

厚生労働科学研究費補助金(臨床研究等 ICT 基盤構築・人工知能実装研究事業)
総括研究報告書

機械学習を活用した診療情報の体系的な把握・分析に基づく、疾患との新たな関連性を発見するための研究

研究代表者 金谷 泰宏 国立保健医療科学院・健康危機管理研究部

研究要旨

稀少疾患は、症例が少ないが故に疾患概念を構築することが難しい。わが国においては平成13年度より全国規模で稀少疾患に関する患者情報を登録する特定疾患調査解析システムより症例の集積が行われてきた。本研究においては、これらデータベースを用いて人工知能による機械学習を試みることで、症例数が少なく、臨床所見、画像診断、遺伝子診断を総合的に組み合わせることで正確な診断が得られる多系統萎縮症(MSA)、脊髄小脳変性症(SCA)を取り上げ、人工知能による診断プロセスの妥当性について検証を試みた。また、腎臓疾患についても初診時における臨床所見と検査所見から重症化の可否との推計が人工知能で可能となるかについて検証を試みた。人工知能によるMSAの診断結果に関する検証では、SND及びOPCAはほぼ人工知能の診断結果と専門医の判定は一致することが認められた。一方で、SDSはSNDとOPCAに含まれるものがそれぞれ15%ずつ認められた。SCAに関しては、孤発性および痙性対麻痺でほぼ全数が専門医の診断と一致したが、常優性では80%、他遺伝性および常劣性は全数で専門医の診断の一致が認められず、孤発、常優性及び痙性対麻痺に分類される等、診断プロセスにおける課題が示唆された。腎臓疾患については、データベースの構築を行い、慢性化の指標を明らかにしたところである。

分担研究者

市川 学	国立保健医療科学院健康危機管理研究部 主任研究官
江藤亜紀子	国立保健医療科学院健康危機管理研究部 上席主任研究官
富田奈穂子	国立保健医療科学院国際協力研究部 主任研究官
佐々木秀直	北海道大学大学院神経内科学教室 特任教授
眞野 訓	順天堂大学革新的医療技術開発研究センター 准教授

A . 研究目的

本研究は、厚生労働省が管理する難治性疾患データベースを活用し、人工知能を用いて診断基準の妥当性、診断基準と

の関連性が高い項目を明らかにする。特に、早期における診断が困難とされる神経疾患、腎臓疾患を取り上げ、専門医による診断と人工知能による診断との乖離を検証するものである。これらの検証を踏まえ、専門医以外により提供された診療情報から確実に対象となる疾患を絞り込めるプログラムを開発し、都道府県等での実装を目指す。

B . 研究方法

B.1 研究計画

初年度においては、厚生労働省における特定疾患治療研究事業によって収集されたデータベースを活用し、機械学習プログラムを利用して神経難病のうちMSAを、腎臓疾患として顕微鏡的多発血管炎を対象として調査票における登録項目と確定診断との関係性について検討を行う(市川)。さらに、機械学習によって得

られた各項目間の関連性について共分散構造分析を用いて検証を行う（江藤）。欧米の登録項目を参考に神経疾患及び腎臓疾患の病態把握に向けたミニマムデータセットを作成する（富田）。北海道大学病院においては、Hokkaido Rare diseases Consortium for MSA (HoRC-MSA)を用いて人工知能による診断プログラムの妥当性の検証を行う（佐々木）。また、ミニマムデータセットを用いた腎臓領域における症例データベースの構築を行う（眞野）。

2年目においては、初年度における機械学習の結果を踏まえ、神経疾患（14疾患）及び腎臓疾患（3疾患）を対象とした診断プログラムの構築を行う（市川、江藤）。さらに、機械学習により登録項目の妥当性を検証し、神経疾患及び腎臓疾患に対するミニマムデータセットを構築する（富田、江藤）。北海道大学病院及び順天堂大学革新的医療技術開発研究センターにおいて機械学習による診断アルゴリズムの検証を行う（佐々木、富田・眞野）。

3年目においては、2年次に設計した神経疾患及び腎臓疾患の診断プログラムの改修と検証を行う（市川、江藤）。2年次までの研究で明らかになった予後因子をアウトカムとして登録項目との関連性を共分散構造分析により明らかにする（江藤）。北海道大学病院及び順天堂大学革新的医療技術開発研究センターにおいてミニマムデータセット、診断プログラムの実装を行う（佐々木、富田・眞野）

B.2 研究資料

研究資料については、厚生労働省に登録された特定疾患治療研究事業・臨床調査個人票データベースを活用する。なお、当該データの利用については、厚生労働省健康局難病課より利用承認を得ている（健疾発 0708 第 1 号。平成 22 年 7 月 8 日）。また、地域における神経疾患のデータベースとして平成 27 年度より

AMED の研究支援を受けて開始された Hokkaido Rare diseases Consortium for MSA (HoRC-MSA) との連携を図る。

（倫理面への配慮）

疫学研究に関する倫理指針を踏まえ、各機関において承認を得た。

C . 研究結果

C.1 神経難病データベースの構築

神経難病に関しては、2004～08 年度まで厚生労働省特定疾患調査解析システムに登録のあった脊髄小脳変性症（SCD）の新規登録症例 7,073 例、多系統萎縮症（MSA）の新規登録症例 4,957 例のデータクリーニングを行い解析用のデータベースを構築した。

C.2 人工知能エンジンの設計

機械学習用ライブラリ Chainer (<https://chainer.org>) を用いて 3 層構造からなるニューラルネットワークを構築した。

C.3 人工知能による神経難病の解析

ミニマムデータ・セットとして、神経分野に関して試行的に 初発症状、発病様式・経過、神経学的初見、画像所見、生活状況の各項目（NeuroI Med Chir, 2017）を選択し、ニューラルネットワークによる解析を試みた。多系統萎縮症としてデータが揃っている 3,577 例（SND894 例, SDS377 例, OPCA2,106 例）を用いて、各 10 例をテスト用データとして残る 3,347 例で機械学習を行った。この中でデータ欠損値が多い画像診断の有無が診断一致率（専門医の診断と AI の診断が一致）に与える影響を評価できた。画像診断無しの場合、一致率は SND64%、SDS 0%、OPCA85%であったが、有りの場合、SND と OPCA は 90%、SDS は 70%まで向上した。このように MSA では画像診断の有無が大きく診断結果に影響することが認められた。一方、SCD に関しては、画像診断の結果の有無にかかわらず、常劣型は一致率が 0%にとどまるなど、疾患毎で大きく異なる結果となっ

ている。この背景として機械学習可能な症例数が少ないことがあげられる。

C.4 腎疾患データベースの構築

腎臓分野については、特定疾患調査解析システムよりループス腎炎の定義を満たす1,905例（18歳未満130例、18歳以上の大人1,775例）を抽出し、データセットとして、性別、登録時・発症時の年齢、臨床症状、検査所見（尿タンパク、赤血球尿、顆粒円柱、血清Cr値）、治療状況、合併症を選択した。

D. 考察

D.1 診断基準

診断基準とは、臨床所見、検査所見、画像所見、鑑別診断等の様々な客観的な視点から一定の共通する要素を持ち合わせた疾患の集団を捉えるために研究者の合意に基づいて設定される。このため、新たな検査法の出現に伴い、診断基準は見直される。とりわけ、数が少ない稀少疾患においては、数が少ないことから対象とする集団の全容を把握することは難しく、発症からの時間経過、何らかの介入により臨床病態は大きく修飾される。これが結果として確定診断を遅らせる原因ともなっている。そこで、本研究では、専門医が現行の診断基準で「ほぼ確定」以上の判断を下し、国としてデータ登録を行ってきた特定疾患治療研究事業対象疾患を対象に、登録時の臨床調査個人票データを用いて、機械的に過去の診断プロセスを学習させることで、神経難病の多系統萎縮症（MSA）及び脊髄小脳変性症（SCA）の診断が可能かについて検証を試みた。

D.2 人工知能の設計と課題

人工知能としてChainerを用いたが、当初2層で構築したものの回答を導くことが困難であり、最終的には3層構造とした。加えて、機械学習を行う上で、欠損の多いデータは除外するか、欠損した情報を補完する必要がある。また、今回の検証では、臨床調査個人票のデータ

を用いたが、北海道大学におけるMSAのデータベース（HoRC-MSA）に使用されている患者レコードの構造、使用されている各項目と必ずしも1対1で対応しておらず、事後、本システムの普及を図る上で、患者レコードの標準化が不可欠であると考えられた。さらに、今回使用したデータベースにおいて、臨床調査個人票において、開眼時立位能力の項目として、「支持なしで立位可能な場合」と「自力立位不可能な場合」の双方を選択しているケースが相当数散見された。このようなケースをそのまま機械学習にかけた場合、誤った解釈を招くこととなるため、論理的な個人票の設計が求められる。

D.3 MSA 及び SCA の解析結果

MSA 及び SCA の解析に介して、解析に適さない症例が、MSA で 4,949 例中 1372 例、SCA で 7,073 例中 2,241 例が認められる等、今後の人工知能の実装における課題である。今回の検証で、とりわけ欠損値が多い画像情報を外した場合、診断一致率は 70%にも満たない等、データによって大きく影響するもの、そうでないものの判別を進める必要がある。我々が設計した人工知能による MSA の診断結果に関する検証では、SND 及び OPCA はほぼ人工知能の診断結果と専門医の判定は一致することが認められた。一方で、SDS は SND と OPCA に含まれるものがそれぞれ 15%ずつ認められた。SCA に関しては、孤発性および痙性対麻痺でほぼ全数が専門医の診断と一致したが、常優性では 80%、他遺伝性および常劣性は全数で専門医の診断の一致が認められず、孤発、常優性及び痙性対麻痺に分類される等、診断プロセスにおける課題が示唆された。

D.4 腎臓領域への応用

腎臓に関しては、神経難病と異なり、診断そのものを機械学習するのではなく、

長期的な予後観察（本研究では観察期間は最大3年間）で人工透析に至った症例を1として、状況が維持されているものをその他0として機械学習にかけることで、重症化するケース、そうでないケースを予測することが可能となる。

E. 結論

稀少疾患は、症例が少ないが故に疾患概念を構築することが難しい。このため、わが国においては平成13年度より全国規模で稀少疾患に関する患者情報を登録する特定疾患調査解析システムより症例の集積が行われてきたところである。そこで、本研究においては、これらデータベースを用いて人工知能による機械学習を試みることで、症例数が少なく、臨床所見、画像診断、遺伝子診断を総合的に組み合わせることで正確な診断が得られる多系統萎縮症(MSA)、脊髄小脳変性症(SCA)を取り上げ、人工知能による診断プロセスの妥当性について検証を試みた。また、腎臓疾患についても初診時における臨床所見と検査所見から重症化の可否との推計が人工知能で可能となるかについて検証を試みた。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

- 1) Chang S, Ichikawa M, Deguchi H, Kanatani Y. Optimizing the Arrangement of Post-Disaster Rescue Activities: An Agent-Based Simulation Approach. JACIII 2017,21:1202-1210
- 2) Chang S, Ichikawa M, Deguchi H, Kanatani Y. A General Framework of Resource Allocation Optimization and Dynamic Scheduling. JCMSI 2017,10:77-84
- 3) Yoshida K, Kuwabara S, Nakamura K, Abe R, Matsushima A, Beppu M, Yamanaka Y, Takahashi Y, Sasaki H, Mizusawa H; Research Group on Ataxic Disorders. Idiopathic cerebellar ataxia (IDCA): Diagnostic criteria and clinical analyses of 63 Japanese patients. J Neurol Sci. 2018; 15(384): pp30-35.
- 4) Yaguchi H, Takeuchi A, Horiuchi K, Takahashi I, Shirai S, Akimoto S, Satoh K, Moriwaka F, Yabe I, Sasaki H. Amyotrophic lateral sclerosis with frontotemporal dementia (ALS-FTD) syndrome as a phenotype of Creutzfeldt-Jakob disease (CJD)? A case report. J Neurol Sci. 15(372):444-446, 2017.
- 5) Kanatani Y, Tomita N, Sato Y, Eto A, Omoe H, Mizushima H. National Registry of Designated Intractable Diseases in Japan: Present Status and Future Prospects. Neurologia medico-chirurgica, 2017;57(1): pp1-7.
- 6) Tomita N, Mano S, Nakagawa Y, Kanatani Y. Severity of Kidney Disease In Systemic Lupus Erythematosus. Value in Health 2017; 20(9): A492.
- 7) 金谷泰宏、市川学. 医療白書 2017-2018年版 AIが創造する次世代型医療. 東京. 独立行政法人情報処理推進機構 AI白書編集委員会. 2017, pp34-39.

2. 学会発表

(発表誌名巻号・頁・発行年等も記入)

- 1) Kanatani Y. Perspectives in satellite and simulation technologies for disaster response. World Bosai Forum IDRC 2017 in Sendai, Miyagi, 2017
- 2) Kanatani Y. Medical responses to CBRNe accidents. Non-

Conventional Threat(NCT) Asia
2017 and the 8th
SISPAT(Singapore International
Symposium for Protection Against
Toxic Substances) conference and
exhibition, Singapore, 2017

(予定を含む。)

1. 特許取得
なし
2. 実用新案登録
なし
3. その他
なし

H . 知的財産権の出願・登録状況