

인공지능 기반 평가 도구를 이용한 한의사의 체질 진단 평가 및 활용 방안에 대한 연구

박무순·황민우¹·이정윤²·김창엽^{3*}·권영규^{4*}

한국한의학연구원 한의약데이터부, 1: 경희대학교 한의과대학 사상체질과, 2: 부산대학교 한의학전문대학원 임상의학4교실 사상체질과, 3: 가천대학교 한의과대학 생리학교실, 4: 부산대학교 한의학전문대학원 양생기능의학교실

Research on the Evaluation and Utilization of Constitutional Diagnosis by Korean Doctors using AI-based Evaluation Tool

Musun Park, Minwoo Hwang¹, Jeongyun Lee², Chang-Eop Kim^{3*}, Young-Kyu Kwon^{4*}

KM Data Division, Korea Institute of Oriental Medicine,

1: Department of Sasang Constitutional Medicine, College of Korean Medicine, Kyung Hee University,

2: Department of Sasang Constitutional Medicine, Division of Clinical Medicine 4, School of Korean Medicine, Pusan National University,

3: Department of Physiology, College of Korean Medicine, Gachon University,

4: Division of Longevity and Biofunctional Medicine, School of Korean Medicine, Pusan National University

Since Traditional Korean medicine (TKM) doctors use various knowledge systems during treatment, diagnosis results may differ for each TKM doctor. However, it is difficult to explain all the reasons for the diagnosis because TKM doctors use both explicit and implicit knowledge. In this study, an upgraded random forest (RF)-based evaluation tool was proposed to extract clinical knowledge of TKM doctors. Also, it was confirmed to what extent the professor's clinical knowledge was delivered to the trainees by using the evaluation tool. The data used to construct the evaluation tool were targeted at 106 people who visited the Sasang Constitutional Department at Kyung Hee University Korean Medicine Hospital at Gangdong. For explicit knowledge extraction, four TKM doctors were asked to express the importance of symptoms as scores. In addition, for implicit knowledge extraction, importance score was confirmed in the RF model that learned the patient's symptoms and the TKM doctor's constitutional determination results. In order to confirm the delivery of clinical knowledge, the similarity of symptoms that professors and trainees consider important when discriminating constitution was calculated using the Jaccard coefficient. As a result of the study, our proposed tool was able to successfully evaluate the clinical knowledge of TKM doctors. Also, it was confirmed that the professor's clinical knowledge was delivered to the trainee. Our tool can be used in various fields such as providing feedback on treatment, education of training TKM doctors, and development of AI in TKM.

keywords : Korean medicine, Decision-making process, Sasang Constitutional Medicine, Knowledge extraction, Random Forest

서론

한의학은 임상현장에서의 결과가 오랜 기간 축적되어 다양한 질병과 질환에 대한 치료의 유효성이 입증된 한국의 전통 의학이다. 많은 한의학 이론 체계 중에서 사상의학은 한국에서 유래한 한의학 이론으로, 체질이라는 개념을 사용하여 진료를 한다. 체질은 사람이 태어날 때부터 가지는 개개인의 고유한 성질을 의미하며, 다른 한의학 이론들과는 다르게 사상의학에서는 체질을 적극적으로 활용한다. 사상의학을 활용하는 한의사들은 얼굴 생김새, 음성, 체형, 성격 등의 다양한 증상을 개개인의 고유한 성질로 파악하고 이를 근거로 진료를 하고 있다. 특히, 체질 별로 나타나는 생리적 증상인 소증은 사상의학 이론 체계에서 사용되는 개념으로, 사상의학의 지식 체계를 대표할 수 있는 지표 중 하나라 할 수 있다. 한의사들이 다양한 소증을 이용하여 체질을 진단하는 것은 환자의 특성을 정밀하게 고려하여 개인별 맞춤 치료를 할 수 있다는 장점이 있다. 하지만, 참고하는 증상이 너무 다양할 뿐더러 한의사들마다 고려하는 증상의 비중이 다른 경우도 있기 때문에 같은 환자를 진료하더라도 종종 다른 체질 진단을 내리기도 한다. 한의학이 가지고 있는 장점이라 할 수 있는 개인별 맞춤치료보다 단점이 될 수도 있는 진단의 불일치에 더 주목하게 된다면, 한의 임상의 유효성과 재현성에 문제가 제기될 수 있다. 이를 극복하기 위해서는 한의학 지식 체계가 가지고 있는 보편성과 더불어 한의사 개인이 사용하는 합리적인 지식

체계를 확인하는 방법이 필요할 것이다. 즉, 한의사가 어떠한 지식을 사용하여 임상적 판단을 내리는 지 보여줌으로써 한의학의 전반적인 진료 체계 내에서 한의사가 임상 현장에서 내리는 합리적인 판단을 명시적으로 보여주어야 하는 것이다. 한의사들이 이러한 임상 의사결정에 사용하는 지식은 명시적 지식¹⁾과 암묵적 지식²⁾으로 나누어진다³⁾. 한의사들은 진료의 근거로 사용한 암묵적 지식을 온전하게 명시적으로 표현하기 어렵기 때문에 본인이 사용한 임상 의사결정의 근거를 전부 설명하기 어렵다. 그렇기 때문에, 한의사의 암묵적 지식을 도출하는 방법에 대해 설명할 수 있어야 임상을 하는 한의사의 합리적인 지식 체계를 확인할 수 있을 것이다.

인공지능(Artificial intelligence, AI)은 지능을 가진 기계를 만드는 과학이나 공학 분야의 기술을 의미한다. 과학 분야에 사용되는 AI는 인간의 지능적 행동을 이해하는 데 사용하는데⁴⁾ 이러한 특성을 이용하여 인간의 암묵적 지식을 정량적으로 추출하기도 한다⁵⁾. AI 종류의 하나인 지도 학습(supervised learning)은 사람이 쉽게 발견할 수 없는 데이터의 내재된 패턴을 잘 식별하는 것으로 알려져 있다⁶⁾. 환자의 증상으로 변증, 처방을 맞추도록 학습시킨 지도 학습 분류 모델은 한의사가 명시적으로 설명하기 힘든 암묵적인 진료 패턴을 정량적으로 보여줄 수 있다. 이를 이용한다면, 한의사의 암묵적 지식을 도출할 수 있기 때문에 임상 의사결정에 대한 차이를 정량적으로 확인할 수 있을 것이다.

최근에 한의사들의 암묵적 지식을 도출할 수 있는 AI 기반의

Chang-Eop Kim, College of Korean Medicine, Gachon University, 1342 Seongnamdaero, Sujeong-gu, Seongnam-si, Republic of Korea

·E-mail : eopchang@gachon.ac.kr ·Tel : +82-31-750-5493

Young-Kyu Kwon, Division of Longevity and Biofunctional Medicine, School of Korean Medicine, Pusan National University, 49, Busandaehak-ro, Mulgeum-eup, Yongsan-si, Gyeongsangnam-do, Korea

·E-mail : kwon@pusan.ac.kr ·Tel : +82-51-510-8471

·Received : 2021/12/14 ·Revised : 2022/03/01 ·Accepted : 2022/03/16

© The Society of Pathology in Korean Medicine, The Physiological Society of Korean Medicine

pISSN 1738-7698 eISSN 2288-2529 <http://dx.doi.org/10.15188/kjopp.2022.04.36.2.73>

Available online at <https://kmpath.jams.or.kr>

평가 도구가 제안되었다⁷⁾. 해당 연구에서는 한방병원에 내원한 알레르기 비염 환자의 임상 및 설문 데이터를 이용하여 AI 기반의 임상 의사결정을 추출하는 도구를 구축했으며, 한의사 8명의 암목적 지식을 추출하였다. 하지만 알레르기 비염 환자 데이터 세트 하나만을 이용하였기 때문에, 다른 한의학 임상 데이터를 사용해도 평가 도구가 제대로 작동하는 지 확인할 수 없다는 문제가 있다. 또한, 변증과 처방이 아닌 다른 한의학적 임상 판단에도 이러한 평가 방법을 사용할 수 있는 지 확인할 수 없다는 문제가 있었다.

한의학자들의 암목적 지식을 도출하였던 선행 연구⁷⁾의 결과에서 교육 환경에 따라 전수되는 암목적 지식이 달라 한의사가 사용하는 암목적 지식에 차이가 생길 수 있다는 결과를 확인하였다. 연구에 참여한 한의사 8명 중 7명은 같은 한방병원에서 수련을 받았고 1명만 다른 한방병원에서 수련을 받았는데, 다른 한방병원에서 수련 받은 한의사의 임상 의사결정이 나머지 한의사들이 사용하는 암목적 지식과 다른 것을 확인하였다. 하지만 선행 연구에서는 경향성만을 확인할 수 있다는 한계가 있었기 때문에 추가 연구가 필요하였다.

그렇기 때문에 본 연구에서는 선행 연구⁷⁾에서 제안한 평가 도구를 다른 한의학 임상 데이터에 적용하고, 한의사 간에 전수되는 암목적 지식을 확인하여 구체적인 활용 방안을 제시하고자 하였다. 체질 진단 설문의 결과와 선행 연구에서 제안한 AI 기반의 평가 도구를 변형하여 해당 도구가 사상의학의 지식 체계에서도 활용될 수 있는 지 확인하였다. 또한, 임상 교수와 전공의 사이의 암목적 지식을 비교함으로써 AI 기반의 분석 도구가 임상 지식의 전달을 평가할 수 있는 지 확인하였다.

연구 방법

1. 연구 데이터 수집

평가 도구 구축에 사용한 데이터는 강동경희대한방병원 사상체질과에 내원한 178명의 환자를 대상으로 수집되었다. 대면으로 체질 진단을 실시한 한의사가 10개월 이상의 워시아웃 기간을 가진 후 소증 정보만을 이용하여 비대면으로 다시 체질 진단을 하였고, 178명 중 대면 진단과 비대면 진단이 일치하는 환자 106명을 분석 대상으로 하였다. 평가 도구 구축에 사용한 데이터는 부산대 IRB에서 심의하여 승인을 받았다(PNU IRB/2017_56_HR).

2. 한의사의 답변 작성

본 연구에는 2년 이상의 체질 진단 경력을 가진 한의사 4명이 참여하였다(Table 1). 먼저, 명시적 지식의 계산을 위해 4명의 한의사들에게 체질 진단에 사용된 소증^{8,9)}(첨부파일 1) 인구통계학적 변수(나이, BMI)의 중요도 점수를 0~100점 사이로 작성하도록 하였다. 다음으로 암목적 지식의 계산을 위해 연구에 참여한 한의사들에게 선정된 환자 106명의 소증과 인구통계학적 변수 정보를 이용하여 환자의 체질을 진단하도록 하였다.

Table 1. Characteristics of Korean Medical Doctors who Participated in the Constitution Determination

Characteristics	
A	임상 연차가 오래된 E 한방병원의 사상체질과 교수 (임상 연차 : 22년, 교수 연차 : 12년)
B	A 한의사에게 수련을 받은 F 한방병원의 사상체질과 신입교수 (임상 연차 : 11년, 교수 연차 : 3년)
C	A 한의사에게 지도를 받고 있는 3년차 전공의
D	B 한의사에게 지도를 받고 있는 2년차 전공의

3. 명시적 지식과 암목적 지식의 정의 및 계산

한의학자들이 기록한 변수(소증, 인구통계학적 변수) 중요도 점수와 체질 진단 결과를 이용하여 한의사의 명시적 지식과 암목적 지식을 정의하였다. 명시적 중요도 점수는 연구에 참여한 4명의 한의사가 작성한 46개 변수의 중요도 점수를 이용하여 계산되었는데, 한의사마다 작성한 값이 다르기 때문에 비교를 위해 중요도 점수의 합계가 100점이 되도록 정규화 하였다. 본 연구에서는 정규화 된 명시적 중요도 점수를 명시적 지식으로 정의하였다.

암목적 중요도 점수는 중요도 점수 계산에 적합한 AI 모델을 이용하였다. 암목적 중요도 점수 계산에 적합한 AI 모델은 해석이 가능해야 한다. 평가 도구를 제안하는 이유는 한의사의 암목적인 임

상 지식을 명시적으로 제시하기 위함이다. 연결주의를 기반으로 한 대부분의 AI는 작동 원리를 알 수 없는 블랙박스(black box)이기 때문에, 사람과 유사한 결론을 도출할 수 있지만 결론이 나온 이유를 설명할 수 없다¹⁰⁾. 그러므로 임상 지식이 도출되는 근거를 제시할 수 있는 AI를 사용해야 한다. 또한, 임상 지식을 도출하는 AI 모델은 다양한 변수 유형을 학습할 수 있어야 한다. 임상현장에서 생산되는 데이터는 주로 범주형 데이터(설문 데이터)와 연속형 데이터(인구통계학적 데이터, 혈액 수치 등의 실험실 데이터 등)로 이루어져 있기 때문에 두 유형의 데이터를 모두 사용할 수 있는 AI 모델이어야 한다. 그래서 위의 특징들을 만족하는 AI 모델인 RF 모델을 암목적 점수 중요도 계산을 위한 평가 도구 모델로 사용하였다^{11,12)}. RF 모델도 다른 AI 모델처럼 블랙박스의 특징을 가지고 있지만, 모델을 구축하는 데 사용된 증상 중요도 점수(feature importance)를 계산할 수 있어 다른 AI 모델과 비교했을 때 증상과 결과 사이의 상관성을 보여줄 수 있다는 장점이 있다. 또한, RF 모델은 다양한 변수 타입을 사용하여 학습시킬 수 있기 때문에 RF 모델을 이용하여 연구를 진행하였다. 본 연구에서는 환자의 소증과 한의사의 체질 진단 결과를 학습시킨 RF 모델에서 도출된 증상 중요도 점수를 암목적 지식으로 정의하였다.

RF 모델은 격자 검색 알고리즘(grid search algorithm)을 이용해 가장 좋은 정확도를 나타내는 것으로 선택되었다. 한의사 개인별로 각각 다른 초매개변수(hyperparameter)를 가지는 트리가 생성될 수 있기 때문에, 한의사 별로 각각 10만 개의 트리를 생성하고 정확도(accuracy) 평균값이 가장 높게 나오는 초매개변수를 선택하여 개인 별 맞춤 RF 모델을 생성하였다. 정확도 계산을 위하여 환자 데이터를 8:2로 나누었으며, 데이터의 80%는 모델의 학습에, 20%는 정확도 계산에 사용하였다. 본 연구에서는 RF 모델을 예측 모델로 사용하는 것이 아닌, 지식을 추출하기 위한 용도로 사용하였기 때문에 추가 검증(validation) 작업은 거치지 않았다. 개인 별 맞춤 RF 모델을 이용하여 환자의 소증과 한의사들의 진료 결과를 학습시킨 10만 개의 트리를 만들고, 10만 개의 분류기에서 추출된 증상 중요도 점수의 평균을 사용하여 암목적 중요도 점수를 계산하였다. 계산된 명시적 중요도 점수의 순위와 암목적 중요도 점수의 순위는 첨부파일(첨부파일 2)로 제시하였다.

RF 모델 구축과 증상 중요도 점수 계산은 python과 scikit-learn library¹³⁾를 사용하였으며, 중요도 점수의 시각화는 matplotlib library와 Cytoscape software¹⁴⁾를 사용하였다.

4. 평가 도구 구축 및 평가 항목 선정

본 연구에서는 중요도 점수를 이용해 한의사들의 임상 지식을 평가할 수 있는 도구를 본 연구에 맞게 변형하였다. 평가 도구는 4개의 평가 점수로 구성되었으며(Table 2), 5가지의 평가 항목으로 구성되었다(Table 3). 기존 연구에서 제시된 도구에서 중복된 정보를 제공하는 네트워크 평가 항목은 제외하였고, 변증과 처방에 대한 중요도 점수를 제공하는 항목은 체질 진단에 대한 중요도 점수로 변경하였다.

Table 2. Types of Importance Scores

Types of importance scores	Calculation method
1 한의사 개인의 명시적 중요도 점수	환자의 소증과 인구통계학적 증상의 중요도에 한의사가 직접 중요한 정도를 점수로 작성하고, 다른 한의사와의 비교를 위해 개인 별로 증상 점수의 총합을 100점으로 정규화한 점수
2 한의사 개인의 암목적 중요도 점수	한의학자의 체질 진단 결과를 이용해 학습한 RF 모델이 계산한 10만 개의 분류기 중요도 점수의 평균 점수
3 한의사 그룹의 명시적 중요도 점수	연구에 참여한 한의사들의 개인 명시적 중요도 점수의 평균 점수
4 한의사 그룹의 암목적 중요도 점수	연구에 참여한 한의사들의 개인 암목적 중요도 점수의 평균 점수

Table 3. Types of Evaluation Items

Evaluation items	
Main 1 (M1)	한의사 개인 및 그룹의 명시적, 암목적 중요도 점수 비교
Main 2 (M2)	한의사 개인의 명시적 중요도 점수와 암목적 중요도 점수의 상관관계 분석
Main 3 (M3)	레이더 차트 분석
Sub 1 (S1)	한의사 개인과 한의사 그룹의 명시적 중요도 점수 비교
Sub 2 (S2)	한의사 개인과 한의사 그룹의 암목적 중요도 점수 비교

5. 평가 도구를 활용한 암묵적 임상 지식 유사도 분석

제안한 평가 도구를 활용하여 교수의 암묵적 임상 지식이 도제식 교육으로 전공의에게 잘 전달되고 있는 지 확인하기 위한 분석을 시행하였다. 한의사가 체질 진단 시 가장 많이 고려하는 소증이 일치하는 지 확인하기 위하여, 한의사별로 암묵적 중요도 점수에서 상위 25%에 속하는 12개의 소증을 한의사가 체질 진단 시 선호하는 소증으로 정의하였다.

한의사 간의 선호 소증의 유사도는 자카드 계수(Jaccard coefficient)를 이용하여 계산하였다¹⁵⁾. 자카드 계수는 두 집합 사이의 유사도를 계산하는 방법으로, 두 집합이 완전히 동일하면 1의 값을 가지고, 두 집합 사이에 공통 원소가 하나도 없으면 0의 값을 가진다.

$$Jaccard\ score(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

결 과

1. 환자들의 일반적 특성

본 연구에 참여한 환자들은 남자 45명, 여자 61명으로 여자 환자가 남자 환자보다 많았다. 연령대는 여자 환자가 더 높으며, 특히 50세~65세 사이의 여자 환자들이 동 나이의 남자 환자들에 비해 연구에 많이 참여하였다. 한편, 남자 환자가 여자 환자보다 과체중(BMI >25)인 비율이 많았다(Table 4).

체질 환자의 소증 정보를 보고 체질을 진단한 4명의 한의사들은 모두 소양 체질로 많이 진단하였다(Table 5). 체질 진단 데이터를 이용하여 학습시킨 RF 모델은 50.1~78.9%의 일치율로 무작위로 학습한 RF 모델보다 좋은 일치율을 보였다(Table 6).

Table 4. General Characteristics of Patients Included in the Sasang Constitution Patient Data

	Category	Participants	Male	Female
Age	15세 미만	8(7.5%)	5(11.1%)	3(4.9%)
	15세~29세	24(22.6%)	12(26.7%)	12(19.7%)
	30세~49세	37(34.9%)	17(37.8%)	20(32.8%)
	50세~65세	26(24.5%)	6(13.3%)	20(32.8%)
	66세 이상	11(10.4%)	5(11.1%)	6(9.8%)
	평균	40.24±18.52	35.67±19.02	43.61±17.40
BMI	<18.5	15(14.2%)	5(11.1%)	10(16.4%)
	18.5~25	60(56.6%)	22(48.9%)	38(62.3%)
	>25	31(29.2%)	18(40.0%)	13(21.3%)
	평균	23.90±4.44	23.67±3.84	22.67±4.78
	계	106(100.0%)	45(42.5%)	61(57.5%)

Table 5. Distribution of Constitution Chosen by Korean Medical Doctors Using Sasang Constitution Data

	So-yang	Tae-um	So-um
A	70	32	4
B	49	40	17
C	47	47	12
D	39	36	31

Table 6. Classification Accuracy of Random Forest Classifier

	Accuracy	Weighted F1-score
A	0.7890±090	0.7510±110
B	0.6170±095	0.5750±105
C	0.6740±091	0.6370±098
D	0.5010±100	0.4880±109

2. 암묵적 지식 평가 도구 구성

본 연구에서는 중요도 점수를 이용하여 5개의 평가 항목을 구성하였다(첨부파일 3). Main 1은 4개 평가에 사용된 모든 중요도 점수의 경향성을 한 눈에 파악하기 위한 평가 항목이다(Fig. 1). X 축은 한의사 그룹의 명시적 중요도 점수 평균값을 내림차순으로 정렬한 설문 문항이며, 체질 데이터에서 도출된 4개의 중요도 점수(개인-명시, 개인-암묵, 그룹-명시, 그룹-암묵)를 정렬된 설문 문항을 기준으로 정렬하여 선 그래프로 나타내었다.

Main 2에서는 명시적 중요도 점수와 암묵적 중요도 점수 간의 상관관계를 이용해 명시적 지식과 암묵적 지식을 비교, 평가할 수 있게 시각화하였다(Fig. 2). 두 개의 그림은 산점도로 표현되어 자료를 보는 한의사가 경향성을 쉽게 볼 수 있게 제시하였고, 상관관계

계수와 p-value를 제시하여 명시적 중요도와 암묵적 중요도 간의 관계를 확인할 수 있게 하였다. 왼쪽 그림은 한의사 개인의 중요도 점수를(Fig. 2A), 오른쪽 그림은 연구에 참여한 한의사 그룹의 중요도 점수 평균값을 산점도로 나타낸 것이다(Fig. 2B).

Comparison of the individual and average explicit and implicit importance scores

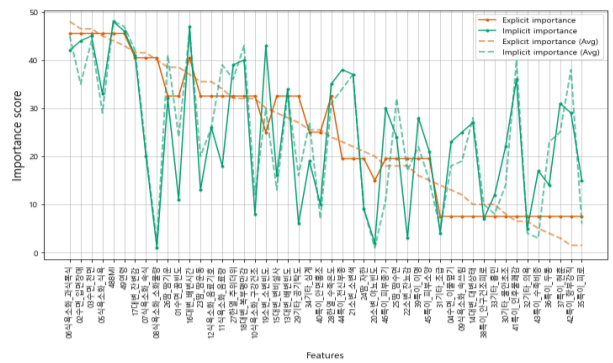


Fig. 1. Representative example of Comparison of the individual and average explicit and implicit importance scores using Sasang constitution data. The explicit importance score and implicit importance score are visualized in orange and cyan. Individual results from each doctor and the average of all doctors' results were expressed as a solid line and a dotted line, respectively. The graph was sorted in descending order according to the average explicit importance score. Avg, Average.

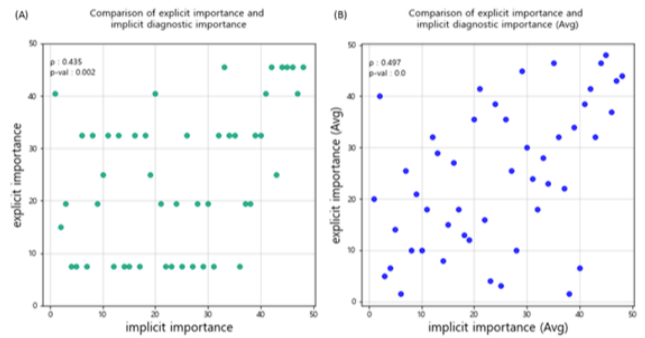


Fig. 2. Representative example of comparison of explicit and implicit importance scores using Sasang constitution data. Both graphs are sorted and visualized in ascending order based on the implicit importance score. The correlations were calculated using Spearman's rank correlation coefficient. (A) Intraindividual comparison of explicit importance score and implicit knowledge score; (B) Comparison of average explicit importance score and average implicit knowledge score.

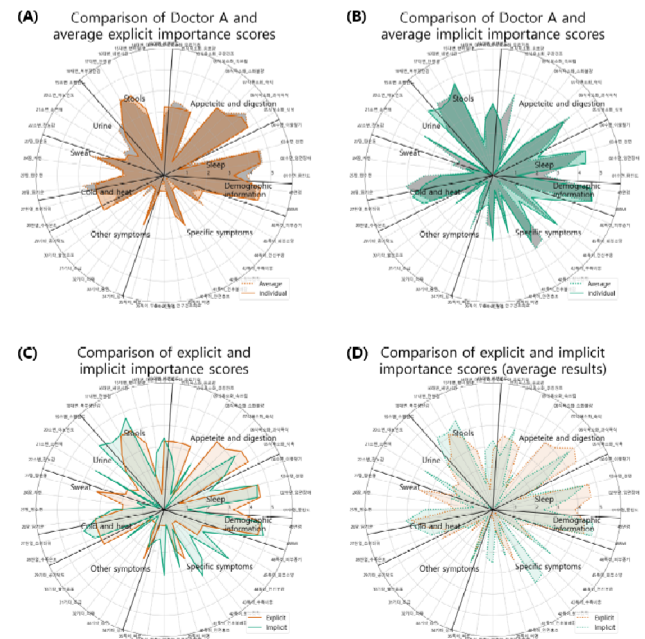
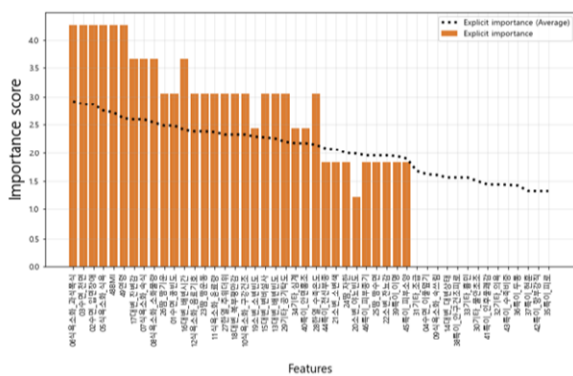


Fig. 3. Representative example of radar chart analysis scores using Sasang constitution data. Explicit importance score and implicit importance score are visualized in orange and cyan. Individual results from each doctors and average of all doctors' results were represented as a solid line and a dotted line, respectively. The radar chart was composed of four graphs: (A) Comparison of explicit importance score between individual and average results; (B) Comparison of implicit importance score between individual doctor and average results; (C) Comparison between explicit and implicit importance score results; (D) Comparison between explicit and implicit importance score (average results). In (A) and (B), the average score is visualized in grey.

Main 3에서는 레이더 차트를 이용하여, 분석에 사용된 모든 변수들을 인구통계학적 정보, 수면, 식욕-소화, 대변, 소변, 땀, 한열, 기타증상, 특이증상의 9개 카테고리로 분류하여 시각화하였다(Fig. 3). 임상 현장에서 한의사가 환자의 상태를 진단할 때, 개별 증상을 중요하게 파악할 뿐만 아니라 증상을 종합적인 상태로 파악하는 경우도 있기 때문에 증상을 카테고리 분류하여 분석을 진행하였다. 증상들의 중요도 점수를 레이더 차트에 나타내었으며, 한의사들이 어떤 증상을 선호하는 지를 카테고리 별로 파악할 수 있게 구성하였다. 다만, BMI의 중요도 점수가 너무 커서 다른 증상의 중요도 점수를 확인하기 어렵기 때문에, 중요도 점수를 순위로 바꾸어 나타내었다.

Sub 1, 2에서는 한의사 그룹의 중요도 점수의 평균값과 한의사 개인의 중요도 점수를 비교하였다(Fig. 4). Sub 1, 2는 각각 명시적 중요도 점수, 암묵적 중요도 점수 데이터를 사용하여 시각화되었다. X축은 한의사 그룹의 중요도 점수의 평균을 내림차순으로 정렬한 설문 문항이며, 평균 점수는 선 그래프(점선)로, 개인 점수는 막대 그래프로 제시하였다. 스피어만 순위 상관계수를 이용하여 한의사 그룹의 중요도 평균 점수와 개인 중요도 점수 간의 상관관계를 제시하여 통계적 유의성을 확인하였다. Sub 2에서는 증상 선호도의 평균값과 비교할 수 있게 암묵적 중요도 점수의 평균값인 baseline을 추가하여, 특정 증상을 얼마나 선호하고 있는 지 확인할 수 있게 하였다.

(A) Comparison of explicit importance score between individual and average results



(B) Comparison of implicit importance score between individual and average results

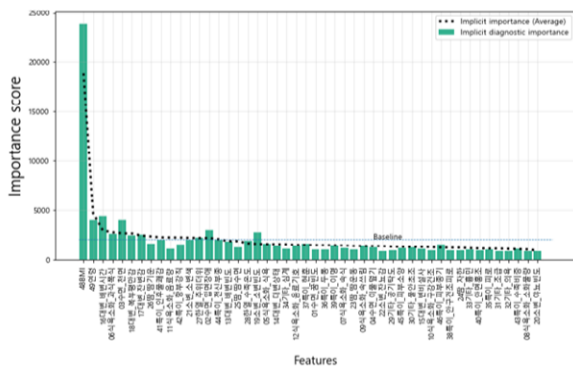


Fig. 4. Comparison of importance score between individual and average results using Sasang constitution data (representative example). Explicit importance score and implicit importance score are visualized in orange and cyan. Individual results from each doctors and average of all doctors' results were expressed as a bar graph and a dotted line, respectively. The graphs were sorted in descending order according to the importance score of average results. Baseline is the average value of implicit importance score of individual results. (A) Comparison of explicit importance score between individual and average results; (B) Comparison of implicit importance score between individual and average results.

3. 평가 도구를 사용한 암묵적 임상 지식 유사도 분석 결과

이번 분석에서는 한의사들이 체질 진단 시 많이 사용하는 선호 소증을 비교하여 암묵적 지식의 전달 여부를 확인하고자 하였다. 두 명의 한의사가 각각 선호하는 소증들 간의 자카드 계수 값을 계산함으로써, 두 한의사의 임상 지식의 유사도를 확인하였다.

비교 결과, A 한의사에게 교육을 받은 기간이 오래된 B 한의사가 비교적 교육 기간이 짧은 C 한의사보다 A 한의사와의 자카드 계수가 높은 것을 확인하였다. 또한, D 한의사는 직접 지도를 받고 있는 B 한의사와의 유사도가 A 한의사와의 유사도보다 높은 것을

확인하였다(Table 7). 이는 교수가 전공의를 지도하는 과정에서 암묵적인 임상 지식이 전달된다는 사실을 뒷받침하는 결과이다. Fig. 5에서 C 한의사는 A 한의사와 9개의 주요 소증이 일치하지만, C 한의사는 A 한의사가 중요하게 생각하는 소변 카테고리의 소증을 주요 소증으로 생각하지 않는 것을 확인하였다. 이는 아직 소변 카테고리의 소증에 대한 임상 지식이 전달되지 않은 것으로, 이러한 결과를 이용한다면 전공의가 본인의 임상 의사결정 과정을 정량적으로 피드백 할 수 있는 근거로 사용할 수 있다.



Fig. 5. Comparison of major symptoms using Sasang constitution data. Major symptoms are extracted using the Sasang constitution data. The major symptoms are those with a higher implicit importance score than the third quartile. The major symptoms are visualized in red color. The x-axis represents the major symptoms, and y-axis represents doctors who participated in this study.

Table 7. The Jaccard Coefficient of Major Symptoms

	Jaccard coefficient	Number of matching symptoms
A & B	0.714	10
A & C	0.6	9
A & D	0.412	7
B & D	0.6	9

고찰

본 연구에서는 기존에 제안된 한의사의 임상 지식을 추출하는 AI 기반의 도구가 사상의학 지식 체계에서도 잘 작동하는 지 확인하고자 하였다. 기존에 제안된 도구를 사상의학 전문가의 임상 지식을 비교할 수 있게 변형하였으며, 이를 이용하여 사상의학 전문가의 체질 진단 의사결정을 잘 추출하는 것을 확인하였다. 또한, 연구에 참여한 교수와 전공의들의 암묵적인 임상 지식을 평가 도구를 이용해 비교하여 임상 지식의 전달을 확인하였다.

기존에는 한의학 변증 및 진단 패턴의 분류에 관한 많은 연구에서 문헌 분석이나 델파이 분석이 주로 활용되었다^{16,17}. 그러나 기존의 연구 방법은 주관적 의견에 대한 의존도가 높기 때문에 임상 의사결정 과정에 대한 정량적인 지식을 도출하기 어렵다는 문제가 있었다. 본 연구에서 제안한 방법은 임상 데이터에서 RF 모델을 이용해 한의사의 임상 지식을 도출하였다. 기존 연구에서도 RF 모델을 이용하여 한의학의 변증을 이해하고자 하는 시도가 있었다¹⁸. A. Maeda-Minami 등은 6개의 병원에서 데이터를 모아 허증과 실증을 감별하는 RF 모델을 구축하였다. 위의 연구에서도 한의사의 임상 지식을 확인하기 위해 RF 모델의 지니계수(Mean decrease Gini)를 확인하여 각 병원별로 허증-실증 감별에 중요하게 생각하는 증상을 도출하였다. 기존 연구와 같이 본 연구에서도 RF 모델로부터 한의사들의 임상 지식을 확인하고자 하였다. 그러나 기존 연구와는 다르게 본 연구에서는 RF 모델에서 도출한 결과를 한의사의 암묵적 지식으로 사용했으며, 연구에 참여한 한의사들에게 명시적 지식으로 사용할 수 있는 데이터를 확보하여 RF 모델의 결과와 비교하였다는 차별점이 있다.

본 연구에서는 사상의학을 전공한 한의사들의 명시적 지식과 암묵적 지식의 차이를 확인하였다. 연구에 참여한 한의사들은 식욕, 소화, 수면 카테고리에 포함된 소증과 BMI, 연령을 체질 진단의 주요 소증으로 생각하였으며, 암묵적 지식 분석 결과에서도 비슷한 결과를 보여주었다(Fig. 1). 또한, 명시적 중요도 점수와 암묵적 중요도 점수 사이에 약한 양의 상관관계가 있는 것을 확인하여 한의사들의 명시적 지식과 암묵적 지식의 중요도가 크게 다르지 않음을 확인하였다(Fig. 2B). 그러나 몇몇 소증들은 '한의사 그룹의 명시적 중요도 점수'와 '한의사 그룹의 암묵적 중요도 점수'에서 차이가 나는 것을 확인하였다(Fig. 1). 소화불량(8번 소증)은 높은 명시적 중

요도 점수를 가지나, 암목적 중요도 점수는 매우 낮았다. 꿈을 꾸는 빈도(1번 소증), 운동 시 나는 땀(23번 소증), 구강의 건조(10번 소증), 대변의 무르기(15번 소증) 등의 소증들도 명시적 중요도 점수에 비해 암목적 중요도 점수가 낮았다. 반대로, 소변의 빈도(19번 소증), 부종(44번 소증), 소변색(21번 소증), 인후의 불쾌감(41번 소증), 항부강직(42번 소증) 등은 명시적 중요도 점수에 비해 암목적 중요도 점수가 높았다. 또한, 레이더 차트를 이용한 카테고리 분석(Fig. 3D)에서는 수면과 식욕/소화 카테고리가 명시적 중요도 점수에 비해 암목적 중요도 점수가 낮은 것을 보여주었으며, 한열증상과 기타 특이증상 카테고리가 명시적 중요도 점수에 비해 암목적 중요도 점수가 높은 것을 확인하였다. 이와 같은 정량적인 접근은 한의사 본인의 임상 진료과정에 도움이 될 것이다. 합의를 통해 만들어진 소증 설문 문항이기 때문에 모든 설문 문항이 의미를 가지고 있을 것이지만, 체질 진단 과정에서 선호하는 소증은 한의사마다 다를 것이다. 이렇게 한의사 본인이 어떤 소증을 암묵적으로 중요하게 생각하고 있는 지 확인할 수 있다면, 본인의 명시적 지식을 갱신하거나 의식적으로 특정 소증에 대한 비중을 조정하여 임상적 판단을 할 수 있는 근거가 마련될 수 있다. 예를 들어, Fig. 4B에서 BMI는 RF 모델을 구축할 때 가장 기여를 많이 한 변수로, 한의사들이 체질 진단을 할 때 우선적으로 고려되고 있다. 이는 BMI라는 변수가 체질 집단, 특히 임상에서 태음인과 태음인이 아닌 사람을 구분할 때 가장 확실하게 사용될 수 있는 지표임을 보여준다. 또한, 한의사 그룹의 BMI 기여도는 평균 18.88%인데 비해, Fig. 4B에 예시로 참여한 한의사 개인은 BMI 기여도가 23.93%인 것으로 나타났다. 이는 해당 한의사가 다른 한의사들보다 BMI 지표를 더 선호하는 결과로, 이를 이용하여 임상에서 본인의 BMI 지표에 대한 비중을 조절하거나 다른 한의사를 교육할 때 사용할 수 있을 것이다. 이렇게 본 연구에서 제안한 도구는 임상 발전과 한의사 본인의 발전에 큰 기여를 할 수 있을 것이다.

본 연구가 가지는 중요한 의의는 선행 연구에서 제안된 평가 도구의 구체적인 활용 방안을 제시한 것이다. 분석 도구를 활용하여 교수의 임상 지식이 전공의에게 전달이 얼마나 되었는지 보여주었다(Fig. 5). Figure 5에서 현재 사제 관계인 A 한의사-C 한의사와 B 한의사-D 한의사간에 암묵적으로 중요하게 생각하고 있는 소증의 중요도 차이를 확인할 수 있다. 음료를 마시는 양(11번 소증)과 수면 중에 흘리는 땀(25번 소증)은 B 한의사와 D 한의사만 중요하게 생각하고 있고, 반대로 전신의 부종(44번 소증)은 A 한의사와 C 한의사만이 중요하게 생각하고 있다. 또한, 직접적인 사제 관계가 아닌 A 한의사-D 한의사와 B 한의사-C 한의사가 유의하게 암묵적으로 중요하게 생각하는 소증은 없었다. 이는 직접적인 사제 관계에서 지식의 전달이 일어나는 것을 보여주는 것으로 생각된다. 이를 활용한다면 전공의의 교육이 잘 이루어지고 있는지, 잘 이루어지지 않았다면 어떠한 부분에서 지식 전달이 부족하지 알 수 있는 근거로 사용할 수 있을 것이기 때문에 한의학 임상 교육에 유용하게 사용될 수 있을 것이다.

Fig. 5에서 A 한의사와 B 한의사는 11번, 25번, 44번을 비롯한 몇 가지 주요 소증에서 차이를 보이고 있다. 이러한 결과는 사상의학이라는 한의학의 지식 체계 내에서 한의사가 암묵적으로 본인만의 합리적 지식 체계를 학습하였기 때문으로 생각할 수 있다. 예를 들어, 음료를 마시는 양과 부종은 체내 수분이라는 정보를 공유하여 서로 상관관계가 높은 소증이기 때문에, 생리병리학적 기전이 아닌 증상의 관점에서는 선호도의 차이로 주요 소증의 차이가 있을 수 있다. 즉, 한의사 개인이 학습한 경험과 개인의 합리적인 판단에 따라 암묵적인 증상 선호도의 차이가 날 수 있는 것이며, 본 연구에서 사용한 평가 도구로 이를 보여주었다.

그러나 이번 연구가 가지는 한계점도 있다. 한의사들의 임상 의사결정 과정은 특정 증상을 조건부로 다른 증상과 조합하여 의사결정을 도출하기 때문에 비선형적인 요소를 가지고 있다고 할 수 있다. 예를 들어, 한의학에서는 ‘손이 차갑다’는 증상으로만 한열을 판단할 수 없으며, 손이 차갑다는 증상을 조건부로 다른 증상을 참고하여 한열 변증의 판단이 이루어진다. 하지만, 본 연구에서 제안한 RF 모델의 증상 중요도는 체질과 증상 변수 사이의 정보량을 기반으로 계산된 값이기 때문에 비선형적인 관계로 보기는 어렵다. 그럼에도 불구하고 본 연구는 처음으로 한의사들의 의사결정을 AI을 이

용해 객관화할 수 있는 방법을 제안한 것에 의의가 있다. 추후 연구에서는 한의사들의 증상 조합을 고려하여 임상 의사결정 과정을 평가할 수 있는 분석 도구가 제안되어야 할 것이다.

또한, 연구에 참여한 한의사의 수가 체질 진단에 사용되는 임상적 지식을 일반화할 만큼 충분하지 않다는 한계도 있다. 그렇기 때문에 본 연구에서 나온 결과를 임상진료지침이나 설문지의 개발 등에 적용하기에 무리가 있다. 그러나 본 연구에서 제안한 평가 도구는 한의사 개개인의 지식을 각각 정량적으로 확인할 수 있다는 장점이 있다. 한의학의 지식 체계는 한의사 개인의 관점을 중요하게 생각하는 특성을 가지고 있는데, 이와 같은 방법은 개인의 특수성을 잘 표현할 수 있는 방법으로 사용될 수 있다. 또한, 충분히 많은 체질 진단 전문가가 참여를 하여 임상적 지식을 일반화할 만큼 데이터가 축적된다면 보편적인 결론도 낼 수 있기 때문에, 추후 연구에서 한의학이 가진 보편성과 한의사 개인이 가진 특수성을 모두 고려할 수 있는 연구를 진행할 수 있을 것이라 전망한다.

최근 관심이 높아진 정밀의학은 같은 질병을 가진 환자에게 동일한 치료를 제공하는 가이드라인 기반의 치료와는 달리 개인의 차이를 고려한 새로운 환자 치료 방법이다¹⁹⁾. 한의학에서 이루어지는 변증 과정은 ‘사람은 전일성을 가진 각각의 유기체’라는 전제를 가지고 이루어지기 때문에, 개인 차이를 고려하는 정밀의학에 사용될 수 있다. 특히, 사상의학은 체질이라는 개념을 사용하여 다른 한의학 지식체계보다도 정밀의학에 더욱 많은 아이디어를 제공할 수 있다. 그러나 한의사 본인이 체질 진단을 하는 근거를 명확하게 설명하지 못한다면, 유용한 아이디어를 가지고 있더라도 정밀의학으로 발전하기는 어려울 것이다. 최근 기계학습과 같은 정량적-계산적 접근이 발달함에 따라 한의사들의 임상적 결정에 대한 암묵적 지식의 탐구가 가능하게 되어 치료 근거를 명시적으로 제시할 수 있게 되었다. 본 연구에서는 한의사들의 체질 진단에 사용된 임상 의사결정의 암묵적 지식을 계산하였으며, 이를 이용하여 전수되는 임상에서의 암묵적 지식을 확인하였다. 이는 한의사의 임상적 판단에 도움이 될 뿐만 아니라, 한의학과 사상의학이 추후 정밀의학으로써 발전하는데 큰 기여를 할 것이다.

결론

본 연구에서는 선행 연구에서 제안된 암묵적 지식 평가 도구를 변형하여 한의사들이 체질 진단 과정에서 사용되는 명시적 지식과 암묵적 지식을 평가, 비교하였다. 또한, 평가 도구의 구체적인 활용 방안을 제안하여, 다음과 같은 결론을 도출하였다.

본 연구에서는 사상의학을 전공한 한의사들이 체질 진단 의사결정에서 사용하는 암묵적인 임상 지식을 추출하여 명시적 지식과 비교하였다. 체질 진단에 사용된 한의사의 명시적 지식과 암묵적 지식의 중요도는 비교적 일치하는 결과를 확인했으나, 명시적 지식과 암묵적 지식의 중요도가 일치하지 않는 소증도 있음을 확인했다. 이러한 결과는 연구에서 제안한 도구가 한의사의 체질 진단 및 임상 지식의 습득과 숙달에 도움이 될 것으로 전망된다. 또한, 본 연구에서는 평가 도구를 활용해 교수와 전공의의 임상 지식을 비교하여 교수의 임상 지식이 전공의에게 전달되는 것을 확인했다. 이러한 결과를 통해 전공의의 교육에 사용하는 평가 도구를 임상에 적용할 수 있는 방안을 제안하였으며, 임상현장에서 사용 가능한 도구의 유용성을 확인하였다.

첨부파일

해당 논문의 첨부파일은 다음 주소에서 확인할 수 있음.

<https://drive.google.com/drive/folders/13-7kEjXUqhV-oGOpA2bGohc7QjdA1GO>

감사의 글

이 논문은 2020년 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기본연구사업(No. 2020R1F1A1075145), 그리고 한국한의학연구원 ‘AI 한의사 개발을 위한 임상 빅데이터 수집 및 서비스 플랫폼 구축 (KSN2021110)’

과제의 지원을 받아 수행된 연구임.

References

1. Polanyi M, Sen A. *The tacit dimension*: University of Chicago press; 2009.
2. Davies M. Knowledge (Explicit, Implicit and Tacit): Philosophical Aspects. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*: Academic Press; 2015. p. 74-90.
3. Dummett M. *The logical basis of metaphysics*: Harvard university press; 1991.
4. McCarthy J. *What is artificial intelligence?* 2007.
5. NEȘTIAN AȘ, SilviuMihail T, GUTĂ AL, editors. Incorporating artificial intelligence in knowledge creation processes in organizations. *Proceedings of the International Conference on Business Excellence*; 2020: Sciendo.
6. Christopher M. Nasrabadi N. *Pattern Recognition and Machine Learning*. *Journal of Electronic Imaging*. 2007;16(4):049901.
7. Park M, Kim MH, Park S-Y, Kang M, Choi I, Kim C-E. Development of an explicit and implicit knowledge identification tool for the analysis of the decision-making process of traditional Asian medicine doctors. *MedRxiv*. 2021.
8. Park M, Lee M, Hwang M. Development of Diagnostic Indicator for the Sasang Constitution Exterior-Interior Disease Based on Original Symptom. *Journal of Sasang Constitutional Medicine [Internet]*. 2020 Dec 31;32(4):65-85.
9. Shin S, Kim Y-H, Hwang M-W. Diagnosis and treatment principle in Sasang medicine: original symptom. *Integrative Medicine Research*. 2016;5(2):99-104.
10. Rudin C, Radin J. Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*. 2019;1(2).
11. Vens C, Costa F, editors. *Random forest based feature induction*. 2011 IEEE 11th international conference on data mining; 2011: IEEE.
12. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*. 2009;10:213.
13. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. *Scikit-learn: Machine learning in Python*. *the Journal of machine Learning research*. 2011;12:2825-30.
14. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. *Genome research*. 2003;13(11):2498-504.
15. Jaccard P. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1. *New Phytologist*. 1912;11(2):37-50.
16. Jang E, Lee EJ, Yun Y, Park YC, Jung IC. Suggestion of Standard Process in Developing Questionnaire of Pattern Identification. *Journal of Physiology & Pathology in Korean Medicine*. 2016;30(3):190-200.
17. Park C-H, Lim H-H. The Review on the Studies Using Delphi Method in the Journals of Korean Medicine from 2006 to 2015. *Journal of Korean Medicine*. 2015;25(3).
18. Maeda-Minami A, Yoshino T, Katayama K, Horiba Y, Hikiami H, Shimada Y, et al. Prediction of deficiency-excess pattern in Japanese Kampo medicine: Multi-centre data collection. *Complement Ther Med*. 2019;45:228-33.
19. Hudson K, Lifton R, Patrick-Lake B, Burchard EG, Coles T, Collins R. *The precision medicine initiative cohort program—Building a Research Foundation for 21st Century Medicine*. Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee to the Director, ed. 2015.