様式　８

学位論文内容の要旨

博士の専攻分野の名称：博士（保健科学）　　　　氏名：張　洪健

学位論文題名

Study of Transfer Learning on medical information processing by explainable artificial intelligence

（説明可能な人工知能による医療情報の転移学習に関する研究）

In recent years, with the development of artificial intelligence (AI) technology, and computer hardware advances, AI technology is increasingly used in the medical field to assist diagnosis and treatment, such as medical image recognition, optimal treatment selection and other work. And the deep learning method has demonstrated a level equivalent to or even exceeding the human level. However, at the same time, the black-box characteristics of deep learning-based AI methods make them face ethical and practical challenges in the medical field, which limits the further application and development of AI in the medical field. As a result, the concept of Explainable AI (XAI) has been proposed to increase the trustworthiness of AI and enable better application in fields that require high trustworthiness, such as healthcare. Compared to general text, medical text has features such as mass terminology and contains structured text data and free-text data. In the medical text processing field, AI is mainly used to deal with the tasks of text classification, namely entity recognition and relationship extraction, for extracting useful information and advanced analysis. However black-box design presents an obstacle to validating developed AI algorithms. It is necessary to demonstrate that a high-performance deep learning model actually recognizes the appropriate regions of an image and does not overemphasize unimportant findings. XAI aims to take the basis of insights into the output results given conclusions drawn by models and present them in a way that is understandable to humans. The purpose of this study is to use the XAI method to give explainability to high-precision deep learning models applied to medical texts, in order to use high-performance models for medical texts with explainablity, so that physicians can be informed of the basis for the output results given by the AI models.

In Chapter 3 “Pre-trained ResNet model transfer learning on medical text classification”, ResNet , a high-precision model commonly used for image recognition, transfer learning to the task of medical text classification. In this study, free-text medical texts are word-embedded (converting words into vectors), and each medical texts are preprocessed so that they can be treated as image format (25×25 pixels, 100channels). Based on it various types of CNN methods (commonly used for image recognition) are applied for comparison of classification accuracies, while traditional text classification methods are directly applied to the original medical text, as well as using a 1Dimension CNN model applied to preprocessed 1D text (1×625 pixels,100 channels). The performance of each of these models was compared.

The result was that the best performance was achieved using ResNet applied to the preprocessed text data (weighted recall 90.9%, weighted precision 91.1%, weighted F1 score 90.2%). In addition, comparison was made which are about the effect of whether the model use the pre-training parameters and whether retrained on the current dataset. And the result was that the pre-training parameters and retraining had a significant impact on the performance improvement.

In Chapter 4 “Apply Grad-CAM on text classification for visualization of explainability”, based on the results in Chapter 3 to make the models be explainable. XAI methods usually used for text processing, such as knowledge graphs and rule bases, are able to explain in detail the results given by AI models or even the process of making the decision , but their construction and maintenance require significant labor costs. In contrast, post-hoc XAI methods usually used on image processing such as the Grad-CAM are more generalizable and more intuitive to visualize. This study aims to make the models be explainable without additional modification of model structures and extra database building, by attaching the CNN models with pos-hoc XAI methods. As the result, heat map was generated to display the areas that models paid attention to. Combining the heat map with the original texts , the text paid attention by models was shown highlighted. And it makes the attention which AI models paid on during giving the output results, allows physicians to understand the basis of the model’s conclusions visually and directly.

In this study, though word embedding and convert medical text into an image-like format, a high performance AI model commonly used in image processing transfer learning on the task of medical text processing was developed. Then comparing its performance with that of applying traditional text classifier methods directly to the text format and applying a 1-dimensional CNN for classifying. And the XAI method is attached to the models to realize the explainability of the model output results. While achieving higher accuracy on medical text classification task, the model has more intuitive and simpler explainability than traditional text XAI methods. The limitation of this study is that the explainability provided is relatively qualitative; the physicians are able to be informed of the basis for the results given by the AI model, however, it is still up to the physicians to make them own judgment as to whether or not that basis is reasonable and whether or not the results should be adopted.