

令和 6 年 6 月 14 日現在

機関番号：12501

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K07590

研究課題名（和文）がんのRadiogenomicsデータに対する深層学習による革新的特徴量選択

研究課題名（英文）Innovative Feature Selection Using Deep Learning for Cancer Radiogenomics Data

研究代表者

森 康久仁（Mori, Yasukuni）

千葉大学・大学院情報学研究院・助教

研究者番号：40361414

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究では、深層学習の優れた特徴抽出能力を活用することで医療画像データ・遺伝子発現データ・生化学データなど、様々な医療データをマルチモーダルの適用し、癌の層別化に関わる臨床的要因を特定・選択する統一的方法論を構築することを目的とした。特に、様々なモダリティのデータに対する特徴量抽出の方法論についての研究を集中的に行った。その中でも、深層学習のモデル中に特徴の重要度を学習により推定する特徴選択層を提案し、消化器癌の中でもっとも予後不良の悪性腫瘍である膵臓がん患者の遺伝子発現データに対して適用した。その結果、本研究により提案された方法は、一定の確信度を持って有効であると確認することができた。

研究成果の学術的意義や社会的意義

本研究は、深層学習を用いたマルチモーダルデータの統一的分析方法論を確立することにより、癌の層別化に関わる重要な臨床的要因を効果的に特定・選択する新しいアプローチの提案を目的とするものである。特に、深層学習モデルにおける特徴選択層の提案は、従来の手法では難しかったデータ間の複雑な相互作用を捉えることを可能にし、癌研究や他の医療分野におけるデータ解析の精度向上に寄与すると考えられる。さらに、患者ごとの最適な治療法を見つけるための新たな手段を提供し、治療効果の向上と副作用の軽減に寄与することが期待され、医療コストの削減や早期診断の精度向上にも貢献し、社会全体の医療品質の向上に繋がると考えられる。

研究成果の概要（英文）：This study aims to establish a unified methodology that leverages the superior feature extraction capabilities of deep learning to apply various types of medical data, such as medical imaging data, gene expression data, and biochemical data, in a multimodal manner to identify and select clinical factors involved in the stratification of cancer. Specifically, we focused on methodologies for feature extraction from various data modalities. Among these, we proposed a feature selection layer within deep learning models that estimates feature importance through learning and applied it to gene expression data of pancreatic cancer patients, which is one of the most malignant tumors with poor prognosis among gastrointestinal cancers. As a result, the method proposed in this study was confirmed to be effective with a certain degree of confidence.

研究分野：機械学習

キーワード：特徴選択 深層学習 遺伝子選択 医用画像処理

## 1. 研究開始当初の背景

現在非常に注目を浴びている領域として、機械学習と医療分野との連携がある。機械学習を CT や MRI などの医用画像に応用し、病変部の発見などの診断支援の研究が盛んに行われている。

そのような状況の中、「深層学習」と呼ばれる機械学習の一手法の発展に伴い、質の良いデータが利用できる状況では、非常に精度の良い識別モデルをデータ駆動で構築することができるようになった。しかしながら、これまで深層学習によって構成した識別モデルにおいて、通常のテーブル型の数値データでは、その推論根拠を示すことは難しく、このことは、通常の深層学習では、重視すべき特徴を見出し、意味のない特徴を明確にするなどといった、現実の問題へのフィードバックが非常に難しい。

さらに現在では数値データだけでなく、画像、テキストデータ、さらには時系列データなど様々なデータが同時に計測され、それらが紐付けられて蓄積されていることが多い。一方で、機械学習では、これら複数の多様なデータ情報を同時に扱って処理を行う「マルチモーダル学習」とよばれる方法が使われ始めてきている。そこで、深層学習に、様々なモダリティのデータを組み入れて学習することで、これまでとは異なる観点からの特徴抽出の方法論が創造できると考えられる。

## 2. 研究の目的

本研究の目的は、画像情報だけでなく、テーブル型の数値データやテキスト情報など、多種多様なデータを用いて深層学習によるマルチモーダル学習を行うことである。これにより、それぞれのデータが持つ情報を相互に補完し、複雑な関連性を持つ複数の要因をデータから学習する方法論を提案する。具体的には、画像データの視覚情報、数値データの定量情報、およびテキストデータの言語情報を統合的に処理し、それぞれのデータの特性を活かしながら、より高次の理解と分析を可能にすることを旨とする。

本研究の方法論を実際のデータセットに適用し、マルチモーダル学習の有効性を実証することを目指し、各モダリティのデータの性質を明らかにする。これにより、将来的なマルチモーダル学習で必要となるモデルの構築などに役立てることができ、異なるデータタイプの相互作用を通じて、単一のデータソースでは捉えきれない複雑な関係性やパターンを明らかにすることができると思われる。最終的には、この方法論が様々な分野でのデータ解析において有用であることを示し、マルチモーダル学習の新たな可能性を開拓することを旨とする。

## 3. 研究の方法

深層学習の優れた特徴抽出能力を活用し、様々な形式のデータに対し、当該タスクに有用な特徴の抽出・選択する方法論を検討する。さらに、実際の様々なデータに対して適用し、各モデルの特性やデータの特質を調査する。特に本研究では、典型的なマルチモーダルなデータである医療データに対して機械学習による分類問題を検討し、各データの特質を調査し、がんの層別化に有効な方法の検討を行う。

本研究では、それぞれのモダリティのデータに対して最適な特徴抽出の方法を提案し、その有効性を検証する。例えば、画像データに対しては、畳み込みニューラルネットワーク (CNN) を用いた特徴抽出を行い、数値データに対しては、新たに提案した特徴選択層を組み込んだニューラルネットワークなどにより特徴選択を行う。また、テキストデータに対しては、自然言語処理 (NLP) 技術を駆使して重要なキーワードや文脈の意味を抽出し、それぞれのデータ形式に最適化された処理方法を開発する。

## 4. 研究成果

### (1) 特徴選択層による遺伝子発現データからの遺伝子選択 [1]

提案した特徴選択層とは、入力に対し 1 対 1 の関係で重みを乗算した値を出力する層であり、これにより、入力した特徴の値が、重みにより増加・減少して、次のネットワークに渡されることになる。したがって、この重みの大小により、各特徴量の重要度を測ることができる。図 1 に特徴選択層の概念図を示す。

新たな提案した特徴選択の方法論では、実際のすい臓に関する遺伝子発現データに適用し、膵臓がん組織と正常な膵臓組織を識別するタスクと、1年予後に関して生存・非生存を識別するタスクの2つの問題を考え、それぞれのタスクにおいて、有意に関連する遺伝子の選択問題を行った。

その結果、タスク1では、ACACB, ADAMTS6, NCAM1, CADPS という遺伝子が選ばれ、タスク2では、CD1D, PLA2G16, DACH1, SOWAHA が選ばれた。これらの選択された遺伝子を The Cancer Genome Atlas (TCGA) データベースによって確認したところ、すべての遺伝子が、膵臓癌の予後因子として知られていることが確認された。本成果は、2021年 Scientific Reports 誌に掲載された。

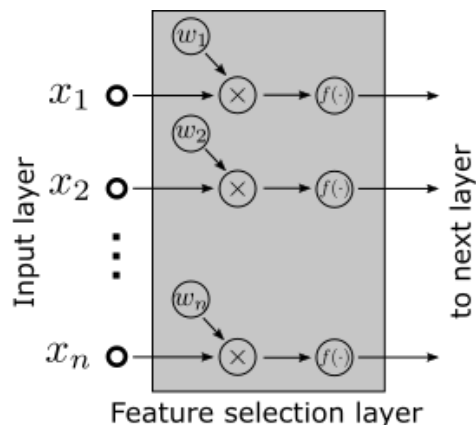


図 1 特徴選択層の概念図

### (2) 読影レポートにおけるテキスト情報を用いた分類問題[2]

本研究は、CT や MRI の読影レポートに含まれる重要所見の見落としを防止するための自動分類手法を提案するものである。連続した読影レポートには、前回と比較して悪化している所見が含まれることがあり、これらの見落としは患者の治療遅延を引き起こす。本研究では、アテンションネットワークを用いて、レポートを文ごとに分割し、それぞれの文の重要性を考慮して分類する方法を提案した。具体的には、BERT を用いて文の特徴ベクトルを抽出し、Coherence Model で文間の関係性を計算、Entailment Model でアテンションを計算する。

実験には千葉大学医学部附属病院で作成された 408 件の読影レポートを使用し、2名の放射線科医が手動で重要所見をラベル付けしたデータを用いた。5 分割交差検証によりモデルを評価し、Accuracy, Recall, Precision を指標とした結果、提案手法は従来手法と比較して Accuracy が 3% 向上した。提案手法では各文のアテンションを計算することで、重要な所見を含むレポートを高精度に分類できることが示された。本研究はがんの層別化には直接関係ないが、マルチモーダル学習におけるテキスト情報を利用する際の Preliminary な研究として行ったものである。

### (3) MR 画像におけるリアルタイム 3 次元腫瘍位置予測[3]

本研究は、放射線治療中にリアルタイムで腫瘍の三次元位置を予測するニューラルネットワークモデルを提案し、その精度を評価するものである。放射線治療においては、呼吸による腫瘍の位置変動を正確に把握することが治療精度向上に不可欠であるが、体内を直接視認できないために困難が伴う。近年、MRI を搭載した放射線治療装置 (MR-linac) が開発され、治療中の腫瘍位置を把握する技術が進展しているが、依然としてリアルタイムでの位置予測には課題が残る。

本研究では、千葉大学医学部附属病院において治療された 6 人の患者から得られた約 78,000 枚の MR 画像を用い、肝臓の三次元位置を予測対象とした。使用するモデルは Non-stationary Transformer であり、不規則な呼吸による位置変動に対して高精度な予測が期待できる。

実験では、冠状面、矢状面、横断面の MR 画像データを用い、それぞれの組み合わせによる予測精度を比較した。冠状面のデータのみを入力した場合と、冠状面に矢状面を加えた場合、および全ての断面を入力した場合の三条件で評価を行った結果、全ての断面を用いた条件が最も高精度な予測を示した。本研究はがんの層別化に関し直接的な関係は薄いものの、画像情報から腫瘍領域の特徴を抽出するための重要な処理に関するものである。

[1] Mori, Y., et al. Deep learning-based gene selection in comprehensive gene analysis in pancreatic cancer. Sci Rep 11, 16521 (2021). <https://doi.org/10.1038/s41598-021-95969-6>.

[2] 村松 他, 読影レポートの見落とし防止に向けた重要所見の自動分類, 電子情報通信学会総合大会, ジュニア&学生ポスターセッション, ISS-SP-049, 2024.

[3] 清水 他, MR 画像誘導放射線治療における照射中リアルタイム 3 次元腫瘍位置予測の試み, 電子情報通信学会総合大会, ジュニア&学生ポスターセッション, ISS-SP-049, 2024.

## 5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Yasukuni Mori, Hajime Yokota, Isamu Hoshino, Yosuke Iwatate, Kohei Wakamatsu, Takashi Uno, Hiroki Suyari	4. 巻 11
2. 論文標題 Deep learning-based gene selection in comprehensive gene analysis in pancreatic cancer	5. 発行年 2021年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41598-021-95969-6	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Yuki Terasaki, Hajime Yokota, Kohei Tashiro, Takuma Maejima, Takashi Takeuchi, Ryuna Kurosawa, Shoma Yamauchi, Akiyo Takada, Hiroki Mukai, Kenji Ohira, Joji Ota, Takuro Horikoshi, Yasukuni Mori, Takashi Uno, Hiroki Suyari	4. 巻 18
2. 論文標題 Multidimensional Deep Learning Reduces False-Positives in the Automated Detection of Cerebral Aneurysms on Time-Of-Flight Magnetic Resonance Angiography: A Multi-Center Study	5. 発行年 2022年
3. 雑誌名 Frontiers in Neurology	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fneur.2021.742126	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計7件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 村松 祐希, 横田 元, 中尾 達司, 堀越 琢郎, 太田 丞二, 宇野 隆, 須鎗 弘樹, 森 康久仁
2. 発表標題 読影レポートの見落とし防止に向けた重要所見の自動分類
3. 学会等名 電子情報通信学会
4. 発表年 2024年

1. 発表者名 清水 涼首, 横田 元, 中尾 達司, 堀越 琢郎, 太田 丞二, 宇野 隆, 須鎗 弘樹, 森 康久仁
2. 発表標題 MR画像誘導放射線治療における照射中リアルタイム3次元腫瘍位置予測の試み
3. 学会等名 電子情報通信学会
4. 発表年 2024年

1. 発表者名 林 千咲子, 森 慎一郎, 石川 仁, 須鎗 弘樹, 森 康久仁
2. 発表標題 DRR 画像から X 線画像への画質変換 - 画像間の位置ずれによる影響を低減するネットワークの提案 -
3. 学会等名 電子情報通信学会
4. 発表年 2024年

1. 発表者名 宮崎拓海, 恒田雅人, 阿部幸太, 横田元, 太田丞二, 堀越琢郎, 宇野隆, 須鎗弘樹, 森康久仁
2. 発表標題 対照学習を用いた教師なし学習によるMR画像から擬似CT画像の生成
3. 学会等名 第123回日本医学物理学学会学術大会
4. 発表年 2022年

1. 発表者名 植野雄大, 横田元, 堀越琢郎, 太田丞二, 須鎗弘樹, 森康久仁
2. 発表標題 DaTScanデータによるパーキンソニズムを呈する疾患の自動分類
3. 学会等名 第194回医用画像情報学会
4. 発表年 2022年

1. 発表者名 吉川耕平, 太田丞二, 橋本拓磨, 横田元, 堀越琢郎, 須鎗弘樹, 森康久仁
2. 発表標題 Triplet Lossを用いた頭部スカウト画像からの患者間違い検出手法に関する検討
3. 学会等名 第194回医用画像情報学会
4. 発表年 2022年

1. 発表者名 熊切俊夫, 森慎一郎, 平井隆介, 橋本綾斗, 立花泰彦, 森康久仁, 須鎗弘樹, 石川仁
2. 発表標題 重粒子線治療高精度化に向けた深層学習によるX線画像腸管ガス領域抽出法の開発
3. 学会等名 第194回医用画像情報学会
4. 発表年 2022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
研究分担者	星野 敢  (Hoshino Isamu)  (10400904)	千葉県がんセンター(研究所)・消化器外科・主任医長   (82504)	
研究分担者	横田 元  (Yokota Hajime)  (20649280)	千葉大学・大学院医学研究院・講師   (12501)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------