대한한의진단학회지 제30권 제1호 (2026년 10월)
The Journal of the Society of Korean Medicine Diagnostics. 2026;30(1):17-32.
[원저 Original Article]

한의학 설진기 기술 발전 및 임상 타당성 체계적 문헌고찰

남동현*
상지대학교 한의과대학 진단생기능의학교실

Development and Clinical Validation of Computerized Tongue Diagnosis Technology: A Systematic Review

Donghyun Nam*
Department of Biofunctional Medicine and Diagnostics, College of Korean Medicine, Sangji University

Abstract

Background : Tongue diagnosis has been a fundamental diagnostic method in Korean Medicine and Traditional Chinese Medicine for over 2,000 years. Subjective interpretation and lack of standardization, however, limit reproducibility. This systematic review analyzes technological development of computerized tongue diagnosis equipment from 1988 to 2026 and evaluates clinical validity evidence.

Methods : Following PRISMA guidelines, we searched PubMed, CNKI, CiNii, and DBpia without date restrictions through August 2026. Of the 337 records identified, 230 were excluded after title/abstract screening; 107 full-text articles were assessed for eligibility, and 48 further excluded (45 reviews or surveys; 3 out-of-scope empirical studies). Fifty-nine empirical studies were included in the final synthesis.

Results : The 59 studies distributed as follows: AI/ML model development 28 (47.5%), image acquisition processing 17 (28.8%), system integration 6, clinical validation 3, health monitoring 2, image segmentation 2, hardware device development 1. Phase distribution: through 2009, 12 studies; 2010-2019, 15; from 2020 onward, 32 (54.2%); single-year peak, 19 in 2025. Key innovations include ultraviolet-wavelength imaging and multi-region color mapping associated with TCM organ theory, with large-scale validation demonstrating statistically significant associations between syndrome symptoms and tongue characteristics. Only 33 of 59 studies (55.9%) achieved quality scores of at least 7/10.

Conclusions : Computerized tongue diagnosis technology has matured in image acquisition and segmentation algorithms, but clinical validation evidence remains extremely limited, with only three observational studies identified; the dominance of post-2020 publications reflects a transition to AI-integration research. Future priorities include multi-center cohort studies, standardized protocols, public datasets, and mobile-integrated real-time systems.

Keywords : tongue diagnosis, artificial intelligence, systematic review, tongue image, development

---

* 교신저자 : DongHyun Nam. Department of Biofunctional Medicine and Diagnosis, College of Korean Medicine, Sangji University. 83, Sangjidae-gil, Wonju-si, Gangwon-do, Republic of Korea.
Tel: +82-10-2500-3146, E-mail: omdnam@naver.com
· 원고접수일 : 2026.08.24 / 심사완료일 : 2026.09.05 / 게재결정일 : 2026.09.11

Ⅰ. 서 론

설진(舌診)은 오랜 기간 한의학의 핵심 진단법으로 자리잡아 왔다. 임상의들은 혀의 색상, 모양, 설태, 습기 및 국소 특징을 종합적으로 평가하여 병리적 패턴을 감별한다. 한의학 이론에 따르면 설질 색은 혈(血) 및 장기 기능 상태를 반영하고[1], 설태 특성은 기(氣)와 소화기능의 상태 및 병인(病因)을 나타내며[2], 혀의 국소 영역은 특정 장기와 밀접한 연관이 있는 것으로 전해진다.

그러나 전통적 설진은 평가자 간 변이, 표준화 부족, 주관에 의존한 진단 등의 해결할 과제를 안고 있다. 서로 다른 의사들이 동일한 혀를 다르게 해석할 수 있으며, 조명 조건, 카메라 사양 또는 정량화 방법에 대한 표준화된 가이드라인이 충분히 확립되지 않았다. "담백설(淡白舌)", "홍설(紅舌)"과 같은 질적 설명에 명확한 수치 기준이 부재하여 재현성을 확보하기 어려웠다[3].

컴퓨터화된 설진 진단 장비 개발이 가능해졌다[4]. 국제적으로는 ISO/TC 249 (전통의약 기술위원회)를 중심으로 한의학 진단기기 표준화 노력이 진행 중이며, 국내에서도 KS 규격 KIS 90057 (설진영상 촬영 방법)과 같이 설진기 표준화를 위한 노력이 이루어지고 있으나, 여전히 임상 현장에서의 광범위한 적용을 위해서는 보다 체계적인 검증이 필요하다.

AI/ML 기반 분석 알고리즘은 딥러닝[5], U-Net 아키텍처[6] 등의 발전으로 정밀도가 크게 향상되었고, CIE-Lab 컬러 스페이스를 활용한 색상 보정 및 분석 방법론도 표준화되어 가고 있다[7,8]. 또한 U-Net 기반 자동 분할 기술[9]은 혀 영역의 정밀한 추출을 가능하게 하여 재현성을 확보하는 데 기여하였다.

설진기 분야는 1980년대 후반부터 현재까지 꾸준한 연구가 수행되어 왔으며, 다양한 기술적 접근법이 제시되었다. 컴퓨터화 설진 및 혀 영상 분석을 다룬 개관·스코핑 리뷰도 이미 다수 발표되었다[10]. 그러나 이러한 선행 리뷰들은 주로 기술 내용의 목록형 개관에 머물렀고[11], 시기별 진화 구조와 임상 타당성 근거의 수준까지 함께 정량적으로 종합한 체계적 문헌고찰은 아직 부족하여, 향후 연구 방향 설정을 위한 체계적 종합이 필요하다.

본 체계적 문헌고찰은 2026년까지 발표된 설진 진단 장비 관련 실증 연구를 포괄적으로 분석하여 기술 발전의 시간적 흐름과 주요 혁신을 매핑하고, 이미지 획득부터 분류에 이르는 기술 구성요소별 발전 과정을 파악하며, 나아가서는 향후 연구 방향을 제시하는 것을 목적으로 한다.

Ⅱ. 연구 방법

1. 가이드라인

본 체계적 문헌고찰은 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 가이드라인을 준수하였다.

1) 정보 출처 및 검색 전략

2026년 8월 10일부터 13일까지, 출판일에 대한 기간 제한 없이 4개 데이터베이스를 검색하였다: PubMed(글로벌)에서 "tongue diagnosis" OR "tongue image*" AND "development"

OR "system" OR "device"로 224편, CNKI(중국)에서 TI='舌诊' AND KY='客观化'로 72편, CiNii(일본)에서 TI='舌診' AND (KY='システム' OR '装置')로 20편, DBpia(한국)에서 TI='설진기' OR TI='설진장치'로 21편을 검색하였다.

2) 선택/배제 기준

두 명의 한의진단학 분야 전문가가 아래의 기준에 따라 독립적으로 검색된 문헌에 대한 선별을 수행하였다.

선택 기준:

(1) 대상: 인간 대상 (건강인 또는 환자)

(2) 중재: 설진 장비 개발, 시스템 설계, 알고리즘 구현 또는 임상 검증을 기술한 연구

(3) 연구 유형: 실증 연구 논문, 학술대회 발표문

배제 기준:

(1) 기존 문헌을 요약하는 종설/서베이 논문 (Review Summary, Technology Review Survey 명시적 제외)

(2) 동물 연구

(3) 장비/알고리즘 개발 없이 혀 미생물군집만 분석한 연구

(4) 초록만 존재하는 논문

(5) CiteSpace/VOSviewer를 활용한 문헌계량분석 (임상 또는 기술 검증 부재)

3) 데이터 추출 및 질 평가

각 포함된 연구에서 다음을 추출하였다: 기본 정보 (저자, 연도, 저널/학술대회, DOI/PMID), 장비 사양 (이미징 모드, 카메라 유형, 조명 시스템, 분광 기능), 알고리즘 구성 요소 (전처리 방법, 분할 알고리즘, 특징 추출 기법, 분류 모델), 임상 초점 (타겟 한의학 증후군 또는 서양 의학적 질환), 성능 지표 (정확도, 민감도, 특이도, AUC, 통계적 유의성), 표본 특성 (표본 수, 데이터 출처, 선택/배제 기준).

질 평가는 문헌 선별을 수행한 두 명의 한의진단학 분야 전문가가 독립적으로 0-10점 척도로 질 점수를 부여한 후, 그 평균값을 반올림하여 최종 확정하였다. 다만 블라인드 처리는 하지 않았고, 독립 평가 → 협의 → 재평가 → 재협의 과정을 여러 차례 반복하는 방식으로 진행되었기 때문에 평가자 간 일치율은 별도로 산정하지 않았다. 평가지표는 다음과 같았다: 표본 수의 크기, 방법론적 엄격성 (블라인딩, 대조군, 외부 검증), 통계 분석 완결성 (효과 크기, 신뢰구간), 투명성 (코드/데이터 공개, 상세 방법).

4) 분석 방법

연구를 다음과 같이 분류하였다: 1) 시간적 기간: Phase 1 (2009년 및 그 이전), Phase 2 (2010-2019년), Phase 3 (2020 년 및 그 이후) — 경계 연도(2009, 2020)의 귀속을 위와 같

이 확정하였다; 2) 기술 구성 요소(최종 7개 카테고리): AI 진단 모델, 이미지 획득·처리, 시스템 통합, 임상 검증, 건강 모니터링, 이미지 분할, 하드웨어 장치 개발; 3) 임상 초점: 특정 변증분형 또는 질환명; 4) 방법론적 접근: 전통 머신러닝, 딥러닝, 다중모달 융합.

메타분석은 동질적인 결과 측정치를 가진 3편 이상의 연구가 어떤 임상 카테고리에서 충분히 확보되면 수행하고자 하였다. 그러나 실제로 식별된 임상 검증 연구는 3편에 불과하였고, 각 연구의 방법론적 이질성이 서로 너무 커서 자료 합성은 불가능하다고 판단되어 메타분석은 별도로 수행하지 않았다.

Ⅲ. 연구결과

1. 문헌 선정

본 체계적 문헌고찰은 PRISMA 가이드라인을 준수하여 문헌 선정 과정을 수행하였으며, PRISMA 흐름도는 Figure 1에 제시하였다. 4개 데이터베이스에서 총 337편의 기록이 식별되었으며, 이중 PubMed 에서 224편으로 가장 많았고, CNKI 72편, CiNii 20편, DBpia 21편이었다. 언어별·권역별로 독립적인 검색을 수행하였으므로 데이터베이스 간 교차 중복은 없었다(중복 제거 0건). 제목·초록 선별 단계에서 주제 불일치 175건, 리뷰 논문 32건, 동물 연구 6건, 기타 17건(원문 미확보 3건 포함) 등 총 230편을 배제하여 107편을 전문 검토 대상으로 선정하였다. 전문 검토에서 종설·서베이 논문 45건과 설진기/혀 영상 분석 범위를 벗어난 실증 연구 3건(얼굴 이미지 기반 폐암 예측, 지식그래프·LLM 기반 처방 생성, 침 로봇)을 배제하여, 최종적으로 59편의 실증 연구를 종합 분석에 포함하였다.

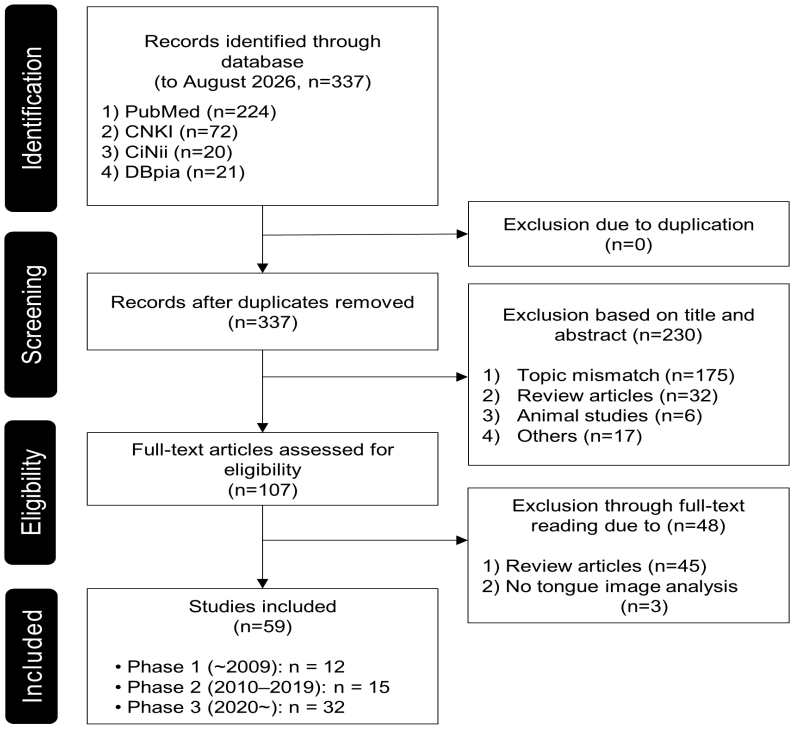


Figure 1. PRISMA flow diagram of study selection.

2. 정량적 분석: 시간적 분포

본 분야는 세 단계의 성장기를 경험하였다. Phase 1에는 12편 (20.3%)[1-4,7,12-18]의 연구가 발표되어 기초 디지털화와 임계값 최적화, Snake 알고리즘, 퍼지 논리 시스템이 주를 이루었다. Phase 2에는 15편 (25.4%)[8,19-32]의 연구로 전통적 방법의 정교화가 진행되었으며 SVM, Random Forest, CIE-Lab 색상 분석, GMM 군집화, UV 파장 이미징 등의 기계학습 기법이 도입되었다. Phase 3에는 CNN 아키텍처 (YOLOv8, U-Net), Transformer 모델, 다중모달 융합 접근법, 대규모 임상 검증 등 최신 기술이 적용된 32편 (54.2%)[5-6,9,33-61]의 연구가 발표되어 세 단계 중 가장 많은 비중을 차지하였다. 총 59편 중 2025년에만 19편이 발표되어 단일 연도 최고치를 기록하였으며, 이는 AI 접목 단계의 본격화를 시사한다 (Figure 2).

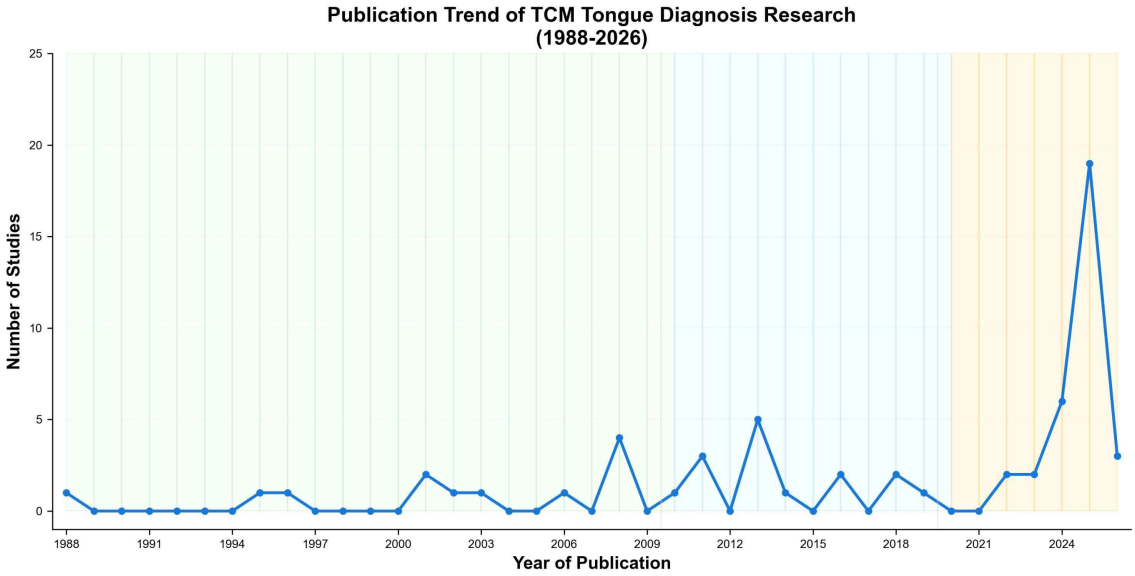


Figure 2. Temporal distribution of published studies. Phase 1 (through 2009): 12 studies (20.3%) - basic digitization and thresholding methods, Snake algorithm, fuzzy logic systems. Phase 2 (2010-2019): 15 studies (25.4%) - SVM, Random Forest, CIE-Lab color analysis, GMM clustering, UV wavelength imaging. Phase 3 (from 2020): 32 studies (54.2%) - CNN architectures (YOLOv8, U-Net), Transformer models, multi-modal fusion approaches, large-scale clinical validation. Single-year peak occurred in 2025 with 19 studies, reflecting rapid growth in AI-integration research.

3. 연구 주제별 분포

포함된 59편의 실증 연구를 분석한 결과, AI 진단 모델 개발 연구가 가장 큰 비중을 차지하였다. 총 28편 (47.5%)이 해당 카테고리에 속하였으며, 주로 간질환, 당뇨병, 암 선별을 위한 CNN 아키텍처 개발, 혀 이미지와 임상 데이터를 결합한 다중모달 융합 프레임워크, Transformer 기반의 세밀한 특징 인식 모델 연구가 수행되었다. 이미지 획득 및 처리 관련 연구는 17편 (28.8%)으로 두 번째로 많았다. 이 분야에서는 CIE D65 표준 광원을 사용한 통

제 조명 환경 구축, 표준 색차트와 신경망 기반 색상 보정 방법, UV/IR 파장 이미징을 통한 설태 시각화 강화 등이 주요 주제였다. 시스템 통합 연구는 6편(10.2%)이었다. 모바일 애플리케이션, 웨어러블 건강 모니터링 시스템, 설진·맥진·안면진단을 통합한 다중모달 진단 플랫폼 개발이 주요 내용이었다. 임상 검증 연구 (설진 정량 지표와 증상/질환 결과 간 연관성을 관찰연구 디자인으로 검증한 연구)는 극히 제한적으로 총 3편(5.1%)만 수행되었다. 이 중 Kim 등은 담적증후군 환자 2,157명을 대상으로 한 대규모 다기관 임상 검증으로, 객관적 설진 정량 지표와 증상 간 통계적 유의성을 확인하였다[36]. 또 Kim 등은 수면장애와 설태 변화의 연관성을 분석하였고[28], Jung 등은 설태 진단 지표를 추출하였다[25]. 혀 영역 분할 알고리즘 연구는 2편(3.4%), 건강 모니터링 연구도 2편(3.4%)이었고, 하드웨어 장치 개발 연구는 1편(1.7%)이었다 (Figure 3).

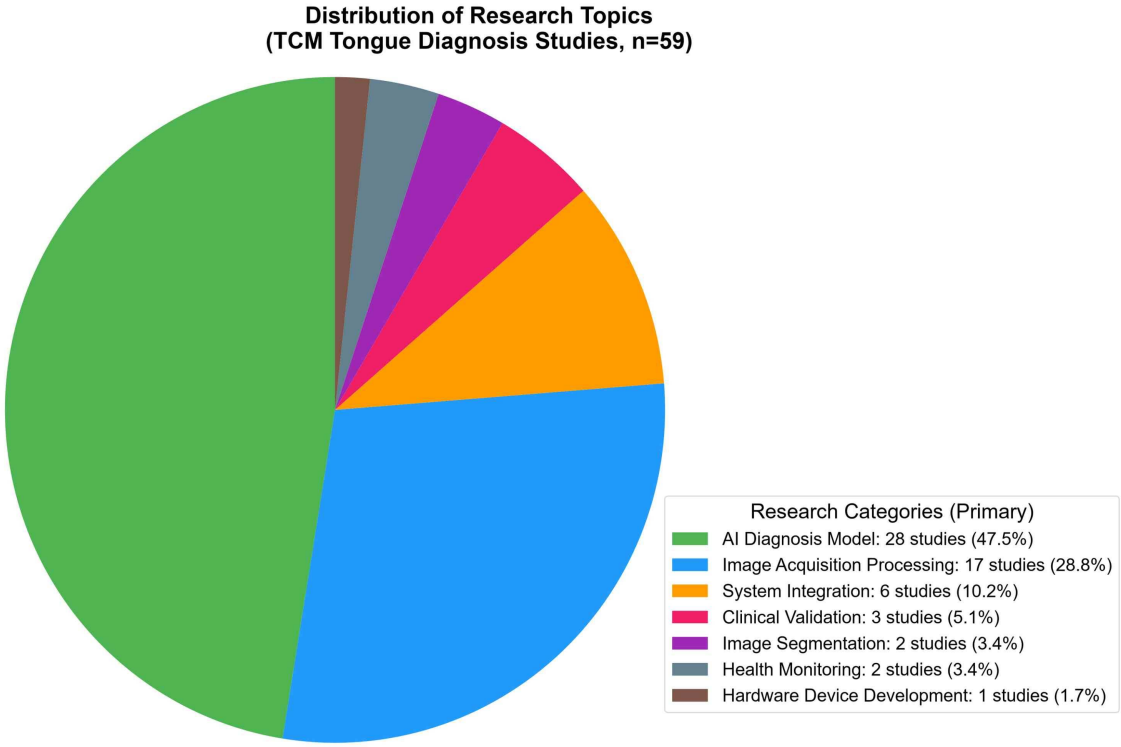


Figure 3. Distribution of research topics among included studies. AI Diagnosis Model (28, 47.5%); Image Acquisition Processing (17, 28.8%); System Integration (6, 10.2%); Clinical Validation (3, 5.1%); Image Segmentation (2, 3.4%); Health Monitoring (2, 3.4%); Hardware Device Development (1, 1.7%).

4. 알고리즘 진화: 기술 심층 분석

1) 이미지 분할 알고리즘

포함된 59편의 실증 연구 중 49편(83.1%)가 다양한 분할 알고리즘 기술을 활용하였다. Phase별 분석 결과, 분할 기술은 명확한 진화 경로를 보이며 2020년 이후 CNN 및 딥러닝 기반으로 급격히 전환되었다.

Phase 1단계에서는 형태학적 처리와 색상 기반 단순 분할이 주를 이루었다. Tian과 Wong은 이전 연구성과를 정리하여 형태학적 연산을 활용한 초기 방법을 제시하였으나 실증적 검증은 제한적이었다[14].

Phase 2단계에서는 전통적 방법의 정교화가 진행되었다. Kim 등은 Snake 알고리즘으로 초기 혀 영역을 검출한 후 가우시안 혼합 모델을 적용하여 CIE-Lab 색상 공간의 a* 채널 분포를 통계적으로 모델링하고, 임계값과 형태학적 연산으로 설태 및 설질 분류를 수행하는 종합적 프레임워크를 제시하였다[26]. Jang 등은 개선된 Snake 알고리즘으로 기존 방법의 경사진 배경에서의 실패 문제를 해결하고 엣지 검출을 결합하여 분할 정밀도를 향상시켰다[29]. 이 시기 Wenjuan 의 연구에서 최초로 인공신경망(ANN) 기반 접근이 도입되기 시작하였다[21].

Phase 3단계에서는 CNN 및 딥러닝으로의 급격한 전환이 이루어져 전체 Phase 3 연구 중 약 93.8%(30/32편)가 분할 기술을 활용하였으며, 그 대부분이 CNN 기반 아키텍처를 사용하였다. Kim 등의 U-Net을 활용한 혀 영역 자동추출 알고리즘 개발은 AI 진단 모델 카테고리에서 딥러닝 기반 분할 기술을 적용한 대표적 사례이다[6]. 이 시기 일부 연구에서는 색상 기반 분할과 형태학적 연산 같은 전통적 방법도 여전히 활용되었으나, CNN 및 딥러닝이 주류 기술로 자리잡았다. Kim 등은 Snake 알고리즘과 색상 분석을 결합하여 임상 검증 연구에서도 분할 기술이 필수 전처리 단계로 통합됨을 보여주었다[36].

Phase 3에서 분할 연구가 급증한 반면 Image_Segmentation 이라는 명시적 카테고리를 가진 독립적 연구는 드물게 확인되었다. 이는 혀 이미지 분할 알고리즘이 독립적 연구 주제보다는 AI 기반 진단 모델의 전처리 단계 또는 하위 구성 요소로서 통합되어 적용되고 있음을 시사한다. 특히 다양한 카테고리에서 이미지 분할 기술이 발견되는 것은 분할이 다학제적 접근의 핵심 요소로 자리 잡았음을 반영한다[47].

2) 색상 분석 모델

CIE-Lab 색상 공간이 표준으로 자리잡았다. 분석된 59편 중 17편(28.8%)에서 CIE-Lab 기반 정량화가 명시적으로 사용되었으며, 특히 설질 및 설태 정량화에 적합하다. 이 색상 공간은 균일한 색차를 제공하며 한의학 개념인 혈어와 명확한 연관성을 보였다. Choi 등은 UV, IR, 가시광선 파장을 비교한 연구로, UV 파장대에서 설태 경계가 가장 명확하게 드러난다는 혁신적 발견을 하였다[22]. Jung 등은 12개 색상 샘플을 활용한 동시 보정 후 CIE-Lab로 변환하여 L* (명도)와 a* (적록색) 채널의 2D 히스토그램 분석을 통해 설태 진단 지표를 추출하였다[25]. Kim 등은 가우시안 혼합 모델 (GMM)을 a* 채널에 적용하여 설태 분류를 수행하는 종합적 프레임워크를 제시하였다[26]. Phase 3에는 CIE-Lab 의 표준화가 더욱 강화되었다. Kim 등은 대규모 임상 검증 연구 (n=2,157)에서 RGB 를 CIE-Lab 로 변환하고 L*와 a* 값을 다중 영역 매핑하여 담적증후군 증상과 통계적 유의성을 확인하였다[36].

HSV/HSI 색상 공간은 약 6편(10.2%)의 연구에서 사용되었으며, 주로 건강 모니터링 시스템에 적용되었다. 인간 지각과 일치하며 계산 효율성이 장점이다. Watsuji 등은 컬러매칭 차트 기반 분석을 수행하여 침 치료 전후의 설색 변화를 정량화하였다[23]. RGB 색상 공간은 원본 데이터 형식으로 널리 사용되었으나 조명 조건에 의존하므로 광범위한 보정이 필요하다.

근적외선 분광법은 질환 판별과 당뇨병·간염 진단 연구에 활용되었다[19-21]. 853.59~1737.26nm 파장대역의 512개 파장점 스펙트럼 데이터를 반사율 정규화로 분석하였으며, PCA-GRNN 모델에서 100% 정확도를 달성한 사례가 보고되었다[20].

주요 색상 보정 방법으로는 12단계 회색도 및 컬러 색상표를 활용한 동시 보정[27], TRM (Topology Refinement and Mapping) 모델을 통한 표준 조건 색도값 복원[24], ICC 색상 특성 파일 기반 보정, 컬러매칭 차트 활용[23] 등이 있다.

3) AI/ML 분류 모델

분류 알고리즘의 발전은 의학 AI 의 더 넓은 동향을 반영한다. 분석된 59편 중 43편에서 분류 모델이 확인되었다. SVM(서포트 벡터 머신)은 약 5.1%의 연구에서 적용되었으며, 설태 분류 및 증상 예측에 활용되었다. Jung 등은 12개 색상 샘플 보정 후 CIE-Lab로 변환하여 L*와 a* 채널의 2D 히스토그램 분석을 통해 25차원 특징 벡터를 추출하고, 최소 제곱법 SVM으로 설태 두께를 분류하였다 (정확도 85.61%, 민감도 85.42%)[25]. 퍼지 논리는 약 4%의 연구에서 사용되었으며, 특히 한열 변증 감별에 효과적이었다. Watsuji 등은 냉/온/냉온복합 증후군을 구분하는 퍼지 규칙 기반 시스템을 개발하여 각 증후군별 점수를 산출하였다[13]. GMM(가우시안 혼합 모델)은 CIE a* 채널의 색상 분포를 통계적으로 모델링하여 설태 분류에 적용되었다. Kim 등은 GMM을 활용한 2단계 분류 프레임워크로, 인구 기반 절대 임계값과 EM 알고리즘을 결합한 하이브리드 접근법을 제시하였다[26]. 신경망 모델은 Wenjuan 등의 연구에서 근적외선 분광법과 결합하여 당뇨병 및 간염 진단에 활용되었다[21]. PCA-GRNN 모델은 512개 파장 점을 8개 주성분으로 축소한 후 100% 정확도를 달성한 사례가 보고되었다[20].

Phase 3에는 CNN 및 딥러닝 아키텍처로 급격한 전환이 이루어졌다. U-Net 기반 분할 모델이 혀 영역 자동 추출에 광범위하게 적용되었는데, Kim 등은 U-Net을 활용한 설진 영상 자동 분할 알고리즘을 제시하였다[6]. 최근에는 다중모달 융합 아키텍처가 주목받고 있다. 임상 데이터, 유전체 마커 또는 미생물군집 특징과 혀 이미지를 결합하는 접근법이 등장하였으며, 시각 및 텍스트 정보 통합을 위한 교차모달 어텐션 모델이 장기 질환 예측에 적용되었고[54], YOLO 기반 검출기를 결합한 세밀한 혀 특징 분류 연구도 수행되고 있다[42]. Liu 등은 전통 한의학 체질 인식 모델을 위해 딥러닝과 이중센터 연구를 결합하여 (전통+딥 특성 융합 MLP 모델로) 학습셋 기준 정확도 89.3%, AUC 0.948(테스트셋 정확도 83.7%, AUC 0.898)을 달성하였다[50].

알고리즘 선택은 연구 목적에 따라 달라졌다. 분할 정밀도가 필요한 경우 CNN 기반 U-Net이 선호되었고, 소규모 데이터셋에서는 SVM과 퍼지 논리가 여전히 유효한 대안으로 남았다. GMM은 색상 분포 모델링에 특화되어 있으며, 근적외선 분광법과 결합 시 PCA와의 연동으로 고차원 스펙트럼 데이터를 효과적으로 처리할 수 있었다.

4) 임상 검증 증거 및 메타분석 가능성

임상 검증 연구는 극히 제한적이었다. 본 문헌고찰은 임상 검증 연구를 '혀 정량 지표와 증상/질환 결과 간의 연관성을 관찰연구 디자인으로 검증한 연구'로 정의하였고, 그 기준을 충

족하는 연구는 단 3편 (5.1%)에 불과하였다. 나머지 실증 연구는 알고리즘 개발 또는 시스템 구축에 중점을 둔 기술 중심 연구였다.

가장 대규모로 이루어진 연구는 Kim 의 관찰연구였는데, 담적증후군 환자 2,157명을 대상으로 한 다기관 임상 검증에서 CIE L*a*b* 색상 공간을 활용한 정량적 설진 지표와 증상 간 통계적 유의성이 확인되었다 (하복부 통증군의 설질 L*=51.95 대 52.85, p=0.035; 소화불량군의 측부 영역 a*=28.24 대 28.79, p=0.045).[36]이 연구는 한의학 장기 매핑 이론을 현대 정량 방법으로 검증한 대표적 사례이나, 질 평가 점수는 중간 수준이었다. Kim 등은 노인 수면장애 및 배변이상과 설색 변화의 연관성을 분석하였으며 (n=454),[28] Jung 등은 설태 두께 분류를 위한 색상 히스토그램 지표를 추출하였다(SVM 정확도 85.61%)[25].

메타분석은 수행되지 않았다. 식별된 3편의 임상 검증 연구는 방법론적 이질성이 너무 커 통계적 통합이 불가능하였다. 첫째, 결과 측정치가 상이하였다: Kim의 연구[36]에서는 CIE-Lab 값에 대한 t-test 직접 비교로, Kim의 연구[28]에서는 설색 정량 지표와 수면장애·배변이상 증상 간 연관성 분석으로, Jung의 연구[25]에서는 분류 정확도만 보고하였다. 둘째, 표본 규모와 보고 형식도 일치하지 않았다: Kim 의 담적연구[36]은 n=2,157의 다기관 상세 데이터를 제공한 반면, 수면장애 연구[28]는 n=454의 수백 명 수준 소규모 데이터였고, 매개변수 추출 연구[25]는 소표본 연구였다. 셋째, 임상 타겟이 다양하였다: 담적증후군, 수면장애·배변이상 등 통일된 진단 질문이 아닌 서로 다른 증상이나 지표를 다루었다. 넷째, 통계적 접근법이 이질적이어서 효과 크기 통합이 불가능하였다. 따라서 본 문헌고찰은 기술적 발전 과정과 주요 혁신을 서술적으로 종합하는 데 초점을 두었으며, 향후 표준화된 프로토콜로 수행될 대규모 다기관 연구가 축적되면 메타분석 가능성을 재검토할 수 있을 것이다.

Ⅳ. 고 찰

지난 38년(1988-2026) 간 발표된 59편의 실증 연구 결과를 체계적으로 검토한 결과, 컴퓨터화 된 설진 기술은 이미지 획득과 분할 알고리즘 분야에서 상당한 성숙도에 도달했음을 확인할 수 있었다. 이미지 획득 분야에서는 CIE D65 표준 광원과 색상 보정 프로토콜을 통한 표준화가 이루어졌으며, 분할 알고리즘은 높은 정확도의 자동 딥러닝 기반 U-Net 아키텍처로 발전하였다. 색상 분석은 한의학 관련 정량화를 위해 주로 CIE-Lab 색공간이 사용되었다. 그러나 충분한 표본 수를 가진 엄격한 임상 검증 연구는 3편에 불과하였고, Kim의 담적연구[36]가 현재까지 가장 큰 표본 수를 가진 연구로 남아있다. 메타분석은 방법론적 이질성으로 인해 불가능하였다.

연구 추세를 재해석하면, 세 단계 중 Phase 3(2020년 이후)가 32편(54.2%)로 가장 많은 비중을 차지하여 AI 접목 단계가 본격화되었음을 보여준다. 2010년대까지 설진기 하드웨어와 기본 기술 개발이 정립되었고, Phase 3에는 이러한 기술 기반 위에 CNN, Transformer, 다중모달 융합 등 AI 기술이 접목되는 방향으로 연구가 전환되었다. 단일 연도 최고치는 2025년 19편으로, 이전 정점이었던 2013년(5편)을 크게 상회한다. 또한 본 연구에서는 기술적 요소를 중심으로 검토하여 엄격한 질환별 임상 검증 연구를 별도로 구분하여 다루었다.

38년간의 연구 분석에서 네 가지 주요 방법론적 혁신이 확인되었다. 첫째, UV 파장 이미징의 발견은 설태 경계를 명확하게 시각화하는 중요한 돌파구였다[22]. 가시광선이나 IR보다 UV

파장에서 설태와 설질의 대비가 훨씬 뚜렷하게 드러나며, RGB 색상 공간에서 균일한 "골짜기" 패턴이 관찰되어 객관적 분할 기준을 제공하였다. 이 연구는 다중 파장 이미징의 잠재력을 최초로 실증적으로 보여주었다. 둘째, 근적외선 분광법의 적용[19-21]은 질병 진단에 새로운 차원을 열었다. 512개 파장점 적외선 스펙트럼 데이터를 반사율 정규화로 분석한 결과, PCA-GRNN 모델을 통해 당뇨병과 간염 진단에서 100% 정확도를 달성하였다[20]. 이는 분광학적 접근이 전통적 색상 분석을 보완할 수 있음을 보여주었다. 셋째, 한의학 장기 이론과 연관된 다중 영역 매핑[36]은 고대 이론을 현대 과학으로 검증한 대표적 사례이다. 혀를 수평 및 수직으로 분할하여 국소 특징 추출이 가능하게 하였으며, 측부 영역 색 변화가 간담 관련 소화불량과 통계적으로 유의하게 연관됨을 입증하였다(p=0.045). 이는 전통 장기 매핑 이론의 과학적 타당성을 지지한다. 넷째, Transformer 기반 세밀한 특징 분류 및 다중모달 융합[54]이 최근 주류로 부상하였다. 임상 데이터, 유전체 마커, 미생물군집 특징과 혀 이미지를 결합하는 접근법이 등장하였으며, 시각 및 텍스트 정보 통합을 위한 교차모달 어텐션 모델이 개발되었다. Liu의 연구는 tongue image 특성 기반 체질 인식 모델(전통+딥 특성 융합 MLP)로 학습셋 정확도 89.3%, AUC 0.948 을 달성하였다[50]. 이러한 발전은 AI 기술의 정교화와 다학제적 접근 강화 추세를 반영한다.

현재 컴퓨터화 된 설진 기술의 임상 적용을 제한하는 방법론적, 기술적, 한의학 통합 측면의 격차가 존재한다. 방법론적 한계로는 소규모 표본이 가장 큰 문제이다. 포함된 연구 중 절반 이상 (32/59)이 100명 미만의 대상자를 포함하였으며, 정상 대조군이 부재하여 많은 연구가 환자 하위군 간 비교만 수행되었고, 블라인딩과 무작위화가 불충분하였다. 조명 조건, 카메라 사양, 색상 보정 프로토콜에 대한 표준화 합의도 부재하여 연구 간 비교 가능성이 제한되었다. 기술적 한계로 공개 데이터셋 부족은 심각하다. TonguExpert[49]를 제외하고는 대규모 주석 데이터셋이 공개되지 않아 대부분의 연구가 독점 데이터를 사용하였다. 외부 검증도 부재하여 몇몇 연구만 독립 다기관 코호트에서 모델을 테스트하였다. 실시간 처리 격차 또한 존재하는데, 대부분의 알고리즘은 모바일/엣지 배포에 부적합한 고성능 컴퓨팅을 필요로 한다. 한의학 통합 측면에서는 불명확한 증후군 매핑 문제가 가장 크다. 디지털 측정치가 전통 한의학 개념과 명확한 대응 관계가 불분명하여 임상 적용이 어려웠다. 또한 자동 시스템과 경험있는 한의사 전문가 사이의 일치도가 정량화되지 않아 전문가 수용성을 검증할 수 없었다.

본 문헌고찰 연구는 기존 스코핑 리뷰들을 확장하고 심화시킨 점에서 차별화된 의미를 지닌다. 단순한 기술 발전의 설명적 요약을 넘어 Phase 별 기술 진화 경로, 각 구성 요소별 구체적 방법론 및 성능 수치를 정량적으로 규명하였으며[10,11], 임상 적용 가능성을 가진 연구와 기술 타당성 검증 단계에 머문 연구를 구분하였다.

컴퓨터화 된 설진 장비는 특히 소화기 장애에서 보조 진단 도구로서의 가능성을 보여주었으나, 광범위한 임상 채택 전에는 표준화된 이미징 프로토콜이 필수적이다. 모바일/웨어러블 통합은 가정 기반 건강 모니터링을 가능하게 하지만 엄격한 검증이 선행되어야 한다. 단기적으로는 표준화된 프로토콜로 다기관 코호트 연구가 수행되어야 하며, 임상 메타데이터가 포함된 대규모 혀 이미지 공개 데이터셋이 구축되어야 한다. 또한 이미징 조건, 색상 보정 방법, 특징 추출 기준에 대한 국제적 합의 노력이 필요하다. 중기적으로는 모바일/웨어러블 플랫폼용 엣지컴퓨팅 호환 알고리즘을 개발하여 실시간 진단 시스템을 구현해야 하고, 질병 발병 또는 진행과 연관되는 혀 특징을 종단적으로 연구하는 예측 모델링이 수행되어야 한다. 또한 설진 기반 개입이 환자 결과를 개선하는지 입증하는 무작위 대조 시험이 필요하다. 장기적으로는 유

전체학, 대사체와 결합한 설진 특징을 활용한 개인 맞춤 진단이 구현되어야 하며, WHO ICD-11 전통의약 코드 연계 및 국제 합의 가이드라인을 통한 글로벌 표준화가 이루어져야 한다. 궁극적으로는 건강 위험 조기 경보 시스템으로서 예방의학 적용이 가능해야 한다.

본 체계적 문헌고찰은 몇 가지 중요한 강점을 지닌다. 아시아 언어 문헌을 포함한 포괄적인 데이터베이스 검색을 수행하였으며, 전체 원문을 추출하고 상세 기술 분석을 통해 59편의 실증 연구를 종합적으로 검토하였다. 투명하고 일관된 기준을 사용하여 임상 검증 증거의 질을 평가하였으며, 구체적인 연구 격차를 식별하고 실행 가능한 권고를 제시하여 향후 연구 방향 설정에 기여하고자 했다.

그러나 본 문헌고찰에는 명확한 한계도 존재한다. 제외된 리뷰 논문들은 실증 연구에 포착되지 않은 귀중한 역사적 맥락을 포함할 수 있으나 분석에서 배제되었다. 언어 장벽으로 인해 비영어 출처의 추출 정확도가 제한될 가능성이 있으며, 전체 원문이 이용 가능함에도 불구하고 중국어, 일본어 문헌의 정확한 해석에 어려움이 있을 수 있다. 또한 질 평가는 두 명의 전문가가 독립적으로 수행했으나 블라인드 처리는 하지 않았고, 독립 평가와 협의를 반복하는 방식이었기 때문에 평가자 간 일치율은 산정되지 않았다. 이러한 한계들은 본 연구 결과 해석 시 고려되어야 한다.

Ⅴ. 결 론

컴퓨터화 된 설진 기술은 기본적인 디지털화에서 2020년 이후 정교한 AI 기반 다중모달 진단 시스템으로 진화하였다. 주요 기술적 성과로는 CIE D65 조명 및 색상 보정을 통한 표준화된 이미지 획득, 높은 자동 분할 알고리즘 개발, 전통 한의학 장기 매핑 이론을 검증하는 다중 영역 분석, 설태 경계를 드러내는 UV 파장 이미징, 담적증후군 증상과 설진 특징 간 통계적으로 유의한 연관성을 입증한 대규모 임상 연구 등이 있었다.

그러나 연구간 비교 가능성을 제한하는 표준화된 프로토콜이 아직 부재하고, 알고리즘 개발 및 벤치마킹을 제한하는 공개 데이터셋이 미비하며, 한의학 임상현장에 통합하고 적용할 수 있는 경로가 부재하였다. 향후 연구는 대규모 다기관 임상연구, 표준화된 이미징 프로토콜, 공개 가능한 데이터셋 및 모바일/웨어러블 플랫폼 통합 실시간 진단 시스템을 우선시해야 한다. 이러한 엄격한 검증을 통해서만 컴퓨터화 된 설진이 광범위한 임상 채택을 달성하고 한의학 변증 감별 및 예방 건강 모니터링을 위한 객관적 정량 도구로서의 잠재력을 실현할 수 있을 것으로 판단된다.

감사의 글

이 연구에서 문헌의 선택과 배제, 질 평가를 함께 수행해 준 대한한의진단학회 대위원인 한경숙 박사에게 감사를 드립니다. 이 연구는 한국연구재단 바이오·의료기술개발사업 (NRF-2022M3A9E4017033) 지원에 의하여 수행되었다.

참고문헌

1. Akira M, Shi Y. Research on the objectification of tongue diagnosis in Traditional Chinese Medicine: about the relationship between tongue color and deep body temperature. Journal of Tianjin College of Traditional Chinese Medicine. 1988;(1):46-47.
2. Fu H. Study on the objectification of tongue diagnosis (1st report): The surface morphology of tongue was observed by microscope. International Journal of Traditional Chinese Medicine. 1995;(3):52-53.
3. Watsuji T, Shinohara S, Arita S. Changes of syndrome with time in tongue diagnosis supporting system using fuzzy theory. Proceedings of the Annual Conference of Biomedical Fuzzy Systems Association. 2002;43-46.
4. Watsuji T, Shinohara S, Arita S. Construction of diagnosis supporting system in Traditional Chinese Medicine using fuzzy theory. Proceedings of the Annual Conference of Biomedical Fuzzy Systems Association. 2003;5(1):23-30.
5. Keita H, Ryuji Y, Takaki T, Akitoshi K, Shu F. Objective assessment of oral hygiene using deep learning with tongue photographs. Kisei Dental Journal. 2022;49(2):81-86.
6. Kim M, Seo J, Nam D. A deep learning algorithm based on U-Net for automatic tongue region segmentation from images acquired through a computerized tongue image acquisition system. Korean Journal of Oriental Physiology and Pathology. 2024;38(6):259-265.
7. Kim T. A study on health monitoring using tongue image analysis. KICS Fall Conference Proceedings. 2008;287-289.
8. Kim KH, Jung CJ, Kim JY. Study of Repeatability Improvement in Capturing Tongue Images. KIEE Summer Conference Proceedings. 2013;539(1):1787-1789.
9. Wu J, Li Z, Cai Y, Liang H, Zhou L, Chen M, et al. A novel tongue coating segmentation method based on improved TransUNet. Sensors. 2024;24(14):4455.
10. Di D, Zhou M, Qin P, Zhu W. Advances on objectification of face and tongue diagnosis in traditional Chinese medicine. Shanghai Journal of Traditional Chinese Medicine. 2012;46(4):89-92.
11. Lin H, Ning Z, Zhang C, Men S, Zhang D. Computerized tongue image analysis for non-invasive disease screening: a review. Chinese Medicine. 2025;20(1):196.
12. Wong W, Huang S. Studies on externalization of application of tongue inspection of TCM. Engineering Science. 2001;3(1):78-82.
13. Watsuji T, Shinohara S, Kitade T, Arita S. Evaluation of coldness-heat in tongue diagnosis supporting system using fuzzy theory. Biomedical Fuzzy

Systems Association. 2001;76-80.

14. Tian Y. The study of the objectification of tongue diagnosis (2): to observe the surface morphology of the tongue by microscope and its correlation with the inquiry of Oriental Medicine. International Journal of Traditional Chinese Medicine. 1996;18(3):62.
15. Kim T, Park B. Preliminary study on health monitoring using tongue image analysis. Journal of the Korea Academia-Industrial cooperation Society. 2006;7(6):1219-1223.
16. Yagura H, Sakaue K. Correlation between Kampo medicine tongue diagnosis and western medicine cardiac and pulmonary Pathologies. Japan Journal of Unhealthy System. 2008;14(2):267-269.
17. Watsuji T, Shinohara S, Arita S. Experiment in the improvement of tongue diagnosis supporting system. Proceedings of the Annual Conference of Biomedical Fuzzy Systems Association. 2008;10(1):25-28.
18. Ling L, Xin X, Gang L. Effect of light resource in spectral method of tongue inspection metrization. Life Science Instruments. 2008;6(5):55-57.
19. Wenjuan Y, Gang L, Ling L. Research on discrimination of tongue diseases with near infrared spectroscopy. Infrared Technology. 2010;32(8):487-490.
20. Wenjuan Y, Gang L, Ling L, Baoju Z, Ying T. Applied research of artificial neural network in tongue diagnosis by near-infrared spectroscopy. Computer Engineering and Applications. 2011;47(27):132-135.
21. Wenjuan Y, Jing Z, Jing Z, Ling L, Gang L. Spectrum identification of hepatitis tongue diagnosis based on principal component analysis and BP neural network. Journal of Tianjin University. 2011;44(4):287-290.
22. Choi C, Lee W, Hong Y, Lee S, Nam D. Optimal tongue image analysis for recognizing a coated tongue in the tongue diagnosis. KSCS Spring Conference Proceedings. 2011;533-534.
23. Watsuji N, Seki S, Shinohara S, Kitakoji H, Yano T. The usefulness and the improvement of the tongue diagnosis and face diagnosis imaging system. Proceedings of the 26th Annual Conference of Biomedical Fuzzy Systems Association. 2013;9-12.
24. Chenghua D, Jianqiang D, Mingfeng Z, Kang Z, Shaoyong Z, Lun N. Exploration of TCM tongue diagnosis research based on image analysis technology. Jiangxi Journal of Traditional Chinese Medicine. 2013;44(11):7-9.
25. Jung C, Kim K. Study of parameter extraction for tongue coating diagnosis using color histogram of tongue image. KIEE Summer Conference Proceedings. 2013;36(1):1809-1811.

26. Kim K, Jung C. Study of tongue coating classification method by using Gaussian mixture model. KIEE Summer Conference Proceedings. 2013;36(1):1812-1814.

27. Ryu I, Siio I. Tongue diagnosis: A tongue diagnosis system for personal health care on smartphones. Proceedings of IPSJ Interaction. 2014;405-407.

28. Kim K, Kim J. Development of a method for diagnosing health condition based on a tongue image. KICS Summer Conference Proceedings. 2016;289-290.

29. Jang M, Lee W. Improved snakes algorithm for tongue image segmentation in oriental tongue diagnosis. The Journal of The Institute of Internet, Broadcasting and Communication. 2016;16(4):125-131.

30. Choi W, Kim K. A Study on repetitive tongue image acquisition method using depth information. KIEE Summer Conference Proceedings. 2018;1648-1650.

31. Kim K, So J, Kim J, Choi W. Suggestion of a method for diagnosing health condition based on a tongue image by using a mobile system. KICS Winter Conference Proceedings. 2018;120-121.

32. Choi W. A study on tongue image acquisition and analysis method for accurate tongue diagnosis. Information and Control. 2019;67-68.

33. Ryo O, Ying D. Estimating stress level from natural face and tongue images. Proceedings of the 84th National Convention of Information Processing Society of Japan. 2022;(2):663-664.

34. Kim K. Development of geometric and color correction methods for tongue image acquisition devices for healthcare. Information and Control. 2023;126-127.

35. Kim K. Proposal of tongue image derivation method using focus discrimination algorithm. ICS 2023 Information and Control Symposium. 2023;57-58.

36. Kim K, Kim J, Choi W. Development of symptom classification technology for phlegm mass syndrome using image-based tongue characteristics. The Institute of Electronics and Information Engineers (IEIE) Conference. 2024;2024(4):26-27.

37. Liu H, Zhang P, Huang Y, Zuo S, Li L, She C, et al. Research on multi-label recognition of tongue features in stroke patients based on deep learning. Sci Rep. 2024;14(1):32144.

38. Yi Z, Lai X, Sun A, Fang S. Tongue feature recognition to monitor rehabilitation: deep neural network with visual attention mechanism. Front Bioeng Biotechnol. 2024 May 9;12:1392513.

39. Chang W, Chen C, Wu H, Hsu P, Lo L, Chu H, et al. Tongue feature dataset construction and real-time detection. PLoS One. 2024;19(3):e0296070.

40. Hao B, Duan M, Zhao X, Wang T, Hu K, Hai R, et al. Objectification and application of tongue diagnosis in insomnia of liver depression and blood

deficiency. World Chinese Medicine. 2025;20(11):1954-1960.

41. Cho E, Choi W, Lim JH, Son JW, Jang SH, Lee SH, et al. Temporal changes in tongue color during immune checkpoint inhibitor therapy in patients with non-small-cell lung cancer: a prospective observational study using digital tongue diagnosis. Oncol Rev. 2025;19:1697252.
42. Yang X, Hao J, Wang Y, Man Y, Yang R, Wu Q. YOLOv13-SwinTongue: tongue coating diagnosis using an enhanced YOLOv13 with Swin transformer. Sensors. 2025;26(1):219.
43. Lu XZ, Liu S, Lin XX, Zeng Y, Chen JH, Ke WP, et al. An AI-powered tongue image model for home-based monitoring of liver fibrosis. NPJ Digit Med. 2025;9(1):67.
44. Ruan T, Wu W, Piao M, Sun Y, Ju X, Liu M, et al. Machine learning-assisted tongue image analysis for the diagnosis of Hashimoto's thyroiditis. Front Med. 2025;12:1673891.
45. Gao J, Chen T, Xu Y, Wu Y, Liu K, Qiu W, et al. Accurate fatty liver disease diagnosis with a multi-source feature fusion model on the segmented tongue image dataset. J Adv Res. 2026;85:507-520.
46. Yu N, Huang J, Liu J, Wang S, Zhang Y, Wu F, et al. Development of a cold-heat syndrome classification model for children with allergic rhinitis based on multimodal data. Transl Pediatr. 2025;14(12):3375-3386.
47. Wang X, Cao Y, Chen Y, Li H, Han A, Tang Y. An improved lightweight tongue segmentation model with self-attention parallel network and progressive upsampling. Sci Rep. 2025;15(1):27622.
48. Dong Z, Zhao L, Fan Y, Ma H, Shao C, Zhang Y, et al. GA-TongueNet: tongue image segmentation network using innovative DiFP and MDi for stable generalization ability. Front Physiol. 2025;16:1617647.
49. Li T, Zuo L, Wang P, Yang L, Liu Z, Wang X, et al. TonguExpert: A deep learning-based algorithm platform for fine-grained extraction and classification of tongue phenotypes. Phenomics. 2025;5(2):109-122.
50. Liu Y, Fan L, Zhao M, Wei D, Zhao M, Dong Y, et al. Study on a Traditional Chinese Medicine constitution recognition model using tongue image characteristics and deep learning: a prospective dual-center investigation. Chin Med. 2025;20(1):84.
51. Sha X, Guan Z, Wang Y, Han J, Wang Y, Chen Z. SSC-Net: A multi-task joint learning network for tongue image segmentation and multi-label classification. Digit Health. 2025;11:20552076251343696.
52. Chen T, Chen Y, Zhou Z, Zhu Y, He L, Zhang J. Deep learning-based automated

tongue analysis system for assisted Chinese medicine diagnosis. Front Physiol. 2025;16:1559389.

53. Yang L, Dong Q, Lin D, Lü X. TongueNet: a multi-modal fusion and multi-label classification model for Traditional Chinese Medicine tongue diagnosis. Front Physiol. 2025;16:1527751.
54. Gan Q, Wang C, Zhong Z, Wu J, Ge Q, Shi L, et al. Cross-modal attention model integrating tongue images and descriptions: a novel intelligent TCM approach for pathological organ diagnosis. Front Physiol. 2025;16:1580985.
55. Liang T, Wang H, Yao W, Yang Q. Tongue shape classification based on IF-RCNet. Sci Rep. 2025;15(1):7301.
56. Tian Z, Zhang J, Fan Y, Sun X, Wang D, Liu X, et al. Diabetic peripheral neuropathy detection of type 2 diabetes using machine learning from TCM features: a cross-sectional study. BMC Med Inform Decis Mak. 2025;25(1):90.
57. Ho SC, Chen Y, Xie YJ, Yeung WF, Chen SC, Qin J. Visceral condition assessment through digital tongue image analysis. Front Artif Intell. 2025;7:1501184.
58. Jin C, Zhang D, Cao X, Zhang Z, Xue C, Zhang Y. Lightweight YOLOv8 for tongue teeth marks and fissures detection based on C2f_DCNv3. Sci Rep. 2025;15(1):1560.
59. Wei J, Song J, Lu W, Men S, Lin C, Zhou P. A portable non-contact tongue imaging system with automated analysis for community and home settings. Digit Health. 2026;12:20552076261455223.
60. Lu CX, Tian CX, Jiao YB, Zhu H, Yu HY, Shu ZX, et al. Tongue image analysis and clinical data fusion: a novel approach for non-invasive diagnosis of metabolic dysfunction-associated fatty liver disease. J Clin Transl Hepatol. 2026;14(4):416-429.
61. Liu C, Luo Y, Chen T, You J, Wang J. TongueNet-GYN: a multimodal deep learning framework for non-invasive gynecological disease screening in digital public health. Front Public Health. 2026;14:1854215.