대한한의진단학회지 제30권 제1호 (2026년 10월)
The Journal of the Society of Korean Medicine Diagnostics. 2026;30(1):107-118.
[원저 Original Article]

황련제 계열 처방 추천을 위한 지식기반 임상 의사결정지원시스템 '준차트'의 시나리오 기반 성능평가

김송이[1], 노의준[2,3], 박진욱[2,3], 김병수[4*]
[1]가천대학교 한의과대학 · [2]정인적방연구소 · [3]교감한의원 · [4]대전대학교 한의과대학 생리학교실

Scenario-based Performance Evaluation of Joonchart, a Knowledge-based Clinical Decision Support System for Hwangryeonje Formula Recommendation

Song-Yi Kim[1], Euy-Joon Ro[2,3], Jinwook Park[2,3], Byoung-Soo Kim[4*]
[1]College of Korean Medicine, Gachon University · [2]Jeongin Jeokbang Institute Inc.
[3]Gyogam Korean Medicine Clinic · [4]College of Korean Medicine, Daejeon University

Abstract

Objectives : To evaluate the consistency of formula recommendations generated by Joonchart, a knowledge-based clinical decision support system, for Hwangryeonje (Coptis-containing formulas), and to assess independent clinicians' agreement with prespecified reference formulas.

Methods : Thirty-nine standardized scenarios representing typical and modified presentations of 16 Hwangryeonje formulas were evaluated. Joonchart ranked candidate formulas from its entire formula knowledge base without restricting recommendations to Hwangryeonje. Top-1 agreement and top-2 and top-3 inclusion rates were calculated with Clopper-Pearson exact 95% confidence intervals (CIs). Separately, three Korean medicine doctors uninvolved in knowledge-base construction or recommendation generation independently selected one of the 16 formulas for each of 18 newly constructed typical scenarios. Agreement with reference formulas and inter-rater agreement using Fleiss' κ were assessed.

Results : Top-1 agreement was 94.9% (37/39; 95% CI, 82.7-99.4%). Both top-2 and top-3 inclusion rates were 100.0% (39/39; 95% CI, 91.0-100.0%). In both top-1 discordant scenarios, Baekduonggagamchoagyo-tang ranked second after Baekduong-tang. Individual clinicians' agreement with reference formulas ranged from 88.9% to 94.4%. At least two clinicians agreed with the reference formula in 17/18 scenarios (94.4%), and all three agreed in 16/18 (88.9%). Inter-rater agreement was Fleiss' κ = 0.96.

Conclusion : Joonchart showed high agreement with prespecified reference formulas in this scenario-based internal evaluation. A separate clinician assessment also showed high reference-formula and inter-rater agreement. These findings support the feasibility of implementing structured formula differentiation criteria in a knowledge-based system. External validation using independent scenarios and real-world clinical data is needed to establish generalizability and clinical utility.

Keywords : Korean medicine; Clinical decision support system; Knowledge-based system; Herbal formula recommendation; Coptis-containing formulas; Inter-rater agreement

---

* 교신저자 : Byoung-Soo Kim. College of Korean Medicine, Daejeon University. College of Korean Medicine, Daejeon University, 62 Daehak-ro, Dong-gu, Daejeon, 34520, South Korea.
Tel: +82-42-280-2616, E-mail: kbsoo25@dju.kr
· 원고접수일 : 2026.09.13 / 심사완료일 : 2026.09.22 / 게재결정일 : 2026.09.29.

Ⅰ. 서 론

한의 임상에서 처방의 선택은 환자가 호소하는 증상뿐 아니라 한열, 허실의 경향, 수면, 소화 및 대변 양상, 정서적 반응, 신체소견 등 다차원적 정보를 종합하는 변증 과정을 통해 이루어진다. 이러한 변증 기반 처방 결정은 개별 환자의 상태를 반영할 수 있다는 장점이 있으나, 수집되는 임상 정보의 범위와 해석 방식, 처방 간 감별 기준의 적용에는 임상의의 지식과 경험이 상당 부분 개입될 수 있다. 또한 실제 진료환경에서는 문진 정보를 포함한 환자의 임상 정보들이 일관된 형식으로 기록되지 않는 경우가 있어 한의 임상지식을 구조화하고 재현 가능한 형태로 활용하기 위한 표준화가 필요하다[1].

‘임상 의사결정지원시스템(clinical decision support system, CDSS)’은 환자별 임상정보를 지식기반 또는 알고리즘과 결합하여 진단, 치료 및 처방 결정을 보조하는 컴퓨터 기반 시스템이다[2]. CDSS는 임상의의 의사결정을 대체하기보다 진료과정에서 필요한 정보를 적절한 시점에 제공하고, 고려해야 할 선택지와 근거를 구조화하여 제시함으로써 의사결정의 일관성과 효율성을 높이는 것을 목적으로 하며, 이는 실제 임상 수행의 향상과 연관되는 것으로 보고되었다[2,3].

한의학 분야에서도 전자의무기록, 변증, 처방 정보를 기반으로 진단 및 처방을 지원하려는 연구가 이루어져 왔다[1,4,5] 이 과정에서 전통의학의 진단 및 처방 개념을 디지털 환경에서 활용하기 위해서는 다양한 한의학적 임상용어와 판단기준을 표준화하고, 한의 처방의 복합적인 감별 기준을 임상적으로 해석 가능한 형태로 구현하고 그 성능을 평가하는 과정이 중요하다[1,3]. 그러나 한의 처방은 하나의 증상과 하나의 처방이 직접 대응하는 방식으로 이루어지지 않으며, 유사한 처방 사이에서도 증상의 정도와 조합, 한열, 허실의 편향, 소화·배변 양상, 수면·정서 증상 및 신체소견에 따라 최종 처방이 달라질 수 있다. 이에 따라 처방 추천시스템은 단순한 증상-처방 대응을 넘어, 유사 처방 사이의 감별 기준과 각 임상정보의 상대적 중요도를 체계적으로 반영하여 후보 처방의 우선순위를 산출할 필요가 있다. 지식기반 규칙 및 가중치 접근은 전문가의 진단 및 처방 논리를 구조화하여 디지털 환경에서 구현할 수 있다는 점에서 한의학적 임상 의사결정지원에 활용 가능성이 있다[4-6].

고방에서 제시된 황련제 계열 처방군은 황련을 중심 약물로 포함하면서도 소화불량, 오심·구토, 설사 또는 변비, 한열 양상, 번조·불면, 흉만·심계, 출혈 및 점막 관련 증상 등에 따라 처방이 세분되는 처방군으로, 황련탕, 반하사심탕, 생강사심탕, 삼황사심탕, 소함흉탕, 황련아교탕 및 백두옹탕 등이 포함된다. 이들은 일부 공통된 임상적 특징을 공유하면서도 특정 증상과 신체소견에 따라 상호 감별이 요구된다. 따라서 황련제 계열은 한의학적 처방 기준을 구조화하고, 구조화된 처방기준이 후보 처방의 선별과 유사 처방 간 우선순위 결정에 적절히 반영되는지를 평가하기에 적절한 초기 검증 대상이 될 수 있다.

‘준차트(Joonchart)’는 문진, 변증 및 신체소견을 구조화된 항목으로 입력 받아, 사전에 구축된 처방 기준과 비교하고, 처방별 적합도를 산정하여 후보 처방을 순위화하여 제시하는 지식기반 임상 의사결정지원시스템이다. 본 연구에서는 준차트에 구축된 황련제 계열 16개 표준처방을 대상으로, 전문가 기준에 따라 구성된 전형적 특성과 일부 임상정보의 변화를 반영한 표준화 시나리오를 구성하고, 사전에 지정된 기준 처방이 준차트의 상위 추천 결과에 포함되는지를 평가하였다. 본 연구의 목적은 실제 환자에서 준차트의 임상적 처방 정확도 또는 치료

효과를 검증하는 것이 아니라, 황련제 계열에 대해 구조화된 처방 감별 기준이 준차트의 추천 로직에서 의도한대로 구현되는지를 시나리오 기반으로 확인하는 데 있다. 이를 통해 유사 처방 간 감별과정에서 보완이 필요한 지식규칙을 파악하고 향후 실제 임상자료를 활용한 후향적 성능평가 및 임상적 타당성 검증을 위한 기초자료를 마련하고자 한다.

Ⅱ. 연구방법

1. 연구 설계 개요

1) 연구 설계

본 연구는 지식기반 임상 의사결정지원시스템인 '준차트(Joonchart)'에 구축된 황련제 계열 처방 추천 로직의 구현 일관성을 평가하기 위한 시나리오 기반 내부 성능평가 연구이다. 본 연구의 검증 대상은 실제 환자의 진료기록이 아니라, 황련제 계열 처방의 전형적인 임상 양상을 반영하여 사전에 구성된 표준화 시나리오이다.

각 시나리오에는 사전에 기준 처방(reference formula)을 지정하였다. 이후 시나리오의 구조화된 임상정보를 준차트에 입력하여 추천 처방 순위를 산출하고, 기준 처방과 시스템 추천 결과의 일치 정도를 평가하였다. 또한 시스템의 지식베이스 구축과 추천 결과 산출에 참여하지 않은 한의사 3명이 준차트 결과를 알 수 없는 상태에서 시나리오 및 기준 처방의 적절성을 독립적으로 평가하였다.

본 연구는 전문가 기준으로 구성된 표준화 시나리오를 대상으로 한 초기 내부 성능평가이므로, 실제 임상 진료기록에 대한 외부 검증 또는 환자 치료결과에 대한 임상적 유효성 평가는 포함하지 않았다.

2. 황련제 계열 처방 지식베이스

준차트의 처방 지식베이스 중 황련제 계열은 준차트 개발자(노의준 원장)가 구축한 표준 처방 목록을 기준으로 추출하였다. 해당 지식베이스는 처방명, 구성 약물 및 용량, 배량 범위, 처방기준 및 환자용 처방명으로 구성되어 있으며, 『상한론』과 『금궤요략』에 제시된 황련제 계열에는 총 16개 표준 처방이 포함되었다(Table 1). 대상 처방은 황련탕-생강, 황련탕-건강, 반하사심탕, 건강황금황련인삼탕, 생강사심탕, 감사거삼탕(甘草瀉心湯去人蔘), 부자사심탕, 삼황사심탕, 대황황련사심탕, 소함흉탕, 황련아교탕, 갈근황금황련탕, 백두옹탕, 백두옹가감초아교탕(白頭翁加甘草阿膠湯), 오매환, 오매환거부자(烏梅丸去附子)로 구성되었다. 각 처방에는 소화불량, 오심·구토, 설사 및 변비 양상, 한열 경향, 수면 및 정서 증상, 흉만·심계, 신체 압통 및 출혈·점막 관련 소견 등 처방 감별에 필요한 임상 항목이 처방기준으로 구조화되어 있다. 본 연구에서는 이러한 기준을 기반으로 유사 처방 간 감별이 요구되는 황련제 계열을 초기 성능평가 대상 처방군으로 선정하였다.

표 1 Table 1. Knowledge base of Joonchart for 16 standard Hwangryeonje formulas

처방군	기준 처방명	핵심 판단 기준 요약
황련탕·사심탕 계열	황련탕-생강	소화불량 또는 식욕부진, 오한·궐냉 또는 수척을 동반한 번적 양상
	황련탕-건강	황련탕-생강의 기본 양상에 한리 경향이 두드러지는 경우
	반하사심탕	소화불량과 한리 우세의 설사, 오심·구토를 동반하는 실열 경향
	건강황금황련인삼탕	반하사심탕보다 실열 경향이 강하고, 비만·오열이 있으면서 오심·구토가 뚜렷하지 않은 경우
	생강사심탕	반하사심탕에 비해 열리 또는 정상변 경향이 있고, 오심·구토·트림이 두드러진 경우
	감사거삼탕	반하사심탕 계열 중 소화불량·식욕부진은 뚜렷하지 않고 점막궤양 또는 설사 경향이 있는 경우
	부자사심탕	변비·복만번조의 사심탕 양상에 오한·궐냉, 불인 또는 침중감이 동반된 경우
	삼황사심탕	변비·복만번조와 소화불량, 비만 또는 오열 경향이 동반된 경우
	대황황련사심탕	변비·복만번조를 중심으로 한 번열 양상
	소함흉탕	심계 또는 흉만이 두드러지고, 변비 또는 정상 대변 양상을 보이는 경우
	황련아교탕	불면·번적 양상에 작약근 긴장·압통, 안진 또는 건조성 피부병·출혈·신체통이 동반된 경우
설사·이질 계열	갈근황금황련탕	열리·오열과 현저한 항강견통을 동반한 경우
	백두옹탕	열리·오열을 보이나 갈근황금황련탕에 비해 항강견통이 뚜렷하지 않은 경우
	백두옹가감초아교탕	백두옹탕 양상에 피부병 또는 출혈 소견이 동반된 경우
오매환 계열	오매환	만성 설사와 한리·오한·궐냉 또는 하복냉이 두드러지고, 성음·안정 경향을 보이는 경우
	오매환거부자	오매환 양상 중 오열 경향이 있어 부자를 제외하는 경우

주. 본 표의 핵심 판단 기준은 준차트 표준 처방 데이터베이스에 기록된 처방기준을 연구목적에 맞게 요약한 것이다. 약물 구성, 용량, 배량 범위 및 세부 항목별 가중치는 지식베이스의 세부 구현 정보에 해당하므로 제시하지 않았다.

3. 준차트 표준화 검증

1) 표준화 검증 시나리오의 구성

검증 시나리오는 대상 처방별 전형 사례와 변형 사례를 포함하도록 구성하였다. 16개 표준 처방에 대해 총 39개 사례(A_001-A_039)가 포함되었으며, 각 사례에는 기본 인적사항, 주요 증상, 변증 관련 정보 및 문진·신체소견 항목이 구조화된 입력 형식으로 기록되었다. 각 사례에는 사전에 기준 처방명이 지정되었고, 처방별 사례 수는 2례에서 4례로 구성되었다.

전형 사례는 해당 처방의 대표적 처방기준을 반영하도록 구성하였고, 변형 사례는 동일 처방 범주 안에서 일부 증상 또는 신체소견이 달라진 경우에도 목표 처방이 유지되는지를 확인하기 위해 구성하였다. 이에 따라 본 검증은 단순한 처방명 확인이 아니라, 황련제 계열 내 유사 처방의 감별 기준이 준차트의 추천 순위에 반영되는지를 평가하도록 설계되었다.

각 사례의 기준 처방은 검증 실행 전에 사전 지정하였으며, 준차트의 추천 결과는 2026년 9월 22일에 산출하였다. 추천 결과에 표시되는 전형·변형 표기는 동일 표준 처방 내의 세부 기준 유형을 의미하며, 성능평가에서는 표준 처방명을 기준으로 일치 여부를 판정하였다.

2) 준차트 입력 및 추천 결과의 산출

각 표준화 검증 시나리오의 문진·변증 및 신체소견 정보를 준차트 입력 양식에 입력하였다. 평가 시나리오의 기준 처방은 황련제 계열 16개 표준 처방으로 한정하였으나, 준차트의 추천 후보 범위는 황련제 계열로 제한하지 않고 실행 당시 전체 처방 지식베이스를 대상으로 하였다. 준차트는 입력 정보와 전체 지식베이스의 처방기준을 비교하여 후보 처방의 적합도를 산정하고, 적합도에 따라 추천 순위를 제시하였다. 각 사례에서 Top-1부터 Top-4까지의 추천 처방을 기록하고, 사전에 지정된 기준 처방이 해당 추천 결과에 포함되는지와 그 순위를 확인하였다.

추천 결과에는 '전형' 또는 '변형'과 같은 세부 유형 및 빈용도 정보가 함께 표시되었다. 그러나 본 연구의 일치도 평가는 세부 유형이 아닌 표준 처방명을 기준으로 수행하였다. 예를 들어 '황련탕-생강(전형/7)'과 '황련탕-생강(변형/7)'은 모두 황련탕-생강으로 분류하였다.

3) 평가 변수

본 연구의 일차 평가변수는 Top-1 일치도였다. Top-1 일치도는 각 검증 시나리오에서 준차트가 가장 높은 순위로 제시한 표준 처방명이 사전 지정된 기준 처방명과 일치한 사례의 비율로 정의하였다.

이차 평가변수는 Top-2 및 Top-3 누적 포함률이었다. Top-2 누적 포함률은 기준 처방이 준차트의 상위 2개 추천 처방 안에 포함된 사례의 비율로 정의하였고, Top-3 누적 포함률은 기준 처방이 상위 3개 추천 처방 안에 포함된 사례의 비율로 정의하였다. 추가로 각 표준 처방별 사례 수, Top-1 일치 사례 수 및 기준 처방이 실제로 제시된 순위를 기술적으로 분석하였다.

4) 통계 분석

Top-1 일치도 및 Top-2·Top-3 누적 포함률은 전체 사례 수를 분모로 하고 각 평가기준을 충족한 사례 수를 분자로 하여 백분율로 산출하였다. 각 비율의 95% 신뢰구간은 이항분포에 근거한 정확 신뢰구간(Clopper-Pearson method)으로 산출하였다.

처방별 성능은 사례 수와 일치 사례 수를 제시하는 기술통계로 분석하였다. Top-1 불일치 사례에 대해서는 기준 처방, 준차트 Top-1 추천 처방 및 기준 처방의 실제 추천 순위를 비교하였다. 기준 처방이 Top-2 또는 Top-3에 포함된 경우에는 이를 별도로 기술하여, 준차트가 처방군 자체를 놓친 경우와 유사 처방 간의 세부 순위가 달라진 경우를 구분하였다.

4. 독립 한의사에 의한 시나리오·기준처방 일치도 평가

1) 평가용 시나리오 구성

준차트의 지식베이스 구축 및 추천 결과 산출에 참여하지 않은 독립 한의사에 의한 시나리오 및 기준 처방의 일치도 평가는 표준화 검증 시나리오에 사용하였던 39개 표준화 검증 시나리오와는 별도로 동일한 처방기준을 바탕으로 전형(典型) 사례만을 대상으로 새롭게 구성한 시나리오 세트를 사용하여 수행하였다. 변형 사례는 전형 사례에 비해 처방기준의 필수 임상항목 수가 적고 판정 기준이 상대적으로 관대하여 제시된 문진·변증 및 신체소견만으로 처방을 감별해야 하는 독립 평가자에게는 판단 근거가 충분하지 않다고 보아 평가 대상에서 제외하였다. 이에 따라 독립 한의사 평가에는 황련제 계열 16개 표준 처방 모두의 전형 시나리오가 포함되었으며 이 중 2개 표준 처방(소함흉탕, 황련아교탕)은 전형 양상이 2가지 하위 유형으로 구분되어 각각 2개 시나리오씩 포함됨에 따라 최종적으로 총 18개 시나리오로 구성하였다.

2) 평가 방법

준차트의 지식베이스 구축 및 추천 결과 산출에 참여하지 않은 한의사 3인이 독립 한의사 평가에 참여하였다. 독립 한의사 평가에 참여한 3인은 남 2명, 여 1명으로 연령의 중앙값은 43세(범위 35-49세), 임상경력 중앙값 4년(범위 0.4-10년), 정인적방 임상 경력 중앙값 4년(범위 0.4-6년)이었다.

평가자는 준차트의 추천 결과를 알 수 없는 상태에서 18개 시나리오 각각의 문진(수면, 심계/흉만, 대변 양상, 소화/허기, 한열양상, 오심/속쓰림, 감정 등), 변증(음양 변증) 및 신체소견(면색, 안진(按診) 정보만을 검토하여 16개 후보 표준 처방 목록 중 가장 적합하다고 판단되는 처방 1개를 독립적으로 선택하였다.

독립 한의사 평가는 전체 처방 지식베이스를 검색하는 준차트 평가와 달리, 황련제 계열 16개 표준 처방 목록 내에서 선택하도록 하였다. 또한 두 평가에는 서로 다른 시나리오 세트를 사용하였으므로, 독립 한의사 평가는 준차트와 임상가의 처방 추천 성능을 직접 비교하기 위한 것이 아니라, 제시된 시나리오에서 기준 처방에 대한 임상가의 일치도와 평가자 간 일치도를 확인하기 위한 평가로 수행하였다.

3) 평가 변수 및 통계 분석

각 평가자가 선택한 처방과 해당 시나리오에 사전 지정된 기준 처방의 일치 여부를 평가자별로 산출하여 개별 일치도로 정의하였다. 또한 3인 중 2인 이상이 기준 처방과 일치하는 처방을 선택한 시나리오의 비율(2인 이상 일치율)과 3인 모두가 기준 처방과 일치하는 처방을 선택한 시나리오의 비율(3인 일치율)을 산출하였다. 개별 일치도, 2인 이상 일치율 및 3인 일치율은 전체 시나리오 수(18개)를 분모로 하고 각 기준을 충족한 시나리오 수를 분자로 하여 백분율로 산출하였다.

3인 평가자 간 처방 선택의 상호 일치도(inter-rater reliability)는 다중 평가자·다중 범주 자료에 적용 가능한 Fleiss' κ로 산출하였다.

Ⅲ. 연구결과

1. 준차트 표준화 검증 결과

1) 전체 성능

총 39개의 표준화 검증 시나리오에 준차트를 적용한 결과, 사전 지정된 기준 처방이 준차트의 Top-1 추천 처방과 일치한 사례는 37례였다. 전체 Top-1 일치도는 94.9%(37/39)이었으며, 95% 신뢰구간은 82.7-99.4%였다.

기준 처방이 준차트의 상위 2개 추천 처방 안에 포함된 사례는 39례로, Top-2 누적 포함률은 100.0%(39/39)였다. 기준 처방의 Top-3 누적 포함률 역시 100.0%(39/39)로 나타났다. Top-2 및 Top-3 누적 포함률의 95% 신뢰구간은 각각 91.0-100.0%였다(Table 2).

황련제 계열 16개 표준 처방 중 15개 처방은 해당 처방에 배정된 모든 시나리오에서 Top-1 일치를 보였다. 백두옹가감초아교탕은 2개 시나리오에서 모두 Top-2로 제시되었다.

Table 2. Overall performance of Joonchart for Hwangryeonje formula recommendation

평가지표	일치 사례/전체 사례	일치도(%)	95% 신뢰구간
Top-1 일치도	37/39	94.9	82.7-99.4
Top-2 누적 포함률	39/39	100.0	91.0-100.0
Top-3 누적 포함률	39/39	100.0	91.0-100.0

주. 95% 신뢰구간은 Clopper-Pearson 정확 이항 방법으로 산출하였다.

2) 처방별 성능 Top-1 불일치 사례

처방별 성능을 분석한 결과, 황련탕-생강, 황련탕-건강, 반하사심탕, 건강황금황련인삼탕, 생강사심탕, 감사거삼탕, 부자사심탕, 삼황사심탕, 대황황련사심탕, 소함흉탕, 황련아교탕, 갈근황금황련탕, 백두옹탕, 오매환 및 오매환거부자 계열의 각 표준 처방은 배정된 모든 검증 시나리오에서 Top-1으로 일치하였다(Table 3).

Table 3. Knowledge base and formula-specific Top-1 performance of Joonchart for 16 standard Hwangryeonje formulas

처방군	기준 처방명	검증 시나리오 수	Top-1 일치, n/N
황련탕·사심탕 계열	황련탕-생강	2	2/2
황련탕-건강	2	2/2	
반하사심탕	3	3/3	
건강황금황련인삼탕	2	2/2	
생강사심탕	3	3/3	
감사거삼탕	3	3/3	
부자사심탕	2	2/2	
삼황사심탕	2	2/2	
대황황련사심탕	2	2/2	
소함흉탕	4	4/4	

황련아교탕	4	4/4	
설사·이질 계열	갈근황금황련탕	2	2/2
백두옹탕	2	2/2	
백두옹가감초아교탕	2	0/2*	
오매환 계열	오매환	2	2/2
오매환거부자	2	2/2	
전체	16개 표준 처방	39	37/39

주) 백두옹가감초아교탕의 2개 시나리오에서 기준 처방은 모두 Top-2로 제시되었다.

3) Top-1 불일치 사례

Top-1 불일치는 백두옹가감초아교탕을 기준 처방으로 한 2개 사례(A_034 및 A_035)에서 발생하였다. A_034의 기준 처방은 백두옹가감초아교탕 전형이었으나, 준차트는 백두옹탕 전형을 Top-1으로, 백두옹가감초아교탕 전형을 Top-2로 제시하였다. A_035에서도 기준 처방인 백두옹가감초아교탕 변형은 Top-2로 제시되었고, 백두옹탕 변형이 Top-1으로 제시되었다(Table 4).

따라서 두 불일치 사례는 백두옹탕 계열이 추천 후보에서 제외된 경우가 아니라, 백두옹탕과 백두옹가감초아교탕 사이의 세부 처방 순위가 달라진 경우였다. 기준 처방은 두 사례 모두 상위 2개 추천 처방에 포함되었으며, 기준 처방이 Top-3 밖으로 제시된 사례는 없었다.

Table 4. Top-1 discordant cases in the scenario-based performance evaluation

사례 ID	기준 처방	준차트 Top-1 처방	기준 처방의 추천 순위
A_034	백두옹가감초아교탕(전형)	백두옹탕(전형)	2
A_035	백두옹가감초아교탕(변형)	백두옹탕(변형)	2

2. 독립 한의사 평가 결과

1) 기준 처방과의 일치율

황련제 계열 16개 표준 처방의 전형 시나리오 18개에 대하여 평가자별로 선택한 처방과 사전 지정된 기준 처방의 일치도는 88.9~94.4% 범위였으며, 3인 중 2인 이상이 기준 처방과 일치하는 처방을 선택한 시나리오는 17개(94.4%), 3인 모두 일치하는 처방을 선택한 시나리오는 16개(88.9%)였다(Table 5).

Table 5. Agreement with the pre-specified reference formula

평가 항목	N/18(%)
3인 모두 기준 처방과 일치	16(88.9)
3인 중 2인 이상 기준 처방과 일치	17(94.4)
평가자 1	17(94.4)

평가자 2	16(88.9)
평가자 3	17(94.4)

2) 평가자간 신뢰도

3인 평가자 간 처방 선택의 상호 일치도는 Fleiss' κ = 0.96으로 나타났다. Landis와 Koch(1977)[7]기준에 따르면 이는 '거의 완벽한 일치(almost perfect agreement, κ = 0.81-1.00)' 수준에 해당한다.

3) 일치 실패 사례

3인 모두가 기준 처방과 일치하지 못한 시나리오는 1개(증례 11)였다. 해당 시나리오의 기준 처방은 '백두옹가감초아교탕'이었으나, 3인 평가자 모두 '백두옹탕'을 선택하였다. 이는 평가자 간에 의견이 갈린 경우가 아니라 3인 모두 동일한 처방을 선택하였음에도 그 선택이 기준 처방과 달랐던 경우이다. 세 평가자가 모두 기준 처방과 다른 동일 처방을 선택한 사례는 기준 처방과의 일치율(88.9%)을 낮추지만, 평가자 간 일치에는 기여한다. 따라서 기준 처방과의 일치도와 평가자 간 일치도(Fleiss' κ = 0.96)는 구분하여 해석해야 한다.

Ⅳ. 고 찰

본 연구는 황련제 계열 처방 추천에 대하여 준차트 자체의 추천 성능과 그 지식베이스 구축에 관여하지 않은 독립 한의사와의 처방 선택 일치도를 함께 평가하였다.

황련제 계열 16개 표준 처방을 대상으로 전문가 기준에 따라 구성된 39개 전형 및 변형 시나리오에서 준차트의 Top-1 일치도는 94.9%(37/39)였으며, 기준 처방은 모든 사례에서 Top-2 이내에 포함되었다. 이는 황련제 계열에서 사전에 구조화된 처방 감별 기준이 준차트의 후보 처방 산출 및 우선순위 결정 과정에 대체로 일관되게 반영되었음을 의미한다.

특히 Top-3 누적 포함률이 100%였다는 점은 준차트가 황련제 계열의 기준 처방을 후보군에서 누락하지 않았음을 보여준다. 한의 처방 선택은 임상의의 최종 판단을 전제로 하며, 처방추천시스템의 역할은 단일 처방을 자동으로 확정하는 데 있기보다 임상적으로 타당한 후보 처방군을 제시하고 감별 판단을 지원하는 데 있다. 이러한 관점에서 본 결과는 준차트가 황련제 계열에 대해 사전에 구조화된 처방기준이 준차트의 추천 로직에 대체로 일관되게 구현되었음을 보여주며 독립 시나리오 및 실제 임상자료를 이용한 후속 검증의 근거를 제공한다.

본 연구에서 발생한 두 건의 Top-1 불일치는 모두 백두옹가감초아교탕 시나리오에서 백두옹탕이 1순위로, 기준 처방인 백두옹가감초아교탕이 2순위로 제시된 경우였다. 이는 준차트가 해당 사례를 백두옹탕 계열로 인식하였으나, 백두옹탕과 백두옹가감초아교탕 사이의 세부 처방 분기에서 기준 처방과 다른 순위를 부여했음을 보여준다. 따라서 향후에는 두 처방을 구분하는 임상 항목과 가중치를 검토하고, 특히 백두옹가감초아교탕의 처방기준을 대표하는 증상 및 신체소견이 충분히 반영되었는지를 점검할 필요가 있다.

다만 본 연구의 준차트 성능평가 결과는 실제 환자 진료기록에 대한 외부 검증 결과가 아니

라, 전문가 기준에 따라 구성된 표준화 시나리오에서의 내부 성능을 나타낸다. 또한 검증 시나리오의 기준 처방과 준차트의 지식베이스가 동일한 전문가의 임상 지식에 기반하였으므로, 본 결과만으로 실제 임상환경에서의 일반화 가능성이나 치료효과를 판단할 수는 없다. 향후에는 실제 임상자료를 이용한 후향적 성능평가가 필요하다.

이러한 한계에도 불구하고, 본 연구는 한의학적 변증에 기반한 처방 선택 과정에서 유사 처방 간에 공통으로 나타나는 증상뿐 아니라 이를 구분하는 감별 기준을 구조화하고, 시나리오에 제시된 문진 및 신체소견을 바탕으로 기준 처방을 일관되게 추천할 수 있는 가능성을 보여주었다는 점에서 한의학적 의의가 있다. 이는 임상가의 처방 판단에 활용되는 지식을 명시적인 기준으로 정리하여 처방 선택을 지원하는 도구로 구현할 수 있음을 시사한다. 현재 준차트는 『상한론』과 『금궤요략』의 처방을 중심으로 구현되어 있으나, 향후 처방별 감별 기준의 정교화와 지식베이스 확충 및 검증을 통해 고방을 넘어 더 폭넓은 후세방으로 적용 범위를 확대할 수 있을 것으로 기대된다.

한편, 16개 표준 처방의 전형 시나리오 18개에 대하여 한의사 3인의 처방 선택은 사전 지정된 기준 처방과 88.9~94.4%의 개별 일치도를 보였으며, 3인 평가자 간 상호 일치도는 Fleiss' κ = 0.96으로 나타났다. 이는 황련제 계열에 대해 구조화된 처방 감별 기준이 준차트라는 알고리즘뿐 아니라, 그 지식베이스 구축에 관여하지 않은 임상가에게도 유사한 방식으로 해석·재현될 수 있을 만큼 명료하게 기술되어 있음을 시사한다.

3인 모두가 기준 처방과 일치하지 못한 유일한 사례(증례 11)는 백두옹가감초아교탕을 기준 처방으로 하였으나 3인 모두 백두옹탕을 선택한 경우였다. 이는 준차트 자체 평가에서 관찰된 두 건의 Top-1 불일치(A_034, A_035)와 동일하게 백두옹탕-백두옹가감초아교탕 사이에서 발생하였다는 점에서 주목할 만하다. 즉 이 불일치는 준차트 알고리즘만의 한계라기보다, 두 처방을 구별하는 임상 항목 자체가 지식베이스 내에서 아직 충분히 변별력 있게 구조화되지 못했음을 인간 평가자와 알고리즘이 공통적으로 드러낸 결과로 해석할 수 있다. 이는 향후 지식베이스 개정의 우선순위를 뒷받침하는 근거가 된다.

이러한 형태의 알고리즘 기반 한의진단 지원도구 평가는 국내에서도 드물지 않다. 사상체질 진단 지식기반 도구인 SCAT(Sasang Constitutional Analysis Tool) β-version은 전문가와의 일치도 69.3%, Kappa 0.510을 보고하였다[8]. 이와 비교하면 본 연구에서 준차트의 처방 감별에 대해 나타난 한의사 평가자 간 일치도(κ = 0.96, 거의 완벽한 일치)는 상대적으로 높은 수준이다. 이러한 차이는 두 과제의 성격 차이에서 비롯되었을 가능성이 있는데, 즉, 사상체질 진단은 체형·안면·음성 등 다차원적이고 연속적인 정보를 종합해 네 가지 범주 중 하나로 판정하는 과제인 반면, 본 연구의 경우 황련제 계열이라는 공통 기반 위에서 소화·한열·수면 등 상대적으로 명시적인 처방기준에 따라 처방군 내부를 감별하는 상대적으로 범위가 더 좁고 구조화된 도구이다. 이는 지식기반 CDSS의 성능이 대상 처방군의 범위와 처방기준의 구조화 정도에 따라 달라질 수 있음을 의미한다.

한편, 규칙기반 지식체계가 아닌 생성형 언어모델(LLM) 접근을 한의진단 영역에 적용한 남 등(2025)[9]의 연구에서는 한의진단학 지식 문항에 대해 RAG 기반 소규모 언어모델이 정답률을 48.57%에서 70.00%로 개선하였음을 보고하였다. 이 연구는 고정된 정답이 있는 지식 문항에 대한 정답률을 다룬 것으로 시나리오 기반 처방 추천과 평가자 간 일치도를 다룬 본 연구와 과제의 성격 및 평가지표가 근본적으로 달라 수치를 직접 비교하기는 어렵다. 그럼에도

불구하고 본 연구에서 준차트가 보인 높은 일치도는 범위가 좁고 명확하게 구조화된 처방군에 규칙기반 접근을 적용했을 때의 강점을 보여주는 사례로 해석할 수 있다.

Ⅴ. 결 론

본 연구는 황련제 계열 처방 추천에 대하여 준차트 자체의 추천 성능과 독립 한의사와의 처방 선택 일치도를 함께 평가하였으며, 두 결과 모두 사전에 구조화된 처방 감별 기준이 대체로 일관되게 적용되고 있음을 보여주었다. 황련제 계열 16개 표준처방을 대상으로 39개 변형 시나리오에서 준차트의 Top-1 일치도는 94.9%(37/39)였으며, 16개 처방의 전형 시나리오 18개에 대한 독립 한의사 3인의 개별 일치도는 88.9~94.4%, 평가자 간 신뢰도는 Fleiss' κ = 0.96이었다. 유일한 불일치 유형(백두옹탕-백두옹가감초아교탕)은 준차트와 독립 한의사 평가 모두에서 동일하게 나타나, 두 평가에서 백두옹탕과 백두옹가감초아교탕 간 불일치가 공통적으로 관찰되어, 두 처방의 감별 기준과 시나리오에 제시된 정보 및 추천 로직을 함께 재검토할 필요가 있음을 시사한다. 본 연구는 전문가 지식에 기반한 표준화 시나리오를 이용한 초기 내부 성능평가라는 의의를 가지며, 실제 임상환경에서의 처방 적합성이나 치료효과를 직접 검증한 것은 아니다. 본 연구는 표준화 시나리오를 이용한 초기 내부 성능평가로 향후 다양한 시나리오와 실제 임상자료를 활용한 외부 검증이 필요하다.

감사의 글

본 연구는 보건복지부의 재원으로 한국한의약진흥원 한의약산업 전주기 지원체계 구축사업의 지원을 받아 수행한 사업임(3234-302)

참고문헌

1. Sung HK, Jung B, Kim KH, Sung SH, Sung AD, Park JK. Trends and future direction of the clinical decision support system in traditional Korean medicine. J Pharmacopuncture. 2019;22(4):260-268. doi:10.3831/KPI.2019.22.035.
2. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;3:17. doi:10.1038/s41746-020-0221-y.
3. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330(7494):765. doi:10.1136/bmj.38398.500764.8F.
4. Kim SK, Jang HC, Kim JH, Oh YT, Kim C, Yea SJ, et al. Traditional Korean medicine diagnosis system based on basic ontology. Korean J Oriental Physiol

Pathol. 2010;24(6):1111-1116.

5. Choi SH. A digital application for implementing the ICD-11 traditional medicine chapter. J Integr Med. 2020;18(6):455-458. doi:10.1016/j.joim.2020.08.003.
6. Shortliffe EH, Buchanan BG. A model of inexact reasoning in medicine. Math Biosci. 1975;23(3-4):351-379. doi:10.1016/0025-5564(75)90047-4.
7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-174. doi:10.2307/2529310.
8. Jang ES, Jin HJ, Do JH, Lee SW, Kim JY. The preliminary study on the coincidence between Sasang Constitutional Analysis Tool β-version and expert of Sasang constitution. J Sasang Constitut Med. 2012;24(2):1-7.
9. Nam D, Jang E, Oh Y. Development and performance evaluation of a small language model for Korean medicine diagnostics based on Command R-plus and retrieval-augmented generation. J Soc Korean Med Diagn. 2025;29(1):59-73.