

博士学位論文 2020（令和2）年度

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

要約版

慶應義塾大学大学院薬学研究科

前田 絢子

## 目次

緒言 .....	1
<b>第一章 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 .....</b>	<b>5</b>
1. 背景と目的 .....	5
2. 方法 .....	5
2.1. 対象者 .....	5
2.2. 患者情報文書 .....	6
2.3. 二人の医師による評価者内信頼性（医師 A と B） .....	8
2.4. 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（医師 A、B と C） .....	8
2.5. 三人の医師による評価者間信頼性（IRR）（医師 A、B と C） .....	9
2.6. 統計分析 .....	9
2.7. 倫理審査 .....	9
3. 結果 .....	10
3.1. 対象者 .....	10
3.2. 二人の医師による評価者内信頼性（医師 A と B） .....	10
3.4. 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（医師 A、B と C） .....	10
3.4. 三人の医師による評価者間信頼性（IRR）（医師 A、B と C） .....	11
4. 考察 .....	11
5. 小括 .....	12
<b>第二章 施設毎の虚実と寒熱の予測モデルの構築 .....</b>	<b>13</b>
1. 背景と目的 .....	13
2. 方法 .....	13
2.1. 対象者 .....	14
2.2. 予測モデルの作成項目 .....	14
2.3. 施設毎の医師の虚実及び寒熱診断状況 .....	14

2.4.	機械学習を用いた施設毎の虚実予測モデルの作成 .....	14
2.5.	機械学習を用いた施設毎の寒熱予測モデルの作成 .....	15
2.6.	統計分析 .....	15
2.7.	倫理審査 .....	16
3.	結果 .....	16
3.1.	対象者 .....	16
3.2.	施設毎の医師の虚実及び寒熱診断状況 .....	17
3.3.	機械学習を用いた施設毎の虚実予測モデルの作成 .....	18
3.4.	機械学習を用いた施設毎の寒熱予測モデルの作成 .....	18
4.	考察 .....	19
5.	小括 .....	20
<b>第三章 複数施設データを用いた虚実と寒熱の balanced 予測モデルの構築 .....</b>		<b>21</b>
1.	背景と目的 .....	21
2.	方法 .....	21
2.1.	対象者 .....	21
2.2.	予測モデルの作成項目 .....	22
2.3.	機械学習を用いた Unbalanced 虚実及び寒熱予測モデルの作成 .....	22
2.4.	機械学習を用いた Balanced 虚実及び寒熱予測モデルの作成 .....	22
2.5.	Unbalanced 予測モデルと Balanced 予測モデルの比較 .....	23
2.6.	虚実予測モデルと寒熱予測モデルの比較 .....	23
2.7.	統計分析 .....	23
2.8.	倫理審査 .....	24
3.	結果 .....	24
3.1.	対象者 .....	24
3.2.	機械学習を用いた Unbalanced 虚実及び寒熱予測モデルの作成 .....	24
3.3.	機械学習を用いた Balanced 虚実及び寒熱予測モデルの作成 .....	25

3.4. Unbalanced 予測モデルと Balanced 予測モデルの比較 .....	25
3.5. 虚実予測モデルと寒熱予測モデルの比較 .....	25
4. 考察 .....	25
5. 小括 .....	28
<b>第四章 複数施設データを用いた虚実中間証を含む虚実の予測モデルの構築 .....</b>	<b>29</b>
1. 背景と目的 .....	29
2. 方法 .....	29
2.1. 対象者 .....	29
2.2. 予測モデルの作成項目 .....	29
2.3. 機械学習を用いた虚実中間証を含んだ虚実予測モデル .....	29
2.4. 全施設と各施設の虚証、虚実中間証、実証のカットオフ値、判別率 .....	30
2.5. 医師の診断した証と予測モデルが予測した証の重要項目の値の検討 .....	30
2.6. 統計分析 .....	31
2.7. 倫理審査 .....	31
3. 結果 .....	31
3.1. 対象者 .....	31
3.2. 機械学習を用いた虚実中間証を含んだ虚実予測モデル .....	31
3.3. 全施設と各施設の虚証、虚実中間証、実証のカットオフ値、判別率 .....	32
3.4. 医師の診断した証と予測モデルが予測した証の重要項目の値の検討 .....	32
4. 考察 .....	33
5. 小括 .....	34
<b>総括 .....</b>	<b>35</b>
<b>参考文献 .....</b>	<b>37</b>
<b>学位対象論文 .....</b>	<b>41</b>
<b>謝辞 .....</b>	<b>42</b>

## 略号

AC	Gwet の一致係数 (agreement coefficient)
BMI	Body Mass Index
FN	False Negative (偽陰性)
FP	False Positive (偽陽性)
ICD	International Classification of Disease
ICD-11	International Classification of Disease 11th Revision
IRR	Inter-rater reliability
ROC 曲線	Receiver Operating Characteristic 曲線
TN	True Negative (真陰性)
TP	True Positive (真陽性)
VAS	Visual Analogue Scale

## 緒言

我が国では近年、漢方薬に関する作用メカニズムの解明<sup>1</sup>および臨床研究の推進によるエビデンスの構築等により、医療現場での使用頻度が増加しており<sup>2-4</sup>、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<sup>4-6</sup>。日本のみならず<sup>3, 7</sup>、中国<sup>7</sup>や韓国<sup>7, 8</sup>などでも伝統医薬は多く使用されている。平成27年に策定された「患者のための薬局ビジョン」では、薬局の強化・充実すべき機能として健康サポート機能があげられた。薬局薬剤師は、医療用医薬品の安全かつ適正な使用に関する助言や支援のみならず、一般用医薬品等に対しても適切に選択できるための助言体制の強化が求められている<sup>9</sup>。また我が国では、平成28年に策定された「アジア健康構想に向けた基本方針」<sup>10</sup>を推進する上で、伝統医学は病気の予防や健康推進、医療の在り方等を考える上で重要であるとされている<sup>11</sup>。国際疾病分類の第11回改訂版（International Classification of Disease 11th Revision; ICD-11）が2019年5月に世界保健総会にて承認され、2022年より運用開始されることとなった<sup>12</sup>。ICD-11には、ICDの歴史で初めて主に中国、韓国および日本を中心に実践されている東アジア伝統医学を含む伝統医学に関する章が追加された<sup>12</sup>。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<sup>13</sup>、今後さらに漢方薬の選択および服薬指導の重要性が高まるだろう。

漢方薬に興味がある薬剤師は7割超であるものの<sup>4, 5, 14</sup>、苦手意識がある者も7割超であり<sup>4-6</sup>、西洋薬よりも服薬指導が難しいとする薬剤師が半数を超えるとの報告もある<sup>4, 5, 14</sup>。また、服薬指導をする上で不足しがちであるが特に必要性が高い情報として、証の見極めが挙げられており<sup>14</sup>、処方意図の把握ができていない者が半数以下であった<sup>5</sup>。一般用漢方製剤については薬剤師自身が利用者に適した漢方薬を選択する必要があるが、適切な漢方薬の選択や情報提供が難しく、同一症状に対してどの漢方薬を選択したら良いか分かり難いといった意見も多い<sup>15</sup>。医師においても、漢方薬処方時に、伝統医学診断に基づき処方している例は少なく、西洋医学的診断を基本に処方している例が約80%に上るとの報告もある<sup>16</sup>。以上のことから、薬剤師は漢方薬の選択や服薬指導、医師は漢方薬の処方選択に難渋している状況が窺える。

漢方の診断から処方決定までのプロセスは、3段階である(図1)。まず、①医師が四診(望診、聞診、問診、切診)を行う。次に、②四診の結果をもとに患者の証を診断する(図2)。Yakuboらは慢性疾患に対して3つのモジュールを乗算することで漢方診断の証を決定する方法を提案した<sup>17</sup>。3つのモジュールとは、虚実、寒熱、気血水である。これらのモジュールの下層にそれぞれ descriptor がある。虚実では、虚証、虚実中間証、実証、寒熱では、寒証、寒熱中間証、熱証、寒熱錯雑証、気血水では、気虚、気うつ・気滞、気逆、血虚、瘀血、水毒、下焦の虚がある<sup>17</sup>。各 descriptor は ICD-11 の 26 章の Traditional medicine

patterns において定義されている<sup>12</sup>(図3)。最後に、③患者の証に基づき、漢方薬を決定する。個々の漢方薬には、それぞれ薬方の証とよばれる証があり、患者の証とあう薬方の証をもつ漢方薬を選択する。薬方の証もまた ICD-11 Traditional medicine patterns に記載がある<sup>12</sup>(図3)。各々の薬方の証の詳細は、ICD-11 には記載はなく各国で作成されているため、漢方薬の薬方の証については日本東洋医学会の用語委員会が作成した。2014 年 10 月から 2019 年 4 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た漢方薬が一剤のみ処方されている患者のなかで、診断された患者の証と処方された薬方の証が虚証、実証、寒証、熱証であった患者を対象とした際の、処方数が多い漢方薬の薬方の証と患者の証を示した(表 1)。漢方専門医が診断し処方決定した際、患者の証と薬方の証が一致していたケースは虚実で 86.3%、寒熱で 97.7%であり、処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。しかし証診断には、医師の五感や経験、暗黙知などに拠るところが多いため、使用経験の少ない医師や薬剤師では困難であることが考えられる。ICD-11 に漢方医学診断が含まれたため、今後使用経験の少ない医師や薬剤師でも簡便かつ適切に証の見極めができるようになることが必要である。

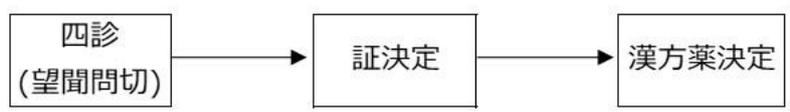


図1 漢方の診断から処方決定までのプロセス

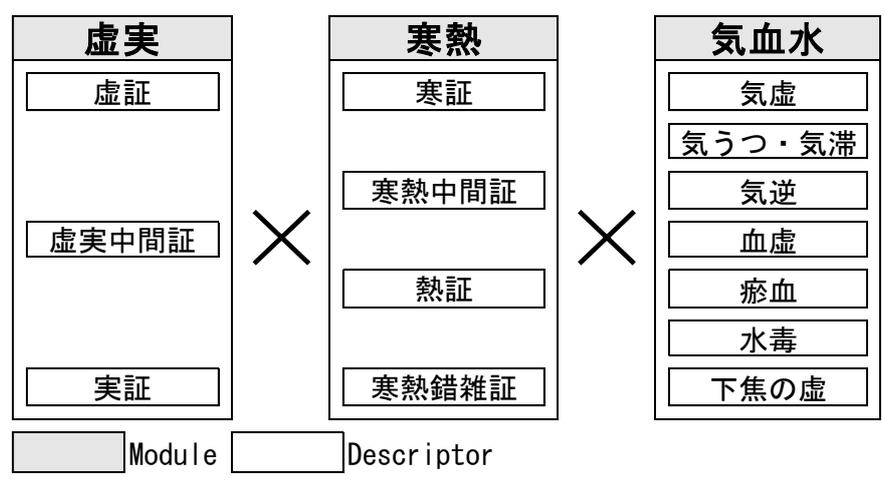


図2 慢性疾患における漢方証診断フロー

Maeda-Minami (2021) Figure を改変

Chapter 26 : Traditional Medicine conditions - Module I	
Traditional medicine disorders	
Traditional medicine patterns	
Principle-based patterns	陰陽/寒熱/虚実/表裏
Body constituents patterns	気血水
Organ system patterns	臓腑
Environmental factor patterns	環境証
Meridian patterns	経絡証
Six stage patterns	六病位
Triple energizer stage patterns	三焦証
Four phase patterns	衛気營血証
Four constitution medicine patterns	四象體質証
Formula patterns	薬方の証

図3 ICD-11 26章 Traditional Medicine conditions - Module Iの概要

\*太字：漢方で使用する証、斜体：本研究に関する証

ICD-11<sup>12</sup>をもとに作成

表1 患者の証と薬方の証が虚証、実証、寒証、熱証であった患者を対象とした際の漢方薬の薬方の証と患者の証 (単位：名)

処方名	薬方の証 (虚実)	患者の証		薬方の証 (寒熱)	患者の証		合計
		虚証	実証		寒証	熱証	
牛車腎気丸	虚証	12	12	寒証	23	1	24
当帰芍薬散	虚証	23	0	寒証	23	0	23
真武湯	虚証	20	0	寒証	20	0	20
八味地黄丸	虚証	13	2	寒証	15	0	15
当帰四逆加呉茱萸生姜湯	虚証	11	0	寒証	11	0	11
人参湯	虚証	10	0	寒証	10	0	10
小建中湯	虚証	9	0	寒証	9	0	9
黄連解毒湯	実証	2	5	熱証	1	6	7
大柴胡湯	実証	0	6	熱証	1	5	6
白虎加人参湯	実証	2	4	熱証	0	6	6

薬方の証は日本東洋医学会用語委員会委員の渡辺賢治氏より提供された資料をもとに作成

使用経験が少なくても証診断が適切に出来るようになるためには、患者の問診結果など、問診以外の四診を用いることなく証の予測ができるシステム開発が有用である。患者の問診結果から虚実と寒熱を予測する先行研究は、虚実では片山ら<sup>18</sup>、谷垣ら<sup>19</sup>、寒熱では谷垣ら<sup>19</sup>、Lee BJら<sup>20</sup>、Lee Jら<sup>21</sup>がある。これらはすべて単施設を対象としているため<sup>18-21</sup>、複数施設での検討が必要である。2008年より、慶應義塾大学漢方医学センターにおいて「漢方自動問診システム」が構築され、情報収集が開始された<sup>22</sup>。漢方自動問診システムとは、患者自身がiPadを用いて、症状に関する問診項目に回答するシステムである。2012年より、千

葉大学、飯塚病院、慶應義塾大学、富山大学、亀田総合病院および自治医科大学の各漢方外来にて、共通の自動問診システムを初診患者に対して使用してきた<sup>22</sup>。複数施設の漢方外来患者レジストリであり<sup>22</sup>、本レジストリを用いて複数施設における検討を行うこととした。

以上のことから、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漢方薬の選択や服薬指導を行うことができ、また処方支援を行うための証予測モデルを開発することを目的とした。簡便に証診断に結びつくよう、患者の問診結果とBMIや血圧などの簡便に得られるバイタルサインのみから証を予測するモデルを開発することとした。まず、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を明らかにした（第一章）。次に虚実と寒熱に焦点をあて、施設毎の予測モデルを構築し、証予測に寄与度が高い重要問診項目の比較を行った（第二章）。この結果を踏まえ、複数施設データを統合したデータから虚実と寒熱の予測モデルを構築した（第三章）。最後に複数施設データを統合したデータから虚実中間証を含む虚実の予測モデルを構築した。

## 第一章 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

### 1. 背景と目的

ICD-11 に伝統医学の章が追加されたため、従来の西洋医学と同様に、伝統医学においても妥当性と信頼性が求められている。評価者間信頼性 (Inter-rater reliability; IRR) は、2人以上の評価者の診断や評価の一貫性およびばらつきを表す指標として定義される。伝統医学の IRR は低い<sup>23</sup>。この原因としては、伝統医学の証が複雑であること、標準化された用語がないこと、統計上の問題があることが挙げられている<sup>23</sup>。これらの課題を克服するために、伝統医学の証診断には信頼性と再現性がある新しい診断システムが求められている。

Popplewellらは、中医学(TCM)の証診断のIRRを評価する際に、Gwetの一致係数(agreement coefficient; AC)の方がFleissの $\kappa$ 係数より適していることを示唆した<sup>24</sup>。伝統医学の証の頻度は均一ではなく、ACは均一に分布していないデータの評価に対応することができるからである<sup>24</sup>。あわせて、ACは偶然の一致をFleissの $\kappa$ 係数より適切に推定することができる<sup>24</sup>。AC1はカテゴリーデータに使用されるが、AC2は順序データに使用され、加重統計に使用できる<sup>24</sup>。

Popplewellらはまた、TCMの証診断のIRRを評価する際に、これらの簡略版であるTCM診断 descriptorを提案した<sup>25</sup>。15の descriptorについて一つずつ該当するか否かを評価者は評価してIRRを算出したところ、通常の方法よりIRRが改善した<sup>25</sup>。IRRをさらに改善するためには、モジュール化する方法がある<sup>26</sup>。

本章では、Yakuboら<sup>17</sup>によって提案されたモジュール化の方法に従い、慢性疾患を対象とし漢方の証診断のIRRを計算することを目的とした。

### 2. 方法

#### 2.1. 対象者

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の外来を初めて受診し、参加に同意した患者とした。登録されたすべての患者に対して、書面によるインフォームドコンセントを行った。上記期間の初診患者は1,286名であった。除外基準は、20歳未満および気血水の descriptor が2つ以上付与されている、あるいは気血水の descriptor が付与されていない患者とし、除外後の人数は638名であった。

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

典型的な気血水の症状を有している 64 名の成人患者を全初診患者から有意抽出した(表 1-1)。患者番号 1-32 は医師 A の初診患者、患者番号 33-64 は医師 B の初診患者から選択した。

本研究に従事した医師は、医師免許を取得してから 10 年以上経過しており、日本東洋医学会の漢方専門医の認定を受けてから 5 年以上の臨床経験がある。

表 1-1 初診時の漢方専門医のモジュール毎の診断数 (単位：名)

		医師 A	医師 B	合計
虚実	虚証	15	14	29
	虚実中間証	12	7	19
	実証	5	11	16
	合計	32	32	64
寒熱	寒証	16	17	33
	寒熱中間証	7	10	17
	熱証	3	4	7
	寒熱錯雑証	6	1	7
	合計	32	32	64
気血水	気虚	5	5	10
	気うつ・気滞	5	5	10
	気逆	2	3	5
	血虚	5	4	9
	瘀血	5	5	10
	水毒	5	5	10
	下焦の虚	5	5	10
	合計	32	32	64

Maeda-Minami (2021) Table 1 を改変

## 2.2. 患者情報文書

医師は、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。この文書には、初診時に患者が自動問診システムを用いて入力した 133 の症状と、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 項目の問診結果が記載されていた(表 1-2)。医師が評価の際に、氏名、現在の病気の病歴、脈診や腹診の結果を含む身体所見などの患者情報を見ることはできないようにしていた。また、初診時から期間が浅い場合、医師が患者情報文書内容を記憶している可能性が考えられるため、初診の 3~5 年後に評価を行った。

表 1-2 質問項目

項目名	項目名
1 食欲がない	72 頭痛 (締め付けられる)
2 食欲が旺盛	73 頭痛 (脈を打つよう)
3 眠れない	74 頭痛 (重い)
4 昼間に眠くなる	75 頭痛の程度
5 夢をよくみる	76 めまい (ぐるぐる回る)
6 起きている間の小便の回数が3回以下	77 めまい (ふわふわする)
7 起きている間の小便の回数が9回以上	78 めまい (立ちくらみ)
8 夜間にトイレに立つ回数	79 めまいの程度
9 小便に関する症状 (排尿困難)	80 視力低下、目が疲れる、かすむ、しょぼしょぼする
10 小便に関する症状 (排尿痛)	81 鼻の症状 (くしゃみ)
11 小便に関する症状 (尿漏れ)	82 鼻の症状 (白くて水っぽい鼻水)
12 小便に関する症状 (残尿感)	83 鼻の症状 (黄色くてどろっとした鼻水)
13 大便の回数が1日に2回以上	84 鼻の症状 (鼻水がのどにおりる)
14 大便の回数が2日以上出ない	85 鼻の症状 (鼻づまり)
15 大便 (下剤を飲んで)	86 鼻の症状 (鼻血が出やすい)
16 大便の性状 (硬い)	87 口が苦い、変な味がする
17 大便の性状 (コロコロしている)	88 味がわからない
18 大便の性状 (軟らかい)	89 つばが多い
19 大便の性状 (水のような下痢)	90 のどが痛む
20 大便の性状 (出にくい)	91 のどがつかえる
21 大便の性状 (残尿感がある)	92 口唇が乾く
22 痔がある	93 耳の症状 (耳なり)
23 脱肛、子宮脱、膀胱脱がある	94 耳の症状 (聞こえにくい)
24 大便に血が混じる	95 咳
25 気分が憂うつになる	96 痰 (白くて水っぽい)
26 ものを忘れる	97 痰 (黄色くてどろっとしている)
27 イライラする	98 息切れ
28 気力がない	99 動悸
29 疲れやすい	100 胸のつまり
30 体がだるい	101 ゲップや胸焼け
31 体が重たく感じる	102 みぞおちがつかえる
32 集中力がない	103 吐き気、嘔吐
33 焦りやすい	104 乗り物酔い
34 朝、起きにくく調子がでない	105 腹が張る
35 ものごとに驚きやすい	106 腹がゴロゴロ鳴る
36 皮膚がかさかさする	107 ガスがよく出る
37 皮膚がかゆい	108 月経痛以外の腹痛 (空腹時)
38 ふけがでやすい、髪が抜けやすい	109 月経痛以外の腹痛 (空腹時) の程度
39 にきび	110 月経痛以外の腹痛 (食後)
40 しみ	111 月経痛以外の腹痛 (食後) の程度
41 目の周りにクマができる	112 月経痛以外の腹痛 (食事とは関係なく突然)
42 爪が割れる・はがれる・根元がささくれる	113 月経痛以外の腹痛 (食事とは関係なく突然) の程度
43 皮下出血がある	114 月経痛以外の腹痛 (排便前に)
44 最も汗をかきやすいところ (頭)	115 月経痛以外の腹痛 (排便前に) の程度
45 最も汗をかきやすいところ (手足)	116 手がこわばる・はばったい
46 最も汗をかきやすいところ (全身)	117 足に力はいらない、ふらつく
47 最も汗をかきやすいところ (寝汗)	118 足がつる
48 最も汗をかきやすいところの程度	119 しもやげができる
49 性欲の減退、インポテンツ	120 閉経している
50 風邪をひきやすい	121 月経周期がいつも24日以内
51 1日の中で気になる症状が変化する	122 月経周期がいつも39日以上
52 最もこりを感じるころ (首から肩)	123 月経周期をピルで調整している
53 最もこりを感じるころ (背中から腰)	124 月経周期で不正出血がある
54 顔胸肩手など上半身の痛み	125 月経周期が毎回4日以上ずれる
55 足腰膝など下半身の痛み	126 出血期間がいつも2日以内
56 顔胸肩手など上半身のしびれ	127 出血期間がいつも8日以上
57 足腰膝など下半身のしびれ	128 出血量が夜用を日中も使用するほど多い
58 顔胸肩手など上半身のふるえ	129 出血量が少ない日の昼用を夜に使用しても大丈夫なほど少ない
59 足腰膝など下半身のふるえ	130 おりものが白っぽくてさらっとしている
60 最も冷えを感じるころ (全身)	131 おりものが黄白色でどろっとしている
61 最も冷えを感じるころ (手足)	132 月経痛
62 最も冷えを感じるころ (足腰)	133 分娩、中絶、流産の合計回数
63 最も冷えを感じるころの程度	134 主訴
64 のぼせや顔のほてり (のぼせやすい)	135 収縮期血圧
65 のぼせや顔のほてり (顔が赤くなる)	136 拡張期血圧
66 のぼせや顔のほてりの程度	137 年齢
67 手足のほてり	138 性別
68 最もむくみを感じるころ (顔)	139 身長
69 最もむくみを感じるころ (手)	140 体重
70 最もむくみを感じるころ (脚)	
71 最もむくみを感じるころの程度	

### 2.3. 二人の医師による評価者内信頼性（医師 A と B）

同一患者に対して同一医師が、「初診時に診断した証」と、限られた情報である「患者情報文書から評価した証」を比較することで、再現性を検討した（図 1-1）。医師 A は患者番号 1-32 の情報文書を、医師 B は患者番号 33-64 を評価した。証の一致率は、初診時の診断と患者情報文書からの評価における各モジュールの一致率とした。虚実は、線形に重み付けされた一致率を、寒熱および気血水は、重み付けされていない一致率を計算した。

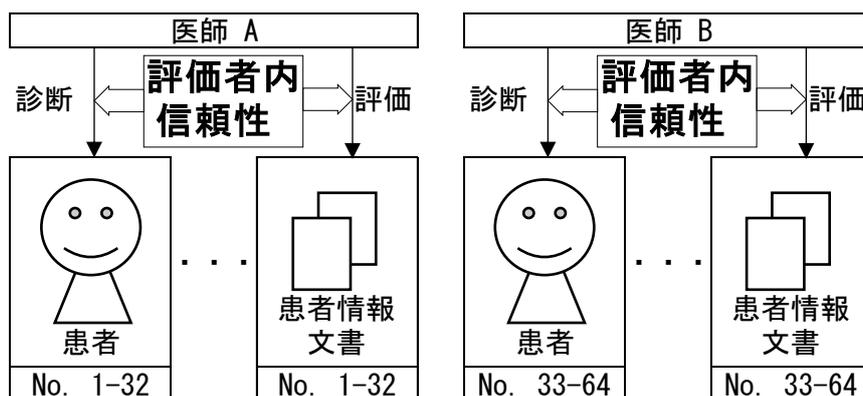


図 1-1 二人の医師による評価者内信頼性

### 2.4. 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（医師 A、B と C）

「初診時に診断した証」と、3 人の医師がそれぞれ「患者情報文書から評価した証」を比較した（図 1-2）。これは、患者情報文書に記載されている患者の問診結果とバイタルサインから証を予測するモデルの判別率の指標となる。3 人の医師（A、B、および C）が 64 人の患者情報文書を使用し、3 つのモジュールを評価した。各々の医師の評価した証と初診時に診断した証の一致率は、2.3. と同様に、虚実は、線形に重み付けされた一致率を、寒熱および気血水は、重み付けされていない一致率を計算した。

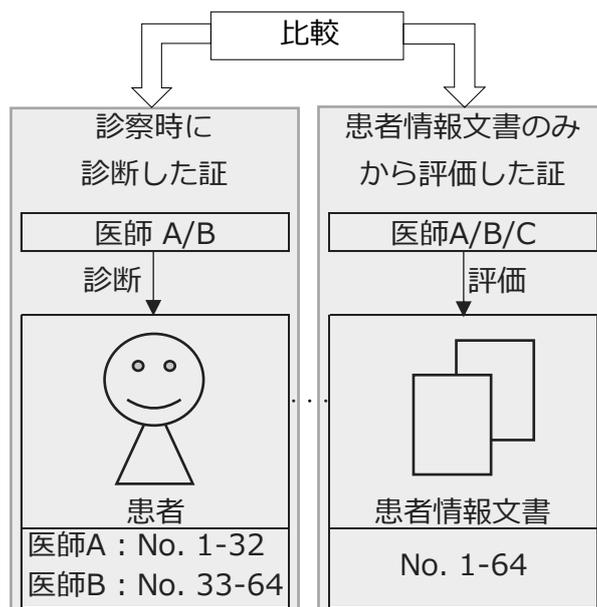


図 1-2 診察時に診断した証と患者情報文書から評価した証の一致率

## 2.5. 三人の医師による評価者間信頼性 (IRR) (医師 A、B と C)

患者情報文書から証を評価する際の IRR を算出した。3 人の医師 (A、B、および C) が 64 人の患者情報文書を使用し、3 つのモジュールを評価した。3 人の評価結果の一致率、虚実では AC2、寒熱と気血水では AC1 を算出した。

## 2.6. 統計分析

すべての解析は R3.5.1 を用いた (<http://www.r-project.org>)。一致率及び AC1、AC2 の算出には R の irrCAC のパッケージを使用した<sup>27</sup>。

## 2.7. 倫理審査

慶應義塾大学医学部倫理委員会の承認を受け実施した (承認番号 20100144 「漢方の特性を利用したエビデンス創出と適正使用支援システムの構築」)。

### 3. 結果

#### 3.1. 対象者

対象者が診断された西洋病名は、寒がり、更年期症候群、高血圧、手足のしびれの順であった。初診時の対象者の証は表 1-1 の通りである。

#### 3.2. 二人の医師による評価者内信頼性（医師 A と B）

初診時に診断した証と患者情報文書を評価した証の虚実、寒熱および気血水的一致率は、医師 A で 85.9%、78.1%、37.5%であり、医師 B では 71.9%、56.3%、34.4%であった（表 1-4）。

表 1-4 患者情報文書をもとに漢方専門医が評価した際の評価者内と評価者間信頼性

	カテゴリー数	評価者内信頼性 (患者と患者情報文書)		評価者間信頼性 (患者情報文書)	
		医師 A No. 1-32 一致率 <sup>a</sup>	医師 B No. 33-64 一致率 <sup>a</sup>	医師 A, B, C No. 1-64 一致率 <sup>a</sup>	AC1/2 <sup>b</sup>
虚実	3	85.9	71.9	85.9	0.708
寒熱	4	78.1	56.3	64.6	0.542
気血水	7	37.5	34.4	35.9	0.254

AC1, Gwet's Agreement Coefficient 1; AC2, Gwet's Agreement Coefficient 2

a 虚実の一致率は、線形で重みづけした一致率、寒熱と気血水的一致率は重みづけなしの一致率である。

b 虚実は順序データであるため AC2 を、寒熱と気血水はカテゴリーデータであるため AC1 を使った。

Maeda-Minami (2021) Table 2 を改変

#### 3.4. 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（医師 A、B と C）

初診時に診断した証と、3 人の医師がそれぞれ患者情報文書から評価した証の虚実、寒熱および気血水的一致率は、医師 A が 80.5%、67.2%、32.8%、医師 B が 78.1%、48.4%、37.5%、医師 C が 81.3%、68.8%、28.1%であった。

表 1-5 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 (単位: %)

	医師A	医師B	医師C
虚実	80.5	78.1	81.3
寒熱	67.2	48.4	68.8
気血水	32.8	37.5	28.1

### 3.4. 三人の医師による評価者間信頼性 (IRR) (医師 A、B と C)

患者情報文書から評価した虚実の IRR は、一致率が 85.9%、AC2 が 0.708 であった (表 1-2)。寒熱の一致率と AC1 は、それぞれ 64.6%と 0.542 であった。気血水の一致率と AC1 は、それぞれ 35.9%と 0.254 であった。

## 4. 考察

慢性疾患を対象とした漢方の証の IRR をモジュール化することで検討した。本研究の結果より、モジュール化することで証の IRR を改善した。これらのモジュールを増やすことで、専門医はより正確に漢方の証診断を行うことができる。本検討は、モジュール化を行って IRR を算出したり、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したり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。IRR は評価者が 2 名で行うことを前提としている指標であるため、IRR を算出した多くの研究は評価者が 2 名であるが<sup>30</sup>、本検討では信頼性および妥当性を高めるために 3 人目の評価者を入れて行った。

気血水の IRR は、虚実や寒熱よりも低かった。原因の一つとして、気血水は虚実や寒熱よりもカテゴリー数が多いことが挙げられる。虚実と寒熱の各 descriptor は排他的であるが、気血水の descriptor は排他的ではなく 7 つのカテゴリーより選択する形式であることも要因の一つであると考えられる。また、モジュール自体が脈診や腹診を含む医師の身体所見の結果に大きく依存している可能性があり、患者情報文書のみからの評価では不十分であったことを示唆している。

寒熱の IRR は虚実よりも低かった。寒熱錯雑証を除いた、寒証、寒熱中間証、熱証は虚実と同様に順序データであるが、寒熱錯雑証があるためカテゴリーデータとして寒熱の一致率を算出した。医師の評価した結果を確認したところ、寒証と熱証を間違えているという例はなく、寒証と寒熱中間証、寒熱中間証と熱証のカットオフ値の幅が医師によって多少異なっている可能性が考えられた。順序データとして算出せずに、カテゴリーデータとして算出することで、一致率が低くなったと考えられる。

本研究の限界は、以下の通りである。IRR を算出する際に、漢方専門医は患者を直接診察しておらず、患者情報文書に基づいて評価した。本研究に貢献した 3 人の漢方専門医は同一施設で働いていたため、結果を一般化することが難しい。したがって、今回得られた結果は多施設共同研究を行うことで検証する必要がある。最後に、モジュール化されていない証診断とモジュール化された診断構造との直接比較は行わなかった。

## 5. 小括

本研究では、慢性疾患を対象とした漢方の証の IRR を算出した。患者情報文書に基づく虚実および寒熱の IRR は、気血水よりも高かった。これらの結果は、descriptor の数を少なくし、証をモジュール化することで IRR が改善する可能性を示唆している。これらのモジュールを増やすことで、専門医は正確に証診断を行うことができると考えられる。

## 第二章 施設毎の虚実と寒熱の予測モデルの構築

### 1. 背景と目的

証の予測モデルを構築した先行研究としては、機械学習とロジスティック回帰分析を使用したものがある<sup>20, 31-38</sup>。これらの研究では、独立変数を診察<sup>32, 33</sup>、聴診と嗅覚による診察<sup>34, 35</sup>、身体測定データ<sup>20</sup>、問診票およびカルテ<sup>20, 36-38</sup>とし、従属変数を医師により診断された証<sup>20, 32-35, 37, 38</sup>、患者への問診結果をスコア化したものにより決定された証<sup>36</sup>としている。ICD-11で説明されている証には様々な種類がある。先行研究で予測された証は、「肝血虚」のように複数の証が組み合わされている<sup>37, 38</sup>。これは、虚実や寒熱だけではなく、気血水、五臓の証を含んでいる<sup>37, 38</sup>。東洋医学会が認定した漢方専門医の臨床現場では、虚実、寒熱、気血水を独立して診断しているため、これらとは異なっている<sup>3, 17</sup>。本研究では、漢方で最も重要な証であり、第一章でIRRが比較的高かった虚実と寒熱に焦点を当てる。

虚実については、舌診の結果から証を予測する研究が存在している<sup>32</sup>。患者アンケート結果から証を予測するものとしては、片山らおよび谷垣らは、慶應義塾大学の初診患者を対象とした虚実予測式を提案し、予測結果と医師の診断との一致率が90%程度となり、BMIが予測に重要な項目であったことを指摘している<sup>18, 19</sup>。片山らおよび谷垣ら以外に問診項目より漢方の証を予測する研究はない。単施設での結果であるため、他施設も交えた複数施設での検討が必要である。

寒熱については、慶應義塾大学の初診患者を対象とした研究<sup>19</sup>と韓国の研究<sup>20, 21, 31</sup>がある。谷垣らは寒熱予測式を提案し、予測結果と医師の診断との一致率が90%程度となり、BMIが寒熱においても予測に重要な項目であったことを指摘している<sup>19</sup>。Lee BJらやLee Jらもまた、BMIが寒熱の予測に最も重要であったことを指摘している<sup>20, 21</sup>。これらの研究も単施設での結果であるため、他施設も交えた複数施設での検討が必要である。

本章の目的は、問診項目とバイタルサインから虚実と寒熱を予測する上で、①日本の漢方外来における虚実と寒熱の診断状況を把握し、②それぞれの施設の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することである。

### 2. 方法

多施設前向き観察研究を行った。

## 2.1. 対象者

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施設の漢方外来にデータ収集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。各施設のデータ収集期間は、千葉大学が2012年3月から2015年1月、飯塚病院が2012年1月から2015年2月、慶應義塾大学が2012年6月から2013年3月、富山大学が2012年6月から2015年2月、亀田総合病院が2012年1月から2013年3月、自治医科大学が2012年1月から2012年3月であった。除外基準は、各種データ欠損（血圧、身長や体重など）、20歳未満、問診項目回答数が20項目未満とした。問診項目回答数が20項目未満の患者では証を予測することが難しかったことが先行研究において報告されている<sup>18</sup>。

## 2.2. 予測モデルの作成項目

問診項目は148項目であった。問診項目に加えて、身長、体重、収縮期血圧、拡張期血圧、年齢、性別を加えた154項目を取得した。148の問診項目は、29のバイナリ項目、119のVisual Analogue Scale (VAS, 0-100mm)の項目を含む。VAS値は、先行研究と同様に患者毎に、それぞれのVAS値をVASの最大値で除すことで正規化した<sup>18, 39</sup>。BMIは、体重(kg) ÷ (身長m)<sup>2</sup>で算出した。慶應義塾大学病院漢方医学センターのデータを用いた片山ら、谷垣らの先行研究では、BMIが虚実を予測する上で最も重要な項目であったため<sup>18, 19</sup>、予測モデルを作成する際に身長と体重はそのまま用いず、BMIを用いた。

## 2.3. 施設毎の医師の虚実及び寒熱診断状況

各施設の医師による虚実及び寒熱の診断状況を確認するため、施設毎に虚実、寒熱の診断数を数えた。虚実は、虚証、虚実中間証、実証の3つのdescriptorがあり、寒熱は、寒証、寒熱中間証、熱証、寒熱錯雑証の4つのdescriptorがある。これらの証の定義はICD-11に基づく。

## 2.4. 機械学習を用いた施設毎の虚実予測モデルの作成

施設毎の虚実予測モデルを構築するために、学習データとして各施設からランダムに40名の対象者（虚証20名と実証20名）をサンプリングした。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。予測モデルを構築

する上で、虚証と実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、千葉大学、飯塚病院、慶應義塾大学、富山大学の 4 施設のみモデルを作成した。

虚実の予測に寄与度が高い重要項目を明らかにするために、各施設のランダムフォレストモデルで最も重要な 30 項目を抽出した。3 つまたは 4 つ全てにて共通にみられた項目を重要項目から抽出した。抽出した項目について、医師に虚証と診断された患者と実証と診断された患者の値を比較した。

## 2.5. 機械学習を用いた施設毎の寒熱予測モデルの作成

施設毎の寒熱予測モデルを構築するために、虚実予測モデルと同様に、学習データとして各施設からランダムに 40 名の対象者（寒証 20 名と熱証 20 名）をサンプリングした。さらに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。予測モデルを構築する上で、寒証および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、千葉大学、飯塚病院、富山大学の 3 施設のみモデルを作成した。

虚実と同様に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、各施設のランダムフォレストモデルで最も重要な 30 項目を抽出した。3 つすべての施設にて共通にみられた項目を重要項目から抽出した。抽出した項目について、医師に寒証と診断された患者と熱証と診断された患者の値を比較した。

## 2.6. 統計分析

すべての解析は R3.5.1 を用いた (<http://www.r-project.org>)。施設毎の虚実と寒熱予測モデルの作成については、ランダムフォレストを用い、R の Randomforests のパッケージを使用した<sup>40</sup>。ランダムフォレストは、決定木をもとに Breiman により構築されたアルゴリズムである<sup>41, 42</sup>。ランダムフォレストのアルゴリズムは、多くの分類木を構築する<sup>42</sup>。本章で使用した分類のアルゴリズムは、未知のデータを評価する際には、各々の分類木の分類結果の多数決が全体の結果となる<sup>42</sup>。このモデルでは、各患者の値は 500 本の決定木の結果となる。値は 0 から 1 までで、0.5 未満は虚証、0.5 から 1 は実証となる。ランダムフォレストモデルを作成することで、各々の項目が予測にどの程度寄与しているかの指標である重要度を求めることができる<sup>42</sup>。重要度は、分類木を構築する際に当該する変数をモデルから除いた場合の Gini 係数の減少 (Mean Decrease Gini) より計算される。値が大きいくほど、当該する項目が診断の予測に高く寄与していることを示す。ランダムフォレストは多くの分野で使用されており、本研究のような証診断を予測するための分類アルゴリズムでは精度がよいことが言われている<sup>18, 19</sup>。虚実および寒熱を予測するために、谷垣らでは他の方法と比較し、

ランダムフォレストの予測精度が一番高かったとしている<sup>19</sup>。

医師に虚証と診断された患者と実証と診断された患者、及び医師に寒証と診断された患者と熱証と診断された患者の値の比較には、Wilcoxon の順位和検定を用いた。P<0.05 を統計学的に有意とみなした。

## 2.7. 倫理審査

慶應義塾大学医学部倫理委員会の承認を受け（承認番号 20100144「漢方の特性を利用したエビデンス創出と適正使用支援システムの構築」）、これに紐づく形で各施設の倫理委員会においても承認を受け実施した。

## 3. 結果

### 3.1. 対象者

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。データ欠損、20歳未満を除いた解析対象患者数は、千葉大学が661名、飯塚病院が695名、慶應義塾大学が660名、富山大学が445名、亀田総合病院が208名、自治医科大学が16名であった。データ欠損の内訳は、問診、西洋病名、漢方病名、薬剤、身長、体重、血圧であった。解析対象患者からさらに20項目未満回答患者を除外し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。

解析対象患者と予測モデル対象患者の年齢 (Mean ± SD) は、千葉大学では、それぞれ 56.2±16.8 歳、55.4±16.6 歳、飯塚病院では、それぞれ 53.3±17.1 歳、51.9±17.2 歳、慶應義塾大学では、それぞれ 56.8±17.6 歳、55.9±17.5 歳、富山大学では、それぞれ 52.4±17.2 歳、51.8±17.1 歳、亀田総合病院では、それぞれ 53.9±15.4 歳、51.3±15.8 歳、自治医科大学では、それぞれ 58.9±16.5 歳、55.3±18.3 歳であった。解析対象患者と予測モデル対象患者の性別 (男:女) は、千葉大学では、それぞれ 170:491、69:343、飯塚病院では、それぞれ 186:509、139:431、慶應義塾大学では、それぞれ 152:508、64:365、富山大学では、それぞれ 141:304、107:260、亀田総合病院では、それぞれ 46:162、14:94、自治医科大学では、それぞれ 9:7、3:5 であった。解析対象患者と予測モデル対象患者の BMI (Mean ± SD) は、千葉大学では、それぞれ 21.9±4.0、21.9±4.1、飯塚病院では、それぞれ 22.1±3.7、

22.1±3.7、慶應義塾大学では、それぞれ21.4±3.3、21.2±3.3、富山大学では、それぞれ22.0±3.5、22.0±3.5、亀田総合病院では、それぞれ22.2±3.7、22.2±3.7、自治医科大学では、それぞれ21.3±3.2、20.3±3.8であった。予測モデル対象患者の患者背景は、解析対象患者と有意差はみられなかった。また、各施設における予測モデル対象患者の患者背景は差は見られなかった。

### 3.2. 施設毎の医師の虚実及び寒熱診断状況

各施設の医師が診断した解析対象患者の虚実の診断状況は、千葉大学では虚証290名、虚実中間証226名、実証145名、飯塚病院では虚証311名、虚実中間証289名、実証95名、慶應義塾大学では虚証211名、虚実中間証310名、実証139名、富山大学では虚証255名、虚実中間証142名、実証48名、亀田総合病院では虚証101名、虚実中間証89名、実証18名、自治医科大学では虚証3名、虚実中間証10名、実証3名であった。各施設の医師が診断した予測モデル対象患者の虚実の診断状況は、千葉大学では虚証194名、虚実中間証135名、実証83名、飯塚病院では虚証259名、虚実中間証238名、実証73名、慶應義塾大学では虚証163名、虚実中間証195名、実証71名、富山大学では虚証217名、虚実中間証113名、実証37名、亀田総合病院では虚証59名、虚実中間証41名、実証8名、自治医科大学では虚証2名、虚実中間証5名、実証1名であった。実証の割合は、すべての施設で虚証の割合よりも低かった。

各施設の医師が診断した解析対象患者の寒熱の診断状況は千葉大学では寒証231名、寒熱中間証135名、熱証178名、寒熱錯雑証93名、寒熱付与なしあるいは2つ以上24名、飯塚病院では寒証253名、寒熱中間証241名、熱証113名、寒熱錯雑証69名、寒熱付与なしあるいは2つ以上19名、慶應義塾大学では寒証346名、寒熱中間証182名、熱証25名、寒熱錯雑証99名、寒熱付与なしあるいは2つ以上8名、富山大学では寒証167名、寒熱中間証174名、熱証63名、寒熱錯雑証37名、寒熱付与なしあるいは2つ以上4名、亀田総合病院では寒証52名、寒熱中間証55名、熱証19名、寒熱錯雑証71名、寒熱付与なしあるいは2つ以上11名、自治医科大学では寒証5名、寒熱中間証9名、熱証0名、寒熱錯雑証2名、寒熱付与なしあるいは2つ以上0名であった。各施設の医師が診断した予測モデル対象患者の寒熱の診断状況は千葉大学では寒証157名、寒熱中間証67名、熱証102名、寒熱錯雑証70名、寒熱付与なしあるいは2つ以上16名、飯塚病院では寒証221名、寒熱中間証182名、熱証87名、寒熱錯雑証62名、寒熱付与なしあるいは2つ以上18名、慶應義塾大学では寒証254名、寒熱中間証63名、熱証16名、寒熱錯雑証89名、寒熱付与なしあるいは2つ以上7名、富山大学では寒証139名、寒熱中間証149名、熱証50名、寒熱錯雑証27名、寒熱付与なしあるいは2つ以上2名、亀田総合病院では寒証30名、寒熱中間証24名、熱証11

名、寒熱錯雑証 42 名、寒熱付与なしあるいは 2 つ以上 1 名、自治医科大学では寒証 3 名、寒熱中間証 3 名、熱証 0 名、寒熱錯雑証 2 名、寒熱付与なしあるいは 2 つ以上 0 名であった。熱証の割合は、すべての施設で寒証の割合よりも低かった。

### 3.3. 機械学習を用いた施設毎の虚実予測モデルの作成

千葉大学、飯塚病院、慶應義塾大学、富山大学の学習データを用いて各施設の虚実予測モデルを構築した。虚実の予測に寄与度が高い重要項目 30 項目を、施設毎に抽出した。4 施設にて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 施設にて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。全施設で BMI が 1 番重要な項目であった。

抽出した項目について、医師に虚証と診断された患者と実証と診断された患者の値を比較した。BMI (虚証 :  $20.4 \pm 3.2 \text{ kg/m}^2$ 、実証 :  $25.2 \pm 3.9 \text{ kg/m}^2$ 、 $p < 0.001$ )、収縮期血圧 (虚証 :  $120.8 \pm 18.6 \text{ mmHg}$ 、実証 :  $129.5 \pm 20.1 \text{ mmHg}$ 、 $p < 0.001$ )、拡張期血圧 (虚証 :  $72.5 \pm 11.9 \text{ mmHg}$ 、実証 :  $78.4 \pm 14.0 \text{ mmHg}$ 、 $p < 0.001$ )、暑がり (虚証 :  $0.241 \pm 0.338$ 、実証 :  $0.343 \pm 0.387$ 、 $p < 0.001$ ) は実証の患者の値の方が虚証の患者の値より有意に大きかった。一方、全身が冷える (虚証 :  $0.242 \pm 0.356$ 、実証 :  $0.101 \pm 0.242$ 、 $p < 0.001$ )、寒がり (虚証 :  $0.527 \pm 0.388$ 、実証 :  $0.352 \pm 0.381$ 、 $p < 0.001$ )、気分が憂うつになる (虚証 :  $0.328 \pm 0.333$ 、実証 :  $0.254 \pm 0.326$ 、 $p < 0.001$ )、目が疲れる (虚証 :  $0.456 \pm 0.373$ 、実証 :  $0.398 \pm 0.364$ 、 $p < 0.001$ )、肩のこり (虚証 :  $0.591 \pm 0.370$ 、実証 :  $0.521 \pm 0.374$ 、 $p = 0.003$ )、朝起きにくく調子がでない (虚証 :  $0.320 \pm 0.373$ 、実証 :  $0.233 \pm 0.330$ 、 $p < 0.001$ ) は、虚証の患者の値の方が実証の患者の値より有意に大きかった。

### 3.4. 機械学習を用いた施設毎の寒熱予測モデルの作成

千葉大学、飯塚病院、富山大学の学習データを用いて各施設の寒熱予測モデルを構築した。寒熱の予測に寄与度が高い重要項目 30 項目を、施設毎に抽出した。3 施設にて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。全施設で BMI が最も重要な項目であった。抽出した項目について、医師に寒証と診断された患者と熱証と診断された患者の値を比較した。BMI (寒証 :  $20.7 \pm 3.2 \text{ kg/m}^2$ 、熱証 :  $24.2 \pm 4.5 \text{ kg/m}^2$ 、 $p < 0.001$ )、収縮期血圧 (寒証 :  $120.5 \pm 19.0 \text{ mmHg}$ 、熱証 :  $128.2 \pm 19.8 \text{ mmHg}$ 、 $p < 0.001$ )、拡張期血圧 (寒証 :  $72.4 \pm 11.9 \text{ mmHg}$ 、熱証 :  $75.8 \pm$

13.6 mmHg、 $p<0.001$ )は熱証の患者の値の方が寒証の患者の値より有意に大きかった。一方、足が冷える(寒証:  $0.486\pm0.402$ 、熱証:  $0.206\pm0.320$ 、 $p<0.001$ )、寒がり(寒証:  $0.574\pm0.379$ 、熱証:  $0.273\pm0.343$ 、 $p<0.001$ )、年齢(寒証:  $54.6\pm17.5$ 歳、熱証:  $51.9\pm17.4$ 歳、 $p=0.028$ )、肩のこり(寒証:  $0.595\pm0.366$ 、熱証:  $0.529\pm0.372$ 、 $p=0.044$ )は、寒証の患者の値の方が熱証の患者の値より有意に大きかった。

#### 4. 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。

本研究では、日本の漢方専門施設で得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。対象とした全6施設で、虚証が実証よりも、寒証が熱証よりも診断された頻度が高かった。千葉大学、飯塚病院、慶應義塾大学、富山大学の4施設のデータを用いて構築した虚実を予測するランダムフォレストモデルを比較することで、共通した重要項目として15項目が明らかとなった。また、千葉大学、飯塚病院、富山大学の3施設のデータを用いて構築した寒熱を予測するランダムフォレストモデルを比較することで、8つの共通した重要項目が明らかになった。この結果は、対象とした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。これらの結果はまた、日本における虚実および寒熱を診断するために重要な項目を示している可能性がある。

各施設のデータセットから構築した虚実予測モデルの重要項目について、BMIがどの施設においても最も重要な項目であった。これは、片山らおよび谷垣らが示した慶應義塾大学を対象とした結果と一致していた<sup>18, 19</sup>。血圧や自覚症状である暑がりなども各施設の重要項目として共通しており、これらの項目の値は実証と診断された対象者の方が有意に高かった。さらに、全身の冷え、寒がりなどの症状も共通して重要な項目に挙げられており、これらの項目の値は、虚証と診断された対象者の方が有意に高かった。

各施設のデータセットから構築した寒熱予測モデルの重要項目についても、虚実同様にBMIがどの施設においても最も重要な項目であった。これは、谷垣らが示した慶應義塾大学を対象とした結果<sup>19</sup>、Lee BJら、Lee Jらの結果<sup>20, 21</sup>とも一致していた。血圧は各施設に共通して重要項目として挙げられており、値は熱証と診断された対象者の方が有意に高かった。さらに、足が冷える、寒がり、年齢も共通してランダムフォレストモデルの重要な項目として挙げられており、これらの項目の値は寒証と診断された対象者の方が有意に高かった。

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。

本研究では6施設のデータを用いた。施設により欠損率に差が見られた。予測モデルを構

築する上で、BMI や血圧を用いたため、BMI や血圧が欠損している者は除外した。身長、体重、血圧が欠損している対象者は 20 歳未満の者が多かった。これは、小児についてはバイタルサインの入力ができない仕様になっていることが原因である。西洋病名、漢方病名、薬剤については、自動問診システム内に医師の診断を医師が入力する必要があり、ヒューマンエラーが原因で欠損したと考えられる。千葉大学は問診データがないものが多かった。これはデータ収集期間が 5 年以上前であるため詳細は不明であるが、無線ネットワークが不良であるとき、患者が問診を入力してもデータに残らないことが原因の可能性もある。亀田総合病院、自治医科大学は成人患者でバイタルサインが欠損していることも見られたが、患者背景を確認したところ、差は見られなかったため研究に用いた。

本研究にはいくつかの限界がある。この観察研究では、各対象者を医師が 1 人で診察したため、虚実および寒熱の診断は他の専門医により検証されていない。本研究に参加した医師は日本東洋医学会専門医で臨床経験が 10 年以上である。医師の診断を標準化する試みはこれまでになされているものの、医師の診断の一致率を算出した研究は検討 1 のみである。検討 1 は、患者情報文書という限られた情報からの一致率を算出している。例えば虚実については、腹診の結果を予測モデルの説明変数に加えると判別率が 10%程度上昇し、約 90%であったことを確認しているため<sup>43</sup>、実際の患者を診断した証の一致率は、患者情報文書のみから評価した証の一致率よりも情報量が増えるため上昇する可能性がある。BMI は小児患者では成人と同様に評価できないため<sup>44</sup>、20 歳未満の対象者を除外した。本予測モデルを小児に適用できるかどうかは不明である。また自覚症状が 19 以下の対象者も除外したため、今回の結果を適応できるか不明である。さらに虚実の定義は、日本、中国、韓国の間で異なる可能性があるため、本研究の結果は、全世界に適応できるかは不明である。この点に関連し、本分析から虚実中間証を除外した。虚実中間証に関して今後検討が必要である。

## 5. 小括

本章において、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。虚実および寒熱を診断するために重要な項目を明らかにした。虚実予測モデルの重要項目においても、寒熱予測モデルの重要項目においても、施設で共通している項目があり、対象とした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が示唆された。また、本章において明らかとなった重要項目は日本における虚実および寒熱を診断するために重要な項目を示している可能性がある。

### 第三章 複数施設データを用いた虚実と寒熱の balanced 予測モデルの構築

#### 1. 背景と目的

虚実および寒熱の統計的予測モデルを質問項目および身体測定データを使用して構築している研究はある<sup>18, 19, 31, 43</sup>。虚実を予測する研究では、予測に最も重要な項目はBMIであることが示されている<sup>18, 19, 31, 43</sup>。さらに第二章において寒熱に関連する症状が虚実を予測する上で重要な要因であることも明らかにした<sup>43</sup>。暑がりの値は、実証と診断された対象者で有意に高かった。一方、全身が冷えると寒がりの値は、虚証と診断された対象者で高くなる傾向があった<sup>43</sup>。BMIはICD-11の虚実の定義に合致するが、寒熱に関連する症状は定義に含まれていない<sup>12</sup>。さらに、寒熱の予測には、寒熱に関連する症状だけでなく、BMIも重要であることが示されている<sup>19-21, 43</sup>。ICD-11では、寒熱の定義に、寒熱に関連する症状は含まれるが、BMIは含まれていない<sup>12</sup>。

第二章において、虚実と寒熱の間に強い関連性があることを明らかにした<sup>43</sup>。学習データのサンプリング時に虚実と寒熱の患者数がアンバランスであるために、虚実と寒熱の予測に重要な項目が重複した可能性がある。第二章の対象患者集団では、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倍であり、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倍であった<sup>43</sup>。したがって、アンバランスな集団から学習データをサンプリングすると、実証データは熱証の患者を多く含み、虚証データは寒証の患者が多く含まれる。先行研究においても、このようなアンバランスな集団からサンプリングしている可能性がある<sup>18-21, 43</sup>。このようなアンバランスなサンプリングを解消し、バランスの取れたサンプリングを行うことで、虚実と寒熱の重要項目の重複が解消されると考える。

本章の目的は、虚実及び寒熱予測モデルを再評価することである。

#### 2. 方法

多施設前向き観察研究を行った。

##### 2.1. 対象者

第二章と同じ6施設のレジストリを使用した<sup>43</sup>。虚実および寒熱を予測する上で重要な項目に共通な項目が多かったため<sup>43</sup>、施設を無視してすべてのデータをプールした。除外基準

は、患者データ（身長、体重、血圧、年齢など）が欠落している、データが重複している、年齢が 20 歳未満、問診回答数が 19 以下、および対象としていない証と診断された患者とした。問診回答数が 20 以上の対象者のみのデータを使用すると、虚実予測の判別率が向上したことの報告がある<sup>18, 43</sup>。さらに虚実中間証、寒熱中間証、寒熱錯雑証の患者を除外した。すべての対象者に対して書面によるインフォームドコンセントを行った。

## 2.2. 予測モデルの作成項目

予測モデルの作成項目は第二章と同様とした。

## 2.3. 機械学習を用いた Unbalanced 虚実及び寒熱予測モデルの作成

これまでの方法と同様な学習データのサンプリング方法で虚実の Unbalanced 予測モデルを構築した。寒熱を無視し、160 名の対象者（虚証 80 名、実証 80 名）をランダムにサンプリングして、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。寒熱の割合がアンバランスなため、このモデルを「Unbalanced」モデルとした。

虚実の割合は考慮せず、寒熱の Unbalanced 予測モデルを構築した。160 名の対象者（寒証 80 名、熱証 80 名）をランダムにサンプリングして、ランダムフォレストアルゴリズムの学習データとした。

虚実と寒熱を予測する重要な項目を明らかにするために、各証の Unbalanced 予測モデルから予測に寄与度が高い重要項目上位 10 項目を抽出した。学習データの残りのデータをテストデータとして使用し、これらのモデルを使用して判別率 (Accuracy)、適合率 (Precision)、再現率 (Recall)、F 値 (F-measure) を計算した。

## 2.4. 機械学習を用いた Balanced 虚実及び寒熱予測モデルの作成

次に、虚実と寒熱の 2×2 の組み合わせから、各証同じ人数になるように学習データをサンプリングした。各証の人数比率のバランスが取れていたため、このモデルを「Balanced」モデルとした。

虚実と寒熱の Balanced 予測モデルを構築するために、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 名ずつランダムにサンプリングし、学習データを取得した。Balanced 虚実予測モデルの学習データとして、160 名の対象者を虚証と実証の 2 つのグループに分けた。同様に、Balanced 寒熱予測モデルの学習データとして、寒証と熱証の 2 つのグループに分けた。

虚実と寒熱を予測するのに寄与度が高い重要な項目を明らかにするために、重要項目の上位 10 項目を抽出した。学習データの残りのデータをテストデータとして使用し、これらのモデルを使用して判別率、適合率、再現率、F 値を計算した。

## 2.5. Unbalanced 予測モデルと Balanced 予測モデルの比較

Unbalanced 虚実、Unbalanced 寒熱、Balanced 虚実、Balanced 寒熱の 4 つの予測モデルを構築した。虚実の Unbalanced 予測モデルと Balanced 予測モデル間、寒熱の Unbalanced 予測モデルと Balanced 予測モデル間で判別率を比較した。続いて、虚実の Unbalanced 予測モデルと Balanced 予測モデル間、寒熱の Unbalanced 予測モデルと Balanced 予測モデル間で重要な 10 項目を比較した。

## 2.6. 虚実予測モデルと寒熱予測モデルの比較

Unbalanced 虚実予測モデルと寒熱予測モデル間、Balanced 虚実予測モデルと寒熱予測モデル間で重要な 10 項目を比較した。

## 2.7. 統計分析

すべての統計分析は R-software version 3.5.1 を使用した (<http://www.r-project.org>)。予測モデルの作成については、第二章と同様に Breiman によって開発されたランダムフォレストアルゴリズムを使用し<sup>40-42</sup>、R の Randomforests のパッケージを使用した<sup>40</sup>。

判別率、適合率、再現率、F 値の算出には混同行列を用いて下記の式から算出した。判別率の比較には、比率の差の検定を用いた。P<0.05 を統計的有意とみなした。

$$\text{判別率} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{適合率} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{再現率} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

表 3-1 混同行列

		真	
		正	負
予測	正	真陽性 (TP)	偽陽性 (FP)
	負	偽陰性 (FN)	真陰性 (TN)

注) TP: True Positive、FP: False Positive、  
FN: False Negative、TN: True Negative

## 2. 8. 倫理審査

慶應義塾大学医学部倫理委員会の承認を受け（承認番号 20100144「漢方の特性を利用したエビデンス創出と適正使用支援システムの構築」）、これに紐づく形で各施設の倫理委員会においても承認を受け実施した。

## 3. 結果

### 3. 1. 対象者

漢方専門施設 6 施設において登録された患者数は、千葉大学 925 名、飯塚病院 791 名、慶應義塾大学 78 名、富山大学 501 名、亀田総合病院 424 名、自治医科大学 59 名であった。適格患者数は、千葉大学 193 名、飯塚病院 211 名、慶應義塾大学 161 名、富山大学 158 名、亀田医科大学 26 名、自治医科大学 1 名であった。データ欠損の内訳は、第二章と同様に、問診、西洋病名、漢方病名、薬剤、身長、体重、血圧であった。

750 名の適格患者の患者背景は、年齢が  $53.8 \pm 17.5$  歳、性別（男：女）が 149:601、BMI が  $21.2 \pm 4.1 \text{kg/m}^2$  であった。漢方専門医が診断した適格患者の虚実と寒熱の診断状況は虚証かつ寒証が 534 名、虚証かつ熱証が 57 名、実証かつ寒証が 48 名、実証かつ熱証が 111 名であった。

### 3. 2. 機械学習を用いた Unbalanced 虚実及び寒熱予測モデルの作成

Unbalanced 虚実予測モデルの学習データを無作為抽出したところ、寒熱の割合がアンバランスで、虚証 80 名のうち寒証が 74 名であるが、実証 80 名のうち寒証が 24 名であった。一方、Unbalanced 寒熱予測モデルの学習データを無作為抽出したところ、寒証 80 名のうち虚証が 74 名であるが、熱証 80 名のうち虚証は 31 名のみであった。Unbalanced 虚実予測モデ

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 で、Unbalanced 寒熱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 であった。

### 3.3. 機械学習を用いた Balanced 虚実及び寒熱予測モデルの作成

決定木の本数は 500 本、特徴量の数は 12 であった。Out-of-bag error は、虚実で 28.75%、寒熱で 27.5% であった。Balanced 虚実予測モデルの判別率、適合率、再現率、F 値は、それぞれ 0.808、0.944、0.828、0.882 で、Balanced 寒熱予測モデルの判別率、適合率、再現率、F 値は、それぞれ 0.749、0.947、0.810、0.873 であった。

### 3.4. Unbalanced 予測モデルと Balanced 予測モデルの比較

虚実の Unbalanced 予測モデルと Balanced 予測モデルで判別率を比較したところ、差はなかった。寒熱の Unbalanced 予測モデルと Balanced 予測モデルの判別率も差はなかった。

虚実の Unbalanced 予測モデルと Balanced 予測モデルの上位 10 項目を比較したところ、共通していた重要項目の中で一番重要度が高い項目は BMI であった。寒熱に関する自覚症状は、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。寒熱の Unbalanced 予測モデルと Balanced 予測モデルの上位 10 項目を比較したところ、BMI は Unbalanced 予測モデルで 1 位、Balanced 予測モデルで 8 位であった。Balanced 寒熱予測モデルでは寒熱に関する自覚症状 6 つが重要項目として挙げられ、Unbalanced 予測モデルよりも Balanced 予測モデルの方で重要度が高かった。

### 3.5. 虚実予測モデルと寒熱予測モデルの比較

Unbalanced 虚実および寒熱予測モデルで共通していた重要項目は 6 項目で、BMI、収縮期血圧、および、足が冷える、寒がり、暑がり、顔のほてりなどの寒熱に関する自覚症状であった。一方、Balanced 虚実および寒熱予測モデルで共通していた重要項目は 2 項目で、BMI、収縮期血圧のみであった。

## 4. 考察

虚実と寒熱の予測モデルを作成する上で、互いの証の影響を考慮したモデルを作成するこ

とができた。虚実と寒熱の Unbalanced 予測モデルと Balanced 予測モデルの判別率を比較したところ、差はなかった。また、第一章で得られた虚実、寒熱の診察時に診断した証と三人の医師が患者情報文書のみから評価した証の一致率の結果と Balanced モデルの判別率を比較しても差はなかった。また Unbalanced モデルであると考えられる先行研究において、谷垣らでは虚実は F 値が 0.85、寒熱は 0.88 であり<sup>19</sup>、LeeBJ らの寒熱の F 値は 0.44-0.87 であったため<sup>20</sup>、本章で作成した Balanced 予測モデルは先行研究と比較して近い値のモデルが構築できたと考えられる。したがって、Balanced 予測モデルを用いることで、虚実と寒熱を予測することができることが明らかとなった。

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<sup>18-21</sup>。虚実の Unbalanced 予測モデルの重要項目には、BMI、収縮期血圧、および寒熱に関する症状が含まれていた。先行研究の結果は、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。

Unbalanced 予測モデルと Balanced 予測モデルの重要項目を比較すると、学習データをサンプリングする際にバランスをとってサンプリングしたことで、2つのモデルの重要項目が大きく変わった。Balanced 予測モデルでは、寒熱に関係する自覚症状が虚実予測モデルの重要項目になかったが、寒熱予測モデルの重要項目には6つあった。Unbalanced 予測モデルの重要項目同士を比較すると、共通な項目が6つあったが、Balanced 予測モデルの重要項目同士を比較すると、共通な項目は2つになり、順番が大きく変わった。

虚実の Balanced 予測モデルの重要項目は、ICD-11 の証の定義と互換できる可能性がある。BMI は虚実の予測において最も重要な項目であった。ICD-11 の虚実の定義は、” a vulnerable/robust body, a weak/strong abdominal wall, feeble/forceful pulse etc.”であった<sup>12</sup>。筋力と正の相関がある骨格筋指数 (SMI) は BMI と正の相関があるとされている<sup>45</sup>。したがって、” a vulnerable/robust body”および” a weak/strong abdominal wall”は筋力と互換できるため、BMI に関連している可能性がある。” feeble/forceful pulse”は血圧に関連している可能性がある。また寒熱に関係する自覚症状は、ICD-11 の虚実の定義には含まれない<sup>12</sup>。寒熱に関係する自覚症状は、Unbalanced 予測モデルでは重要項目に含まれていたが、Balanced 予測モデルでは11位以下に低下した。

寒熱の Balanced 予測モデルの重要項目もまた、ICD-11 の証の定義と互換できる可能性がある。寒熱に関係する自覚症状は、寒熱の Balanced 予測モデルの重要項目に含まれていた。ICD-11 の寒熱の定義は、” cold/fever, aversion to cold/heat, flushed face etc.”を含む<sup>12</sup>。Unbalanced 予測モデルでは BMI が最も重要であったが、Balanced 予測モデルでは8位になった。以上から、寒熱の診断において重要な項目は、BMI ではなく、寒熱に関係する自覚症状である可能性がある<sup>21, 36, 46</sup>。

Balanced モデルを構築するにあたり、アルゴリズム、学習データサンプリング方法、VAS

の正規化について確認を行った。アルゴリズムは、教師あり学習かつ分類のアルゴリズムである線形のサポートベクターマシン、非線形の CART、XGBoost においても検討を行った。虚実の判別率は、Randomforest が 80.8%であったものの、その他は 70%代であった。寒熱の判別率は、Randomforest が 74.9%であったものの、その他は 70%代前半であった。したがって、Randomforest のアルゴリズムを用いた場合の判別率が一番高かったため採用した。これは、先行研究を支持している<sup>19</sup>。学習データサンプリング方法については、本研究では一番患者数が少なかった実証かつ寒証の患者数にあわせて 40 名ずつ各証よりランダムサンプリングを行った。一番患者数が多かった虚証かつ寒証の患者数にあわせてランダムサンプリングを行ったところ、out-of-bug error は、虚実、寒熱ともに 0.72%となり、第二章の IRR 値と大きく差が見られ、過学習していると考えられた。過学習している理由としては、実証かつ寒証や虚証かつ熱証の患者のデータが 10 倍程度になっているからと考えられる。外的妥当性が下がり、漢方的に説明がつかない項目が重要項目として挙がってきたため、今回は各証 40 名で行った。なお、本研究で構築した balanced 予測モデルの out-of-bug error は、虚実で 28.75%、寒熱で 27.5%であったため、第二章の IRR 値と差が見られなかったため、過学習していないと考えられる。本研究では VAS 値を扱ううえで、先行研究<sup>18,39</sup>に倣い正規化した。正規化せずに予測モデルを作成したところ、虚実と寒熱の判別率はそれぞれ 80.3%、58.3%となった。ランダムフォレストでは分岐の際に VAS 値で切るため、個人間で差がない方がよいのではないかと考えられる。また、寒熱の判別率が低かった原因としては、本研究で作成した Balanced モデルの重要項目の上位が自覚症状であるためと考えられる。VAS を正規化せずに寒熱の Balanced モデルを作成したところ、重要項目の上位が BMI、血圧となっており、これは正規分布に近い分布のものが上位にきてしまう可能性が考えられる。

本研究では、残された課題がある。第二章と同様に、BMI が小児には適応できないため<sup>44</sup>、本研究では 20 歳未満は除外した。本研究の予測モデルが小児への適応の可否は不明である。また症状が 19 以下の患者も除外したため、本研究の予測モデルは症状が少ない患者への適応の可否も不明である。寒熱予測モデルでは重要項目の多くが症状であったため、寒熱の予測に特に関係する可能性がある。本研究の予測モデルは患者の症状と BMI 等のバイタルサインのみから構築した。前田らの先行研究では、これらに腹力を追加することで、虚実予測の判別率が 82%から 90%に増加した<sup>43</sup>。前述したように、証の診断には、診察、聴診、および触診も必要であるため、これらを説明変数に追加した方がよい可能性がある。しかし本研究の目的は、漢方専門医でない医師や漢方薬・生薬認定薬剤師でない薬剤師が証診断を簡便に行うために予測モデルを構築することであったため、本研究で得られた結果は重要であると考えられる。本研究では、各対象者を診察した医師は 1 人であったため、虚実および寒熱の診断結果は複数の医師で検証していない。本研究では、1 人の医師の診断結果を教師データとして使用している。本研究の予測モデルのアルゴリズムは教師あり学習であるため、

教師データに依存する<sup>47</sup>。さらに虚実の定義は、日本、中国、韓国の間で異なる可能性があり、現在、虚実と寒熱の世界的な一つの定義はない。日本は気温が低いため、日本の寒熱の患者数は気温が暑い国よりも多い可能性がある。ブラジルと日本の妊婦の寒熱の割合に関する先行研究では、ブラジルの寒証の頻度は日本よりも有意に低かった<sup>48</sup>。したがって、気温が証の頻度に影響を与える可能性がある。本研究で得られた予測モデルは、漢方医学以外の伝統医学には適用できない可能性がある。

## 5. 小括

日本の6つの漢方専門施設の患者の問診結果およびバイタルサインデータからバランスを取ってサンプリングした学習データを用いることで、虚実と寒熱の予測モデルを改訂できた。さらに、虚実と寒熱を診断するために重要な項目を明らかにし、虚実と寒熱のBalanced予測式から抽出できた虚実と寒熱の重要問診項目はICD-11のそれらの定義と一致した。虚実の重要項目はBMIと血圧を含み、寒熱の重要項目は寒熱に関する症状を含む。患者の問診結果およびバイタルサインデータのみから医師が証を評価した際の一致率と差がないモデルを構築できた。

## 第四章 複数施設データを用いた虚実中間証を含む虚実の予測モデルの構築

### 1. 背景と目的

第二章および第三章において残された課題として虚実中間証を含む虚実の予測モデルの構築がある。先行研究では、虚証と実証の患者のみを抽出して虚実の予測モデルを作成しており、虚実中間証は除いていた<sup>18, 19, 43, 49</sup>。第二章の表 2-3 において、虚実中間証の人数はどの施設においても初診患者の 30~40%前後で大きな割合を占めており<sup>43</sup>、虚実を予測する上で虚実中間証を含んだ予測モデルが必要である。虚実中間証は ICD-11 において虚実中間証は虚証と実証の間に位置しているとされているが<sup>12</sup>、虚証と虚実中間証、虚実中間証と実証の境界を明らかにする必要がある。

本章の目的は、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することである。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。

### 2. 方法

多施設前向き観察研究を行った。

#### 2.1. 対象者

第二および第三章と同じ 6 施設のレジストリを使用した<sup>43, 49</sup>。第三章と同様に、除外基準は、患者データ（身長、体重、血圧、年齢など）が欠落している、データが重複している、年齢が 20 歳未満、自覚症状が 19 以下、および対象としていない証と診断された患者とした。さらに、寒熱中間証、寒熱錯雑証の患者を除外した。

#### 2.2. 予測モデルの作成項目

予測モデルの作成項目は第二章と同様とした。

#### 2.3. 機械学習を用いた虚実中間証を含んだ虚実予測モデル

6 施設のデータを全て統合した学習データを使用して虚実中間証を含む虚実予測モデルを

構築し、虚実を予測する上で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。本章では、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、学習データ以外のデータをテストデータとした。学習データを用いてランダムフォレストのアルゴリズムから虚実中間証を含む虚実予測モデルを構築した。目的変数は、実証を1、虚実中間証を0.5、虚証を0とした。説明変数は、第二、三章と同様に153項目とした(表2-1)。虚実予測モデルを用いることで、各テストデータに対して虚実予測値(0~1)が計算される。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。医師が診断した虚実の診断結果と予測モデルにより算出された虚実予測値を比較するために、テストデータの各患者の虚実予測値を、医師が診断した証毎にプロットした。

## 2.4. 全施設と各施設の虚証、虚実中間証、実証のカットオフ値、判別率

2.3で求め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)曲線を作成した。医師の診断した証について、虚証と虚実中間証、虚実中間証と実証の「感度(Sensitivity)-(1-特異度(Specificity))」の値が最大となるカットオフ値(Cutoff)を求め、真陽性(TP)、真陰性(FP)、偽陰性(FN)、偽陽性(TN)、感度(Sensitivity)、特異度(Specificity)、医師の診断との一致率(Accuracy)を算出した。

さらに虚実予測値を用いて、本予測モデルの判別率を計算した。虚実は順序データであるため、重み付けを行い一致率を算出した。医師の診断が虚証の場合、予測モデルが予測した証が虚証である場合は1、虚実中間証である場合は0.5、実証である場合は0とし、医師の診断が虚実中間証である場合、予測モデルが予測した証が虚実中間証である場合は1、虚証または実証である場合は0とし、医師の診断が実証である場合、予測モデルが予測した証が実証である場合は1、虚実中間証である場合は0.5、実証である場合は0とした。

## 2.5. 医師の診断した証と予測モデルが予測した証の重要項目の値の検討

医師の診断した証と予測モデルが予測した証が一致しなかった患者がいたため、原因を検討するために、虚実の診断に寄与度が高い重要項目上位3項目の値を検討した。医師の診断した証毎に、予測モデルが予測した実証、虚実中間証、虚証の患者の3群間で、虚実の診断に寄与度が高い重要項目上位3項目の値を比較した。

## 2.6. 統計分析

すべての解析は R3.5.1 を使用した (<http://www.r-project.org>)。予測モデルの構築については、第二、三章と同様にランダムフォレストを用い、R の Randomforests のパッケージを使用した<sup>42</sup>。本章では第二、三章で用いたランダムフォレスト分類モデルではなく、ランダムフォレスト回帰モデルを用いた。本モデルはランダムフォレスト分類モデルと同様に、多くの分類木を構築する<sup>42</sup>。虚実予測値は、各々の分類木の分類結果を平均して求める<sup>42</sup>。重要度は、分類木の分岐点毎の純度の増分 (Increase in Node Purity) をもとに計算される<sup>42</sup>。近年、ランダムフォレストは様々な分野にて用いられている。これまでに虚実、寒熱の予測をするのにランダムフォレストが最も一致率がよかったとの報告がある<sup>19</sup>。ROC 曲線は R の ROCR のパッケージを用いて作成した<sup>50</sup>。医師の診断した証と予測モデルが予測した証の重要項目の値の比較は、Tukey 検定を用いた。P<0.05 を統計的有意とみなした。

## 2.7. 倫理審査

慶應義塾大学医学部倫理委員会の承認を受け (承認番号 20100144 「漢方の特性を利用したエビデンス創出と適正使用支援システムの構築」)、これに紐づく形で各施設の倫理委員会においても承認を受け実施した。

## 3. 結果

### 3.1. 対象者

6 施設で登録された患者数は、千葉大学 925 名、飯塚病院 791 名、慶應義塾大学 781 名、富山大学 501 名、亀田総合病院 424 名、自治医科大学 59 名であった。適格患者数は、千葉大学 259 名、飯塚病院 308 名、慶應義塾大学 270 名、富山大学 189 名、亀田総合病院 39 名、自治医科大学 3 名であった。適格患者 1068 名の患者背景は、年齢が  $54.0 \pm 17.5$  歳、性別 (男 : 女) が 218 : 850、BMI が  $21.6 \pm 3.9 \text{kg/m}^2$  であった。全施設の医師が診断した虚実の診断状況は、虚証が 591 名、虚実中間証が 318 名、実証が 159 名であった。

### 3.2. 機械学習を用いた虚実中間証を含んだ虚実予測モデル

6 施設すべてのデータを統合したデータから、学習データを各証 40 名ずつ抽出し、虚実

中間証を含んだ虚実予測モデルを構築した。虚証、虚実中間証、実証毎にテストデータの虚実予測値をプロットした。医師の診断した証と予測モデルが予測した証は概ね一致していたが、医師が虚証と診断したが予測モデルが予測した虚実予測値が1に近い患者や、医師が実証と診断したが予測モデルが予測した虚実予測値が0に近い患者が少数見られた。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。

### 3.3. 全施設と各施設の虚証、虚実中間証、実証のカットオフ値、判別率

テストデータの虚実予測値を用い、実証—虚実中間証、虚実中間証—虚証のROC曲線を作成した。実証—虚実中間証のカットオフ値は、全施設が0.539、千葉大学が0.570、飯塚病院が0.517、慶應義塾大学が0.531、富山大学が0.566、亀田総合病院が0.550であった。虚実中間証—虚証のカットオフ値は、全施設が0.425、千葉大学が0.411、飯塚病院が0.460、慶應義塾大学が0.436、富山大学が0.462、亀田総合病院が0.482であった。感度と特異度はいずれも0.7程度であった。医師の診断と虚実中間証を予測した場合との一致率は、実証—虚実中間証が0.748、虚証—虚実中間証が0.688となった。施設間でカットオフ値を比較したところ、全ての施設で実証—虚実中間証のカットオフ値は0.5より大きく、虚証—虚実中間証のカットオフ値は0.5未満であった。また、本予測モデルの判別率は76.8%であった。

### 3.4. 医師の診断した証と予測モデルが予測した証の重要項目の値の検討

医師の診断した証と予測モデルが予測した証が一致しなかった患者がいた原因を検討するため、虚実の予測に寄与度が高い重要項目上位3項目であるBMI、収縮期血圧、拡張期血圧の値を検討した。医師の診断した証に関わらず、予測モデルの予測した証が虚証の患者は実証の患者より、BMIと収縮期血圧値が有意に小さかった。医師が虚証と診断したが、予測モデルが実証と予測した患者は、医師の診断と予測モデルの予測が一致した患者より、BMI値、収縮期血圧値、拡張期血圧値が有意に大きく、医師が実証と診断したが、予測モデルが虚証と診断した患者は、医師の診断と予測モデルの予測が一致した患者より、BMI値、収縮期血圧値、拡張期血圧値が有意に小さかった。医師が虚証と診断した患者のなかで、予測モデルが虚証と予測した患者は、虚実中間証と予測した患者よりBMI値、収縮期血圧値が有意に小さかった。また、予測モデルが実証と予測した患者は、予測モデルが虚実中間証と予測した患者よりBMI値、拡張期血圧値が有意に大きかった。医師が実証と診断した患者のなかで、予測モデルが虚証と予測した患者は、予測モデルが虚実中間証と予測した患者より拡張期血圧値が有意に小さかった。また、予測モデルが実証と予測した患者は、予測モデルが虚実中間証と予測した患者よりBMI値、収縮期血圧値、拡張期血圧値が有意に大きかった。

## 4. 考察

6 施設のデータを統合したデータを使用して、虚実中間証を含んだ虚実予測モデルを構築し、虚実予測に寄与度が高い重要項目を抽出したところ、BMI、収縮期血圧、拡張期血圧となり先行研究と同様であった<sup>18, 19, 43, 49</sup>。さらに予測精度について検討し、全施設の実証－虚実中間証、虚実中間証－虚証のカットオフ値を求めたところ、狭い範囲に収まった。

全施設統合データを用いた実証－虚実中間証のカットオフ値は 0.539、虚実中間証－虚証のカットオフ値は 0.425 となった。施設毎のカットオフ値も狭い範囲に収まり、0.5 をまたぐことはなかった。このカットオフ値において、感度と特異度、医師の診断との一致率は共に 0.7 程度であった。また、本予測モデルの判別率は 76.8% であった。第一章で得られた虚実の初診時に診断した証と、3 人の医師がそれぞれ患者の問診項目とバイタルサインから評価した証の一致率と比較すると、78.1-81.3% であったため、差はなかった。したがって、医師が患者の問診項目とバイタルサインから評価した際と精度はあまり変わらないモデルを構築できた。また第二章の検討において、ランダムフォレストを用いて虚実の予測に寄与度が高い重要項目は、施設間で共通な項目が多かった<sup>43</sup>。本章の検討で施設毎のカットオフ値が狭い範囲に収まったことは、参加施設の医師の虚実診断方法に大きな差異がないことが示唆され、虚証と実証の間に虚実中間証が位置することを数学的に示すことができた。

テストデータの虚実予測値を算出したところ、医師の診断結果と予測モデルの予測結果が大きく乖離するデータが一部存在した。先行研究<sup>18, 19, 43, 49</sup>と同様に、本章の予測モデルにおいても重要項目が BMI、収縮期血圧、拡張期血圧であったため、虚証だが BMI 値が大きい、あるいは実証だが BMI 値が小さい患者については、本章で用いた説明変数のみからは正しく予測できない可能性が高い。医師の診断結果が実証であったが予測結果が虚証、あるいは医師の診断結果が虚証であったが予測結果が実証であったテストデータ 70 名分のうち、医師の診断結果と腹力の診断結果が一致していなかったのは 5 名 (7%) であった (data not shown)。前田らの先行研究では、今回と同様の説明変数のみから構築した虚実中間証を含まない虚実予測モデルの予測結果と医師の診断との一致率が 82.3% であったが、説明変数に腹力を加えた虚実予測モデルの予測結果と医師の診断との一致率が 89.9% に上昇した<sup>43</sup>。しかしながら本研究は、漢方を専門としない医師および漢方薬・生薬認定薬剤師でない薬剤師でも簡便に証を診断し、漢方薬を選択することができるための支援システムを開発することが重要な目的の一つであったため、今回も腹診の結果を含めなかった。

本研究の限界は、以下の通りである。虚実中間証の定義は文献や時代により異なるが、本研究では ICD-11 の定義に従った。虚実中間証という概念を初めて提唱したのは、江戸時代

の和田東郭で、その著書の「導水瑣言」では、浮腫を「実腫」、「虚実間腫」、「虚腫」に分けている<sup>51</sup>。龍野一雄は「虚実論」で、方証相対からの発想で、処方の中には適応症が虚とも実とも言いかねる虚と実の中間的な場合があるとしている<sup>52</sup>。大塚敬節は柴胡剤を虚から実へと並べ、虚実は段階的なものであるとしている<sup>53</sup>。日本東洋医学会用語委員会で虚実の定義が決定され<sup>17</sup>、ICD-11において虚証、虚実中間証、実証が定義された<sup>12</sup>。第二、三章と同様に、小児のBMI値を成人のBMI値に適応するのが難しいため<sup>44</sup>、20歳未満は予測モデルの学習データに含めなかったことから、20歳未満には結果は適応できない。症状数が19項目以下の患者は、本研究で用いた説明変数のみから予測することが難しい<sup>18, 19, 43, 49</sup>。虚実予測に寄与度が高い重要項目はBMI、血圧であるが、これらに問診項目を加えていくことで医師の診断との一致率が向上することを先行研究で確認している。対象者より寒熱中間証、錯雑証を除外したため、さらなる研究が必要である。虚実中間証を含む虚実予測モデルを本章で作成したことにより、虚実中間証が虚証と実証の間に位置することが確認できているため、寒熱中間証についても、第三章のBalanced寒熱予測モデルに寒熱中間証の患者のデータを入ると、voteの結果が寒証と熱証の中間にくることが想定され、容易に予測できる。寒熱錯雑証は、顔は火照っているが手足が寒いという症状の患者の証であり、寒熱錯雑証が中間証の亜型であるのか、寒熱どちらも有しているのか議論がある。一人の医師が各々の患者を診断しているため、診断結果が他の医師によって評価されていない。漢方を専門としている施設のみに絞って検討を行ったため、他の東洋医学に本研究の結果が応用できるか不明である。中医学では、虚実を単独で診断することは少ないが、機械学習を用いて医師の診断を予測するモデルを構築するという研究は多数存在する<sup>31</sup>。

## 5. 小括

日本の漢方を専門としている6施設のデータから、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—虚実中間証、虚実中間証—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。医師が患者の問診項目とバイタルサインから評価した際と精度に差がないモデルを構築できた。

## 総括

本研究では薬剤師がより簡便で適切に漢方薬の選択、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。

第一章では、患者の問診結果と BMI や血圧などの簡便に得られるバイタルサインのみから漢方専門医が証診断した際の一致率を明らかにした。患者情報文書から評価した虚実の AC2 は 0.708、寒熱の AC1 は 0.542、気血水の AC1 は 0.254 であった。虚実および寒熱の IRR は、気血水よりも高かった。これらの結果は、descriptor の数を少なくし、証をモジュール化することで IRR が改善する可能性を示唆している。

第二章では、第一章で AC2 および AC1 が高かった虚実および寒熱に焦点をあて、施設毎の予測モデルを構築し、証予測に寄与度が高い重要問診項目の比較を行った。まず、日本の漢方専門施設の患者分布に同様な傾向があることを確認した。虚実および寒熱予測モデルの重要項目において、複数施設で共通している項目があり、対象とした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が示唆された。また、本章において明らかとなった重要項目は日本における虚実および寒熱を診断するために重要な項目を示している可能性がある。

第三章では、日本の 6 つの漢方専門施設の患者の問診結果およびバイタルサインデータからバランスを取ってサンプリングした学習データを使用して虚実および寒熱の予測モデルを構築した。Balanced 予測モデルの判別率と F 値は虚実がそれぞれ 0.808、0.882、寒熱がそれぞれ 0.749、0.873 であった。Unbalanced 予測モデルの判別率、第一章で得られた患者情報文書のみから医師が証を評価した際の一致率と差がないモデルを構築できた。虚実と寒熱を診断するために重要な項目を明らかにし、虚実と寒熱の Balanced 予測式から抽出できた虚実と寒熱の重要問診項目は ICD-11 のそれらの定義と一致した。虚実の重要項目は BMI と血圧を含み、寒熱の重要項目は寒熱に関する症状を含む。虚実と寒熱の予測モデルを作成する上で、互いの証の影響を考慮したモデルを作成することができた。

第四章では、複数施設データを統合したデータから虚実中間証を含む虚実の予測モデルを構築した。すべての施設の虚証、虚実中間証、実証間のカットオフ値は狭い範囲に含まれていたため、参加施設における虚実診断に大きな差異がないことを示唆するとともに、国際的には認知度の低い虚実中間証について、いずれの施設でも虚証と実証の間に位置していることを数学的に示すことができた。第一章で得られた患者情報文書のみから医師が証を評価した際の一致率と差がないモデルを構築できた。

本研究では、機械学習を用いて虚実及び寒熱予測モデルを構築した。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。

## 参考文献

1. 秋葉哲生. 医療用漢方製剤の歴史. *日本東洋医学雑誌* 2010; 62: 881-888.
2. Katayama K, Yoshino T, Munakata K, et al. Prescription of kampo drugs in the Japanese health care insurance program. *Evid Based Complement Alternat Med.* 2013; 2013: 576973. doi:10.1155/2013/576973
3. Moschik EC, Mercado C, Yoshino T, Matsuura K, Watanabe K. Usage and attitudes of physicians in Japan concerning traditional Japanese medicine (kampo medicine): a descriptive evaluation of a representative questionnaire-based survey. *Evid Based Complement Alternat Med.* 2012; 2012: 139818. doi:10.1155/2012/139818
4. 川添和義. 生薬学の伝統と革新 教育・研究・臨床の立場から 薬剤師にとって「生薬学」とは. *薬学雑誌* 2011; 131: 401-405.
5. 川添和義, 坂本久美子, 伏谷秀治, et al. 現役薬剤師を対象とした漢方薬意識調査 -効果的な漢方教育の実現に向けて- *医療薬学* 2009; 35: 351-359.
6. 草薙みか, 岩政かおり, 門間智恵, et al. 病院薬剤師を対象とした漢方意識調査による効果的な生涯教育の提案. *日本病院薬剤師会雑誌* 2013; 49: 1181-1186.
7. Park HL, Lee HS, Shin BC, et al. Traditional medicine in china, Korea, and Japan: a brief introduction and comparison. *Evid Based Complement Alternat Med.* 2012; 2012: 429103. doi:10.1155/2012/429103
8. Huang CW, Hwang IH, Lee YS, et al. Utilization patterns of traditional medicine in Taiwan and South Korea by using national health insurance data in 2011. *PLoS One.* 2018; 13 (12): e0208569. doi:10.1371/journal.pone.0208569
9. 厚生労働省. 患者のための薬局ビジョン 平成27年10月23日. 厚生労働省; 2015.
10. 首相官邸. アジア健康構想に向けた基本方針. 健康・医療戦略推進本部. Accessed 1, December, 2020. [http://www.kantei.go.jp/jp/singi/kenkouiryou/suisin/ketteisiryou/dai22/siryou22\\_1.pdf](http://www.kantei.go.jp/jp/singi/kenkouiryou/suisin/ketteisiryou/dai22/siryou22_1.pdf)
11. 株式会社メディヴァ. 「アジア健康構想」実現に向けた東洋医学のエビデンス作成に向けた実証可能性等調査報告書. 株式会社メディヴァ; 2020.
12. World Health Organization. ICD-11 Beta Draft. Accessed 20 November, 2018. <http://apps.who.int/classifications/icd11/browse/l-m/en>
13. 厚生労働省. 平成30年薬事工業生産動態統計年報. 厚生労働省医政局; 2019.
14. 橋本加奈, 柴田実香, 玉田実花, et al. 保険薬局における漢方薬の使用と服薬指導の実態調査 -漢方薬の生涯教育の提案のために- *医療薬学* 2017; 43: 373-380.
15. 合田幸弘. 漢方製剤の安全性確保に関する研究「安全に使うための漢方処方の確認票」シートの作成. 厚生労働科学研究費補助金分担研究報告書; 2014.
16. 日本漢方生薬製剤協会. 漢方薬処方実態調査. Accessed 14, April, 2017. <http://www.nikkankyo.org/aboutus/investigation/pdf/jittaichousa2011.pdf>
17. Yakubo S, Ito M, Ueda Y, et al. Pattern classification in kampo medicine. *Evid Based Complement Alternat Med.* 2014; 2014: 535146. doi:10.1155/2014/53514618.

18. Katayama K, Yamaguchi R, Imoto S, Watanabe K, Miyano S. Analysis of questionnaire for traditional medicine and development of decision support system. *Evid Based Complement Alternat Med*. 2014; 2014: 974139. doi:10.1155/2014/974139
19. Tanigaki T, Imai T, Ohe K, Yoshino T, Watanabe K. The study of prescription support model of Japanese Kampo formulas for non-Kampo medical experts. *Iryoujyohoudakkai/Jinkoutinougakkai AMI Goudoukennkyuukaisiryou SIG-AIMED-005-04*. 2018; 4: 1-5.
20. Lee BJ, Lee JC, Nam J, Kim JY. Prediction of cold and heat patterns using anthropometric measures based on machine learning. *Chin J Integr Med*. Jan 2018;24(1):16-23. doi:10.1007/s11655-016-2641-8
21. Lee J, Han SY, Hwang M, Bae N, Chae H. Biopsychological and pathophysiological features of Cold-Heat subgroup of Sasang typology with Sasang Digestive Function Inventory, Sasang Personality Questionnaire and Body Mass Index. *Integr Med Res*. Mar 2018;7(1):68-76. doi:10.1016/j.imr.2017.09.005
22. 有田龍太郎, 吉野鉄大, 堀場裕子, et al. 患者中心の自動問診システムを目指した課題抽出とその解決. *日本東洋医学雑誌* 2018; 69: 82-90.
23. Schnyer RN, Citkovitz C. Inter-Rater Reliability in Traditional Chinese Medicine: Challenging Paradigmatic Assumptions. *J Altern Complement Med*. 2019; 25 (11): 1067-1073. doi:10.1089/acm.2019.0331
24. Popplewell M, Reizes J, Zaslowski C. Appropriate Statistics for Determining Chance-Removed Interpractitioner Agreement. *J Altern Complement Med*. 2019; 25 (11): 1115-1120. doi:10.1089/acm.2017.0297
25. Popplewell M, Reizes J, Zaslowski C. A Novel Approach to Describing Traditional Chinese Medical Patterns: The "Traditional Chinese Medical Diagnostic Descriptor". *J Altern Complement Med*. 2019; 25 (11): 1121-1129. doi:10.1089/acm.2018.0065
26. Schnyer RN, Citkovitz C. Modularization of Pattern Differentiation in East Asian Medicine: Increasing Reliability or Ignoring Confounders? (Response to Yoshino and Watanabe re: ". *J Altern Complement Med*. 05 2020; 26 (5): 439. doi:10.1089/acm.2020.0041
27. Gwet KL. Package 'irrCAC'. Accessed 1 June, 2020. <https://cran.csiro.au/web/packages/irrCAC/irrCAC.pdf>
28. 西浦博. 観察者間の診断の一致性を評価する頑健な統計量 AC1 について. *日本放射線技術学会雑誌* 2010; 66: 1485-1491.
29. Gwet K. *Handbook of Inter-rater Reliability (Second edition)*. Advanced Analytics.
30. Jacobson E, Conboy L, Tsering D, et al. Experimental Studies of Inter-Rater Agreement in Traditional Chinese Medicine: A Systematic Review. *J Altern Complement Med*. 2019;25(11):1085-1096. doi:10.1089/acm.2019.0197
31. Zhao C, Li GZ, Wang C, Niu J. Advances in Patient Classification for Traditional Chinese Medicine: A Machine Learning Perspective. *Evid Based Complement Alternat Med*. 2015;2015:376716. doi:10.1155/2015/376716
32. Watsuji T, Arita S, Shinohara S, Kitade T. Medical application of fuzzy theory to the diagnostic system of tongue inspection in traditional Chinese medicine. *1999 IEEE International Fuzzy Systems Conference Proceedings*. 1999;1:145-148.
33. Kanawong R, Obafemi-Ajayi T, Ma T, Xu D, Li S, Duan Y. Automated Tongue Feature Extraction for ZHENG

- Classification in Traditional Chinese Medicine. *Evid Based Complement Alternat Med.* 2012;2012:912852. doi:10.1155/2012/912852
34. Yan JJ, Wang YQ, Guo R, et al. Nonlinear Analysis of Auscultation Signals in TCM Using the Combination of Wavelet Packet Transform and Sample Entropy. *Evid Based Complement Alternat Med.* 2012;2012:247012. doi:10.1155/2012/247012
  35. Chiu C, Yang M, Lin C. Using fractal dimension analysis on objective auscultation of traditional Chinese medical diagnosis. *Journal of Medical and Biological Engineering.* 2002;22:219-225.
  36. Ryu H, Lee H, Kim H, Kim J. Reliability and validity of a cold-heat pattern questionnaire for traditional Chinese medicine. *J Altern Complement Med.* Jun 2010;16(6):663-7. doi:10.1089/acm.2009.0331
  37. Liu GP, Li GZ, Wang YL, Wang YQ. Modelling of inquiry diagnosis for coronary heart disease in Traditional Chinese Medicine by using multi-label learning. *BMC Complement Altern Med.* Jul 2010;10:37. doi:10.1186/1472-6882-10-37
  38. Li GZ, Yan SX, You M, Sun S, Ou A. Intelligent ZHENG Classification of Hypertension Depending on ML-kNN and Information Fusion. *Evid Based Complement Alternat Med.* 2012;2012:837245. doi:10.1155/2012/837245
  39. Katayama K, Yamaguchi R, Imoto S, Matsuura K, Watanabe K, Miyano S. Clustering for visual analogue scale data in symbolic data analysis. *Procedia Computer Science.* 2011;6:370-374.
  40. Breiman L, Cutler A, Liaw A, Wiener M. Breiman and Cutler's Random forests for Classification and Regression. Package 'randomForest'. Accessed 31 October, 2018. <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
  41. Breiman L. Bagging predictors. *Machine Learning.* 1996;24:123-140.
  42. Breiman L. Random forests. *Machine Learning.* 2001;45:5-32.
  43. Maeda-Minami A, Yoshino T, Katayama K, et al. Prediction of deficiency-excess pattern in Japanese Kampo medicine: Multi-centre data collection. *Complement Ther Med.* 2019;45:228-233. doi:10.1016/j.ctim.2019.07.003
  44. Sugiura R, Murata M. Problems with body mass index as an index to evaluate physical status of children in puberty. *Pediatr Int.* Oct 2011;53(5):634-642. doi:10.1111/j.1442-200X.2010.03312.x
  45. Watarai A, Nakayama T, Mogi J, Mitsube H, Kawamura T. The Relationship between Muscle Mass and Risk Factors of Lifestyle Related Disease in Middle-aged Workers. *Journal of Japanese society of occupational medicine and traumatology.* 2017;65:269-275.
  46. Park YJ, Nam J, Do JH, Jin HJ, Kim JY. Bodily differences between Cold- and Heat-prescription groups in Sasang medicine. *Integr Med Res.* 2016;5(2):118-123. doi:10.1016/j.imr.2016.04.002
  47. Adamson AS, Welch HG. Machine Learning and the Cancer-Diagnosis Problem - No Gold Standard. *N Engl J Med.* 2019;381(24):2285-2287. doi:10.1056/NEJMp1907407
  48. Nakamura S, Ichisato SM, Horiuchi S, Mori T, Momoi M. Pregnant women's awareness of sensitivity to cold (hiesho) and body temperature observational study: A comparison of Japanese and Brazilian women. *BMC Res Notes.* 2011;4:278. doi:10.1186/1756-0500-4-278

49. Maeda-Minami A, Yoshino T, Katayama K, et al. Discrimination of prediction models between cold-heat and deficiency-excess patterns. *Complement Ther Med.* 2020;49:102353. doi:10.1016/j.ctim.2020.102353
50. Sing T, Sander O, Beerenwinkel N, Lengauer T, Unterthiner T, Ernst FGM. Visualizing the Performance of Scoring Classifiers. Package 'ROCR'. Accessed 16, August, 2019. <https://cran.r-project.org/web/packages/ROCR/ROCR.pdf>
51. 大塚敬節, 矢数道明. 近世漢方医学書集成 16 和田東郭二. 名著出版; 1979:359-360.
52. 龍野一雄. 虚実論. 漢方の臨床. 1968;15:383-392,445-457.
53. 大塚敬節, 矢数道明, 清水藤太郎. 漢方診療医典. 南山堂; 1969.

## 学位対象論文

**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).

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.

**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* 2020; **49**: 102353.

**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* 2019; **45**: 228-233.

## 謝辞

本研究を行うにあたり、御指導、御鞭撻賜りました慶應義塾大学薬学部 医療薬学・社会連携センター 医療薬学部門 中村智徳 教授に深く感謝申し上げます。以前より大変興味を持っていた漢方薬の研究を行う機会を与えてくださいます御礼申し上げます。

本研究の審査にあたり、貴重な御意見、御指導を賜りました慶應義塾大学薬学部 医療薬学・社会連携センター 社会薬学部門 山浦克典 教授、慶應義塾大学薬学部 医薬品情報学講座 堀里子 教授に厚く御礼申し上げます。

本研究の遂行ならびに論文作成にあたり、多大なる御指導、御鞭撻を賜りました慶應義塾大学 医学部 漢方医学センター 三村将 センター長、渡辺賢治 客員教授、堀場裕子 医局長、吉野鉄大 特任講師に厚く御礼申し上げます。

本研究の遂行にあたり、御協力ならびに御助言を賜りました東京大学 医科学研究所 ヒトゲノム解析センター 健康医療インテリジェンス分野 片山琴絵 助教、しきのケアセンター 引網宏彰 施設長、富山大学 医学部 和漢診療学講座 嶋田豊 教授、千葉大学大学院 医学研究院 和漢診療学講座 並木隆雄 診療教授、飯塚病院 東洋医学センター 漢方診療科 田原英一 部長、亀田総合病院 東洋医学診療科 南澤潔 部長、自治医科大学 地域医療学センター 東洋医学部門 村松慎一 教授、愛知県がんセンター研究所 システム解析学分野 山口類 分野長、東京大学 医科学研究所 ヒトゲノム解析センター 健康医療インテリジェンス分野 井元清哉 教授、東京医科歯科大学 M&D データ科学センター 宮野悟 センター長、東京大学 大学総合教育研究センター 全学教育推進部門 美馬秀樹 准教授に心より感謝申し上げます。

本研究の遂行にあたり、御支援ならびに御助言を賜りました慶應義塾大学薬学部 医療薬学・社会連携センター 医療薬学部門 鈴木小夜 教授、河添仁 専任講師、地引綾 助教、横山雄太 助教、平賀ゆい 助教に心より感謝申し上げます。

本研究の遂行にあたり、御助言いただきました慶應義塾大学薬学部 医療薬学・社会連携センター 医療薬学部門、慶應義塾大学医学部 漢方医学センターの皆様心より御礼申し上げます。

最後に、長きに渡る学生生活を支えてくれた家族に心より感謝いたします。