

報告番号	甲 乙 第	号	氏 名	前田 絢子
主論文題名 :				
機械学習による漢方薬の処方支援及び服薬指導支援システムの開発に関する研究				
(内容の要旨)				
【緒言】				
<p>医療用、一般用漢方製剤の生産金額はともに上昇しており、病院および保険薬局では8割以上の施設で漢方薬の調剤が行われている。セルフメディケーションも推進されており、今後さらに薬剤師による一般用漢方製剤の選択および服薬指導の重要性が高まると考えられる。この潮流は世界的傾向であり、2019年に世界保健総会にて承認された国際疾病分類第11回改訂版(ICD-11)では、中国、韓国、日本で実践されている東アジア伝統医学の診断分類の章が新たに追加された。</p> <p>個々の漢方薬には、「薬方の証」とよばれる証がある。漢方専門医は漢方薬を選択する際、四診(望診、聞診、問診、切診)の結果をもとに診断した「患者の証」と合致する「薬方の証」をもつ漢方薬を選択する。したがって、漢方薬を決定するためには、患者の正しい証診断が必要である。証の見極めは医師の五感や経験などに拠るところが多いが、今後使用経験の少ない薬剤師であっても、一般用漢方製剤の選択及び証に基づく服薬指導ができるようになることが求められる。</p> <p>以上から、本研究は薬剤師がより簡便で適切にセルフメディケーションにおける一般用漢方薬の選択、服薬指導、処方支援を行うために証診断を支援するシステムを開発することを目的とし、次の4点について検討した。</p> <p>検討1) 患者の情報文書を利用した漢方専門医による証診断の一致率          検討2) 施設別虚実と寒熱予測モデルの構築          検討3) 複数施設統合データを用いた虚実と寒熱の <b>balanced</b> 予測モデルの構築          検討4) 複数施設統合データを用いた虚実中間証を含む虚実予測モデルの構築</p> <p><u>検討1) 患者情報文書を利用した漢方専門医による証診断の一致率</u></p> <p>&lt;目的&gt; 漢方専門医が患者診察時に診断した証と患者情報文書から評価した証の一致率および3人の医師の評価者間信頼性を計算した。</p> <p>&lt;方法&gt; 対象者は、2014年10月から2017年12月に慶應義塾大学病院漢方医学センターを初めて受診し、参加に同意した患者とした。上記期間の初診患者は1,286名からデータ欠損、20歳未満、気血水の descriptor が2つ以上付与されているあるいは気血水の付与がない患者を除外した638名の中から、患者数が少ない気逆を除き、各気血水の6つの descriptor が均一に分布するように、典型的な気血水の症状を有しており、医師AおよびBの初診患者である64名の成人患者を有意抽出した。医師は、初診から3~5年経過後、初診時に取得した133の症状、主訴、収縮期および拡張期血圧、年齢、性別、身長、体重を含む140項目が記載されている患者情報文書のみから、患者の虚実、寒熱、気血水を評価した。診察時に診断した証と患者情報文書より評価した証の一致率および</p>				

3人の医師の評価者間信頼性は、虚実は線形に重み付けし、寒熱と気血水は重み付けせず算出した。評価者間信頼性は虚実では Gwet の一致係数(Agreement Coefficient; AC)2、寒熱と気血水では AC1 を算出した。

<結果>診察時に診断した証と患者情報文書より評価した証の一致率は、虚実が 0.781-0.813、寒熱が 0.484-0.688、気血水が 0.281-0.375 であった。評価者間信頼性は、虚実の一致率と AC2 は 0.859 と 0.708、寒熱の一致率と AC1 は 0.646 と 0.542、気血水の一致率と AC1 は 0.359 と 0.254 であった。

<考察>本検討は、漢方専門医の診断の一致率について初めて算出した研究であり、予測モデルを作成した際の判別率の指標となるデータを収集するためのパイロットスタディという位置づけで行った。漢方専門医が患者診察時に診断した証と患者情報文書から評価した証の一致率は、虚実、寒熱、気血水でそれぞれ 0.8、0.6、0.3 割程度であり、3人の医師の評価者間信頼性は同程度であった。

## 検討 2) 施設別虚実と寒熱予測モデルの構築

<目的>各施設における虚実と寒熱の診断状況を把握し、各々の施設データを用いて虚実と寒熱の予測モデルを作成し、予測に寄与度が高い重要問診項目を抽出し比較した。

<方法>対象者は、千葉大学、飯塚病院、慶應義塾大学、富山大学、亀田総合病院、自治医科大学の6施設の漢方外来に最長2012年1月から2015年2月までの期間に初めて受診した患者で、研究に同意した患者とした。除外基準は、データ欠損、20歳未満、問診項目回答数が20項目未満とした。虚実と寒熱の診断状況については、施設毎に虚実、寒熱の診断数を数えた。予測モデルの説明変数は、148の問診項目、血圧(収縮期、拡張期)、年齢、性別、BMIの計153項目とした。各施設からランダムに40名の対象者(虚実:虚証20名と実証20名、寒熱:寒証20名と熱証20名)をサンプリングし学習データとした。ランダムフォレストのアルゴリズムを使用し、施設毎に40名の対象者を用いた予測モデルを構築した。虚実の場合は虚証と実証の患者が、寒熱の場合は寒証と熱証の患者がそれぞれ20名以上いる施設を対象としたため、虚実は、千葉大学、飯塚病院、慶應義塾大学、富山大学、寒熱は、千葉大学、飯塚病院、富山大学のみモデルを作成した。虚実及び寒熱の予測に寄与度が高い重要項目を明らかにするために、施設毎に重要な30項目を抽出した。虚実では3つまたは4つ全ての施設、寒熱では3つ全ての施設で共通にみられた項目を重要項目から抽出した。

<結果>6施設において登録された初診患者数は、千葉大学が925名、飯塚病院が791名、慶應義塾大学が781名、富山大学が501名、亀田総合病院が424名、自治医科大学が59名であった。予測モデル対象患者数は、千葉大学が412名、飯塚病院が570名、慶應義塾大学が429名、富山大学が367名、亀田総合病院が108名、自治医科大学が8名であった。診断状況は、虚実では全施設で虚証の割合が高く、実証の割合が低かった。また、寒熱では全施設で寒証の割合が高く、熱証の割合が低かった。虚実について、4施設で共通に含まれていた項目は8項目で、BMI、寒がり、拡張期血圧、収縮期血圧、年齢、全身が冷える、目が疲れる、肩のこりであった。3施設で共通に含まれていた項目は上記8項目に加えて7項目で、暑がり、身体が重たく感じる、足がつる、足がふらつく、朝起きにくく調子がでない、ガスがよく出る、気分が憂うつになるであった。寒熱について、3施設で共通に含まれていた項目は8項目で、BMI、収縮期血圧、足が冷える、寒がり、拡張期血圧、年齢、目が疲れる、肩のこりであった。虚実、寒熱ともに全施設でBMIが1番重要な項目であった。

<考察>日本の漢方専門施設の患者は、虚実では虚証の割合が高く、実証の割合が低く、

寒熱では寒証の割合が高く、熱証の割合が低く、患者分布に同様な傾向がみられた。患者問診結果とバイタルサインを用いて虚実と寒熱の診断を予測するために重要な項目を明らかにした。虚実では共通した重要項目が 15 項目、寒熱では共通した重要項目が 7 項目であった。この結果は、対象施設の医師が同様の方法で虚実及び寒熱を診断していることを示唆している。

各施設のデータセットから構築した虚実および寒熱予測モデルにおいて、BMI がどの施設においても最も重要な項目であり、これは先行研究と一致していた。また血圧、寒がりや冷えなども虚実、寒熱両方で重要項目として挙げられた。これらの結果は、虚実と寒熱の間に強い関係があることを示唆している。

### 検討 3) 複数施設統合データを用いた虚実と寒熱の balanced 予測モデルの構築

<目的>検討 2)において、虚実と寒熱の間に強い関係が見られた。関係が見られた要因は 2 点考えられる。1 点目は、虚実と寒熱の定義そのものに強い関係があるということ、2 点目は、学習データでの患者の証の分布が Unbalance であるということだ。1 点目については、ICD-11 の定義では、虚実と寒熱の定義には強い関係はないため考えにくい。2 点目については、検討 2)で得られた患者の証の実態は、虚証かつ寒証の患者数 (534 名) が虚証かつ熱証の患者数 (57 名) の約 9 倍、実証かつ熱証の患者数 (111 名) が実証かつ寒証の患者数 (48 名) の約 2 倍であり、虚実と寒熱の患者数が Unbalance であったために重要項目の重複が見られた可能性が考えられる。学習データの虚実と寒熱のお互いの影響を小さくし、外的妥当性がより高い予測モデルを構築するため、虚実及び寒熱予測モデルを再評価した。

<方法>検討 2)と同じ 6 施設のレジストリを使用した。除外基準は、各種データ欠損、データ重複、20 歳未満、問診回答数 20 項目未満、および対象としない証 (虚実中間証、寒熱中間証、寒熱錯雑証) と診断された患者とした。

まず、先行研究および検討 2)と同様な学習データの抽出方法で虚実および寒熱の Unbalanced 予測モデルを構築した。虚実は、虚証 80 名、実証 80 名、寒熱は、寒証 80 名、熱証 80 名の計 160 名の対象者を無作為抽出し、ランダムフォレストのアルゴリズムの学習データとした。次に、虚実と寒熱の Balanced 予測モデルを構築した。実証かつ熱証、実証かつ寒証、虚証かつ寒証、虚証かつ熱証の対象者をそれぞれ 40 名ずつ無作為抽出した。虚実では、虚証かつ寒証の 40 名と虚証かつ熱証の 40 名をあわせた 80 名を虚証の学習データとし、実証かつ熱証の 40 名と実証かつ寒証の 40 名をあわせた 80 名を実証の学習データとした。なお寒熱も同様に作成した。

各証より学習データを除いたデータをテストデータとし、それぞれのモデルにおいて判別率 (Accuracy)、適合率 (Precision)、再現率 (Recall)、F 値 (F-measure) を算出した。虚実および寒熱の Unbalanced と Balanced 予測モデル間で判別率を比較した。それぞれのモデルにおいて、予測に寄与度が高い重要項目上位 10 項目を抽出した。虚実と寒熱の Unbalanced 予測モデル間、虚実と寒熱の Balanced 予測モデル間、虚実および寒熱の Unbalanced と Balanced 予測モデル間で重要項目を比較した。

<結果>適格患者数は 750 名 (虚証かつ寒証 : 534 名、虚証かつ熱証 : 57 名、実証かつ寒証 : 48 名、実証かつ熱証 : 111 名) であった。

虚実では、Unbalanced 予測モデルの判別率、適合率、再現率、F 値は、それぞれ 0.819、0.933、0.828、0.877、Balanced 予測モデルでは、それぞれ 0.808、0.944、0.828、0.882 であった。寒熱では、Unbalanced 予測モデルの判別率、適合率、再現率、F 値は、それぞれ 0.773、0.851、0.951、0.898 で、Balanced 予測モデルでは、それぞれ 0.749、

0.947、0.810、0.873 であった。虚実、寒熱ともに Unbalanced と Balanced 予測モデルで判別率に差は見られなかった。

Unbalanced 虚実および寒熱予測モデルの重要項目上位 10 項目のうち、共通していた重要項目は、BMI、収縮期血圧、および寒熱に関係する自覚症状 4 つの計 6 項目であった。Balanced 虚実および寒熱予測モデルの重要項目上位 10 項目のうち、共通していた重要項目は、BMI、収縮期血圧の 2 項目のみであった。虚実の Unbalanced と Balanced 予測モデルの重要項目のうち、一番重要度が高い項目はともに BMI であった。寒熱に関係する自覚症状は、Unbalanced 予測モデルでは重要項目として挙げられたが、Balanced 予測モデルでは 11 位以下になった。寒熱において、Balanced 予測モデルでは重要項目として寒熱に関係する自覚症状 6 つが挙げられ、Unbalanced 予測モデルよりも上位であった。BMI は Unbalanced 予測モデルで 1 位であったが、Balanced 予測モデルでは 8 位に下降した。

＜考察＞Balance を取ってサンプリングした学習データを用いた虚実と寒熱の予測モデルを構築した。虚実と寒熱の Unbalanced と Balanced 予測モデルの判別率に差は見られなかった。検討 1) で得られた診察時に診断した証と患者情報文書より評価した証の一致率と Balanced 予測モデルの判別率に差はなかった。また、Unbalanced モデルであると考えられる先行研究の結果と比較して近い値であった。

Unbalanced 予測モデルの結果は、先行研究の虚実及び寒熱の予測に重要な項目と一致していたため、先行研究では虚実および寒熱の患者比率を無視してサンプリングした可能性がある。虚実の Balanced 予測モデルの重要項目である BMI、血圧、および寒熱の Balanced 予測モデルの重要項目である寒熱に関係する自覚症状もまた、ICD-11 の証の定義と互換できる可能性がある。Balanced 予測モデルは、より外的妥当性の高いモデルとなっていることが期待される。

#### 検討 4) 複数施設統合データを用いた虚実中間証を含む虚実予測モデルの構築

＜目的＞問診項目とバイタルサインから虚実中間証を含んだ虚実予測モデルを構築した。また施設毎のテストデータを構築した予測モデルにあてはめてカットオフ値を計算し、施設間で比較し、判別率を算出した。

＜方法＞検討 2) と同じレジストリを使用した。データ欠損、データ重複、20 歳未満、問診回答数 20 項目未満、対象でない証（寒熱中間証、寒熱錯雑証）の患者を除外した。6 施設のデータを統合した学習データを用いて虚実中間証を含む虚実予測モデルを構築し、虚実予測に寄与度が高い重要度項目を抽出し、テストデータを用いて虚実予測値を算出した。虚証寒証、虚証熱証、虚実中間証寒証、虚実中間証熱証、実証寒証、実証熱証の 6 つの証の患者をそれぞれ 40 名ずつ無作為抽出して学習データとし、それ以外をテストデータとした。虚実を予測する上で寄与度が高い重要項目上位 3 項目を抽出した。テストデータ毎の虚実予測値を用いて、全施設及び各施設の虚証と虚実中間証、虚実中間証と実証の Receiver Operating Characteristic(ROC) 曲線を作成しカットオフ値を求め、判別率を算出した。

＜結果＞適格患者数は 1068 名（虚証：591 名、虚実中間証：318 名、実証：159 名）であった。重要項目上位 3 項目は、BMI、収縮期血圧、拡張期血圧であった。医師の診断結果をもとに ROC 曲線からカットオフ値を求めたところ、全施設の虚実予測値のカットオフ値は虚証と中間証が 0.5 以下、実証と中間証が 0.5 以上であった。判別率は 0.768 であった。

＜考察＞虚実中間証を含む虚実予測モデルを構築することができた。すべての施設の実

証一虚実中間証、虚実中間証一虚証のカットオフ値は狭い範囲に収まっていることが明らかとなった。検討 1)で得られた結果と判別率を比較したところ差はなかった。

#### 【結語】

本研究では、薬剤師がより簡便で適切に一般用漢方製剤の選択、服薬指導、医師の処方支援を行うために、証診断を支援するシステムを開発した。本研究で構築した予測モデルは、漢方薬の選択に最も重要な、虚実と寒熱の見極めという、使用経験の少ない薬剤師にとって難しい段階を予測可能であることを示すことができた。本予測モデルを用いることで、薬剤師は、利用者の証にあう薬方の証をもつ一般用漢方製剤の選択、医療用漢方製剤の処方監査、患者の証にあった服薬指導や生活指導、医師の医療用漢方製剤の処方支援をより簡便かつ適切に行うことが可能となる。

#### 【主論文に関する原著論文】

- 1) **Maeda-Minami A**, Yoshino T, Horiba Y, Nakamura T, Watanabe K. Inter-rater Reliability of Kampo Diagnosis for Chronic Diseases. *Journal of Alternative and Complementary Medicine*. (in press).
- 2) **前田絢子**, 吉野鉄大, 片山琴絵, 堀場裕子, 引網宏彰, 嶋田豊, 並木隆雄, 田原英一, 南澤潔, 村松慎一, 山口類, 井元清哉, 宮野悟, 美馬秀樹, 三村將, 中村智徳, 渡辺賢治. 虚実中間証を含む虚実証の予測モデル. *日本東洋医学雑誌* 2020;71 (4):315-325.
- 3) **Maeda-Minami A**, Yoshino T, Katayama K, Hikiami H, Shimada Y, Namiki T, Tahara E, Minamizawa K, Muramatsu S, Yamaguchi R, Imoto S, Miyano S, Mima H, Mimura M, Nakamura T, Watanabe K. Discrimination of prediction models between cold-heat and deficiency-excess patterns. *Complementary Therapies in Medicine*. Mar 2020; 49: 102353.
- 4) **Maeda-Minami A**, Yoshino T, Katayama K, Hikiami H, Shimada Y, Namiki T, Tahara E, Minamizawa K, Muramatsu S, Yamaguchi R, Imoto S, Miyano S, Mima H, Mimura M, Nakamura T, Watanabe K. Prediction of deficiency-excess pattern in Japanese Kampo medicine: multi-centre data collection. *Complementary Therapies in Medicine*. Aug 2019; 45: 228-233.