DOI：10.3969/j.issn.1004-4949.2026.17.030

# 面向黄褐斑及色素沉着性皮肤病图像识别的公开数据集分层使用路径研究

张 茜，付 敏，郑 娟，王堃梅
（乐山市中医医院，四川 乐山 614000）

[摘 要]**目的** 评价公开皮肤病图像数据集用于色素沉着性皮肤病图像识别研究的可用性，明确可直接用于监督学习、外部验证和可解释性分析的数据来源。**方法** 以“皮肤病图像、色素沉着、黄褐斑、hyperpigmentation、melasma、dermatology image dataset、skin tone”等为关键词，检索公开数据集平台、数据集官方网站及原始论文，纳入提供皮肤病图像、疾病标签或皮损概念标签，且可公开获取或按许可申请获取的数据集。提取数据集图像规模、标签结构、色素沉着相关标签、肤色/肤型元数据、标注来源、许可方式和图像识别研究适用性等指标；采用描述性统计分析数据集构成，并对具有明确类别数量的Mendeley SDC数据集进行类别不平衡分析。**结果** 共纳入SCIN、Mendeley SDC、Fitzpatrick17k和SKINCON 4个公开可获取数据集。Mendeley SDC数据集含9548张图像，其中色素沉着类700张，占7.33%；5类样本最大/最小类样本量比为4.52。SCIN含10 408张自愿贡献皮肤病图像，具备皮肤科医生标签、症状和肤色/肤型信息；Fitzpatrick17k含16 577张临床皮肤病图像和114类疾病标签，并提供Fitzpatrick肤型标注；SKINCON在Fitzpatrick17k和DDI子集上提供48个皮损临床概念标注。**结论** 现有公开数据集中，Mendeley SDC数据集最适合作为色素沉着识别任务的初始监督训练数据；SCIN和Fitzpatrick17k更适合用于外部验证及肤色分层评估；SKINCON适合用于模型可解释性和错误分析。公开数据集可为色素沉着性皮肤病图像识别研究提供可复现的数据基础，但用于黄褐斑临床辅助诊断前仍需进行疾病标签二次核验和本地人群外部验证。

[关键词] 色素沉着；黄褐斑；皮肤病图像；公开数据集；人工智能；图像识别
[中图分类号] R751; TP391.41 [文献标识码] A [文章编号] 1004-4949（2026）17-0123-06

## Research on Hierarchical Use Pathway of Public Datasets for Image Recognition of Melasma and Hyperpigmentation Disorders

**ZHANG Qian, FU Min, ZHENG Juan, WANG Kunmei**
(Leshan Hospital of Traditional Chinese Medicine, Leshan 614000, Sichuan, China)

**[Abstract]Objective** To evaluate the availability of publicly available skin disease image datasets for image recognition research of hyperpigmentation disorders, and identify data sources that can be directly applied to supervised learning, external validation and interpretability analysis. **Methods** Public dataset platforms, official dataset websites and original publications were searched using keywords including dermatological image, hyperpigmentation, melasma, dermatology image dataset and skin tone. Datasets with publicly available or accessible upon license application that provide dermatological images, disease labels or lesion concept labels were included. Indicators such as dataset image scale, label structure, hyperpigmentation-related labels, skin tone/skin type metadata, annotation sources, licensing modes and applicability for image recognition research were extracted. Descriptive statistics were used to analyze dataset composition, and class imbalance analysis was performed on the Mendeley SDC Dataset with clear category quantities. **Results** A total of four publicly available datasets were included: SCIN, Mendeley SDC Dataset, Fitzpatrick17k and SKINCON. The Mendeley SDC dataset contained 9548 images, among which 700 were hyperpigmentation-related images,

**第一作者**：张茜（1989.4-），女，四川仁寿县人，本科，中级，主要从事医疗美容工作

医学美学美容

accounting for 7.33%. The ratio of the maximum sample size to the minimum sample size among 5 categories was 4.52. The SCIN dataset contained 10,408 voluntarily contributed dermatological images with dermatologist-assigned labels, symptoms and skin tone/skin type information. The Fitzpatrick17k dataset contained 16,577 clinical dermatological images and labels of 114 disease categories, and provided Fitzpatrick skin type annotations. SKINCON provided 48 clinical concept annotations of skin lesions based on subsets of Fitzpatrick17k and DDI. **Conclusion** Among existing public datasets, the Mendeley SDC dataset is most suitable as initial supervised training data for hyperpigmentation recognition tasks. SCIN and Fitzpatrick17k are more appropriate for external validation and skin tone-stratified evaluation. SKINCON is applicable for model interpretability and error analysis. Public datasets can offer a reproducible data foundation for image recognition research of hyperpigmentation disorders. Secondary verification of disease labels and external validation on local populations are still required before their application in clinical auxiliary diagnosis of melasma.

**[Key words]** Hyperpigmentation; Melasma; Dermatological image; Public dataset; Artificial intelligence; Image recognition

色素沉着性皮肤病是皮肤科和医学美容门诊常见问题，黄褐斑、炎症后色素沉着及其他获得性色素增加性疾病均可表现为不同程度的褐色或黑褐色斑片。黄褐斑是一种慢性、获得性面部色素增加性皮肤病，发病机制涉及遗传、日光、性激素、黑素合成增加、血管改变、炎症反应和皮肤屏障受损等因素[1]，临床诊断主要依据病史、皮损特征和无创检测技术[2]。由于皮损颜色、边界、分布、复发倾向及肤色背景差异明显[3]，单纯依靠肉眼观察容易受到光照、拍摄设备和经验差异影响。近年来，皮肤病图像识别在皮肤肿瘤、炎症性皮肤病和常见皮损分类中发展较快[4]，但色素沉着性皮肤病公开数据基础相对薄弱。与黑色素瘤或皮肤肿瘤数据集相比[5, 6]，黄褐斑及广义色素沉着类疾病常被归入上位标签，缺乏标准化分型、严重程度评分和长期随访信息。若直接使用未经核验的公开图像训练模型，可能造成标签混杂、肤色偏倚和临床外推风险。因此，在开展具体模型训练或临床转化前，有必要先对公开皮肤病图像数据集进行系统梳理，明确哪些数据集可用于监督学习、哪些适合外部验证、哪些可用于解释模型错误。本研究以公开可获取皮肤病图像数据集为对象，围绕色素沉着性皮肤病图像识别任务进行二次资料分析，从数据源、标签粒度、肤色/肤型信息、标注可靠性和复现便利性等方面建立可执行的数据路径，为后续开展黄褐斑及相关色素沉着性疾病的辅助识别研究提供依据。

## 1 资料与方法

1.1 数据来源与检索策略 2026年5月6日检索。数据来源包括Mendeley Data、Hugging Face、GitHub、数据集官方网站、JAMA Network Open、arXiv及相关论文索引平台。中文和英文检索词包括：“皮肤病图像”“色素沉着”“黄褐斑”“医学图像识别”“公开数据集”“dermatology image dataset”“skin disease dataset”“hyperpigmentation”“melasma”“skin tone”“Fitzpatrick skin type”“SKINCON”“SCIN”等。本研究使用公开可获取数据集说明文档和已发表文献进行二次资料分析，不涉及研究者直接接触患者、采集人体样本或获取可识别个人身份信息。数据集原始图像的使用应遵循各数据集发布方许可协议和所在机构伦理规范，所有作者声明无利益冲突。

1.2 纳入与排除标准 纳入标准：数据集收录完整皮肤病原始图像资料；配套结构化标注信息，包含适用于图像识别研究的元数据；具备公开获取渠道，支持免费下载或依据官方许可协议申请使用；配套官方说明文档、原始研究文献，完整披露样本总量、标签体系、数据来源等核心信息。排除标准：仅收录皮肤肿瘤或皮肤镜图像、与色素沉着性疾病无直接关联的数据集；为已纳入数据集的完全重复或镜像，未提供额外标签或独立信息的数据集。依据上述纳入、排除标准完成数据集筛选流程，整体筛选与后续分析流程见图1。

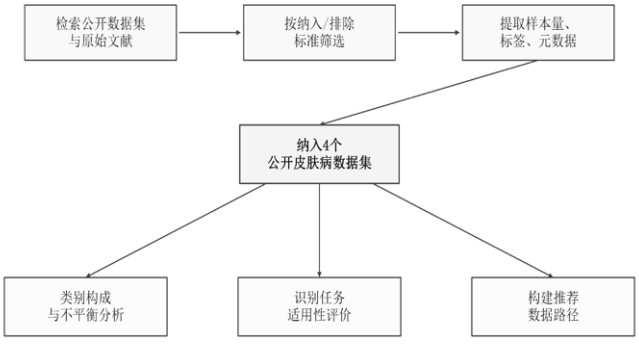



图1 公开皮肤病图像数据集筛选与分析流程

医学美学美容

1.3 指标提取 由研究者从数据集说明文档和原始论文中提取以下指标：数据集名称、发布或论文年份、图像数或病例数、疾病标签或概念标签数量、是否包含色素沉着相关标签、是否包含肤色/肤型信息、是否有皮肤科医生或领域专家标注、数据许可方式、适用的图像识别研究场景。对具有明确类别数量的Mendeley Skin Disease Classification Dataset（Mendeley SDC数据集）[7]，进一步计算各类别构成比和类别不平衡程度。

1.4 统计学方法 采用描述性统计分析数据集规模、类别构成和标签特点。计数资料以例数或图像数表示，构成比以%表示。类别不平衡程度采用最大类别样本量与最小类别样本量之比表示。为直观比较不同数据集在色素沉着性皮肤病图像识别任务中的可用性，本研究建立结构化评价矩阵：0分表示公开说明中未见或明显不足，1分表示部分具备，2分表示具备或较强；评价维度包括色素沉着标签直达性、样本规模、肤色/肤型信息、临床/专家标注、概念可解释性、获取与复现便利性。

## 2 结果

2.1 纳入数据集总体情况 本研究最终纳入4个公开可获取数据集：SCIN（Skin Condition Image Network）[8, 9]、Mendeley SDC[7]、Fitzpatrick17k[10]和SKINCON[11, 12]。Mendeley SDC数据集专门设置色素沉着分类标签，适配模型初始训练；SCIN依托真实线上患者图像，配套完整人口学与肤色相关元数据；Fitzpatrick17k覆盖疾病谱较广，标准化肤型分类体系完善；SKINCON侧重皮肤形态概念标注，多用于模型结果溯源与错误分析。4套数据集图像规模、标签结构、色素沉着相关性、肤色/肤型信息、标注特点及主要用途汇总见表1。各数据集图像样本数量直观对比见图2。

表 1 纳入公开皮肤病图像数据集基本特征

| 数据集 | 图像规模 | 标签结构 | 色素沉着相关性 | 肤色 / 肤型信息 | 标注特点 | 主要用途 |
|---|---|---|---|---|---|---|
| Mendeley SDC | 9548 张 | 5 类疾病标签 | 含色素沉着类 700 张 | 公开说明未见 | 医院 / 在线来源；标签较粗 | 色素沉着识别初始训练 |
| SCIN | 10 408 张；5033 份贡献 | 皮肤科医生标签；症状 / 人口学信息 | 可映射筛选相关图像 | 含 FST 与 Monk 肤色标签 | 自愿贡献；有隐私与质量控制 | 外部验证、肤色分层 |
| Fitzpatrick17k | 16 577 张 | 114 类皮肤病标签 | 可筛选色素性疾病 | 含 Fitzpatrick 肤型标签 | 来自公开皮肤病图谱；标签需核验 | 跨肤型泛化与公平性 |
| SKINCON | 3886 张概念标注图像 | 48 个皮损临床概念 | 可关注颜色 / 色素概念 | 具备肤型 / 肤色背景 | 领域专家密集概念标注 | 可解释性与错误分析 |

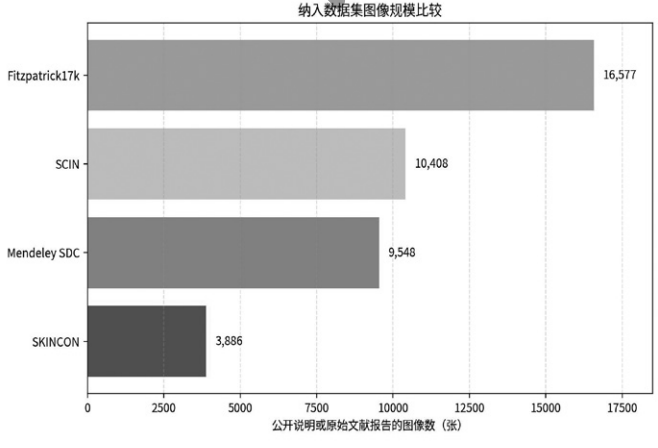



图 2 纳入数据集图像规模比较

2.2 色素沉着相关监督标签的可用性 Mendeley SDC数据集是4个数据集中与色素沉着性皮肤病图像识别最直接相关的数据集。该数据集共9548张图像，包含痤疮、白癜风、色素沉着、甲银屑病和Stevens-Johnson综合征/中毒性表皮坏死松解症（SJS-TEN）5类。其中色素沉着类700张，占7.33%。各类别样本量差异明显，最大类别为SJS-TEN，最小类别为色素沉着，最大/最小类样本量比为4.52，各类别样本占比明细见表2。从识别任务角度看，该数据集可构建“色素沉着与其他常见皮肤病”的多分类模型，也可重构为“色素沉着/非色素沉着”的二分类任务。但需要注意，hyperpigmentation是上位标签，可能同时包含黄褐

医学美学美容

斑、炎症后色素沉着及其他色素增加性表现，若研究目标聚焦黄褐斑，仍需进行皮肤科医师二次阅片和标签细化。

表 2　Mendeley SDC 数据集各类别样本构成

| 类别 | 英文标签 | 图像数（张） | 构成比（%） |
|---|---|---|---|
| 痤疮 | acne | 1148 | 12.02 |
| 白癜风 | vitiligo | 2016 | 21.11 |
| 色素沉着 | hyperpigmentation | 700 | 7.33 |
| 甲银屑病 | nail psoriasis | 2520 | 26.39 |
| SJS–TEN | SJS–TEN | 3164 | 33.14 |
| 合计 | – | 9548 | 100.00 |

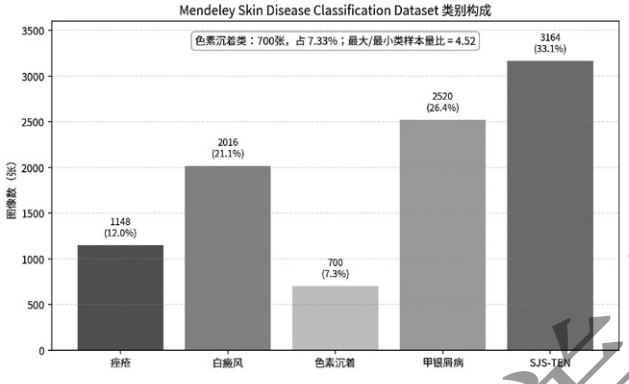



图 3　Mendeley SDC 数据集类别构成及色素沉着类占比

2.3　肤色/肤型与外部验证价值　色素沉着性皮肤病的图像识别易受肤色背景、光照条件和拍摄设备影响。SCIN和Fitzpatrick17k的优势在于提供肤色或肤型相关信息。SCIN同时包含受试者自报Fitzpatrick肤型、皮肤科医生评估Fitzpatrick肤型和Monk肤色估计标签，并记录年龄、性别、症状和病程等信息，有助于开展肤色分层、图像质量和临床背景信息对识别结果影响的研究。Fitzpatrick17k包含16 577张临床皮肤病图像和Fitzpatrick肤型标签，可用于评估模型在不同肤型中的泛化表现。二者在色素沉着疾病识别中的局限也较明确：SCIN并非专门为黄褐斑或色素沉着性疾病构建，需要通过疾病标签和临床文本进行二次筛选；Fitzpatrick17k图像来源于公开皮肤病图谱，疾病标签未必均经过病理或临床金标准确认，存在诊断清晰度和误标风险。因此，二者更适合作为外部验证或偏倚评估数据，不宜直接替代经过皮肤科医师核验的本地临床数据。

2.4　可解释性与错误分析价值　SKINCON数据集支持皮损层级的概念标注，涵盖斑块、鳞屑、糜烂及各类颜色、形态等皮肤科临床常用描述词。针对色素沉着性皮肤病识别任务，该概念标注可用于解析模型决策依据，明确模型是否聚焦皮损的颜色深浅、边界特征、分布模式与表面改变，而非背景、标尺、文字水印、拍摄部位等伪相关特征。在模型开发中，SKINCON可用于：①利用概念激活向量或概念瓶颈模型评价模型是否学习到临床相关视觉概念；②对错分样本进行细粒度错误分析，区分颜色混淆、形态混淆和图像质量问题；③在向临床使用者展示模型结果时提供可解释说明，提高辅助识别工具的可理解性。

4套数据集在标签直达性、样本规模、肤色/肤型信息、临床/专家标注、概念可解释性、获取与复现便利性6项维度的综合适配评价结果见表3，可视化评价矩阵见图4。

2.5　色素沉着性皮肤病图像识别推荐数据路径　公开数据集用于色素沉着性皮肤病图像识别时宜采用“主训练集—外部验证集—可解释性数据集”的分层路径。Mendeley SDC数据集提供明确的hyperpigmentation类别，可作为初始监督训练数据；SCIN和Fitzpatrick17k可在标签映射和二次筛选后用于外部验证、肤色分层和来源差异评估；SKINCON可作为概念级解释和错误分析专用数据集。整套分层数据使用流程见图5。该路径可降低单一数据集带来的来源偏倚，提升模型实验可复性与临床解释价值。

表 3　数据集在色素沉着性皮肤病图像识别任务中的适用性评价

| 数据集 | 标签直达性 | 样本规模 | 肤色 / 肤型 | 临床 / 专家标注 | 概念可解释性 | 获取与复现便利性 |
|---|---|---|---|---|---|---|
| Mendeley SDC | 2 | 1 | 0 | 0 | 0 | 2 |
| SCIN | 1 | 2 | 2 | 2 | 0 | 1 |
| Fitzpatrick17k | 1 | 2 | 2 | 1 | 0 | 1 |
| SKINCON | 1 | 1 | 2 | 2 | 2 | 1 |

注：0=公开说明中未见或明显不足；1=部分具备；2=具备或较强。该矩阵用于数据源选择，不代表模型性能。

医学美学美容

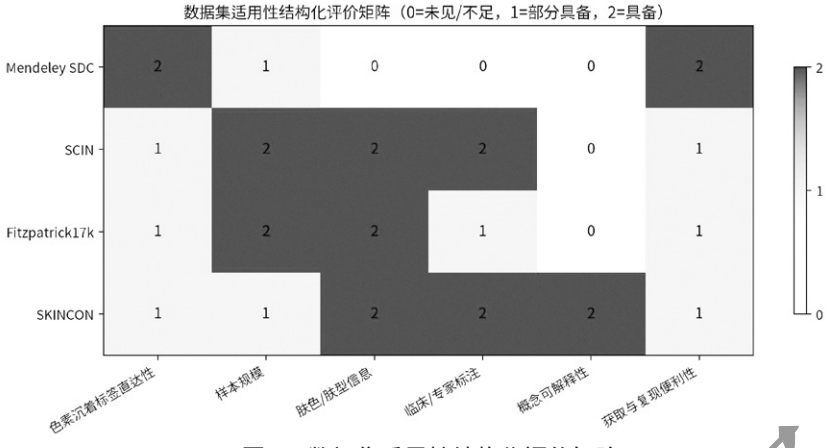



图 4 数据集适用性结构化评价矩阵

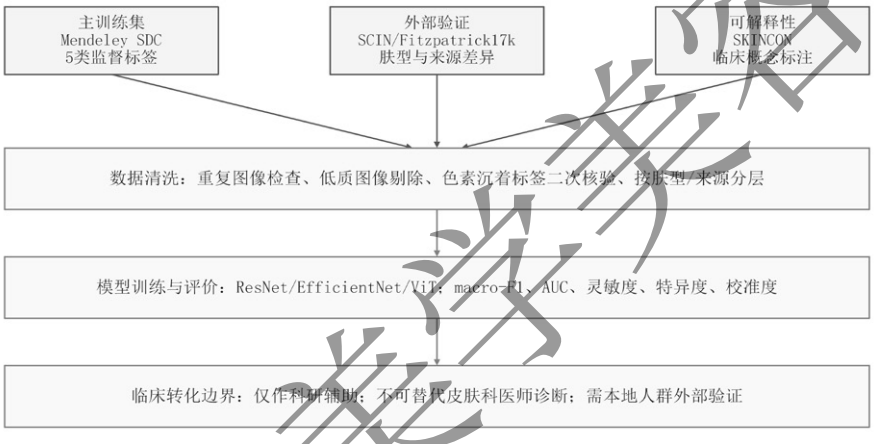



图 5 色素沉着性皮肤病图像识别推荐数据路径

## 3 讨论

现有公开皮肤病图像数据集可用于色素沉着性皮肤病识别，但需依据研究目标分层选用。若旨在构建可复现的初始分类模型，优先选用带有hyperpigmentation标签的数据集[7]；若目标是评价模型在不同肤色中的稳定性，则需引入带有Fitzpatrick肤型或Monk肤色标签的数据集[8-10]；若目标是提高临床可解释性，则需概念级标注数据集[11, 12]。单一数据集难以同时兼顾监督标签、肤色分层、专家标注与可解释性等要求。黄褐斑虽属于色素沉着性疾病，但不能直接等同于通用的“色素沉着”标签。公开数据集中的hyperpigmentation仅适合作为上位筛选标签，用于识别颜色加深类皮损，无法替代黄褐斑专属诊断标签。对于以黄褐斑为目标的模型开发，较稳妥的策略是先利用公开数据集完成图像预处理、类别不平衡处理和模型结构验证，再由皮肤科医师对本地临床图像开展黄褐斑特异性标注，结合MASI评分、标准化面部图像和无创检测结果，构建更高质量的临床数据集。

Mendeley SDC数据集中色素沉着类仅占7.33%，是5类中样本量最少的一类，类别不平衡比达到4.52。在多分类模型中，若仅报告总体准确率，模型可能倾向于识别样本量较大的类别，从而掩盖对色素沉着类识别不足的问题。因此，色素沉着识别任务不宜只采用准确率作为核心指标，应同时报告macro-F1、每类灵敏度、每类特异度、ROC曲线下面积和混淆矩阵。可从三方面处理类别不平衡问题：①数据层面，采用分层抽样、少数类样本增强与图像去重，规避训练集与验证集间的信息泄露；②算法层面，引入类别权重、焦点损失[13]或重采样策略，提升模型对少数类样本的关注度；③评价层面，借助外部验证集与肤色分层分析，防止模型性能仅在单一数据源或特定肤型人群中表现优异。

色素沉着性皮肤病的视觉特征与肤色背景高度相关。浅肤色人群中轻度褐色斑片易于辨识，同等黑素增量在深肤色人群中则难以通过常规图

医学美学美容

像识别。若训练集以浅肤色样本占主导，模型在深肤色人群上的灵敏度可能下降[14]，甚至将正常肤色差异误判为病理性色素沉着。Fitzpatrick17k和SCIN提供肤色/肤型变量，使研究者能够在模型评价中报告不同肤型分层结果，这是色素沉着识别研究区别于一般皮肤图像分类研究的关键。国内临床应用还面临本地人群外推问题。公开数据集多来源于国外互联网用户、在线图谱或国外医院资料，年龄结构、肤色构成、拍摄设备和就医行为与国内中医美容或皮肤科门诊人群不同。模型即使在公开数据集上表现良好，也不能直接作为临床诊断工具。合理的做法是将公开数据集作为方法学起点，最终仍需在本地真实世界图像中完成外部验证。

医学图像识别模型应用于高风险临床场景时，不应仅输出分类概率，还需提供判断依据。色素沉着性皮肤病对模型可解释性要求较高，斑片颜色、边界、分布、炎症背景及皮肤屏障状态均可影响临床判读。SKINCON提供的皮损概念标注可帮助研究者检查模型是否学习了临床相关特征，而非依赖背景、水印、尺标、拍摄角度等非医学信息。在黄褐斑及医美相关研究中，可解释性还便于对接客观检测设备结果。若整合标准化面部图像、MASI评分、皮肤检测参数与生活质量量表，模型除判别有无色素沉着外，还可分析色斑范围、色素深度、面积变化及治疗动态趋势。该方向相较于单纯优化分类准确率，更契合临床随访与疗效评估的实际需求。

公开皮肤病图像数据集可为色素沉着性皮肤病图像识别研究提供可复现的数据基础，但不同数据集适用场景不同。Mendeley SDC数据集最适合作为色素沉着识别任务的初始监督训练数据；SCIN和Fitzpatrick17k适合用于外部验证、肤色/肤型分层和泛化能力评价；SKINCON适合用于模型可解释性和错误分析。面向黄褐斑的临床辅助识别研究不应直接将公开数据集中的hyperpigmentation标签等同于黄褐斑诊断，应在公开数据集方法验证基础上，进一步开展本地临床图像、标准化评分和专家二次标注研究。

## [参考文献]

[1]中国中西医结合学会皮肤性病专业委员会色素病学组,中华医学会皮肤性病学分会白癜风研究中心,中国医师协会皮肤科医师分会色素病工作组.中国黄褐斑诊疗专家共识(2021版)[J].中华皮肤科杂志,2021,54(2):110-115.

[2]Passeron T,Picardo M.Melasma,a photoaging disorder[J].Pigment Cell&Melanoma Research,2018,31(4):461-465.

[3]Desai SR,Alexis AF,Elbuluk N,et al.Best practices in the treatment of melasma with a focus on patients with skin of color[J].J Am Acad Dermatol,2024,90(2):269-279.

[4]Esteva A,Kuprel B,Novoa R A,et al.Dermatologist-level classification of skin cancer with deep neural networks[J].Nature,2017,542(7639):115-118.

[5]Wen D,Khan SM,Ji Xu A,et al.Characteristics of publicly available skin cancer image datasets:a systematic review[J].Lancet Digit Health,2022,4(1):e64-e74.

[6]Tschandl P,Rosendahl C,Kittler H.The HAM10000 dataset,a large collection of multi-source dermatoscopic images of common pigmented skin lesions[J].Sci Data,2018,5:180161.

[7]Khushbu SA.Skin Disease Classification Dataset[DS/OL].V1.Mendeley Data(2024-07-31)[2026-05-06].https://data.mendeley.com/datasets/3hckgznc67/1.

[8]Ward A,Li J,Wang J,et al.Creating an Empirical Dermatology Dataset Through Crowdsourcing With Web Search Advertisements[J].JAMA Network Open,2024,7(11):e2446615.

[9]Google Research.SCIN dataset:Skin Condition Image Network[DS/OL].GitHub,2024[2026-05-06].https://github.com/google-research-datasets/scin.

[10]Groh M,Harris C,Soenksen L,et al.Evaluating Deep Neural Networks Trained on Clinical Images in Dermatology with the Fitzpatrick 17k Dataset[DS/OL].arXiv:2104.09957,2021[2026-05-06]. https://arxiv.org/abs/2104.09957.

[11]Daneshjou R,Yuksekgonul M,Cai Z R,et al.SkinCon:A skin disease dataset densely annotated by domain experts for fine-grained model debugging and analysis[DS/OL].arXiv:2302.00785,2023[2026-05-06]. https://arxiv.org/abs/2302.00785.

[12]SKINCON Dataset Team.SKIN Concepts (SKINCON) Dataset[DS/OL].2023[2026-05-06].https://skincon-dataset.github.io/.

[13]Lin TY,Goyal P,Girshick R,et al.Focal Loss for Dense Object Detection[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV),2017:2980-2988.

[14]Daneshjou R,Vodrahalli K,Novoa R A,et al.Disparities in dermatology AI performance on a diverse,curated clinical image set[J].Sci Adv,2022,8(32):eabq6147.

收稿日期：2026-7-18　　编辑：刘雯

医学美学美容