



MIA-Prognosis: A Deep Learning Framework to Predict Therapy Response

Jiancheng Yang^{1,2,3}, Jiajun Chen³, Kaiming Kuang³, Tiancheng Lin¹,
Junjun He¹, and Bingbing Ni^{1,2,4}(✉)

¹ Shanghai Jiao Tong University, Shanghai, China
nibingbing@sjtu.edu.cn

² MoE Key Lab of Artificial Intelligence, AI Institute,
Shanghai Jiao Tong University, Shanghai, China

³ Dianei Technology, Shanghai, China

⁴ Huawei Hisilicon, Shanghai, China

Abstract. Predicting clinical outcome is remarkably important but challenging. Research efforts have been paid on seeking significant biomarkers associated with the therapy response or/and patient survival. However, these biomarkers are generally costly and invasive, and possibly dissatisfactory for novel therapy. On the other hand, multi-modal, heterogeneous, unaligned temporal data is continuously generated in clinical practice. This paper aims at a unified deep learning approach to predict patient prognosis and therapy response, with easily accessible data, *e.g.*, radiographics, laboratory and clinical information. Prior arts focus on modeling single data modality, or ignore the temporal changes. Importantly, the clinical time series is asynchronous in practice, *i.e.*, recorded with irregular intervals. In this study, we formalize the prognosis modeling as a **multi-modal asynchronous** time series classification task, and propose a *MIA*-Prognosis framework with **Measurement**, **Intervention** and **Assessment** (*MIA*) information to predict therapy response, where a Simple Temporal Attention (SimTA) module is developed to process the asynchronous time series. Experiments on synthetic dataset validate the superiority of SimTA over standard RNN-based approaches. Furthermore, we experiment the proposed method on an in-house, retrospective dataset of real-world non-small cell lung cancer patients under anti-PD-1 immunotherapy. The proposed method achieves promising performance on predicting the immunotherapy response. Notably, our predictive model could further stratify low-risk and high-risk patients in terms of long-term survival.

Keywords: Asynchronous time series · Prognosis · Immunotherapy

1 Introduction

Modeling patient prognosis is a challenging but important topic in clinical research, where researchers analyze and predict clinical outcomes including

J. Yang and J. Chen—Contributed equally

© Springer Nature Switzerland AG 2020

A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12262, pp. 211–220, 2020.

https://doi.org/10.1007/978-3-030-59713-9_21

response to certain therapy (*e.g.*, radiotherapy, chemotherapy, surgery, immunotherapy for oncology), patient progression-free survival (PFS) and overall survival (OS). Research efforts have been paid on seeking significant biomarkers, *e.g.*, EGFR mutation for EGFR-TKI therapy [23], PD-L1 expression and tumor mutational burden (TMB) for immunotherapy [5]. However, these biomarkers are generally costly and invasive, and possibly dissatisfactory for novel therapy, *e.g.*, anti-PD-1 and anti-PD-L1 immunotherapy [14]. With more novel revolutionary therapy (including combination therapy [11]) available, a unified analytic framework for modeling patient prognosis is urged.

We address this issue via emerging deep learning technology by mining clinical data, *e.g.*, electronic health records (EHR) [13]. Specifically, we focus on a unified approach to model patient prognosis under certain therapy. Prior arts are generally developed on a single data modality [10, 16]. Besides, only a few studies [18] take into account the temporal/serial information. In clinical practice, multi-modal temporal data is continuously generated with numerous kinds of sensors and records. It is remarkably valuable to mine the easily accessible information to develop the prognosis prediction system, *e.g.*, radiographics, laboratory and clinical information. We formalize the prognosis modeling as a **multi-modal asynchronous** time series classification task, and propose a *MIA-Prognosis* framework with **Measurement, Intervention and Assessment** (MIA) information, where Measurement and Intervention information are treated as inputs of multi-modal asynchronous time series to predict the Assessment as ground truth (details in Sect. 2.1).

An algorithmic challenge is how to effectively and efficiently process multi-modal asynchronous time series like clinical information. Binkowski *et al.* [1] propose a gated CNN for asynchronous time series analysis, where asynchronous time intervals are regarded as input features. This approach might not be suitable for the clinical scenario since it is not essentially asynchronous; data is needed to learn representation for time intervals. What we need for real-world clinical data processing is a natively asynchronous model, which is flexible and light-weight to learn from expensive clinical data. Inspired by recent advances in natural language processing (NLP), *e.g.*, attention transformers [3, 17, 21] and relative position encoding [15], we propose a Simple Temporal Attention (SimTA) module to process asynchronous time series, where attention matrix is learned simply from the time intervals of asynchronous time series (details in Sect. 2.2).

The SimTA module is proven to be superior to standard RNN-based approaches in a synthetic asynchronous time series prediction dataset. Moreover, we experiment the proposed *MIA-Prognosis* framework on an in-house retrospective dataset of real-world non-small cell lung cancer (NSCLC) under anti-PD-1 immunotherapy. Our predictive model achieves promising performance on predicting the immunotherapy response after 90 days. Notably, this model could further stratify low-risk and high-risk patients in terms of long-term survival.

2 Methods

2.1 MIA Prognosis: The Framework

Categorizing Clinical Information. In clinical practice, data of a numerous variety of modalities is collected. Most medical data is unaligned in time steps, which means that it has varying intervals between adjacent steps in time series. Such limitations call for a unified framework that integrates asynchronous data of different modalities. To address this issue, we first divide clinical information into three categories according to data sources: **measurement**, **intervention** and **assessment**, which defined our *MIA-Prognosis* framework. Measurement data comes from medical examinations such as imaging data (computed tomography, ultrasound, X-ray), laboratory and genetic tests. Measurement is the main information in our *MIA-Prognosis* framework. Interventions include actions such as injections and operations. Assessment evaluates the effectiveness of interventions, *e.g.*, Response Evaluation Criteria in Solid Tumors (RECIST) [4], or 1-year overall survival rates. In this study, we use RECIST to obtain the “ground truth”¹ of therapy response, where complete response (CR), partial response (PR), stable disease (SD) are regarded as response (R), and PD (progressive disease) is regarded as non-response (non-R). Note that measurement, intervention and assessment can also be categoried into either serial or static data, which depends on its status over time. These categorizations are the basis of our framework’s capability of integrating heterogeneous multi-modal data.

Model Overview. We propose a framework that integrates multi-modal data in asynchronous time series, named *MIA-Prognosis*. Figure 1 gives an overall description of our framework. Due to the fact that clinical data of different modalities is usually unaligned in time, it is impractical to simply concatenate these vectors together and pad zeroes at the time step where a certain modality is missing. Therefore, we process each modality independently in our framework. We pass serial data of each modality through its own SimTA module, which outputs a summary vector. The summary vector is added with a temporal encoding (adapted from position encoding [17]) of time intervals between the assessment time and the last time stamp. Static data goes through a multi-layer perceptron (MLP) that encodes it in high-dimensional embedding. We then concatenate summary vectors of serial data with static embedding, and input the concatenated vector into another MLP to give the final prediction of therapy response (R: response/non-R: non-response) after an unobserved period.

Existing deep sequential models, such as recurrent neural network (RNN), assume that time series data is synchronous in nature. However, this assumption does not hold in the context of clinical practice. Here we introduce a new module to help us process asynchronous data, named SimTA. Inspired by recent advances in natural language processing (NLP), *e.g.*, attention transformers [3, 17] and relative position encoding [15], SimTA utilizes time interval information of asynchronous series to generate attention matrix, capturing temporal relationships

¹ RECIST is not theoretically perfect. We refer to “ground truth” in a clinical sense.

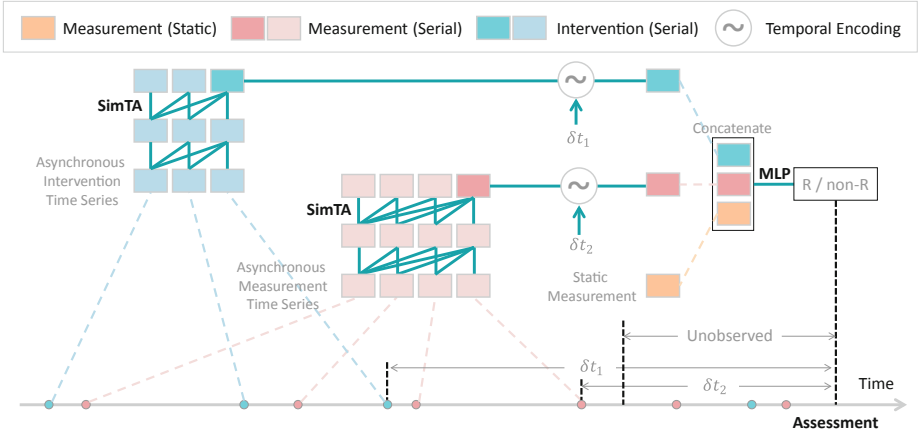


Fig. 1. The MIA-Prognosis framework, with Measurement, Intervention and Assessment information. The asynchronous time series is encoded by the proposed Simple Temporal Attention (SimTA) module into a summary vector. The summary vector is further added with a temporal encoding of time intervals between the assessment time and the last time stamp (δt_1 and δt_2). Together with static information, these features predict the therapy response (R/non-R) after an unobserved period.

between asynchronous time steps. It is worth noting that the latest steps in time series of different modalities are not likely to coincide with each other. In such cases, we use temporal encoding to make use of this information.

2.2 Simple Temporal Attention for Asynchronous Time Series

Simple Temporal Attention (SimTA) is the key factor that enables our framework to process asynchronous times series. Let $X \in \mathbb{R}^{T \times C}$ be an asynchronous time series of length T , and $\tau = [\tau_1, \tau_2, \dots, \tau_{T-1}] \in \mathbb{R}^{T-1}$ be the time interval vector between any adjacent time steps. A general formula of a single SimTA module can be described as:

$$\text{SimTA}(X, \tau) = \text{softmax}(A)\sigma(f(X)), \quad (1)$$

where σ denotes an activation function of our choice, and f is a fully-connected layer. $A = S(X, \tau) \in \mathbb{R}^{T \times T}$ is the attention matrix that encodes relations between any two time steps. Matrix A can be calculated using any edge-aware attention mechanism, *e.g.*, multi-head self attention [17] with relative position encoding [15]. In this study, we use an extremely simplified version that only encodes linearly time intervals, which is validated effective in our experiments:

$$A = S(\boldsymbol{\tau}) = \begin{bmatrix} 0, & -\infty, & -\infty, & \dots -\infty \\ -\lambda\tau_1 + \beta, & 0, & -\infty, & \dots -\infty \\ -\lambda(\tau_1 + \tau_2) + \beta, & -\lambda\tau_2 + \beta, & 0, & \dots -\infty \\ \vdots & \vdots & \vdots & \ddots \vdots \\ -\lambda\sum_1^{T-1}\tau_i + \beta, -\lambda\sum_2^{T-1}\tau_i + \beta, -\lambda\sum_3^{T-1}\tau_i + \beta, \dots & 0 \end{bmatrix}, \quad (2)$$

where $\lambda \in \mathbb{R}^+$ and $\beta \in \mathbb{R}$ are trainable parameters that apply a linear transformation on $\boldsymbol{\tau}$. The simplicity of this attention mechanism can help us cope with overfitting as well, considering we only have limited amount of data. The formula is based on the assumption that the more recent time steps should have stronger correlations with the current time step than further ones. Complex temporal information can be captured by stacking multiple SimTA modules. The complete SimTA model pipeline is a SimTA block of one or multiple SimTA modules, which outputs a summary vector, followed by an MLP that outputs the final prediction with softmax activation. To the best of our knowledge, the proposed SimTA is the first study to introduce attention mechanism to asynchronous time series analysis with proven effectiveness.

2.3 Counterpart Approaches

To demonstrate SimTA's capability of learning asynchronous temporal relations, we bring in LSTM (Long Short-term Memory) [7,9] as comparison. LSTM is a special type of RNN designed to capture relations over extended time intervals in sequences. In our experiments, we use LSTM models with a comparable size and the same training configuration as SimTA. Let X_i be the value of the time series at time step t_i . Three LSTM approaches are tested, which differ in their input: (1) Only X_i ; (2) X_i and time intervals τ_i ; (3) X_i and time stamps t_i . We mark them as LSTM, LSTM(i) and LSTM(s), respectively.

3 Experiments

3.1 Proof of Concept on Synthetic Dataset

We first validate the superiority of the proposed SimTA over RNN on asynchronous time series using a synthetic dataset.

Dataset and Experiment Settings. The synthetic dataset consists of the summation of N trigonometric functions with random periods. Each time series is computed as:

$$X_t = \sum_{j=1}^N [\alpha_j \sin(\omega_j \pi t + b_j) + \beta_j] + \eta \epsilon \quad (3)$$

where N is the number of trigonometric functions involved. ϵ is a white noise following the standard normal distribution, whose magnitude is controlled by a

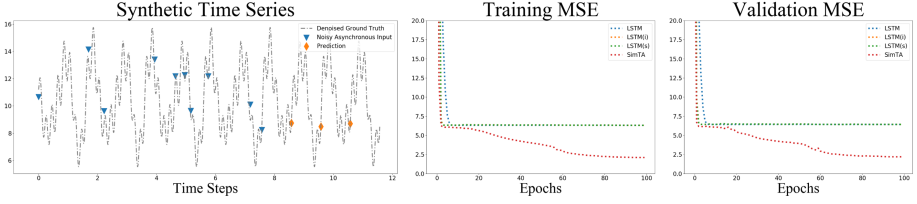


Fig. 2. The synthetic time series and MSE loss curves of LSTM, LSTM(i), LSTM(s) and the proposed SimTA. **Left:** The illustration of a data sample. **Middle:** Training MSE loss curves. **Right:** Validation MSE loss curves. We clip the y axis in both loss curves for the sake of better visualization.

constant η . In our experiments, we choose $N = 10$ and $\eta = 0.5$. We generate 10,000 such asynchronous series, which are split in 80/20 for training/validation, respectively. The training/validation data is sampled once from the predefined distribution and fixed throughout the experiment.

During training, 10 different X_t are randomly sampled from each series. The time intervals between any two adjacent points follow a uniform distribution between 0 and a maximum interval level I . The model is tasked to predict the next 3 points (+1, +2, +3) following the last one in the input. Figure 2 shows a sample series of the synthetic dataset. In our experiments, we compare performances of SimTA and three LSTM models, which are LSTM, LSTM(i) and LSTM(s). For LSTM(i) and LSTM(s), X_t and the time information are concatenated into one vector. SimTA follows the model described in Sect. 2.2.

Results. We train all four models for 100 epochs. Figure 2 shows the training and validation mean squared error over the training phase. SimTA outperforms all three LSTM models on both training and validation data. It achieves significantly lower MSE (2.197 on SimTA and 6.427 on all LSTM approaches) compared with LSTM. The time information does help LSTM(i) and LSTM(s) to converge faster than vanilla LSTM, but all three end up with errors at the same level. It is worth noting that the LSTM model variants underfit the training set. From the observations above, we conjecture that the proposed SimTA outperforms existing standard sequential models such as LSTM for asynchronous time series.

3.2 Predicting Response to Anti-PD-1 Immunotherapy for Non-Small Cell Lung Cancer (NSCLC)

Background. Lung cancer is the most commonly diagnosed cancer worldwide. According to [2], lung cancer accounts for 18.4% of the global cancer deaths in 2018. NSCLC makes up 80%–85% of these cases. Deep learning has shown its potential in precision medicine for lung cancer [19, 22, 24, 25]. Recently, immunotherapy has been proven to remarkably increase the overall survival and the life quality of patients with a variety of cancers, including NSCLC. However, only a small percentage of patients benefit from immunotherapy and show lasting responses. There has been research on the identification of response predictors,

whereas most of the effort are focused on biopsy analyses and serum biomarkers, *e.g.*, PD-L1 expression and tumor mutation burden (TMB) for first-line immunotherapy. These methods are expensive, invasive, and not always consistently associated with tumor responses. Furthermore, no biomarker is available for predicting second-line NSCLC immunotherapy outcome so far. Such limitations emphasize the necessity for convenient, economical and non-invasive indicators, especially for second-line immunotherapy treatment.

Dataset and Experiment Settings. In this retrospective study, 99 patients with advanced or metastatic stage IIIB and IV NSCLC under second-line immunotherapy are included. The dataset includes 793 CT scans, 1335 laboratory blood tests, 99 clinical data, and 320 response evaluations as per RECIST1.1 [4]. All data is further categorized into serial data and static data.

The CT scans are labelled by one radiologist with 8 years of experience, by manually segmenting the volume of interest (VOI) of the target lesion in each scan. An oncologist with 30 years of experience reviewed and confirmed the segmentation. CT volumes and segmentation masks are resampled to uniform spacing (1 mm \times 1 mm \times 1 mm), with B-spline interpolation for CT volumes and nearest-neighbor interpolation for VOI masks. We use radiomics features [6] to represent the radiological features due to limited number of samples. With large data available, a fully end-to-end CNN could also be used as the feature extractor. 107 radiomics features are extracted from each VOI using PyRadiomics [8]. Radiomics features are treated as serial data unless there is only one CT examination. The serial blood test features are in 22 dimensions and static clinical information features are in 18 dimensions. All categorical features are encoded in one-hot vectors. Numeric features are normalized by removing the mean and scaling to unit variance to ensure stable training and faster convergence. Intervention information is one-hot encoded (*i.e.*, a binary flag at a time step). Serial radiomics, laboratory blood test and intervention is asynchronous in time.

In our experiments, models are tasked to output binary predictions of R (response) or Non-R (non-response) of each response evaluation, with static data and all serial data before 90 days prior to time of response assessment. We use binary cross entropy (BCE) as the loss function. SimTA model is optimized with Adam optimizer [12]. We split the 99 patients into 3-fold (33 patients in each fold), and perform 3-fold cross validation for evaluating our method. Hyperparameters of model structure and training configuration are chosen using a grid search with bootstrap on the training dataset in each cross validation fold. To verify the effectiveness of SimTA on asynchronous time series, we include LSTM, LSTM(i) and LSTM(s) in our ablation study as comparison. LSTM models of comparable parameters are trained under similar setting.

We further associate model prediction with clinical survival benefits, specifically, overall survival (OS) and progression-free survival (PFS). A cutoff value of 0.5 is used for stratifying patients into high-risk and low-risk groups. The serial data before 90 days prior to time of response assessment is used as input to the trained model. Kaplan-Meier analysis and log-rank test for the survival analysis validate the effectiveness of our method in terms of patient survival.

Table 1. Model performance of on predicting immunotherapy response, including standard RNN approaches (LSTM, LSTM(i), LSTM(s)) instead of the proposed SimTA, and our methods with multi-modal and single-modal inputs.

Methods	LSTM	LSTM(i)	LSTM(s)	Ours	Ours w/o radiomics	Ours w/o lab
AUC	0.71	0.71	0.70	0.80	0.47	0.58

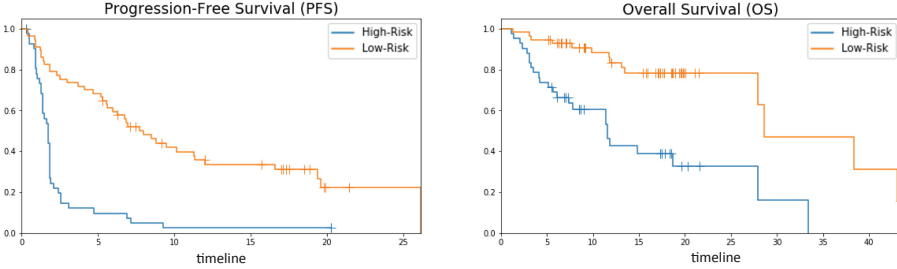


Fig. 3. Model performance on predicting NSCLC patient survival under anti-PD-1 immunotherapy. **Left:** Patient survival curve visualized by Kaplan-Meier (K-M) plot of progression-free survival (PFS), p-value of log-rank test to high/low-risk groups is < 0.01 . **Right:** K-M plot of overall survival (OS), with $p < 0.01$.

Results. As depicted in Table 1, promising results are observed in predicting immunotherapy outcome using the proposed framework. The area under curve (AUC) of receiver operating characteristic (ROC) curve is 0.80 with our *MIA*-Prognosis framework, whereas vanilla LSTM, LSTM(i) and LSTM(s) are achieving 0.71, 0.71 and 0.70 AUC respectively. The LSTM counterparts does not totally fail in this case because the patients are taking CT scans and blood tests on a fairly regular schedule, the interval variance is mostly smaller than seven days. Still, SimTA outperforms LSTM by a large margin in this “mildly” asynchronous serial data modelling task. We also validate the necessity of multi-modal input. Without radiomics feature or laboratory blood test results, our framework reaches very low AUC of 0.47 and 0.58 respectively, suggesting the significance of multi-modal model in this task. Moreover, as shown in Fig. 3, the p-values for Kaplan-Meier analysis are significant in both PFS and OS tests. Therefore, our predictive model could further stratify the low- and high-risk patients in terms of patient survival.

4 Conclusion and Further Work

In this paper, we focus on a unified deep learning framework to predict therapy response, with easily accessible clinical data. The proposed framework named *MIA*-Prognosis utilizes clinical information including Measurement, Intervention and Assessment to model patient prognosis. We also propose a Simple

Temporal Attention (SimTA) module to process the asynchronous time series. The proof-of-concept experiments validate the superiority of SimTA over standard RNN approaches in asynchronous time series analysis. Moreover, our method is proven effective on an in-house dataset on predicting response to anti-PD-1 immunotherapy for real-world non-small cell lung cancer (NSCLC) patients. Importantly, our predictive model is associated with long-term patient survival in terms of progression-free survival (PFS) and overall survival (OS).

In future studies, it is valuable to apply the proposed *MIA-Prognosis* framework on other therapy and diseases. On the other hand, it is also important to design efficient and effective non-linear temporal attention module to enhance temporal relation learning of SimTA. Besides, a fully end-to-end model with CNN-based Radiomics [20] to encode the signature of radiographic features is worth exploring. Furthermore, it is interesting to explain what the *MIA-Prognosis* models from data-driven approaches.

Acknowledgment. This work was supported by National Science Foundation of China (61976137, U1611461). Authors would like to appreciate the Student Innovation Center of SJTU for providing GPUs.

References

1. Binkowski, M., Marti, G., Donnat, P.: Autoregressive convolutional neural networks for asynchronous time series. In: ICML (2018)
2. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**(6), 394–424 (2018)
3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT (2019)
4. Eisenhauer, E.A., et al.: New response evaluation criteria in solid tumours: revised recist guideline. *Eur. J. Cancer* **45**(2), 228–47 (2009)
5. Gibney, G.T., Weiner, L.M., Atkins, M.B.: Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol.* **17**(12), e542–e551 (2016)
6. Gillies, R.J., Kinahan, P.E., Hricak, H.: Radiomics: images are more than pictures, they are data. *Radiology* **278**(2), 563–577 (2016)
7. Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B., Schmidhuber, J.: Lstm: a search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **28**, 2222–2232 (2017)
8. van Griethuysen, J.J., et al.: Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**(21), e104–e107 (2017)
9. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997)
10. Hosny, A., et al.: Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS Med.* **15**(11), e1002711 (2018)
11. Jain, R.K.: Normalizing tumor vasculature with anti-angiogenic therapy: a new paradigm for combination therapy. *Nature Med.* **7**, 987–989 (2001)
12. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
13. Rajkomar, A., et al.: Scalable and accurate deep learning with electronic health records. *NPJ Digit. Med.* **1**(1), 18 (2018)

14. Sacher, A.G., Gandhi, L.: Biomarkers for the clinical use of pd-1/pd-l1 inhibitors in non-small-cell lung cancer: a review. *JAMA Oncol.* **2**(9), 1217–22 (2016)
15. Shaw, P., Uszkoreit, J., Vaswani, A.: Self-attention with relative position representations. In: *NAACL-HLT* (2018)
16. Sun, R., et al.: A radiomics approach to assess tumour-infiltrating cd8 cells and response to anti-pd-1 or anti-pd-l1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol.* **19**(9), 1180–1191 (2018)
17. Vaswani, A., et al.: Attention is all you need. In: *NIPS* (2017)
18. Xu, Y., et al.: Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin. Cancer Res.* **25**(11), 3266–3275 (2019)
19. Yang, J., Deng, H., Huang, X., Ni, B., Xu, Y.: Relational learning between multiple pulmonary nodules via deep set attention transformers. In: *ISBI* (2020)
20. Yang, J., Fang, R., Ni, B., Li, Y., Xu, Y., Li, L.: Probabilistic radiomics: ambiguous diagnosis with controllable shape analysis. In: Shen, D., et al. (eds.) *MICCAI 2019*. LNCS, vol. 11769, pp. 658–666. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_73
21. Yang, J., et al.: Modeling point clouds with self-attention and gumbel subset sampling. In: *CVPR*, pp. 3323–3332 (2019)
22. Yang, Y., Yang, J., Ye, Y., Xia, T., Lu, S.: Development and validation of a deep learning model to assess tumor progression to immunotherapy. *ASCO* **37**, e20601–e20601 (2019)
23. Yu, H.A., et al.: Analysis of tumor specimens at the time of acquired resistance to egfr-tki therapy in 155 patients with egfr-mutant lung cancers. *Clin. Cancer Res.* **19**(8), 2240–2247 (2013)
24. Zhao, W., et al.: Toward automatic prediction of egfr mutation status in pulmonary adenocarcinoma with 3d deep learning. *Cancer Med.* **8**(7), 3532–3543 (2019)
25. Zhao, W., et al.: 3d deep learning from ct scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas. *Cancer Res.* **78**(24), 6881–6889 (2018)