令和二年度　筑波大学工学システム学類卒業研究論文

Applying Convolutional Neural Network to Predict Pathology of Postchemotherapy Retroperitoneal Nodal Masses in Germ Cell Tumors

学籍番号　201711099

氏名　岩野　善允

指導教員　掛谷　英紀

Applying Convolutional Neural Network to Predict Pathology of Postchemotherapy Retroperitoneal Nodal Masses in Germ Cell Tumors

胚細胞腫瘍における化学療法後の後腹膜リンパ節腫瘤病理予測への畳み込みニューラルネットワークの応用

Yoshimasa IWANO
岩野　善允
（指導教員 掛谷 英紀）

要約 – 精巣腫瘍の一種である非セミノーマの治療の基本は化学療法と後腹膜リンパ節の残存腫瘤の外科的切除である。残存腫瘤には、奇形腫と壊死組織がある。奇形腫は切除の必要があるが、壊死組織には過剰治療である。従来の精巣腫瘍バイオマーカーでは、奇形腫を同定できず、手術前に病理結果を予測することは困難である。そこで、本研究は、CT 画像を畳み込みニューラルネットワークで学習することで、病理結果を予測することを目的とする。さらに、臨床データによる予測とニューラルネットワークによる予測を組み合わせることで、奇形腫の判定精度向上も試みる。全 65 人の患者から 155 例の関心領域を抽出し、それぞれの病理結果を予測する Resnet50 と呼ばれるニューラルネットワークを構築した。関心領域の CT 画像をスライスしていき、一枚一枚の画像の判定確率を掛け合わせることで最終的な判定を出した。3 分割クロスバリデーションで汎化性能を測定し、既存研究と同等程度あるいはそれ以上の予測精度が得られることを確認した。さらに、画像認識と臨床データを組み合わせた識別も試みたが、大きな結果の改善は見られなかった。

1. Introduction

Although testicular cancer is one of the most common cancers among young adult men, it accounts for only 1% of all cancers in men [1]. However, the number of cases has been increasing in the world.

Testicular cancers can be subdivided into two groups: seminoma and non-seminoma. The method of treatment is determined by the type of testicular cancer (seminoma / non-seminoma) and the presence or absence of lymph-vascular invasion. Radical inguinal orchidectomy is recommended for all patients with suspected testicular cancers to allow accurate diagnosis.

On the other hand, testicular cancer is classified as metastatic disease, and treatment methods differ between seminoma and non-seminoma if a lymph-vascular invasion is located. Seminoma is treated by chemotherapy and radiotherapy and is almost completely cured, but non-seminoma is treated only by chemotherapy and retroperitoneal lymph node dissections (RPLND) for residual tumors.

Residual tumors are categorized into teratoma and necrosis. Teratoma requires RPLND, but necrosis does not. RPLND is overtreatment for patients with necrosis because RPLND is associated with some risks of retrograde ejaculation and infertility. The conventional treatment cannot be differentiated between teratoma and necrosis before RPLND. Therefore, the purpose of this research is to develop a system that brings a more accurate diagnosis by using machine learning of computed tomography (CT) images and clinical data.

One of the prominent methods in machine learning applicable to medical data is neural networks, including deep learning. The concept of neural networks has been studied since the 1960s. From late 1980s to early 1990s, multilayered neural networks with error back-propagation learning [2] were applied in various fields, which soon shrank because of its limited performance.

After the dark age of neural networks, researchers at the University of Toronto made a major breakthrough in image recognition, proposing deep learning, which was applied in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 [3]. The development of deep learning in image recognition has been striking. In the field of segmentation tasks in medical imaging, 2D U-Net [3] and 3D U-Net [4] are most widely used and have achieved high performance.

In the segmentation task of abdominal organs, 3D U-JAPA-Net [5] was proposed to deal with the disadvantage of 3D U-Net, where only local

information is taken into account.

Although much research has been conducted on image recognition and medical imaging so far, only a few studies have applied machine learning to the diagnosis of germ cell tumors because they are rare, while the following two prior studies are available [6][7].

The first research [6] employed a support vector machine with radial basis function on 77 patients' data (ROI = 102), where the accuracy was 72%, sensitivity was 56.2%, and specificity was 81.9%. The second study [7] applied gradient-boosted trees on 80 patients' data (ROI = 204), where the accuracy was 81%, sensitivity was 88%, and specificity was 72%.

In this paper, we propose a new method that predicts the pathology of residual masses after chemotherapy by applying a convolutional neural network (CNN) to CT images. We adopt Resnet50 [8], which has attained high performance in the field of image recognition. We also utilize clinical data, which is gained from the treatment process of testicular cancer.

We obtain CT images of 65 patients with 155 ROIs (teratoma: 63; necrosis: 92) from the department of urology at the University of Tsukuba Hospital, which is a core clinical institute in the area and has abundant cases of RPLND, accepting patients with testicular cancers.

This paper is organized as follows. The following section explains the summary of data preparation and concrete methods used in this research. Section 3 describes the results of experiments in detail. Discussion on the results is given in Section 4 and the paper is concluded in Section 5.

2. Materials and Methods

2.1. Patients

The data set in this study includes 66 patients diagnosed with advanced germ cell tumors, who underwent pc-RPLND (post-chemotherapy RPLND), collected from 2005 to 2019.

We collected clinical data in addition to CT images, including the pathology of primary tumors, tumor marker level, CT images, a maximum diameter of retroperitoneal lymph nodes before and after chemotherapy, and pathological results from pcRPLND (teratoma/necrosis).

A primary tumor is the first lesion to develop a tumor and is one of the most important guidelines in deciding medical treatment. Primary tumors have six types: seminoma, embryonal carcinoma, yolk sac tumor, chorionic carcinoma, teratoma, and burned out.

Similarly, tumor markers are lactate dehydrogenase (LDH), α-fetoprotein (AFP), and human chorionic gonadotropin (HCG). Tumor markers are substances contained in the blood and generated by cells that are associated with the presence of a particular tumor.

2.2. Data Arrangement

CT images have three-dimensional information. However, we slice CT images in the axial direction and treat them as two-dimensional images so that the number of parameters may not increase when CT images are applied to the CNN. We extract tumor images from each slice according to the labeled images, which indicate the region of lymph nodal masses. Note that each region of interest (ROI) has multiple 2D image slices. Moreover, some patients have multiple ROIs. Accordingly, we utilize 155 ROIs with 5372 CT images.

The problem is that we need a constant image size to be applied to the CNN despite the variety of tumor sizes (Fig. 1). In preliminary experiments, the image size of 100 × 100 produces high performance, containing the entire region of most tumors. In addition, the images fed to the network are preprocessed so that the center of the region of interest may be located in the center of images.

The range of HU in CT images is wide enough to lose decisive and distinct texture, making images unclear. Clipping the range of HU is one of the most crucial problems when dealing with medical images for machine learning, because the distribution of HU is different depending on the part of body. We need to determine the best range of HU for the task. The distribution of HU in tumor images is shown in Fig 2. We used the range of [-150, 150] in this study

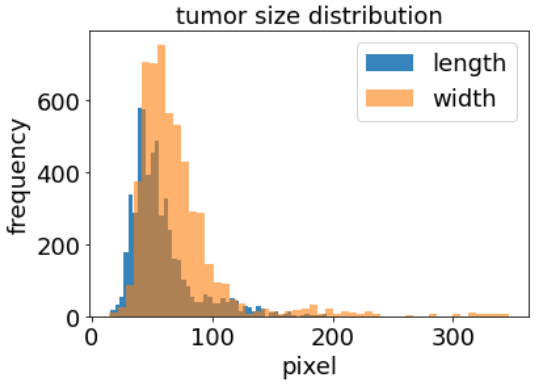

Fig 1: Tumor size distribution.

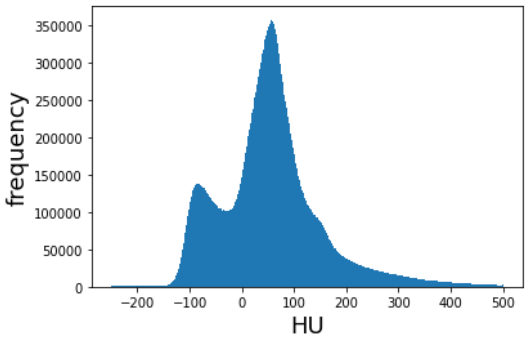

Fig 2: HU distribution.

2.3. Resnet

Resnet was proposed at ILSVRC 2015 [8] and won the 1st place on the classification task. The most distinguished feature of Resnet is its residual networks, which enable deeper convolutional neural networks and easier optimization. Resnet has been applied to many tasks of medical imaging [9][10]. In this research, we use Resnet for this reason.
for this reason.

As a preliminary experiment, we searched for the best Resnet model in this classification problem. To be concrete we tried three types of Resnet. The results of Resnet34, Resnet50, and Resnet101 are shown in Fig 3. The accuracy and standard deviation of Resnet34 is 77.5 ± 2.0 %, those of Resnet50 is 78.1 ± 2.5%, and those of Resnet101 is 77.9 ± 1.6%. We decided to apply Resnet50 to this research from the results of the trial.

The simple structure of Resnet50 is shown below (Fig 4). In the original paper, the output of the fully connected layer (fc layer) is 1000, which means the CNN classifies 1000 kinds of objects. In this research, however, the output of the fc layer is two because the purpose of this study is to classify teratoma and necrosis. Finally, the prediction probability of whether the ROI is teratoma or necrosis is calculated through softmax function.

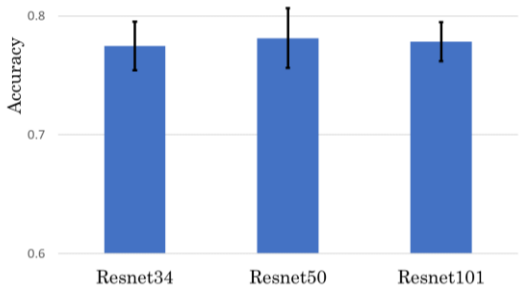

Fig 3: Comparison of Resnet.

2.4. Judgment Criteria

Resnet50 determines the classification of tumor images on each slice. Thus, we need to aggregate predictions of all images included in the ROI. The final decision of ROI is conducted as follows:

$$Y_i = \sum_{j=1}^{N} \log X_{i,j}$$
$$(i = 0 : teratoma, 1 : necrosis).$$

Here X is a vector of the slice prediction probability and Y is a vector of the final decision criterion; N is the number of slices in the ROI. Final decision is given by comparing the values of Y_0 and Y_1.

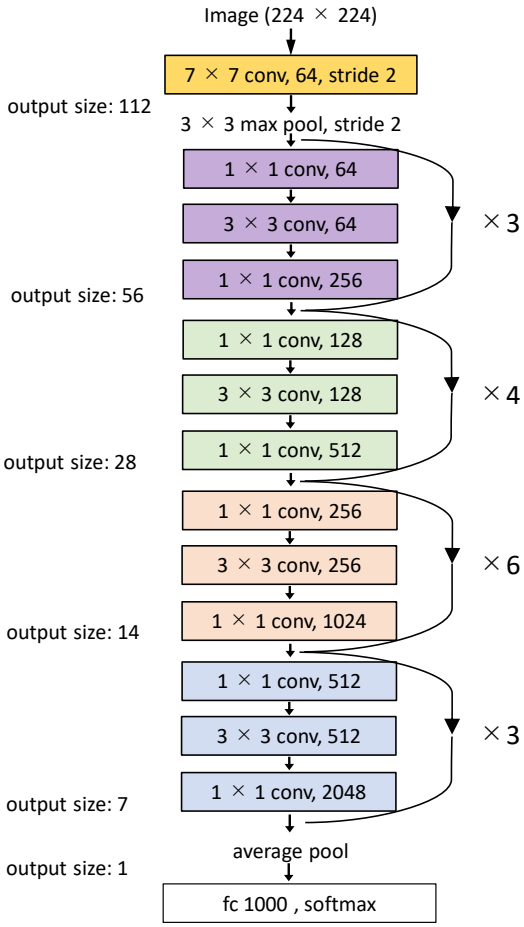

Fig 4: Resnet50.

We evaluate model performance by accuracy, sensitivity, and specificity based on ROI, using a confusion matrix (Table 1). The accuracy, sensitivity, and specificity are given by

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN'}$$

$$sensitivity = \frac{TP}{TP + FN},$$

$$specificity = \frac{TN}{TN + FP},$$

respectively.

Table 1: Confusion matrix.

		Prediction result: Positive	Prediction result: Negative
Label	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Accuracy is the ratio of data that can be predicted correctly, which measures the validity as a whole. In this study, however, the proportion of necrosis is 93/155 (59%), and that of teratoma is 63/155 (41%), where accuracy is not enough to evaluate the model performance, for the accuracy of a wild model would be 59 % even if predicted all ROI as necrosis. That is why we consider sensitivity and specificity. Sensitivity is the ratio of positive data that can be predicted correctly, which excludes the outcome of negative data prediction. On the other hand, specificity is the ratio of negative data that can be predicted correctly, which excludes the outcome of positive data prediction. We can evaluate the model performance appropriately by considering these indices.

2.5. Cross Validation

Cross validation is a statistical method to evaluate the versatility of the model, splitting data into train and test. K-fold cross validation is common in machine learning because it can be applied to both classification and regression problems.

Stratified 3-fold cross validation was used to enhance generalization performance. A restricted data prevents 10-fold cross validation from ensuring enough validation data, so we employed 3-fold cross validation to this research (Fig 5). Furthermore, the pathological result of residual tumors has an imbalanced distribution (Fig 6). We applied stratified cross validation to this research so that each validation set may have a similar proportion of tumor features as shown in Fig. 5.

2.6. Clinical Data

Through the medical treatment of testicular cancer, clinical data plays an important role in estimating the existence of metastatic cancers and the progress of the lymph-vascular invasion. We conjecture that clinical data will contribute to the improvement of model accuracy. Thus, we concatenate the feature values extracted from CT images and those extended from clinical variables (Fig 7).

We create 50 feature values from clinical variables by utilizing the 50-fc layer and then concatenate these features with the Resnet50 output before pouring them into the final fc layer.

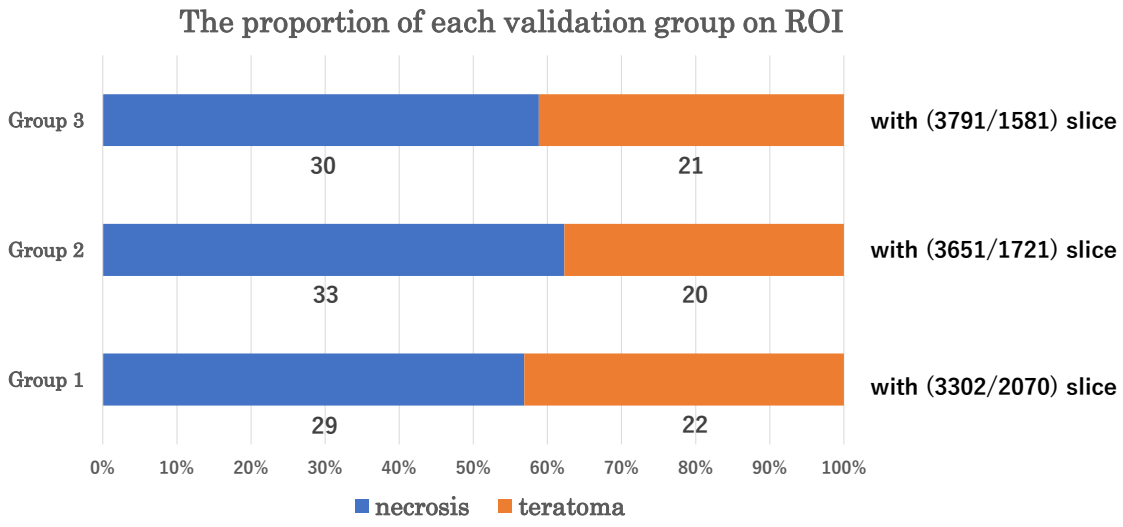

Fig 5: Cross Validation

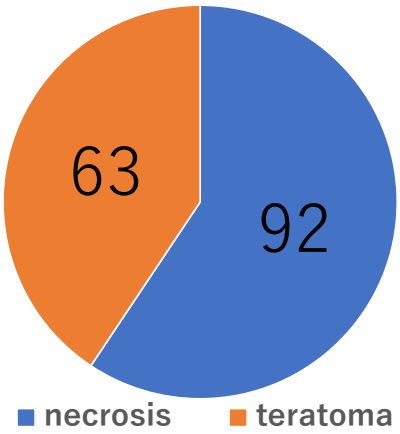

Fig 6: Tumor feature distribution.

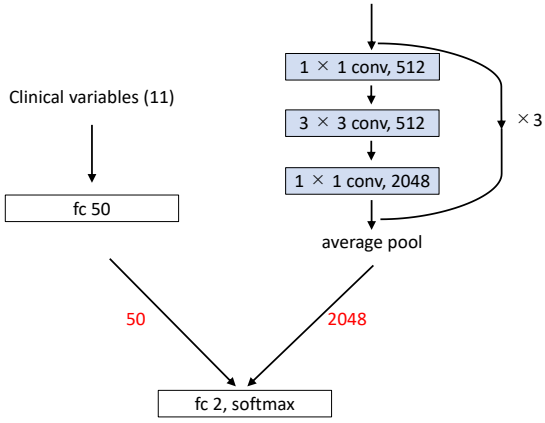

Fig 7: Concatenate Resnet50 + clinical variables.

3. Results

3.1. Configuration

The parameters including batch size, epoch number, loss function, and optimizer to complete the process of classifier used in Resnet50 in this study is shown in Table 2.

Table 2: Configuration of Resnet.

Model	Resnet50
Batch Size	32
Loss Function	Cross Entropy Loss
Optimizer	Stochastic Gradient Descent(SGD)
Epoch	20

Firstly, the batch size and the epoch number are subject to the amount of data in general. Conventionally, the batch size is a power of 2. The number epochs is fixed to 20 after some trials, where the loss function converges and the overfitting does not occur for this task.

Secondly, we adopt the cross-entropy loss, which is a loss function for classification problems. The cross-entropy loss is calculated from the prediction probability of the true label:

$$Loss = -\sum_{i=0}^{1} t_i \log y_i$$

Here t is a classification label (teratoma/necrosis), and y is its probability.

As in the original paper on Resnet, we use stochastic gradient descent (SGD) with a learning rate of 0.001 and a momentum of 0.9 as an optimizer.

Cross-entropy loss and SGD are often used together because of their good compatibility.

3.2. Image Recognition

We built a classifier only based on CT images and set a configuration of Resnet50 as mentioned above. We obtained the following results.

The classifier achieved a mean accuracy of 78.1 ± 2.5% (standard deviation) in 10 repetitions of the stratified 3-fold cross validation, which corresponds to a sensitivity of 61.3 ± 5.4%, specificity is 89.7 ± 3.6% (Figs. 8, 9, and 10).

3.3. Image and Clinical Data

We built a classifier based on CT images and clinical data as explained in Subsection 2.6 and obtained the following results.

The classifier achieved a mean accuracy of 78.3 ± 1.7% in 10 repetitions of the stratified 3-fold cross validation, which corresponds to a sensitivity of 63.0 ± 4.2%, specificity is 88.8 ± 3.7% (Figs. 8, 9, and 10).

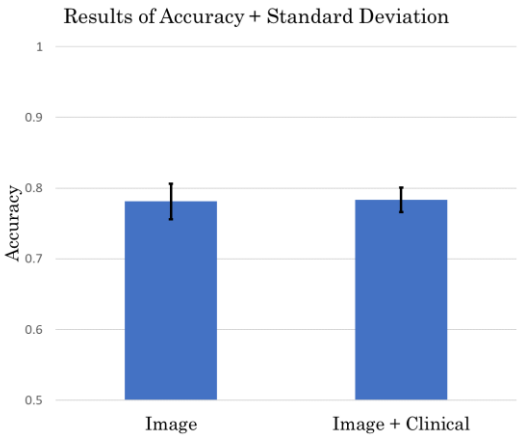

Fig 8: Results of accuracy.

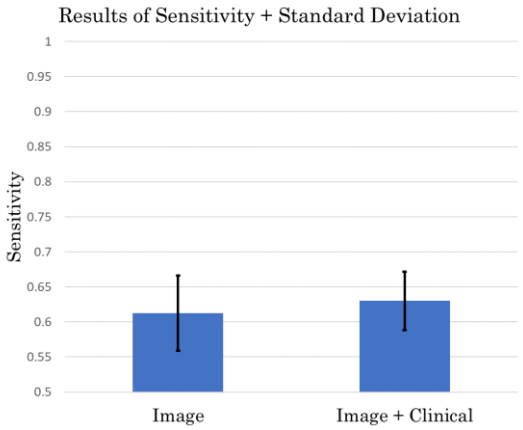

Fig 9: Results of sensitivity.

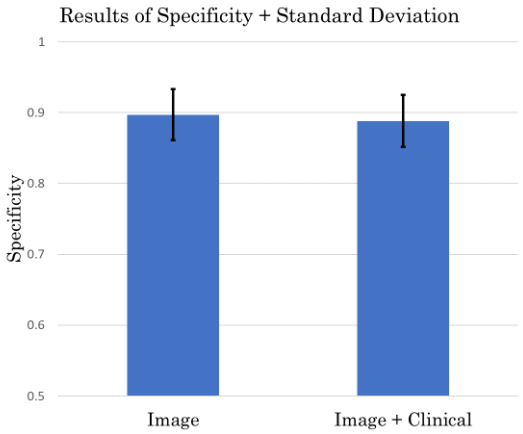

Fig 10: Results of specificity.

The confusion matrix given by machine learning only with CT image and that given by machine learning with image and clinical data combined are shown in Tables 2 and 3 respectively, where the averages of cases after 10 trials are shown.

Table 2. Confusion matrix given by training with CT image only.

		Prediction	
		teratoma	necrosis
Label	teratoma	38.6	24.4
	necrosis	9.5	82.5

Table 3. Confusion matrix given by training with CT image and clinical data.

		Prediction	
		teratoma	necrosis
Label	teratoma	39.8	23.2
	necrosis	10.3	81.7

4. Discussion

When the results given by two classifiers are compared, the accuracies are almost the same, which means that the clinical data do not contribute to attain higher accuracy.

As far as sensitivity is concerned, however, it has a tendency to increase slightly by adding clinical data to the classifier, although it is the lowest of the three indices (accuracy, sensitivity, and specificity) in this experiment. The aim of this research is to identify the presence of teratoma, which is a malignant tumor and needs pcRPLND. The improvement of sensitivity matches with the goal to avoid unnecessary surgeries or pcRPLND.

One of the possible reasons why clinical data do not contribute to higher accuracy is the disparity in the number of features between image and clinical data (2048 and 50). It is hard to reduce the number of image features to be fed to the fully connected multilayer neural network.

In this paper, we concatenate features of image and clinical data as the input to the multilayer neural network. Instead, we can build independent two models for image and clinical data and give the final decision based on the two outputs.

To improve the results, further study is needed. To begin with, we need to investigate a tendency of accuracy according to the lymph node size since smaller tumor images include more information surrounding the tumor, while the larger ones lose information of the tumor edge. In the prior research [6], the accuracy of classification was higher when different classifiers were applied depending on the size of ROI, though the amount of training data is decreased within each data set.

The cases of testicular cancers are rare among adult men, and available data of patients with lymph-vascular invasion and pcRPLND are limited. Data augmentation is one approach to tackle this problem.

Also, fine-tuning or transfer learning are effective methods because these methods supplement data scarcity by transferring the model or the parameters of the model trained with a larger dataset, including other cancer data and semantic segmentation.

Another possible approach is use of 3D CNN instead of Resnet50 to capture the 3D structure of CT images. Resnet50 is compatible with 2D images and dismiss the vertical relevance of each slice. The location information and shape of residual tumors are relevant to the pathological results.

5. Summary

We have applied Resnet50 to predict the pathology of retroperitoneal nodal masses in germ

cell tumors after chemotherapy based on the CT images of patients. We have also tried to utilize clinical data, which are obtained from the treatment process of testicular cancer, combined with the CT image data. With cross-validation trials, we have achieved a mean accuracy of 78.3%, a mean sensitivity of 63.0%, and a mean specificity of 88.8%. Further research is needed to increase the accuracy of prediction.

Acknowledgment

I express my deepest gratitude to all the people who have supported me in many ways. First and foremost, I would like to express my sincere thanks for Prof. Kakeya's support. He has been mentoring me for a year and has shared his views not only about the expertise, but also a wide range of knowledge acquired through his career and life. I also appreciate Dr. Nitta and Dr. Okada for supporting me to learn the background of this study and to acquire the skills of programming needed to deal with medical images. Besides, I would like to thank all members of Visual Media Laboratory, the University of Tsukuba, for helping me a lot. Best of all, I appreciate all members of my family, who have raised me since my birth and have supported my 4-year campus life. I would like to thank you all involved in my graduate research.

Preference

[1] Omar Khan, Andrew Protheroe: Testis cancer. Postgraduate medical journal, 2007

[2] David E. Rumelhart, Geoffrey E. Hinton, Ronald J. Williams.: Learning representations by backpropagating errors. Nature 323(6088), 533–536, 1986.

[3] Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton: Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems,* pp. 1097-1105, 2012.

[4] Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer Assisted Intervention. Springer International Publishing, 2015.

[5] Özgün Çiçek and Ahmed Abdulkadir and Soeren S. Lienkamp and Thomas Brox and Olaf Ronneberger (2016) 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin S., Joskowicz L., Sabuncu M., Unal G., Wells W. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, vol 9901. Springer, Cham. https://doi.org/10.1007/978-3-319-46723-8_49

[6] Hideki Kakeya, Toshiyuki Okada, Yukio Oshiro. 3D U-JAPA-Net: Mixture of Convolutional Networks for Abdominal Multi-organ CT Segmentation. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018

[7] Jeremy Lewin, Padraig Warde, Peter Chung, Lynn Anson Cartwright, Joan Sweet, Aaron R. Hansen, and Philippe L. Bedard, Michael A.S. Jewett, Robert J. Hamilton, Madhur Nayan, Ricardo Leão, Aaron R. Hansen, and Philippe L. Bedard, Paul Dufort, Jaydeep Halankar, Martin O'Malley, and Ur Metser, and Abha Gupta, Armando Lorenzo, and Jeffrey Traubici: Applying Radiomics to Predict Pathology of Postchemotherapy Retroperitoneal Nodal Masses in Germ Cell Tumors. JCO Clin Cancer Inform. 2018; 2:1-12

[8] Bettina Baessler, Tim Nestler, Pia paffenholz, David Pfister, and Axel Heidenreich, Daniel Pinto Dos Santos, Vikram Zeuch: Radiomics allows for detection of benign and malignant histopathology in patients with metastatic testicular germ cell tumors prior to postchemotherapy retroperitoneal lymph node dissection. Eur Radiol. 2020; 30:2334-2345

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016; pp. 770.

[10] Qiao Zhang, Zhipeng Cui, Xiaoguang Niu, Shijie Geng , Yu Qiao : (2017) Image Segmentation with Pyramid Dilated Convolution Based on ResNet and U-Net. In: Liu D., Xie S., Li Y., Zhao D., El-Alfy ES. (eds) Neural Information Processing. ICONIP 2017. Lecture Notes in Computer Science, vol 10635. Springer, Cham. https://doi.org/10.1007/978-3-319-70096-0_38

[11] A. Sai Bharadwaj Reddy and D. Sujitha Juliet, Karunya: Transfer Learning with

ResNet-50 for Malaria Cell-Image Classification. 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0945-0949, doi: 10.1109/ICCSP.2019.8697909.