OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Automated classification of clinical T staging pulmonary oncology from chest CT reports: a natural language processing framework for real world data utilization
Author(s)	西嶌, 伸郎
Citation	大阪大学, 2026, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/107028
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

論　文　内　容　の　要　旨
Synopsis of Thesis

氏　　名 Name	西嶌　伸郎
論文題名 Title	Automated classification of clinical T staging pulmonary oncology from chest CT reports: a natural language processing framework for real world data utilization （胸部CTレポートからの肺癌臨床的T分類自動推定：リアルワールドデータ利活用を目指した自然言語処理フレームワーク）

論文内容の要旨

〔目　的(Objective)〕

臨床的TNM分類（cTNM）は予後や治療効果など癌評価に必要な情報であり、リアルワールドデータ（RWD）として、重要な患者背景情報である。しかしながら、cTNMはしばしば画像診断レポートに直接の記載がない、あるいはフリーテキストとして関連する情報が記述されるのみであり、二次利用が困難なことが多い。そのため、専門医が目視でstagingに必要な情報を抽出の上で分類する必要があり、その労力が課題となる。この課題を克服するため、我々は、肺癌患者の胸部CTレポートを対象とし、臨床的T分類（cT）を推定する自然言語処理（NLP）システムを開発した。

〔方法ならびに成績(Methods/Results)〕

本システムは、深層学習による言語理解と肺癌取り扱い規約第9版に基づくルールベースを組み合わせたものである。本システムは、最初に胸部CTレポートからcT分類に必要な情報が記載されるレポートを抽出し、次にcT分類の推定を行う。大阪大学医学部附属病院で肺癌と診断され、がん登録された患者の胸部CTレポートを対象に、284件の訓練セット（2019年1月から2020年12月に撮像）と165件のテストセット（2021年1月から2021年12月に撮像）に分割した。cT分類に必要な情報が記載されるレポートの割合は訓練セットとテストセットでそれぞれ83%、85％であり、加重F1値として0.96、0.93の性能で抽出された。cT分類に必要な情報が記載されると正しく判定されたレポートに対し、cT分類推定の加重F1値は、訓練セット、テストセットそれぞれで0.97、0.95であった。

〔総　括(Conclusion)〕

cT分類に必要な情報が記載されるレポートを抽出し、cT分類の推定を行う深層学習とルールベースを組み合わせた自然言語処理システムを開発した。本システムが優れた性能を示したことから、cT分類を解析対象としたRWD研究への応用が期待される。

論文審査の結果の要旨及び担当者

（申請者氏名）　西嶌　伸郎

	（職）	氏　　名	
論文審査担当者	主　査　大阪大学教授	武田 理宏	署名
	副　査　大阪大学教授	川崎 良	署名
	副　査　大阪大学教授	新谷 康	署名

論文審査の結果の要旨

臨床的TNM分類は予後や治療効果など癌評価に使用される重要な患者背景情報であるが、リアルワールドデータ（RWD）では、フリーテキストから専門医が目視でstagingに必要な情報を抽出する必要があり、利活用に課題がある。そこで、肺癌患者の胸部CTレポートを対象とし、臨床的T分類（cT）を推定する自然言語処理システムを開発した。本システムは、深層学習による言語理解と肺癌取り扱い規約第9版に基づくルールベースを組み合わせたもので、最初に胸部CTレポートからcT分類に必要な情報が記載されるレポートを抽出し、次にcT分類の推定を行う。165件のテストセットのうち、cT分類に必要な情報が記載されるレポートの割合は85％であり、加重F1値として0.93の性能で抽出された。cT分類に必要な情報が記載されると正しく判定されたレポートに対し、cT分類推定の加重F1値は、テストセットで0.95であった。優れた性能を示した本システムは、cT分類を解析対象としたRWD研究への応用が期待され、この成果は学位に値すると考える。