

令和 3 年 6 月 23 日現在

機関番号：84404

研究種目：若手研究(A)

研究期間：2017～2019

課題番号：17H05078

研究課題名（和文）医療ビッグデータと人工知能を用いた新次元のアプローチによる心筋梗塞発症予測

研究課題名（英文）Precision medicine of myocardial infarction by a new dimension approach using medical bigdata and artificial intelligence

研究代表者

中尾 葉子（Nakao, Yoko）

国立研究開発法人国立循環器病研究センター・オープンイノベーションセンター・室長

研究者番号：90752824

交付決定額（研究期間全体）：（直接経費） 9,400,000円

研究成果の概要（和文）：本研究の目的は、循環器ビッグデータを用いて機械学習およびコグニティブコンピューティング等の人工知能応用技術を利用し、あらゆる診療情報をもとに、心筋梗塞発症の精緻な予測を目指すことである。

電子カルテや部門システムからのデータを自動抽出することに成功し、大規模データベースを構築し、妥当性の検証を行った。本研究により、人工知能による情報の自動抽出、因子間の関係性の探索、予測精度向上という、最小の予測因子の違いで良好な結果を予測可能にするアルゴリズムを提案することができた。

研究成果の学術的意義や社会的意義

患者1人1人の診療録内にある膨大な臨床情報を深く分析することで、医療ビッグデータを用いた循環器領域におけるプレジジョン医療の実現につながる。

研究成果の概要（英文）：The purpose of this study is to precisely predict myocardial infarction by using artificial intelligence application technologies such as machine learning and cognitive computing using cardiovascular big data.

We succeeded in automatically extracting data from electronic medical records and other systems, built a big cardiovascular database, and verified its validity. Through this research, I can propose a prediction algorithm that applies electronic medical record information and artificial intelligence technology.

研究分野：循環器疫学

キーワード：心筋梗塞 医療ビッグデータ 人工知能 予測 疫学 プレジジョン医療

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

プレジジョン医療とは、特定の疾病群の罹患性について患者をサブグループに分け、そのグループ毎に適切な治療法や予防法の開発を目指すものである。米国では特に癌や希少疾患領域でのプレジジョン医療が促進されており、ゲノム情報等を含めたさまざまな医療情報からサブグループを同定し、そこから治療法の確立や疾患発症予防を目指している。循環器領域においては、フラミンガム研究をはじめとしたコホート研究の大きな功績である喫煙、高血圧や脂質異常症等の古典的リスク因子、そして我が国でのナトリウム利尿ペプチドの発見以降、疾患特異的な新規マーカーや単一遺伝子変異が同定されるに至っていない。さらに、循環器疾患は遺伝的素因、生活習慣、環境因子等さまざまな要因が、継時的に変化しながら複雑に関係し合うため、循環器領域におけるプレジジョン医療の実現は難しいと考えられてきた。しかしながら近年、医療ビッグデータの活用と人工知能技術の応用、機械学習を含めた新規解析手法を組み合わせることで、心血管プレジジョン医療の実現可能性が見いだされてきた。

そこで本研究では、電子カルテに含まれる莫大な診療情報1つ1つを「データ」として捉え、診療録、オーダリングシステム、画像システム等さまざまなシステムを超えて統合したデータを元に大規模なデータベースを構築する。そして、新規統計的機械学習法を用いて疫学的に深く分析することにより、循環器疾患の新しい分類体系を確立し、精緻な予後予測アルゴリズムの開発を行う。

2. 研究の目的

本研究の目的は、循環器ビッグデータを用いて機械学習およびコグニティブコンピューティング等の人工知能応用技術を利用し、さまざまな診療情報をもとに、心筋梗塞発症の精緻な予測を目指すことである。

3. 研究の方法

当院心臓血管内科へ入院した全患者を対象とし、電子カルテ内のあらゆる診療情報、心臓超音波データ、心臓カテーテル検査データ等を、1入院を1レコードとした院内疾患登録システムを構築、時系列データを抽出し、データベースを構築した。このデータの自動抽出では、テンプレート等で統一的に入力された項目は含まれないため、例えば症状等の临床上重要な情報は入手できていない。そこで、IBM Watson により、Watson Knowledge Studio (WKS)/Watson Explorer (WEX)の両方を組み合わせ、機械学習により自然言語解析を行った。医学論文の自然言語データを学習させるため、文章中の言葉(単語)の識別を行い、それぞれがどの種別に属するかを学習させ、各言葉に対して同一性の指定を行った。そして、単語間の「関係」を教え、これらの学習と機械学習を繰り返すことで、自然言語の識別と関係の識別が自動的に可能となった。抽出データの妥当性の検証では、実臨床データのほかに、既報、循環器疾患診療実態調査報告書、ナショナルデータベースオープンデータ等を参照した。さらに、人工知能技術を応用した予測アルゴリズムの開発を行った。

4. 研究成果

当院心臓血管内科へ受診・入院した全患者を対象とし、電子カルテ内の診療情報、生理機能検査データ、心臓カテーテル検査データ等を、データとして自動抽出するシステムを開発した。抽出項目は、米国で実施されている循環器疾患のレジストリ等を参照し、我が国の実情に合わせて改変した上で決定した。

電子カルテ内の抽出対象となったデータには、データとしての抽出が容易な数値データ、カテゴリデータの他に、テキスト形式で書かれた自然言語データが存在する。前者に関しては、自動抽出のシステムを、後者に関しては、循環器疾患の非構造化データの分析に関する基盤整備を行った。

心筋梗塞の診断名に関し、自動抽出するためのアルゴリズムを開発した。傷病名として、ICD-10の3桁分類および傷病名コードを用いた。疑い病名は除いた。傷病名と緊急入院に関する加算の組み合わせにより複数の病名決定ロジックを作成した。疾患特定の感度を重視した定義および疾患特定の特異度を重視した定義である。

| ICD-10 | 心筋梗塞 1 | 心筋梗塞 2 |
|--------|--------|--------|
| I20 | X | ○ |
| I21 | | |
| I22 | | |
| I23 | | |
| I24 | X | ○ |

| | | |
|-----|---|---|
| I25 | X | X |
| 備考 | | 上記病名かつ緊急入院関連コード ・ I20：「不安定狭心症」のみ採用 ・ I23.8「梗塞後狭心症」、I24.1「心筋梗塞後症候群、ドレ ッスラー症候群」の3病名は除く |

IBM Watsonのうち、IBM WKSを用いて、機械学習により自然言語解析を行い、電子カルテに記述された自然言語から該当箇所をデータとして抽出するプログラムを開発した。因子定義と辞書チューニングを繰り返し、心筋梗塞病名その他、特にシステムティックレビューおよびメタアナリシスに必要な主要項目（研究デザイン、症例数、血圧、曝露群、比較群、統計量、信頼区間、p値）に関する機械学習を応用した自然言語解析を行い、通常のレビューワーとの一致率と同程度となった。

| 項目 | カテゴリ | 使用ツール | Precision 精度 | Recall カバー |
|---------------------|------|-------|-----------------|---------------|
| 研究デザイン | 定性 | WEX | 1 | 1 |
| 症例数 | 定量 | WKS | 0.96 | 0.93 |
| 収縮期/拡張期血圧・ 内服下血圧 | 定量 | WEX | 0.82 | 1 |
| 血圧 | 定量 | WEX | 0.82 | 0.96 |
| 曝露群(Exposure) | 定性 | WEX | 0.56 | 0.67 |
| 比較群(Control) | 定性 | WEX | 0.56 | 0.67 |
| 統計量 | 定量 | WEX | 0.79 | 0.92 |
| P値 | 定量 | WEX | 1* | 1 |
| 信頼区間 | 定量 | WEX | 1* | 0.9 |

自然言語分析の追加により、電子カルテからの必要抽出項目の1/3程度が補足可能となった。

診断名の妥当性の検証においては、心筋梗塞に関しては97.5%一致し、別病名は2.5%で、その内訳は、心停止、蘇生に成功した心停止、遅れの心筋梗塞・心不全であった。心不全に関しては、一致率は73.9%であり心筋梗塞より低値であった。誤分類の内訳は、冠動脈疾患、高血圧、心房細動、大動脈弁狭窄症等であった。

また、所属機関の建替・移転に伴い電子カルテシステムが新しく入れ替えられ、移転後の新システムにおいてデータを抽出するための見直しが必要となった。新システムにおいても日常の臨床で得られたデータの自動抽出を可能にするため、旧システムと新システムとのマッピングを行い、プログラムの修正を行った。これらの統合データベースより、予測アルゴリズムの開発を行った。

本研究により、電子カルテからの自動抽出および人工知能による自然言語分析による情報の自動抽出、因子間の関係性の探索、予測精度向上という、最小の予測因子の違いで良好な結果を予測可能にするアルゴリズムを提案することができた。引き続き学習用データを増やし、より汎用的な機械学習モデルを作成予定である。

5. 主な発表論文等

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

| | |
|---|-----------------|
| 1. 著者名 中尾（舛方）葉子 | 4. 巻 - |
| 2. 論文標題 医療ビッグデータや人工知能を用いた疾病対策および先端医療の研究開発 人生100年時代のデジタルヘルス | 5. 発行年 2021年 |
| 3. 雑誌名 臨床雑誌『内科』 | 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） なし | 査読の有無 無 |
| 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 中尾（舛方）葉子，中尾一泰 | 4. 巻 127 |
| 2. 論文標題 女性の胸痛 | 5. 発行年 2021年 |
| 3. 雑誌名 臨床雑誌『内科』 | 6. 最初と最後の頁 1091-1094 |
| 掲載論文のDOI（デジタルオブジェクト識別子） なし | 査読の有無 無 |
| 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 - |

〔学会発表〕 計3件（うち招待講演 3件 / うち国際学会 0件）

| |
|--|
| 1. 発表者名 中尾葉子 |
| 2. 発表標題 医療ビッグデータを用いた本邦循環器診療の現状・質の把握、その到達点と課題 日英でのデータ解析の経験から |
| 3. 学会等名 第57回日本循環器予防学会学術集会パネルディスカッション（招待講演） |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 Yoko M. Nakao |
| 2. 発表標題 Gender differences of Diagnostic and Prognostic Role for Coronary Artery Calcification in Patients with Suspected Coronary Artery Disease : Current Evidence and Insights from the NADESICO study. |
| 3. 学会等名 第83回日本循環器学会AHA-JCS Joint Symposium Emerging Advance for Heart Disease in Woman（招待講演） |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 中尾葉子 |
| 2. 発表標題 心不全予防における生活習慣改善 - 高まる管理栄養士の専門性と役割 - |
| 3. 学会等名 日本臨床栄養学会シンポジウム（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 |
|-------|-----------------------------------|-----------------------|----|
| 研究協力者 | 宮本 恵宏 (Miyamoto Yoshihiro) | | |
| 研究協力者 | 安田 聡 (Yasuda Satoshi) | | |
| 研究協力者 | 安齊 俊久 (Anzai Toshihisa) | | |
| 研究協力者 | 野口 暉夫 (Noguchi Teruo) | | |
| 研究協力者 | 平松 治彦 (Hiramatsu Haruhiko) | | |

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 |
|-------|-----------------------------------|-----------------------|----|
| 研究協力者 | 草野 研吾 (Kusano Kengo) | | |
| 研究協力者 | 西村 邦宏 (Nishimura Kunihiro) | | |
| 研究協力者 | 泉 知里 (Izumi Chisato) | | |
| 研究協力者 | 穴戸 稔聡 (Shishido Toshiaki) | | |
| 研究協力者 | 住田 陽子 (Sumita Yoko) | | |

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 |
|---------|---------|
| | |