Survival prediction using two-dimensional multi view interim PET images and clinical data in diffuse large B cell lymphoma patients

Dong-Eon Kim
Master's student, Department of Mathematics and Statistics
Chonnam National University,
Gwangju 61186, Republic of Korea
ehddjs0717@gmail.com

So-Yeon Park
Master's student Department of Mathematics and Statistics
Chonnam National University,
Gwangju 61186, Republic of Korea
syeonon1512@gmail.com

Su Hoon Choi
Ph.D. student Department of Mathematics and Statistics
Chonnam National University,
Gwangju 61186, Republic of Korea
magafand@naver.com

Min Soo Kim
Professor, Department of Statistics
Chonnam National University,
Gwangju 61186, Republic of Korea
kimms@channam.ac.kr

ABSTRACT

Diffuse large B-cell lymphoma (DLBCL) is the most common type of non-Hodgkin's lymphoma in Korea and worldwide. This study aims to improve the prognostic performance when using medical imaging data of DLBCL patients compared to using clinical data alone. The analysis target is 449 patients diagnosed with DLBCL at Chonnam National University Hwasun Hospital (CNUHH). Interim F-18 fluorodeoxyglucose positron emission tomography (interim PET) images and clinical data were used as input data for modeling. The interim PET image was reconstructed in the standard uptake value (SUV) method. For analysis, it was converted into a two-dimensional image rotated by 18° through the maximum intensity projection (MIP) method. Image features were extracted through a convolution neural network (CNN) model. Clinical data were divided into Before Tx and During Tx according to the absence or presence of the Deauville score. By combining image features and clinical data, a survival estimation model was used to predict patient survival times. As a result of the analysis, in Before Tx, when images were added, the performance improvement was insufficient. In the During Tx, concordance index (C-index) 0.01 and mean absolute error (MAE) 110 days were improved in a Single view compared to using only clinical data. In Multi view, C-index 0.02, MAE 340 days of improvement. In conclusion, it seems that the presence or absence of the Deauville score variable had a significant effect on the reflection of tumor information in the model.

KEYWORDS

Convolutional neural network, CoxCC, Diffuse large B-cell lymphoma, Positron emission tomography, F-18 fluorodeoxyglucose.

1 INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) is the most common type of non-Hodgkin's lymphoma in Korea and worldwide, accounting for 43% of malignant lymphomas in Korea [1]. DLBCL is a rapidly progressing, aggressive lymphoma. F-18 fluorodeoxyglucose (FDG) positron emission tomography/computed tomography (PET/CT) imaging has been actively used for staging and evaluating the therapeutic response of aggressive lymphomas [2].

There are several predictors for predicting survival in DLBCL patients. Among them, the Deauville score, which is a visual scoring system for FDG PET avidity of lymphoma (Table 1), evaluated during treatment, is being studied usefully [3]. However, there are still limitations in prognostic prediction. Therefore, we introduced a deep learning technique using a two-dimensional MIP image projected on the coronal plane with the maximum value of the PET image. Through this, we wanted to better extract PET image features to improve prediction performance.

We constructed a survival prediction model at different time points; before treatment and during treatment. We compared the prognostic performance of the model using a single view MIP image and the model using multi views in each prediction model.

CNN is excellent at extracting features from images [4]. Therefore, it is expected that the tumor information can be found in the interim PET image containing the tumor information of the patient's whole body and its characteristics can be reflected in the model [5, 6]. We analyze survival time by extracting tumor characteristics from images and combining information such as the patient's stage or individual characteristics. This is to perform a more accurate prognostic analysis than prognostic prediction using only clinical data.

Table 1: Criteria for the Deauville score

score	Criteria
1	no FDG uptake
2	FDG uptake ≤ mediastinum
3	FDG uptake > mediastinal but ≤ liver
4	FDG uptake > liver at any site
5	FDG uptake > liver and new sites of disease
X	new areas of FDG uptake unlikely to be related

2 Related works

In a study related to lung cancer, there was one that used medical imaging data to predict the survival time of patients [7]. Inspired by this paper, we conducted a study to predict the survival time of patients with DLBCL using CNN and a multi-model combining survival estimation model and clinical data. A study was conducted to predict the survival time of DLBCL patients using multiple models combined with data.

3 Proposed Method

3.1 Interim PET Image preprocessing

The analysis data are interim PET images and clinical data of 449 DLBCL patients who visited CNUHH from 2004 to 2019.

All images were sized to the same size as the tallest patient using zero padding. The characteristics of 3D images may have various problems such as difficulty in visualization, increase in learning time, increase in data size, etc., so conversion to 2D is required. Therefore, to reflect the image information as much as possible, a two-dimensional image projected on the coronal plane with the maximum value through the MIP method was used [8]. In addition, when the 3D interim PET image is MIP-converted since it is a projection on the front of the tumor, there is a risk of loss of information about the depth of the tumor and the tumor at the same location. Therefore, the image projected on the coronal plane by rotating the interim PET scan data by 18° was used as input data. And to find out whether the tumor depth information is reflected in the model, it was analyzed by dividing it into a single view and multi view, plotted in Fig. 1. Finally, a uniform distribution of the feature map was induced so that the gradient update process runs smoothly though Min-Max normalization for each image.

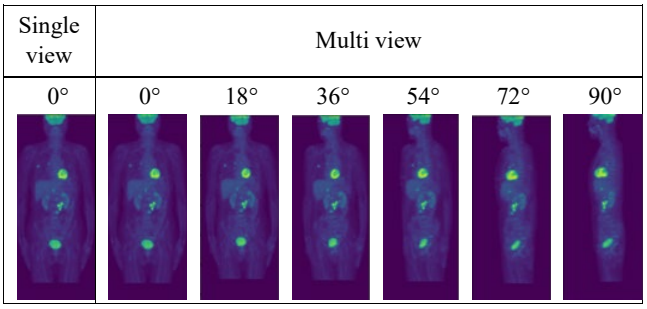

Figure 1: Single view and Multi view image

3.2 Clinical Data

Clinical data including age, sex, ECOG performance score, B-symptom, lactate dehydrogenase (LDH), extranodal involvement status, Ann Arbor stage, bone marrow involvement, revised international prognostic index (R-IPI), International prognostic index score (IPI score), Bulky and Deauville score were used for overall survival prediction. And to find out whether the Deauville score affects the reflection of tumor information in the model, analysis was conducted by dividing it into two models, Before Tx and During Tx. In the Before Tx model, the clinical data were analyzed by removing the Deauville score variable, and in the During Tx model, all clinical data were used.

3.3 Analysis Model

DenseNet, EfficientNet, and ResNet were used as CNN models to extract risk-related features from PET images [9-11]. The image features extracted from the CNN models were combined with clinical data, and the survival time of each patient was predicted using CoxCC, a survival estimation model. As shown in Fig. 2. [12]. Five-fold cross-validation was performed to compare reproducibility and performance. C-index and MAE were used as the performance parameter.

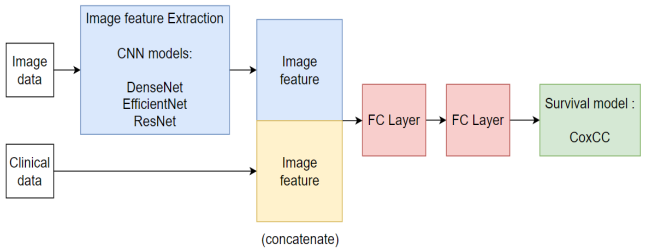

Figure 2: Model structure

4 Analysis Result

4.1 Before Tx

In the Single view (combining single view and clinical data), the best results were obtained when extracting images through EfficientNet, and in Multi view (combining multi view and clinical data), when extracting images through ResNet, the best results

were obtained. The C-index showed almost the same results when analyzing only clinical data (Clinical only, Single view, and Multi view. It improved by 200 days (Table 2).

Table 2: Experiment Results (Before Tx)

Models	MAE	C-index
Single view		
Clinical+DenseNet	656±45	0.6982±0.0051
Clinical+EfficientNet	637±98	0.7085±0.0059
Clinical+ResNet	714±89	0.6813±0.0227
Multi view		
Clinical+DenseNet	473±46	0.6844±0.0220
Clinical+EfficientNet	467±16	0.6907±0.0107
Clinical+ResNet	469±41	0.7099±0.0211
Clinical only	675±37	0.7026±0.0354

4.2 During Tx

In the Single view, the best results were obtained when extracting images through DenseNet, and in multi view, when extracting images through ResNet, the best results were obtained. The C-index increased by 0.015 in Single view and 0.023 in Multi view compared to when only clinical data was used. And MAE improved by 340 days in Multi view (Table 3).

Table 3: Experiment Results (During Tx)

Models	MAE	C-index
Single view		
Clinical+DenseNet	687±72	0.7388±0.0111
Clinical+EfficientNet	740±150	0.7253±0.0196
Clinical+ResNet	749±189	0.7358±0.0143
Multi view		
Clinical+DenseNet	492±44	0.7347±0.0333
Clinical+EfficientNet	467±26	0.7426±0.0286
Clinical+ResNet	459±34	0.7465±0.0173
Clinical only	799±190	0.7233±0.0314

5 CONCLUSIONS

Before Tx, there was no difference in survival prediction performance between single view and multi view. However, during Tx, there was an improvement in C-index and MAE in multi view than in single view. It seems that the model reflects the tumor depth information through the multi-view image when the Deauville score variable is used.

In the pre-treatment stage where the Deauville score cannot be used, the multi-view model does not seem to have much benefit compared to the single view. However, it seems that the multi-view model can be developed and utilized when predicting the middle of treatment by adding the Deauville score.

Future studies aim to better predict the prognosis by further utilizing staging PET data to reflect the attenuation of tumor size or amount in intermediate PET.

ACKNOWLEDGMENTS

This work was supported in part by the Bio & Medical Technology Development Program of the National Research Foundation(NRF) funded by the Korean Government(MSIT) under Grant NRF-2019M3E5D1A02067961, and in part by the BK21 FOUR(Fostering Outstanding Universities for Research, NO.5120200913674) funded by the Ministry of Education(MOE, Korea) and National Research Foundation of Korea(NRF)

REFERENCES

[1] Yi HG, Kim JS, Suh C, Kim WS, Kwak JY, Lee JS, Kim YS, Joo YD, Min YH, Lee HG, Yoon SS, Won JH, Park S, Kim HC, Kim CS. 2013. Clinical features and survival outcomes of patients with diffuse large B-cell lymphoma: analysis of web-based data from the Korean Lymphoma Working Party Registry. Blood Res. 2013 Jun;48(2):115-20. DOI: https://doi.org/10.5045/br.2013.48.2.115. Epub 2013 Jun 25. PMID: 23826580; PMCID: PMC3698396.

[2] Zhang, X., Chen, L., Jiang, H. et al. 2022. A novel analytic approach for outcome prediction in diffuse large B-cell lymphoma by [^{18}F]FDG PET/CT. Eur J Nucl Med Mol Imaging 49, 1298-1310 (2022). https://doi.org/10.1007/s00259-021-05572-0

[3] M. Meignan, A. Gallamini, M. Meignan, A. Gallamini & C. Haioun. 2009. Report on the First International Workshop on interim-PET scan in lymphoma. Leukemia & lymphoma 50.8 (2009), 1257-1260. DOI: https://doi.org/10.1080/10428190903040048

[4] A. Krizhevsky, I. Sutskever & G. E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, 25. DOI: https://doi.org/10.1145/3065386

[5] Burggraaff, C.N., de Jong, A., Hoekstra, O.S. et al. 2019. Predictive value of interim positron emission tomography in diffuse large B-cell lymphoma: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging* 46, 65-79 (2019). https://doi.org/10.1007/s00259-018-4103-3

[6] K. G. Chee, D. V. Nguyen, M. Brown, D. R. Gandara, T. Wun & P. N. Lara. 2008. Positron emission tomography and improved survival in patients with lung cancer: the Will Rogers phenomenon revisited. Archives of Internal Medicine 168.14 (2008). 1541-1549. DOI: https://doi.org/10.1001/archinte.168.14.1541

[7] S. Oh, J. Im, S. -R. Kang, I. -J. Oh & M. -S. Kim. 2021. PET-Based Deep-Learning Model for Predicting Prognosis of Patients With Non-Small Cell Lung Cancer. IEEE Access 9 (2021). 138753-138761. DOI: https://doi.org/10.1109/access.2021.3115486

[8] Eliot K. Fishman, Derek R. Ney, David G. Heath, Frank M. Corl, Karen M. Horton, Pamela T. Johnson. 2006. Volume Rendering versus Maximum Intensity Projection in CT Angiography: What Works Best, When, and Why. *Radiographics* 26.3 (2006): 905-922. DOI: https://doi.org/10.1148/rg.263055186

[9] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang & Y. Kluger. 2018. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC medical research methodology 18.1 (2018), 1-12. DOI: https://doi.org/10.1186/s12874-018-0482-1

[10] M. Tan, Q. Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning. PMLR 97 (2019), 6105-6114. Available from https://proceedings.mlr.press/v97/tan19a.html

[11] K. He, X. Zhang, S. Ren, J. Sun. 2016. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. (2016). pp. 770-778. DOI: https://doi.org/10.1109/cvpr.2016.90

[12] K. Havard, Ø. Borgan, & I. Scheel. 2019. Time-to-event prediction with neural networks and Cox regression. arXiv preprint arXiv:1907.00825 (2019)