

令和 2 年 6 月 15 日現在

機関番号：14301

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K08740

研究課題名(和文) デジタル病理画像の機械学習による肺腺癌の予後予測モデルの構築

研究課題名(英文) Construction of prediction model of lung adenocarcinoma by machine learning

研究代表者

吉澤 明彦 (Yoshizawa, Akihiko)

京都大学・医学研究科・准教授

研究者番号：80378645

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：第一段階として、多重解像度解析・改良kNN法によるデジタル病理画像解析を行った。まずはアノテーションを行った関心領域からラベル付きパッチ画像を教師データとして抽出、ウェーブレット変換しクラスター分析で数値化した。次に、このパッチ画像全を訓練データ、テストデータとしてSVMを用い、多重解像度解析した。結果、腫瘍、非腫瘍の判別に関しては、0.7777の正答率が得られた。結果の改善のために新規アルゴリズムを共同で開発し検討を進めた(Soft switch FCN法)。提案手法で、腫瘍、非腫瘍判別の正解率は0.95と改善した。一方、各種別の正解率は、満足な結果とは行ならなかった(0.61-0.93)。

研究成果の学術的意義や社会的意義

病理デジタル画像においても深層学習を用いた判別の可能性が示された。しかしながら、教師データであるWSI画像は放射線画像と比較し大きく、そのまま、入力することはできない。そのため、パッチ画像として入力することになるが、組織構築の判別を行う際はより広い視野をパッチとしなければならないためその画質は落とさざるを得ないジレンマがある。そこで我々はU-Netに、最適な視野領域の組み合わせを予測できるSoft switch法を組み合わせたSoft switch FCN法を共同開発し、検討に用いた。肺癌の予後予測に関しては検討途中であるが、組織構造の分類には有用である可能性が示され、その意義は大きいと考える。

研究成果の概要(英文)：As the first step of the study, we performed digital pathological image analysis by multi-resolution analysis and improved kNN method. After extracting labeled patch images (128x128 pixels) from the annotated region of interest as training data, they were wavelet transformed and digitized by cluster analysis (11 variables). Next, multi-resolution analysis was performed using a support vector machine. As a result, it was possible to distinguish between tumor and non-tumor with an accuracy rate of 0.7777. To improve the result, we developed a novel deep learning method (Soft switch FCN method) and conducted a study using the system. With the proposed method, the accuracy rate for the determination between tumor and non-tumor was 0.95. On the other hand, the accuracy rate for each class was not satisfactory results. We are keeping to examine the proposed algorithm with the total dataset to construct Ca-CHS that can predict the prognosis of the patients with resected lung adenocarcinoma.

研究分野：病理診断

キーワード：肺癌 腺癌 予後予測 機械学習 深層学習

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

肺癌の中で、肺腺癌は壺型として最も多い。切除可能な肺腺癌は切除後、特徴的な5つの組織パターン(肺胞置換型、乳頭型、腺房型、微小乳頭型、充実型)を元にさらに分類が行われるが(Comprehensive Histologic Subtyping: CHS) その認識に関して客観性に問題があることが指摘されている。世界的に CHS を元に肺腺癌の組織壺分類を行う提案がなされてからしばらくたつが、客観性の問題は未だ残っている。

2. 研究の目的

本研究では、肺腺癌の病理組織標本をデジタル化し(Whole Slide Imaging: WSI)し、機械学習を用いて、より客観的で、予後を予測できるようなモデル(Computer-aided CHS: Ca-CHS)を構築することを目指す。

3. 研究の方法

対象は、研究者の所属施設にて2001年から2015年の間に外科的に完全切除された肺腺癌症例で、単発、3 cm以下の症例とした。基本情報として、個人情報に配慮しながら、年齢、性別、喫煙歴、病期、予後、病理診断に加え、病理医が顕微鏡を用いて行ったCHS(HE染色にて判定)を収集した。機械学習を行うため、主病巣のHE染色をWSI化した。WSI取得後、機械学習のための教師データを作成するため、1)腫瘍領域の抽出のためのアノテーション、2)腫瘍領域における5つの組織型分類を行うためのアノテーション、を2人の病理医で行った。機械学習の方法としては、多重解像度解析・改良KNN法および深層学習を用いた。評価は、小区画画像(パッチ画像)での正解率、あるいは病理医が判定したCHSあるいは臨床像(予後)と対比することで、より客観的で、予後を予測できるようなモデル(Ca-CHS)を構築することを目指した。尚、研究者の施設で倫理審査を行っている。

4. 研究成果

検討第一段階として、多重解像度解析・改良KNN法によるデジタル病理画像解析を行った。この時点では、対象症例を20例とした。アノテーションを行った関心領域(ROI)からラベル付きパッチ画像(128x128 pixel)を教師データとして抽出、ウェーブレット変換しクラスター分析で数値化(変数11個)した。次に、このパッチ画像全30250枚を訓練データ20167枚、テストデータ10083枚として、サポートベクターマシン(SVM)を用い、多重解像度解析した。結果、腫瘍、非腫瘍の判別に関しては、0.7777の正答率が得られた。また、同じデータをデータオーギュメンテーションにて約4倍のデータ量を確保し、深層学習による検討を行った(Tensorflowを利用)。正答率は0.7624と優位な改善はみられなかった。詳細を検討するため、パッチ画像を評価したところ、特に5つの組織パターンのうち肺胞置換型は正常肺胞構造に類似し、判別が正確に行えていないことがわかった。また、乳頭状と微小乳頭状の判別も不正解が多く、これはパッチの大きさが問題であることがわかった。そこで、深層学習のうちでも医療用の画像処理で用いられるアルゴリズムの一つ全層畳み込みネットワーク(ここではU-netモデル)と最適な視野領域の組み合わせを予測できるSoft switch法(ここではXceptionモデル)を組み合わせ、領域の大きさに重みをつけた深層学習での検討を行った(Soft switch FCN法)。学習データは、29例、パッチサイズは256x256としながらも画像のスケールを変えている(0.5, 0.25, 0.12)。学習データに用いるパッチ画像は、非腫瘍7063枚、肺胞置換型2400枚、乳頭型/腺房型5563枚、微小乳頭型5687枚、充実型4486枚とした。提案手法で、腫瘍、非腫瘍は0.95の正解率であった。一方、各種別の正解率は、肺胞置換型0.93、乳頭型/腺房型0.61、微小乳頭型0.63、充実型0.75とクラスによっては、決して満足な結果とは行ならなかった。これまでの検討は、病理医

が完全にアノテーションした“強い教師データ”を用いて各クラス判別を検討してきたが、現在、取得してあるクラス比率のデータを“弱教師データ”として用い、検討を進めている。最終的には、提案手法をよりおおきなデータを用い検討し、予後推定可能な Ca-CHS の構築を進め、報告する予定である。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 1件 / うち国際学会 0件）

1. 発表者名 寺本 祐記, 吉澤 明彦, 石橋 雄一, 羽賀 博典
2. 発表標題 Wavelet変換及びsupport vector machineを用いた肺腺癌の組織パターン解析(Lung adenocarcinoma classification: a computerized approach using wavelet support vector machine)
3. 学会等名 第107回日本病理学会総会
4. 発表年 2018年

1. 発表者名 吉澤明彦
2. 発表標題 近未来の病理診断 人工知能と肺癌病理診断
3. 学会等名 第60回日本肺癌学会学術集会（招待講演）
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考