

令和 4 年 6 月 20 日現在

機関番号：32643

研究種目：基盤研究(B)（一般）

研究期間：2019～2021

課題番号：19H01806

研究課題名（和文）個別化医療のための人工知能・深層学習に対する統計学的評価法の開発

研究課題名（英文）Development of statistical evaluation method for artificial intelligence and deep learning for personalized medicine

研究代表者

松浦 正明 (Matsuura, Masaaki)

帝京大学・公私立大学の部局等・教授

研究者番号：40173794

交付決定額（研究期間全体）：（直接経費） 13,300,000円

研究成果の概要（和文）：本研究では、個別化医療の実現および加速化のために、医学において最近利用され始めている人工知能・深層学習を用いたデータ解析が広まりつつある。しかし、どのように解析結果が導出されたか解析内容がブラックボックス的であるとの批判がある。そこで「深層学習の各特徴量の寄与度」に関して統計学的視点からの評価を加えるための研究を行い、特に各種の実データを用いた具体的問題に関して研究を行ない、これらの経験を基に簡便的な新規手法の提案を行った。

研究成果の学術的意義や社会的意義

これまでの人工知能・深層学習を用いたデータ解析には、主に画像解析の画素をデータとして取り扱う分野の発展が著しいが、医学的臨床情報を用いる分野では各種臨床検査値やオミックスデータの情報を変数として取り扱う分野の発展が望まれてきた。しかし、後者の場合の解析結果は、最終結果のみが出力され、医学で重要な解析で使用した各種因子や説明変数の寄与度が明らかでない点が問題であった。本研究では、実データを用いて研究を推進した。

研究成果の概要（英文）：In this research, data analysis using artificial intelligence and deep learning, which has recently begun to be used in medicine, is becoming widespread in order to realize and accelerate personalized medicine. However, there are criticisms that the analysis content is black box-like as to how the analysis results were derived. Therefore, we conducted research to add an evaluation from a statistical point of view regarding "contribution of each feature of deep learning", and in particular, conducted research on specific problems using various actual data, and based on these experiences. We proposed a simple new method.

研究分野：統計学、疫学、バイオインフォマティクス

キーワード：統計 個別化医療

1. 研究開始当初の背景

(1) 医学領域において人工知能・深層学習アルゴリズムの活用が進展してきた。特に画像処理の解析で大きな進展が見られるが、一点では意味を持たない画素データを対象としており、画素の集合を特徴量として利用する事により成功を収めている。一方、医学データでは、患者のゲノムなど多層オミックスデータが収集されるようになり、患者の各種臨床検査情報を含めた表形式のテーブルデータの解析の必要性が高まっている。しかし、ブラックボックス的判定結果のみを出力しただけでは結果に対する批判が強い場合が多く、重要な臨床情報や何のゲノムが強く関連しているのかの情報の出力されなければ、因果関係の推測が不可能である。特に医療機器開発において、特許取得や機器の認可容易化のために、オミックスデータを含む医学的臨床情報、臨床検査項目などの利用した解析において、各変数の寄与度を数値化することが望まれてきた。

(2) 医療・社会現象を解析するにあたり、電子計算機に蓄積されたビッグデータの活用が注目を浴び、「ビッグデータ」の用語が使用され始めてきた。しかし、ビッグデータを統計解析するにはデータの不均一性が問題になり、この問題を克服しないと予測問題は解決しない。既に、前回採択の基盤研究(B)の成果として「ビッグデータ解析方法」としての「サブタイプ検出アルゴリズムと関連バイオマーカー抽出」の特許申請を行ったので、このアルゴリズムの高精度化が望まれていた。

2. 研究の目的

(1) 医療機器認可容易化のため画像解析の画素データではなく、深層学習によるブラックボックス的判定結果に対して、各変数の寄与度を数値化できないか？この視点から深層学習の解析可能性を事前にデータから評価できないか？特に個別化医療の早期実現に向けては、各種のオックス情報と臨床情報を結び付け、患者の状態予測を行うため、深層学習からの結果に対して因果関係を推測するための予測医学の実現のための方法が必要である。

(2) 前回採択の基盤研究(B)の成果として特許出願した「ビッグデータに対するサブタイプ検出アルゴリズムと関連バイオマーカー抽出」に対して高精度化できないか？

3. 研究の方法

(1) 深層学習を用いた解析において、画像解析の画素データではなく、各行が患者データ、各列が各種検査項目・オミックスデータなどの各種変数を示す表形式のテーブルデータを扱う。まずは、解析に使用される深層アルゴリズムを検討し、特徴量の寄与度の評価を行う過去の文献調査を行うと共に、実データを用いて深層学習の解析経験を積み、テーブルデータの各変数の寄与度を算出するアイデアを回帰分析の視点より検討した。特に初期アイデアとして深層学習で得られた予測結果と、多重ロジスティック回帰などの古典的予測手法を用いた予測結果の比較を基に、各変数の寄与度を算出する方法を検討する。なお、本研究に適した実データの探索も同時に行った。

(2) 特許出願した内容は、最初にビッグデータの不均一性を仮定し、データを相関の強さを基に信頼度の高いデータを用いて部分集団に分け、部分集団ごとに関連マーカーとなる変数を自

動抽出する。これら抽出した変数を用いた予測方法を整理し、多変量の場合にも対応できる高精度化の検討を行う。

4. 研究成果

(1) 初年度(2019年度)は、深層学習の各層特徴抽出結果に対する各変数による統計学的性能評価法については、表形式型データを用いた2値予測に対する深層学習アルゴリズムの検討と、特徴抽出器として使用することができるソフトウェアに対して調査を行った。最終結果に対する各変数の寄与度を調べる最も単純な方法は、その変数を削除して同様の深層学習を行った場合に、どれだけ最終結果が変化するかを調べることで、その特定の変数の最終結果に対する寄与度が調査可能である。しかし、最終結果を導出した変数の数が膨大であると、その変数の個数だけ深層学習の解析を行う必要があり、膨大な時間がかかる欠点を克服する必要があり、これらについて調査・検討を進めた。

(2) 初年度の計画のもう一つは、前回の基盤研究(B)(松浦代表)で特許出願したビッグデータに対する解析方法・評価としての「不均一性抽出アルゴリズムとサブグループ発見およびサブグループ関連マーカーの抽出方法」に対する高精度化と再評価を計画し、この研究を推進した。本アルゴリズムに対しては、既に日中米国に特許出願した内容を、2018年国際計量生物学会(バルセロナ)で発表し特許開示を行っている。アルゴリズムの概略は、まず相関係数を基に信頼度の高い変数ペアを抽出し、これらの信頼度の高い変数を基に信頼度の高いサンプル検体を取り出す。次に、取り出したサンプル検体に基づいて順次集団のサブクラスターを作成し、各サブクラスターを規定するバイオマーカーの抽出を行う。この際、抽出したバイオマーカーを用いて、各検体がどのサブクラスターに属するかの予測確率を計算する必要があるが、この部分の改良を行った。改良の内容は各変数での判定の平均を用いる形式から、多重ロジスティック回帰を用いて、解析に使用した全変数の線形結合量を基に判別のROC曲線を構成し、感度や特異度で定義される適切なカットオフ値を基に多変量予測判定を行う方式である。さらに、この予測判定に関して、生存時間データへの応用を検討し、その内容を2020年度の国際計量生物学会(韓国ソウル)にて報告するため登録を行い、発表に関して学会側の了承を得る事ができた。ただし、コロナにより対面での学会は中止となったが、延期して開催された2020年度の国際計量生物学会(ソウル)にてオンライン報告を行った。本研究の成果として、これまで2値判別での解析は多重ロジスティック回帰を用いて予測モデルを作成し、交差検証または検証研究を行って予測性能を評価してきたが、この解析手法が生存時間データに適用できる点は、データ解析の方法を広い分野に拡大し大きな進展を与えたと考える。

(3) 「深層学習の各層特徴抽出」については、最近では画像データに対して画素のセットを特徴量として抽出する方法も提案されている事が調査で判明した。しかし、今後の個別化医療への応用を考えた場合は、各種検査・医療情報などのテーブルデータに対して、各変数およびそれらの変数セットの組み合わせと最終結果との関連性を寄与度として評価することが、因果関係を解明する点で重要である場合がある点も、事例研究により明らかになった。これに対応する方法として「SHAP」のアルゴリズムが存在している事が調査で判明した。SHAP(Shapley Additive exPlanations)は、結果に対する各特徴量の貢献度を定量化している。特定の特徴量が無い場合とある場合で、予測がどのように変わったかを全組合せについて算出して平均を取ることで算出される。ゲーム理論で有名なShapleyのゲームの参加者の貢献度計算理論を応用し、予

測結果である値を特徴量の貢献度の和として定量的に分解したものの対応している。

(4) 最終年度(2021年度)は、本研究の知見を基に、SHAPを用いた実データ解析を複数行い、ロジスティック回帰における予測性能との比較に関して各種実データを用いて行った。特に本研究代表者は、家族性乳癌関連遺伝子BRCA1を世界で初めてクローニングした三木義男教授(東京医科歯科大学)とAMEDの共同研究「遺伝性腫瘍原因遺伝子に見いだされるVUSの相同組換え活性を指標とした機能的アノテーションに関する基盤研究」の分担研究者を担当し、本研究の知見を応用した。具体的には、遺伝子変異機能評価を行うため、7種の既存in silico変異機能予測プログラムを利用し、乳癌遺伝子変異データベースのClinvarからPathogenicとBenignが判明している変異1265例を抽出し、Shapを用いた深層学習法で解析した結果、10-交差検証法で感度0.974、特異度0.967、正答率0.971の成績を示した。この予測モデルを用いて、未知のVUSの判定を行ったところ、三木先生グループは新規HR活性法で評価した結果とほぼ一致してPathogenicの変異を抽出でき、本研究はAMED研究との相乗効果を示す事ができた。

また、これらの研究を基に、新規の各変数の寄与度を評価する方法として、深層学習で算出した予測スコアを全変数の組み合わせで算出し、深層学習の全予測スコア値を重回帰分析の結果変数として用い、各変数の標準化偏回帰係数を算出する事により、簡便的な変数の寄与度の算出が可能となる手法を開発した。

(5) 前回の基盤研究(B)(松浦代表)で特許出願したビッグデータに対する解析方法・評価としての「不均一性抽出アルゴリズムとサブグループ発見およびサブグループ関連マーカーの抽出方法」に関しては、2017年4月27日に共同開発の島津製作所と特許申請(特願2017-088365)を行っていたが、2022年4月13日に特許第7057913号として、日本での特許登録がなされた。他に中国および米国にも出願しているため、これらも近い将来特許取得の予定である。

<引用文献>

Scott M. Lundberg, Su-In Lee. A Unified Approach to Interpreting Model Predictions 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. <https://arxiv.org/abs/1705.07874>

5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 8件 / うち国際共著 0件 / うちオープンアクセス 5件）

1. 著者名 Konishi Y, Nakata Y, Nemoto A, Ushijima M, Matsuura M.	4. 巻 30
2. 論文標題 The preoperative risk factors of postoperative self-extubation in elderly patients.	5. 発行年 2019年
3. 雑誌名 Int J Risk Saf Med. 2019;30(1):9-18.	6. 最初と最後の頁 9-18
掲載論文のDOI (デジタルオブジェクト識別子) 10.3233/JRS-180011.	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Konishi KI, Yoshida M, Nakao A, Tsuchiya K, Tsurisawa C, Ichiki K, Takeda T, Hisaeda Y, Hirota A, Amagata S, Odashima T, Yokomori K, Ishida K, Matsuura M, Ito S, Fujiwara M, Nakahara S.	4. 巻 54
2. 論文標題 Prostaglandin E-major urinary metabolite as a noninvasive surrogate marker for infantile necrotizing enterocolitis.	5. 発行年 2019年
3. 雑誌名 J Pediatr Surg.	6. 最初と最後の頁 1584-1589
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jpedsurg.2018.08.044.	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Hirai T, Nemoto A, Ito Y, Matsuura M.	4. 巻 181
2. 論文標題 Meta-analyses on progression-free survival as a surrogate endpoint for overall survival in triple-negative breast cancer.	5. 発行年 2020年
3. 雑誌名 Breast Cancer Research and Treatment.	6. 最初と最後の頁 189-198
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s10549-020-05615-4	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Yoshida R, Hagio T, Kaneyasu T, Gotoh O, Osako T, Tanaka N, Amino S, Yaguchi N, Nakashima E, Kitagawa D, Ueno T, Ohno S, Nakajima T, Nakamura S, Mik, Y, Hirota T, Takahashi S, Matsuura M, Noda T, Mori S.	4. 巻 112(3)
2. 論文標題 Pathogenicity assessment of variants for breast cancer susceptibility genes based on BRCAness of tumor sample.	5. 発行年 2021年
3. 雑誌名 Cancer Sci.	6. 最初と最後の頁 1310-1319
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/cas.14803.	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 Nishiya N, Oku Y, Ishikawa C, Fukuda T, Dan S, Mashima T, Ushijima M, Furukawa Y, Sasaki Y, Otsu K, Sakyo T, Abe M, Yonezawa H, Ishibashi F, Matsuura M, Tomida A, Seimiya H, Yamori T, Iwao M, Uehara Y.	4. 巻 112(5)
2. 論文標題 Lamellarin 14, a derivative of marine alkaloids, inhibits the T790M/C797S mutant epidermal growth factor receptor.	5. 発行年 2021年
3. 雑誌名 Cancer Sci.	6. 最初と最後の頁 1963-1974.
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/cas.14839.	査読の有無 無
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 Osako, T., Matsuura, M., Yotsumoto, D., Takayama, S., Kaneko, K., Takahashi, M., Shimazu, K., Yoshidome, K., Kuraoka, K., Itakura, M., Tani, M., Ishikawa, T., Ohi, Y., Kinoshita, T., Sato, N., Tsujimoto, M., Nakamura, S., Tsuda, H., Noguchi, S., Akiyama, F.	4. 巻 128(10):
2. 論文標題 A prediction model for early systemic recurrence in breast cancer using a molecular diagnostic analysis of sentinel lymph nodes: A large-scale, multicenter cohort study.	5. 発行年 2022年
3. 雑誌名 Cancer	6. 最初と最後の頁 1913-1920.
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/cncr.34144.	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 Hirai T, Suzuki A, Yamori T, Matsuura M.	4. 巻 40(1)
2. 論文標題 Reforms of regulatory pathways for approval of new antineoplastic drugs in Japan from 2004 to 2019 and accompanying changes in pivotal clinical trial designs.	5. 発行年 2022年
3. 雑誌名 Invest New Drugs	6. 最初と最後の頁 142-150
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s10637-021-01165-8.	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Gu YH, Zhang Q, Guo J, Wang F, Bao Y, Qiu Z, Zheng P, Ushijima M, Matsuura M, Xie X, Zhang T.	4. 巻 68, 126814
2. 論文標題 Higher serum homocysteine and lower thyroid hormone levels in pregnant women are associated with neural tube defects.	5. 発行年 2021年
3. 雑誌名 J Trace Elem Med Biol.	6. 最初と最後の頁 1-9
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jtemb.2021.126814.	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 Mizutani S, Takahashi K, Matsuura M.	4. 巻 64(1):e14868
2. 論文標題 Situation analyses of caregivers' knowledge on infant liquid formula in Japan: a crosssectional questionnaire-based study.	5. 発行年 2021年
3. 雑誌名 10.1111/ped.14868.	6. 最初と最後の頁 1-7
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/ped.14868.	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3件(うち招待講演 0件/うち国際学会 2件)

1. 発表者名 Masaaki Matsuura, Tasuku Yoshida and Shuji Ando
2. 発表標題 Evaluation of prediction performance for Cox proportional hazards model
3. 学会等名 International Biometric Society 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Masaaki Matsuura, Tasuku Yoshida and Shuji Ando
2. 発表標題 Evaluation of prediction performance for Cox proportional hazards model
3. 学会等名 International Biometric Society 2020 (国際学会)
4. 発表年 2019年

1. 発表者名 顧 艶紅、松浦正明、鄭 萍、牛嶋 大、張 霆
2. 発表標題 神経管閉鎖障害児の胎盤における甲状腺ホルモン脱ヨード酵素の一塩基多型に関する症例対照研究
3. 学会等名 第32回日本疫学会学術総会
4. 発表年 2022年

〔図書〕 計1件

1. 著者名 室井誠, 馬島哲夫, 牛嶋大, 松浦正明	4. 発行年 2021年
2. 出版社 羊土社	5. 総ページ数 11
3. 書名 「オミックス解析: 表現型・薬物活性とリンクする分子経路の探索」 マウス・ラット モデル作製・解析・プロフェッショナル 先端モデル動物支援プラットフォーム (AdAMS) / 編	

〔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	生亀 清貴 (Iki Kiyotaka) (30711593)	日本大学・経済学部・講師 (32665)	
研究分担者	安藤 宗司 (Andou shuji) (40803226)	東京理科大学・工学部情報工学科・助教 (32660)	
研究分担者	牛嶋 大 (Ushijima Masaru) (60328565)	公益財団法人がん研究会・有明病院 臨床研究・開発センター・研究員 (72602)	
研究分担者	小森 理 (Komori Satoshi) (60586379)	成蹊大学・理工学部・准教授 (32629)	
研究分担者	松井 茂之 (Matsui Shigeyuki) (80305854)	名古屋大学・医学系研究科・教授 (13901)	

6. 研究組織（つづき）

	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	顧 艶紅 (Ko Enkou) (30470595)	大阪医科薬科大学・医学部・講師 (34401)	
研究分担者	江口 真透 (Eguchi Shinto) (10168776)	統計数理研究所・数理・推論研究系・教授 (62603)	

	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
研究協力者	梶原 茂樹 (Kajihara Shigeki)	島津製作所・基盤技術研究所・研究員	

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