

# 令和元年度 情報工学コース卒業研究報告要旨

森 研究室	氏 名	鈴 村 悠 輝
卒業研究題目	読影レポート解析を利用した 医用画像解析 AI 学習法に関する初期検討	

本論文では、読影レポート解析を利用した医用画像解析 AI 学習法について述べる。近年、CT や MRI を用いた 1 日あたりの検査件数は増加している。それにより医師が読影する画像数は増加しており、医師の読影負担は大きい。厚生労働省の NDB (National Database) の第 1 回および第 4 回 NDB オープンデータの画像診断項目によると、CT 撮影の総検査実施件数は平成 26 年度の約 190 万件、平成 29 年度は約 200 万件となっている。この問題を解決するために、コンピュータ支援診断 (CAD) システムの研究が多くなされている。CAD システム作成には、近年様々な画像認識や画像解析のタスクで高い精度を出しているディープラーニング (DL) をはじめとした機械学習が用いられている。しかし、DL などの機械学習においてアノテーションは必須である。医用画像に対しては、医師のような専門家でないアノテーションすることは難しい。また、DL はデータが多いほど精度が上がるが多いため、医師が膨大な量のアノテーションをすることになり、DL を用いた CAD システムを用いたとしても医師の負担は存在する。本研究では、医用画像に付随している読影レポートを利用し、画像アノテーションを行うことなく、特定の疾患の画像ベースでの自動分類を行う機械学習アルゴリズムの学習法を提案する。

本研究では、自然言語処理を用いることで、負担の大きい手動の画像アノテーションを必要とせず、特定の疾患を画像ベースで自動分類する分類器を構築する。提案手法では、所見文から情報を抽出することで、画像分類用のアノテーション付きデータセットを作成する。まず、所見文、画像などが含まれているデータベースから所見文に特定の疾患名を含んでいるデータを取得する。次に、取得したデータの中から、1 文中に疾患名と否定ワードが入っているデータを除外する。最後に、除外したデータの中から 1 文中に疾患名と「著変なし」のような疾患名とその存在を否定していないデータを追加する。また、特定の疾患名を含まないデータも同様の処理により取得して正常データとする。作成したデータセットを用いて CNN を学習することで特定の疾患を分類する分類器を構築する。

実験では国立情報学研究所の医療画像ビッグデータクラウド基盤にあるデータに提案手法を適用し、「間質性肺炎」と「くも膜下出血」の分類器を構築した。クラウド基盤から取得したデータ数は、「間質性肺炎」画像が 337,663 枚、「正常肺」画像が 6,660 枚、「くも膜下出血」画像が 151,295 枚、「正常脳」画像が 27,639 枚であった。取得したデータセットを学習データセットと評価データセットに分割し、「間質性肺炎」の有無を分類する「間質性肺炎」分類器と、「くも膜下出血」の有無を分類する「くも膜下出血」分類器を学習したところ、「間質性肺炎」分類器の精度が 64.2%、「くも膜下出血」分類器の精度が 89.1% であった。「間質性肺炎」(図 1)、「くも膜下出血」(図 2) と正しく分類できた画像を以下に示す。この実験結果から、手動の画像アノテーションなしで疾患分類器は構築可能であることが示された。精度が低い原因はデータ数が少ないことであると考えられる。今後の課題としては、データ数の増強、様々な疾患に対しての分類器を作成することが挙げられる。

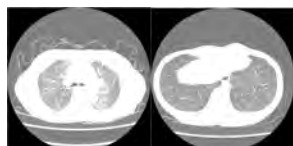


図 1: 左「間質性肺炎」に分類された画像  
右「正常肺」に分類された画像

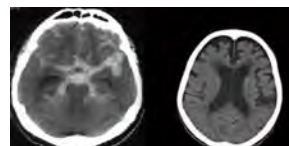


図 2: 左「くも膜下出血」に分類された画像  
右「正常脳」に分類された画像