

# 学位論文の要旨

氏名

和田 清隆

学位論文題目

ディープラーニングを用いた放射線皮膚炎グレード判定システムに関する研究

本論文は、放射線治療を受ける患者の90%以上に発症する放射線皮膚炎の有害事象共通用語規準(Common Terminology Criteria for Adverse Events: CTCAE)に基づくグレード判定について、ディープラーニング (Deep convolutional neural network: DCNN) を用いた放射線皮膚炎グレード判定システム開発研究をまとめたものである。

第1章は、がん治療における放射線治療の有害事象について、その放射線皮膚炎の評価は人間の判定で生じる曖昧な基準や個人差によって生じる不確定なグレード判定であることを述べた。先行研究では、大菌らの評価統一を目的に人間が学習する手法、Zendaらの視覚的評価基準となるグレーディングアトラスを作成した手法が報告されている。しかし、人間の評価は継続した学習が必要であり、グレーディングアトラス作成後の評価は行われてないため、評価統一への課題は残されたままであると述べた。

第2章では、近年、医用画像分野でも応用されているディープラーニングを用いた先行研究について述べた。Fujisawaらは、14種類の皮膚腫瘍分類にDCNNを用いた結果、識別率は76.5%であったと報告している。放射線皮膚炎のグレード判定にDCNNを用いたシステム開発の必要性とDCNNを用いたシステムの実現に4つの課題を提示した。

第3章では、データセットを工夫して従来のデータ拡張手法であるData Augmentation (DA) と人工症例画像を混合したハイブリッド生成によるディープラーニングを用いた放射線皮膚炎グレード判定システム (Hybrid generation method Radiation dermatitis grading support system: Hyb-RDGS) を作成した。課題に対して、放射線皮膚炎画像選定プロトコルを策定し、DCNNの学習画像の品質を向上した。

さらに収集したデータ数の不均衡の解消のため、アンダーサンプリング、データ拡張法

によるオーバーサンプリングを実施した。極少数データであるグレード4の症例検知には、ポアソン合成を用いた人工症例画像を生成した。これらの成果より、Hyb-RDGSの正答率は、85.1%を達成し混同行列を用いた性能評価は、グレード4の判定では、感度93.3%、適合率84.7%、F1値88.5%であった。これより、ハイブリッド生成法が有効であることを実証した。しかし、本章では学習画像から除外しているため、曖昧な判定を受けた複数（グレード）判定画像に対する判定手法は、開発されていない。

第4章では、この課題に対して複数の重み付けの異なるモデルにベイズ推定を適用して、最終グレード判定を出力するシステム開発を行った。さらに、Hyb-RDGSの精度を実証した上で効率的な学習モデルとされているEfficientNetモデルの適応について、精度評価を行った結果について述べた。EfficientNetモデルは、従来のニューラルネットワークで行う手動チューニングに代わり、複合スケール法を特徴とする効率的なモデルである。データ拡張にEfficientNetモデルで使用実績のあるRA法を採用し、効率性の高いシステム実現を目指した。解像度とデータ拡張手法の工夫により、最適なモデル作成検証を実施した。その結果、EfficientNet-B4、384画素（データセット：DA）のモデルにおいて、正答率は86.4%を達成した。また、課題の一つであるグレード4の検知における評価（confusion matrix）は、sensitivity ( $p < 0.05$ )、F1-measure (n. s.) は、低下したが、precisionにおいてはHyb-RDGSを有意に上回る性能であった ( $p < 0.05$ )。これより、第3章で作成したHyb-RDGSを上回る性能を持つ効率的な放射線皮膚炎のグレード判定システムの可能性を実証した。さらに、複数（グレード）判定画像に対して、複数の高精度EfficientNetモデルを選定し、ベイズ推定により最終判定結果を提案できることを述べた。評価者とシステムの評価結果の関係について分析し、評価者の半数以上において95%で一致する結果を得た。また、評価者の判定傾向が新たな知見として得られた。

第5章では、放射線皮膚炎の評価統一目的および類似するDCNNを用いた病変識別の先行研究と比較し、本研究で作成したシステムを考察し、今後の課題について述べた。放射線皮膚炎の評価統一目的における研究は、それぞれのアプローチは異なるが、先行研究は人間が行う評価に変わりはなく、主観的評価のままである。これに対して、我々の作成した放射線皮膚炎グレーディングシステムは、客観的評価であるため個人差のない評価が可

能であることを示した。また、DCNNを用いた先行研究に対して、分類クラスは少ないものの先行研究を上回る正答率であったことを示した。

第6章では、本研究で開発した放射線皮膚炎グレード判定システムは、放射線皮膚炎のグレード判定に有効なシステム実現の可能性を示した。さらに、放射線皮膚炎の評価目的だけでなく、様々な病変診断にも応用可能であることを述べた。本研究で開発したEfficientNetモデルを用いた放射線皮膚炎グレード判定システムは、本研究の成果として評価者の判定を補助するセカンドオピニオン、サードオピニオンとなる可能性となることを総括した。

# Summary of Doctoral Dissertation

Title of Doctoral Dissertation:

Study on Radiation Dermatitis Grading System using Deep Learning

Name: WADA Kiyotaka

This study summarizes the development of a radiation dermatitis grading system using deep convolutional neural network (DCNN) for Common Terminology Criteria for Adverse Events (CTCAE) that occurs in more than 90% of patients receiving radiation therapy.

In Chapter 1, we discussed that for adverse events in the case of radiation therapy for cancer treatment, the evaluation of radiation dermatitis is an uncertain grading assessment owing to ambiguous criteria and individual differences caused by human evaluations. Previous studies have reported a method for learning to unify the evaluation of Ozono et al. and create a grading atlas as a visual evaluation standard for Zenda et al. However, human evaluations require continuous learning and evaluations have not been performed since the creation of a grading truss; hence, the issue of unifying evaluations persists.

In Chapter 2, we described previous research based on deep learning, which has been employed in the field of medical imaging recently. Fujisawa et al. achieved a discrimination rate of 76.5% using a DCNN to classify 14 types of skin tumors. We discussed the importance of developing a system using the DCNN for grading radiation dermatitis and four issues associated with realizing a system using the DCNN.

In Chapter 3, we created a data set and hybrid generation method radiation dermatitis grading support system (Hyb-RDGS) using deep learning by employing the hybrid generation method, which is a mixture of data augmentation (DA), a conventional data expansion method, and artificial case images. In response to the problem, we developed a radiation dermatitis image selection protocol and improved the quality of DCNN learning images. Furthermore, undersampling and oversampling using the data expansion method were performed to eliminate the imbalance in the number of collected data. Artificial case images using Poisson image editing were generated for grade 4 case detection, which comprises a very small number of data. Based on these results, the overall accuracy of Hyb-RDGS was determined to be 85.1%. Moreover, the *sensitivity*, *precision*, and *F1-measure* were determined to be 93.3%, 84.7%, and 88.5% in the grade 4 assessment based on the performance evaluation using the confusion matrix. Therefore, the hybrid generation method is effective. However, because this chapter excludes it from the training image, a judgment method for multiple (grade) assessment images that have received ambiguous judgment has not been developed.

In Chapter 4, we developed a system for the final grade assessment by applying the Bayesian inference to multiple models with different weights. Furthermore, the results of evaluating the accuracy of the adaptation of EfficientNet, which is considered an efficient learning model, are described. The EfficientNet model is a model that uses a composite scaling model instead of manual tuning performed using a conventional neural network. For data expansion, we used the Rand augmentation method, which has been used in the EfficientNet model to realize a highly efficient system. The resolution and data expansion method was used to perform the optimal model creation verification. Therefore, the correct answer rate was 86.4% in the case of EfficientNet-B4, comprising a model with 384 pixels (data set: DA). Moreover, the evaluation (confusion matrix) in grade 4 detection, which is an issue, *sensitivity* ( $p < 0.05$ ), and *F1-measure* (n.s.) decreased; however, in *precision*, he significantly exceeded Hyb-RDGS ( $p < 0.05$ ). Therefore, we demonstrated the possibility of an efficient radiation dermatitis grading system that outperforms the Hyb-RDGS developed in Chapter 3. Furthermore, multiple high-precision EfficientNet models can be selected for multiple (grade) assessment images and the final assessment results can be proposed using the Bayesian estimation. The relationship between the evaluator and the system's evaluation result was analyzed, and more than half of the evaluators obtained a 95% consensus result. Additionally, the evalu-

ator's assessment tendency was reported as a new finding.

In Chapter 5, we compared the unified purpose of evaluating radiation dermatitis to a previous study of lesion identification using a similar DCNN by considering the system developed in this study and described future issues. Although each approach is different in research for unification purposes, the evaluation performed by humans in previous research remains the same as the subjective evaluation. However, the proposed radiation dermatitis grading system affords an objective evaluation, allowing for an evaluation without individual differences. Moreover, the overall accuracy obtained herein was higher than that obtained in the previous study, although the classification class was smaller than that of the previous study using the DCNN.

In Chapter 6, the radiation dermatitis grading system using DCNN developed in this study showed the possibility of realizing an effective system for radiation dermatitis grade determination. Furthermore, this grading system can be applied not only to the evaluation of radiation dermatitis but also to the diagnosing various lesions. The radiation dermatitis grading system based on the EfficientNet model developed in this study can become the second or third opinion for assisting the evaluator's assessments.