

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)
分担研究報告書

人工知能を用いた診断アルゴリズムの設計に関する研究

分担研究者 市川 学 芝浦工業大学システム理工学部環境システム学科 准教授

研究要旨

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。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、Shy-Drager syndrome (SDS)における診断一致率は70%台と低く、原因としてシェロング試験に関する情報が欠けており、これが影響したと考えられた。そこで、これらの情報を補うとともに診断制度の向上に向けたモデルの検証と診断に大きく影響する項目について検証を試みた。

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。最終年度においては、すでに構築した診断プログラムの精度向上に向けた検証を試みた。

B. 研究方法

B.1 研究計画

初年度に解析を行ったMSAデータについて、シェロングテストデータを補い、以下の手順に従い、診断結果を予測するモデルを作成し、検証を行う。

1. データの欠損の確認
2. 欠損値の補完
3. 各データの相関係数の導出
4. 決定木学習による予測モデルの作成
5. 予測モデルの検証
6. 予測モデルの精度向上

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)。(倫理面への配慮)

該当なし

C. 研究結果

C.1 データの欠損の確認

データ分析を開始する前に、使用できるデータを選択するため欠損値の確認を行った。図1に欠損値があったデータ項目と欠損率一覧を示す。

表1 各項目別の欠損値一覧

	欠損数	%
収縮期前(臥位)	1883	56.495650
拡張期前(臥位)	1884	56.525653
収縮期後(立位)	1898	56.945695
拡張期後(立位)	1898	56.945695
収縮期前後差	1901	57.035704
拡張期前後差	1901	57.035704
呼吸障害	97	2.910291
その他神経所見	6	0.180018
異常の有無	35	1.050105
1.小脳萎縮	43	1.290129
2.脳幹萎縮	43	1.290129
3.橋の異常信号(十字サイン)	51	1.530153
4.線条体萎縮又は異常信号	57	1.710171
5.第3脳室拡大	56	1.680168
6.大脳萎縮	57	1.710171
7.大脳白質病変	56	1.680168

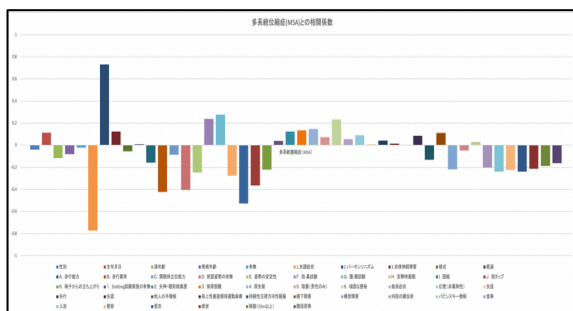
C.2 欠損値の補完

欠損が発生しているデータに対して、どのような補完を行うかを設計する。データの欠損が著しく多い場合は、当該変数の削除も検討する。発病年齢は欠損率 25.8%であるが、重要な指標であることから、中央値による補完を行う。その他の欠損値は、症状がないと仮定して、0 を代入した。

C.3 各データの相関係数の導出

補完したデータを用いた相関関数を以下に示す。

図1 各データと診断結果との相関



MSA (多系統萎縮症) に関するもののみ (相関係数が $|0.3|$ 以上の項目) の相関関係を示す。

相関の強い項目

- | | |
|-------------|--------|
| 1. 失調症状 | -0.773 |
| 2. パーキンソニズム | +0.731 |
| B. 歩行異常 | -0.422 |
| D. 前屈姿勢の有無 | -0.404 |
| I. 固縮 | -0.529 |
| J. 指タップ | -0.364 |

上記の項目が多系統萎縮症(MSA)の診断に強く影響を与えていることがわかる。

C.4 決定木学習による予測モデルの作成

以下の2つのモデルを実装し、予測を選定したプロセスを可視化した(図2)。

- A 全ての変数を使用し作成したモデル
- B 相関係数が $|0.3|$ 以上の変数を用いたモデル

図2a 全ての変数を用いたモデル



図2b 相関係数が $|0.3|$ 以上の変数を用いたモデル



C.5 予測モデルの検証

学習した決定木の予測モデルを評価するために、Leave One Out (以下 loo) を用いて交差検証を行った。

A 全ての変数を用いたモデル

モデル正答率; all_feature 0.8149

B 相関係数が $|0.3|$ 以上の変数を用いたモデル

モデル正答率; select_feature 0.8473

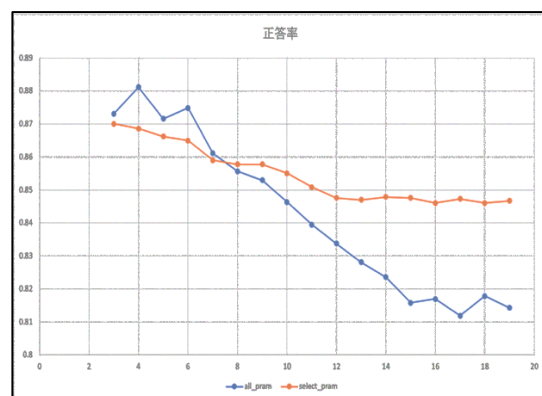
相関の強い変数のみを使用したものの方がより精度の高いモデルであることが示された。

C.6 モデルの予測精度向上

説明変数が多い場合、予測モデルが過学習となることで予測精度が悪くなる可能性が考えられる。そこで、決定木におけるハイパーパラメータを設定した。決定木におけるハイパーパラメータとして、「木の深さ」のみを対象に、max_depth3~20の区間でチューニングを実施した。

「木の深さ」を増やすことで過学習に陥り、判定が正答率を下回ることがわかった。正答率が最も高くなるモデルは、全てのパラメータを使用し、max_depth=4としたもので正答率は88.2%であった(図3)。

図3 Max_depth と正答率



D. 考察

本研究においては、MSA の症例データを用いて専門医による診断と機械による診断結果との一致率の向上に向けた検証を行った。これまでの検証の中で、MSA のうち SND と OPCA については80%を超える診断精度が得られたが、SDS では診断精度は70%台と低く、その原因として SDS に特異的に認められるシェロング試験データを補うことで診断精度の向上が得られるか検証を行った。結果として、診断精度はMSA 全体として80%を超える診断一致率が得られた。

しかしながら、MSA の診断と各項目との相関を検証した結果、失調症状 (-0.773)、パーキンソニズム (0.731)、歩行異常 (-0.422)、前屈姿勢の有無 (-0.404)、固縮 (-0.529)、指タップ (-0.364) の6項目が高い相関を示したが、シェロング試験についてはデータ欠損率も高いこともあり、相関は得られなかった。逆に、診断精度の向上には、相関の高い項目 (相関係数が|0.3|以上の項目) を選択的に用いたほうが、より高い正答率を示した。

一方で、決定木におけるハイパーパラメータとして、本研究では「木の深さ」を用いたが、max_depth を4に設定した場合、all_feature は、select_feature と比較して、88.2%の正答率を示した。図3に示すように、all-feature は max-depth を大きくするに従い正答率が下がる反面、select feature における正答率はmax-depth を増大しても一定レベルを維持することが示された。すなわち、今後の個人票の見直しを検討する際に、本疾患においてはより項目を絞ることで診断精度

を高めることができることを示唆した。

E. 結論

人工知能によるMSAの診断精度の向上に向けて、失調症状、パーキンソニズム、歩行異常、前屈姿勢の有無、固縮、指タップの6項目が診断と大きく相関する傾向が示された。また、診断に必要なとされる項目を増やしても逆に過学習となることで診断精度の低下を招くことが示された。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表
なし
2. 学会発表
(発表誌名巻号・頁・発行年等も記入)
なし

H. 知的財産権の出願・登録状況

(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