segmentation. *Computers in Biology and Medicine* **109**, 290 – 302 (2019). <https://doi.org/https://doi.org/10.1016/j.compbiomed.2019.04.042>

21. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (eds.) *Computer Vision – ECCV 2018*, pp. 3–19. Springer International Publishing, Cham (2018)
22. Zhu, X., Yao, J., Huang, J.: Deep convolutional neural network for survival analysis with pathological images. In: *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1, pp. 544–547. IEEE (2016). <https://doi.org/10.1109/BIBM.2016.7822579>