

# 基于皮肤镜图像智能分析的早期蕈样肉芽肿诊断模型构建

刘兆睿<sup>1</sup>, 张漪澜<sup>2</sup>, 谢凤英<sup>2</sup>, 刘 洁<sup>1</sup>

<sup>1</sup> 中国医学科学院 北京协和医学院 北京协和医院皮肤科 疑难重症及罕见病国家重点实验室  
国家皮肤与免疫疾病临床医学研究中心, 北京 100730  
<sup>2</sup> 北京航空航天大学宇航学院图像处理中心, 北京 100191

通信作者: 刘 洁 电话: 010-69151536, E-mail: Liuji04672@pumch.cn  
谢凤英 电话: 010-61716929, E-mail: xfy\_73@buaa.edu.cn

**【摘要】目的** 比较基于皮肤镜图像的卷积神经网络 (convolutional neural network, CNN) 二分类模型在蕈样肉芽肿 (mycosis fungoides, MF) 与炎症性疾病鉴别诊断中的应用价值。**方法** 回顾性纳入 2016 年 1 月至 2020 年 12 月北京协和医院皮肤科门诊确诊的早期 MF 患者和临床表现与之相似的炎症性皮肤病患者, 并按 4 : 1 的比例随机分为训练集和测试集。使用训练集患者的皮肤镜图像对 6 种经典网络结构利用迁移学习进行训练, 以构建 CNN 二分类模型。同时, 在测试集中随机挑选每例患者 1 幅图像, 并结合皮损的临床图像, 由 13 名皮肤科医师对疾病归类进行判读。比较 CNN 二分类模型与皮肤科医师对测试集病例早期 MF 与炎症性疾病鉴别诊断的性能, 结果以曲线下面积 (area under the curve, AUC)、灵敏度、特异度、Kappa 值等表示, 并采用受试者工作特征 (receiver operating characteristic, ROC) 曲线进行可视化分析。**结果** 共纳入 48 例早期 MF 患者 (皮肤镜图像 402 幅) 和 96 例炎症性皮肤病患者 (皮肤镜图像 557 幅), 其中训练集 117 例 (皮肤镜图像 772 幅), 测试集 27 例 (皮肤镜图像 187 幅)。测试集中, 皮肤科医师鉴别诊断早期 MF 与炎症性皮肤病的灵敏度和特异度分别为 70.19% (95% CI: 59.68%~80.70%) 和 94.74% (95% CI: 91.77%~97.71%), Kappa 值为 0.677 (95% CI: 0.566~0.789)。按图像分类时, CNN 二分类模型对早期 MF 与炎症性皮肤病鉴别诊断的 AUC 为 0.87 (95% CI: 0.84~0.89), 灵敏度和特异度分别为 75.02% (95% CI: 70.19%~79.85%) 和 82.02% (95% CI: 79.30%~84.87%), Kappa 值为 0.563 (95% CI: 0.507~0.620); 按病例分类时, CNN 二分类模型对早期 MF 与炎症性皮肤病鉴别诊断的 AUC 为 0.97 (95% CI: 0.95~0.99), 灵敏度和特异度分别为 87.50% (95% CI: 78.55%~96.45%) 和 93.85% (95% CI: 88.93%~98.77%), Kappa 值为 0.920 (95% CI: 0.884~0.954)。ROC 曲线显示, 按病例分类时网络结构为 EfficientNet-B0 的 CNN 二分类模型诊断早期 MF 的 AUC 为 0.99, 灵敏度和特异度分别为 88.9% 和 100%, 且 13 名皮肤科医师诊断的灵敏度和特异度均值对应点位于曲线右下方。**结论** 基于皮肤镜图像智能分析的 CNN 二分类模型可实现对早期 MF 与炎症性皮肤病的精确分类, 对二者的鉴别诊断能力优于皮肤科医师的平均水平。

**【关键词】** 早期蕈样肉芽肿; 炎症性皮肤病; 皮肤镜图像; 卷积神经网络

**【中图分类号】** R751; TP29 **【文献标志码】** A **【文章编号】** 1674-9081(2021)05-0689-09

**DOI:** 10.12290/xhyxzz.2021-0496

## Early Diagnosis Model of Mycosis Fungoides Based on Intelligent Analysis of Dermoscopic Images

LIU Zhaorui<sup>1</sup>, ZHANG Yilan<sup>2</sup>, XIE Fengying<sup>2</sup>, LIU Jie<sup>1</sup>

<sup>1</sup>Department of Dermatology, State Key Laboratory of Complex Severe and Rare Diseases, National Clinical Research Center for Dermatologic and Immunologic Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China

刘兆睿、张漪澜对本文同等贡献  
基金项目: 国家自然科学基金 (61871011, 62071011, 61771031, 82173449); 中国医学科学院中央级公益性科研院所基本科研业务费专项资金 (2019XK320024); 北京市自然科学基金 (4192032)  
引用本文: 刘兆睿, 张漪澜, 谢凤英, 等. 基于皮肤镜图像智能分析的早期蕈样肉芽肿诊断模型构建 [J]. 协和医学杂志, 2021, 12 (5): 689-697. doi: 10.12290/xhyxzz.2021-0496.

<sup>2</sup>Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China

Corresponding authors: LIU Jie Tel: 86-10-69151536, E-mail: Liujie04672@pumch.cn

XIE Fengying Tel:86-10-61716929, E-mail:xfy\_73@buaa.edu.cn

**【Abstract】 Objective** To compare the application value of the binary classification model based on dermoscopic images of convolutional neural network (CNN) in the diagnosis of mycosis fungoides (MF) and inflammatory dermatosis. **Methods** Patients diagnosed with early MF or inflammatory dermatosis with similar clinical manifestations in the dermatology clinic of Peking Union Medical College Hospital from January 2016 to December 2020 were retrospectively included. The patients were divided into the training set and the test set at a ratio of 4 : 1. Six classical network structures were trained by using the dermoscopic images of patients in the training set, and the CNN binary classification model was constructed by using transfer learning. At the same time, in the test set, 1 image of each patient that was randomly selected, together with clinical images of the skin lesions, was interpreted by 13 dermatologists. Compare the CNN binary classification model with dermatologists in the differential diagnosis of early MF and inflammatory dermatosis in the test set. The results were expressed in terms of area under the curve (AUC), sensitivity, specificity, Kappa coefficient, etc., and receiver operating characteristic (ROC) curve was used for visual analysis. **Results** A total of 48 patients with early MF (402 dermoscopic images) and 96 patients with inflammatory dermatosis (557 dermoscopic images) were included. Among them, there were 117 cases in the training set (772 dermoscopic images), and 27 cases in the test set (187 dermoscopic images). In the test set, the sensitivity and specificity of dermatologists in the differential diagnosis of early MF and inflammatory dermatosis were 70.19% (95% CI: 59.68%–80.70%) and 94.74% (95% CI: 91.77%–97.71%) respectively, and the Kappa coefficient is 0.677 (95% CI: 0.566–0.789). When classified by the single image, the AUC of the CNN binary classification model for the differential diagnosis of early MF and inflammatory dermatosis was 0.87 (95% CI: 0.84–0.89); the sensitivity and specificity were 75.02% (95% CI: 70.19%–79.85%) and 82.02% (95% CI: 79.30%–84.87%), respectively; the Kappa coefficient was 0.563 (95% CI: 0.507–0.620). When classified by cases, the AUC of the CNN binary classification model for the differential diagnosis of early MF and inflammatory dermatosis was 0.97 (95% CI: 0.95–0.99); the sensitivity and specificity were 87.50% (95% CI: 78.55%–96.45%) and 93.85% (95% CI: 88.93%–98.77%), respectively; the Kappa coefficient was 0.920 (95% CI: 0.884–0.954). The ROC curve showed that the AUC of the CNN binary classification model with EfficientNet-B0 for diagnosing MF was 0.99 when classified by cases, the sensitivity and specificity were 88.9% and 100%, and the corresponding point of the average diagnostic sensitivity and specificity of 13 dermatologists were at the lower right of the curve. **Conclusions** The CNN binary classification model based on the intelligent analysis of dermoscopic images can accurately classify early MF and inflammatory dermatosis, and its ability of differential diagnosis is better than the average level of dermatologists.

**【Key words】** early mycosis fungoides; inflammatory dermatosis; dermoscopic image; convolutional neural network

**Funding:** National Natural Science Foundation of China (61871011, 62071011, 61771031, 82173449); The Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2019XK320024); Beijing Natural Science Foundation (4192032)

*Med J PUMCH*, 2021,12(5):689–697

蕈样肉芽肿 (mycosis fungoides, MF) 是最常见的皮肤 T 细胞淋巴瘤 (cutaneous T-cell lymphoma, CTCL), 为皮肤科疑难重症及罕见疾病, 晚期患者预后差、治疗手段有限, 早期识别及诊治可改善患者预

后, 具有重要的临床意义。早期 MF 可表现为红斑鳞屑性皮损, 与银屑病、慢性湿疹等良性炎症性皮肤病难以鉴别。近年来随着皮肤影像学的发展, 皮肤镜图像在皮肤病的鉴别诊断中发挥了重要作用<sup>[1]</sup>。鉴于

早期 MF 及炎症性皮肤病的预后存在较大差异，目前尚缺乏有效的无创指标进行早期诊断，而通过构建基于图像智能分析技术的早期诊断模型可高效识别皮肤镜图像并进行二分类，有望实现 MF 的早期诊断。本研究以早期 MF 患者和炎症性皮肤病患者为研究对象，比较基于皮肤镜图像的智能分析与皮肤科医师对二者的鉴别诊断能力，并对错误分类的皮肤镜图像进行分析，以提高早期 MF 的诊断效率，为构建 MF 早期诊断模型奠定基础。

1 资料与方法

1.1 研究对象及分组

1.1.1 研究对象

回顾性纳入 2016 年 1 月至 2020 年 12 月北京协和医院皮肤科门诊确诊的早期 MF 患者和临床表现与之相似的炎症性皮肤病（银屑病、慢性湿疹、脂溢性皮炎、玫瑰糠疹、扁平苔藓）患者。纳入标准：（1）早期 MF 均经组织病理学、免疫组化、T 细胞受体基因重排检查等确诊，并符合欧洲癌症治疗研究组织（the European Organization of Research and Treatment of Cancer, EORTC）以及国际皮肤淋巴瘤学会（International Society for Cutaneous Lymphoma, ISCL）提出的早期 MF 诊断标准<sup>[2]</sup>；（2）根据临床表现、皮肤镜、组织病理学（部分病例）、治疗反应等，由 2 名经培训且经验丰富的皮肤科医师共同作出炎症性皮肤病的诊断；（3）入组前 1 个月内未接受系统或局部治疗；（4）均行皮肤镜检查，且选取早期（斑片期及斑块期）皮损进行分析。排除标准：（1）仅毛发、甲、黏膜等特殊部位受累的病例；（2）皮肤镜检查图像不清晰者。

本研究已通过北京协和医院伦理审查委员会审批（审批号：JS-2003）。

1.1.2 分组

按 4：1 的比例，将患者随机纳入训练集和测试集。

1.2 研究方法

1.2.1 皮肤镜图像采集及诊断特征

采用奥地利 MoleMax 公司（MoleMax lid, Digital Imaing Systems）皮肤镜系统进行图像采集。根据皮损表面情况，采用偏振光浸润式、偏振光非浸润式或非偏振光浸润式进行拍摄，用 75% 乙醇作为镜头与皮损之间的浸润液体，放大倍数为 20 倍、30 倍或 40 倍。根据既往临床经验、文献<sup>[1-5]</sup>及预实验，MF 的

皮肤镜特征（图 1A）为：暗红色背景，均匀分布的点状血管，“精子样”血管，橘黄色斑片状区域，片状分布的白色鳞屑，多角形色素网、色素点（见于皮肤异色病样 MF）。炎症性皮肤病的皮肤镜特征如下：（1）银屑病（图 1B）：亮红色背景，规则分布的点/球状血管，片状白色鳞屑；（2）慢性湿疹（图 1C）：不均匀分布的点状血管，海绵状水疱，片状分布/弥漫分布的黄色鳞屑，暗红色背景；（3）扁平苔藓（图 1D）：Wickham 纹，多种血管结构，蓝灰色、黄棕色色素结构，毛囊角栓；（4）脂溢性皮炎（图 1E）：不均匀分布的分支状血管和线状弯曲血管，毛囊周围黄色或白色无结构区，蜂窝状色素网；（5）玫瑰糠疹（图 1F）：外周分布的白色鳞屑（“领圈”征），黄色背景，不规则分布或簇集分布的点状血管和线状血管。

1.2.2 模型构建及性能验证

采用五倍裁剪法对训练集进行数据扩充，在像素为 256×256 皮肤镜图像的左上、左下、右上、右下角以及中心剪裁出像素为 224×224 的图像，使图像数目扩展为原来的 5 倍。使用训练集病例皮肤镜图像对 6 种经典网络结构利用迁移学习进行训练，以构建卷积神经网络（convolutional neural networks, CNN）二分类模型。6 种网络结构分别为 AlexNet<sup>[3]</sup>、VGG16<sup>[4]</sup>、ResNet18<sup>[5]</sup>、DenseNet121<sup>[6]</sup>、SENet<sup>[7]</sup> 以及 EfficientNet-B0<sup>[8]</sup>，均为 2012 年以来的经典 CNN 结构，常用于皮肤镜图像的分类。二分类模型的构建采用 Logistic 回归法，使用交叉熵函数作为损失函数。采用 5 折交叉验证法对 CNN 二分类模型进行训练。训练过程中使用迁移学习法将在 ImageNet 数据集中预训练的模型参数在训练开始时对 CNN 二分类模型进行初始化，以降低因数据量较小导致的过拟合现象，提高 CNN 二分类模型的泛化能力。

采用测试集数据对 CNN 二分类模型的性能进行验证。6 种网络结构均选取经 5 折交叉验证后的最佳模型（每种网络结构均有 1 个最佳模型），并取 6 个最佳模型的均值为各项指标的最终结果：（1）对每幅皮肤镜图像进行鉴别诊断；（2）对所有病例的皮肤镜图像进行判读，将网络分类层输出归一化后取均值，对每例病例进行疾病归类。

1.2.3 皮肤科医师诊断标准

从测试集每例病例的皮肤镜图像中随机挑选 1 幅图像，并结合皮损的临床图像制成网络问卷，由 13 名有经验的皮肤科医师在未知 CNN 二分类模型分类、病理结果及其他临床特征的情况下对测试集病例的皮



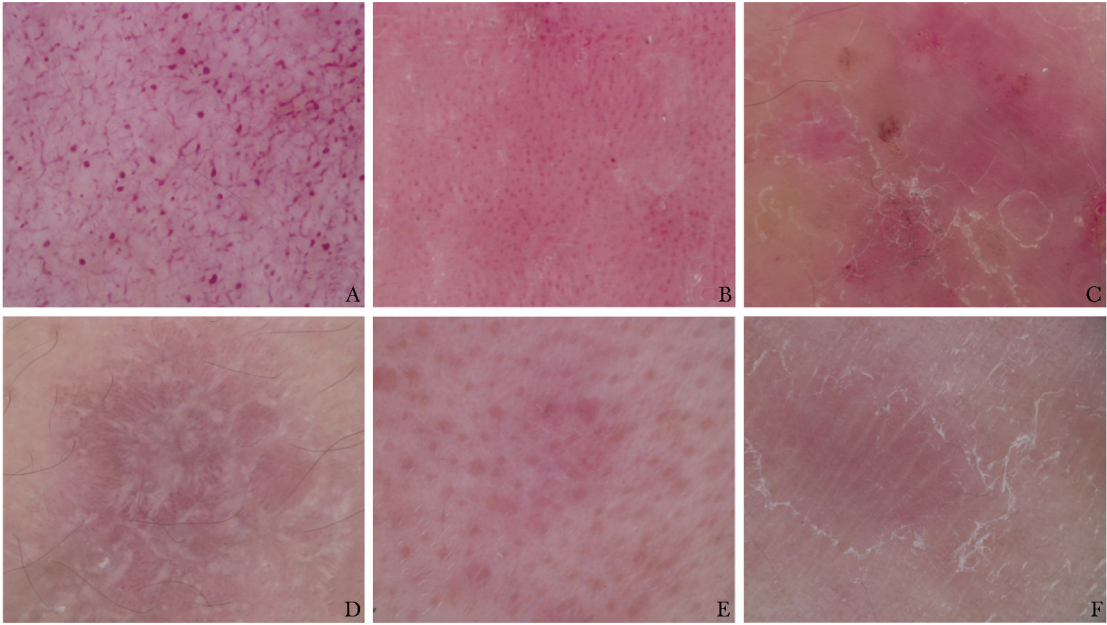


图 1 早期蕈样肉芽肿及炎症性皮肤病皮肤镜图像

A. 蕈样肉芽肿; B. 银屑病; C. 慢性湿疹; D. 扁平苔藓; E. 脂溢性皮炎; F. 玫瑰糠疹

肤镜图像进行诊断。13 名皮肤科医师中，高级职称 2 名，中级职称 2 名，初级职称 9 名。所有参与诊断的皮肤科医师均经过皮肤镜培训且通过考核，同时具有 1 年以上皮肤镜判读经验。

1.3 统计学处理

采用 SPSS 23.0 软件进行统计学分析。CNN 二分类模型交叉验证评估结果以均数±标准差表示。计算 CNN 二分类模型、皮肤科医师的诊断准确度、灵敏度、特异度、阳性似然比、阴性似然比等指标，并采用独立样本 *t* 检验进行组间比较。采用 Kappa 值评估 CNN 二分类模型与皮肤科医师组内诊断结果的一致性。其中 Kappa 值>0.75 表示组内诊断结果的一致性较高，0.40~0.75 表示一致性中等，<0.40 表示一致性较差<sup>[9]</sup>。采用受试者工作特征（receiver operating characteristic，ROC）曲线评估 CNN 二分类模型与皮肤科医师在早期 MF 与炎症性皮肤病中的鉴别诊断价值。其中曲线下面积（area under the curve，AUC）越接近于 1，表明模型对二者鉴别诊断的效果越好。取双侧检验，以 *P*<0.05 为差异具有统计学意义。

2 结果

2.1 一般临床资料

共纳入 48 例早期 MF 患者（皮肤镜图像 402

幅）和 96 例炎症性皮肤病患者（皮肤镜图像 557 幅）。其中训练集 117 例（皮肤镜图像 772 幅），测试集 27 例（皮肤镜图像 187 幅）。训练集中，MF 患者 40 例，炎症性皮肤病患者 77 例（银屑病 20 例、慢性湿疹 20 例、扁平苔藓 17 例、脂溢性皮炎 10 例、玫瑰糠疹 10 例）；测试集中，MF 患者 8 例，炎症性皮肤病患者 19 例（银屑病 5 例、慢性湿疹 5 例、扁平苔藓 4 例、脂溢性皮炎 3 例、玫瑰糠疹 2 例）。研究流程见图 2。

2.2 卷积神经网络二分类模型交叉验证评估结果

训练集 117 例病例的 772 幅皮肤镜图像，经五倍剪法扩展共获得 3860 幅皮肤镜图像。表 1 为训练集 6 种基于 CNN 的网络结构经 5 折交叉验证法训练后对早期 MF 和炎症性皮肤病进行二分类的统计指标，可知 EfficientNet-B0 网络在交叉验证中获得了较低的阴性似然比，较高的阳性似然比、灵敏度、特异度以及准确度，在早期 MF 的识别中有较好的分类性能。

2.3 卷积神经网络二分类模型及皮肤科医师诊断结果比较

测试集中，皮肤科医师鉴别诊断早期 MF 与炎症性皮肤病的灵敏度和特异度分别为 70.19% 和 94.74%，Kappa 值为 0.677。按图像分类时，CNN 二分类模型对早期 MF 与炎症性皮肤病鉴别诊断的 AUC 为 0.87，灵敏度和特异度分别为 75.02% 和 82.02%，

Kappa 值为 0.563。按病例分类时，CNN 二分类模型对早期 MF 与炎症性皮肤病鉴别诊断的 AUC 为 0.97，灵敏度和特异度分别为 87.50%和 93.85%，Kappa 值为 0.920，见表 2。

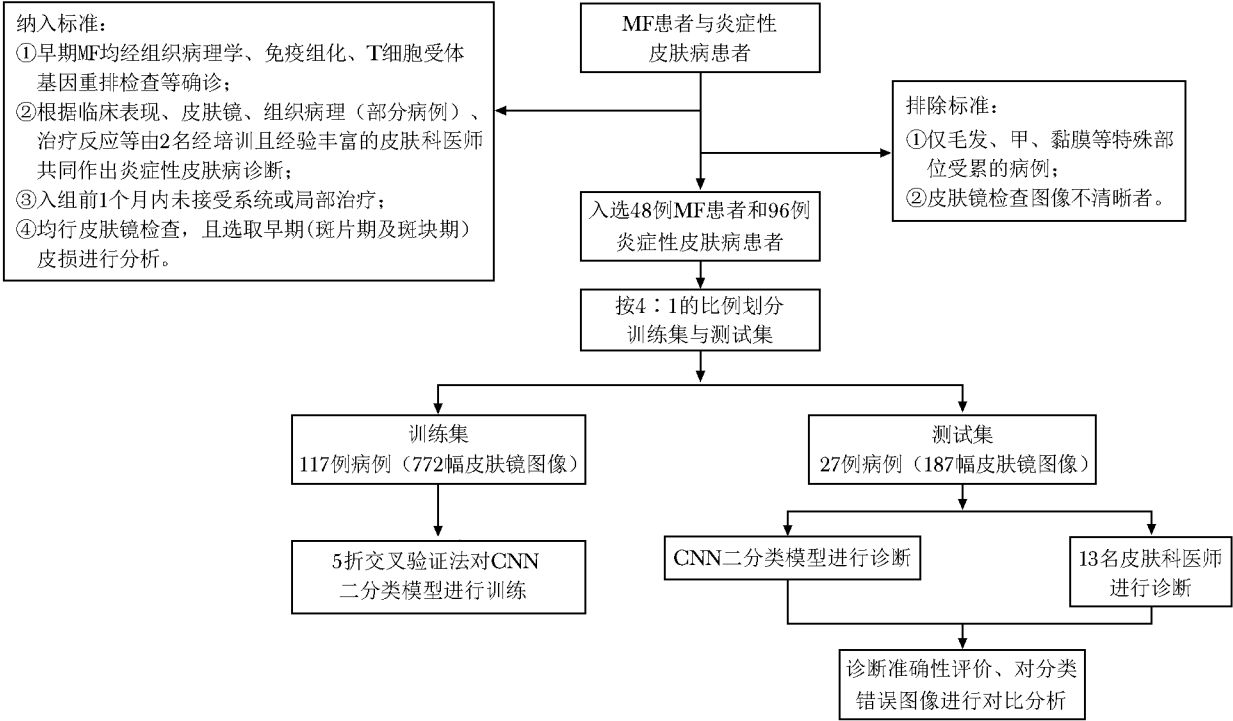


图 2 研究流程图  
MF：蕈样肉芽肿；CNN：卷积神经网络

表 1 CNN 二分类模型交叉验证评估结果 (x̄±s)

网络结构	阳性似然比	阴性似然比 (%)	灵敏度 (%)	特异度 (%)	准确度 (%)	阳性预测值 (%)	阴性预测值 (%)
AlexNet	4.59±2.62	40.51±5.84	66.72±3.62	82.60±6.52	75.82±4.59	75.25±6.69	76.42±4.19
VGG16	3.59±1.06	38.55±8.77	69.58±6.12	79.36±5.74	75.02±4.60	72.26±4.51	77.30±5.78
ResNet18	4.03±2.16	32.25±8.45	74.76±7.49	77.92±8.80	76.60±4.44	73.06±7.10	80.36±5.36
SENet	3.41±1.14	52.24±11.18	57.36±8.17	81.94±5.06	71.26±5.04	70.76±6.90	71.68±5.45
DenseNet121	3.00±0.80	44.21±8.04	65.88±8.31	76.40±8.66	71.98±3.16	68.76±4.01	74.94±4.50
EfficientNet-B0	4.48±2.40	33.47±5.84	72.82±5.54	80.96±6.96	77.48±3.57	75.18±6.54	79.68±3.97

CNN：同图 2

表 2 CNN 二分类模型与皮肤科医师诊断结果比较 [均值 (95% CI)]

指标	皮肤科医师 (n=27)	CNN 二分类模型	
		按图像分类 (n=187)	按病例分类 (n=27)
阳性似然比	NA	4.32 (3.61~5.02)	NA
阴性似然比 (%)	31.87 (20.46~43.28)	30.52 (24.56~36.48)	17.54 (8.34~26.77) <sup>#</sup>
AUC	/	0.87 (0.84~0.89)	0.97 (0.95~0.99)
灵敏度 (%)	70.19 (59.68~80.70)	75.02 (70.19~79.85)	87.50 (78.55~96.45) <sup>#</sup>
特异度 (%)	94.74 (91.77~97.71)	82.02 (79.30~84.87) <sup>#</sup>	93.85 (88.93~98.77)
准确度 (%)	87.46 (83.32~91.60)	79.52 (76.87~82.16)	91.98 (88.52~95.44)
Kappa 值	0.677 (0.566~0.789)	0.563 (0.507~0.620)	0.920 (0.884~0.954) <sup>#</sup>
阳性预测值 (%)	85.83 (77.57~94.09)	70.60 (67.09~74.11) <sup>#</sup>	87.43 (78.55~96.45)
阴性预测值 (%)	88.68 (85.07~92.29)	85.23 (82.77~87.70) <sup>#</sup>	94.93 (91.43~98.44)

CNN：同图 2；AUC：曲线下面积；NA：诊断结果中存在特异度为 100%的情况，未能估计阳性似然比；/：无 AUC；<sup>#</sup>与皮肤科医师诊断结果比较差异有统计学意义

2.4 卷积神经网络二分类模型与皮肤科医师诊断早期蕈样肉芽肿的效能比较

选取 6 种 CNN 二分类模型中对早期 MF 诊断效能最好的 EfficientNet-B0 网络结构进行可视化分析。ROC 曲线显示,按病例分类时,CNN 二分类模型诊断早期 MF 的 AUC 为 0.99,灵敏度和特异度分别为 88.9%和 100%,且 13 名皮肤科医生的诊断灵敏度和特异度均值对应点位于曲线右下方,提示该模型对早期 MF 的分类效果较好,且优于皮肤科医师的平均水平,见图 3。

2.5 分类错误的图像分析

图 4 为 CNN 二分类模型误诊的 1 例早期 MF 患者(男性,30 岁)的皮肤镜图像及临床图像。在采用 CNN 二分类模型进行单幅图像鉴别诊断时,4 幅图像中的 3 幅(图 4A~4C)诊断错误,1 幅(图 4D)诊断正确。皮肤科医生选择图 4B 进行诊断,其诊断准确率为 76.92%。

皮肤科医师对 6 例早期 MF 病例的诊断准确度较低(0~76.92%,对其余患者的诊断准确度为 92.31%~100%)。其中对 1 例早期 MF 病例的诊断准确度为 0(图 5A),3 例为 69.23%(图 5B~5D),2 例为 76.92%。6 种 CNN 二分类模型对上述 6 例病例的平均诊断准确度分别为 100%(皮肤科医师:0)、100%(皮肤科医师:69.23%)、83.33%(皮肤科医师:69.23%)、100%(皮肤科医师:69.23%)、50.00%(皮肤科医师:76.92%)、100%(皮肤科医师:76.92%)。

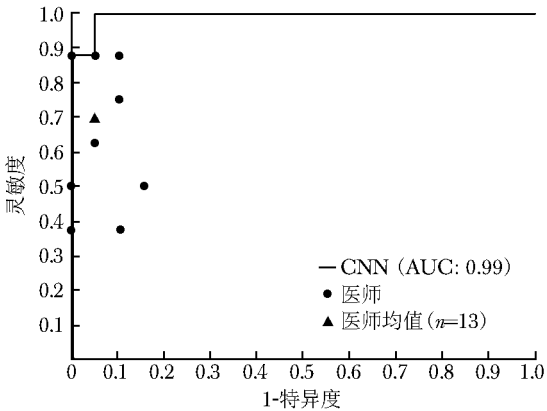


图 3 CNN 二分类模型诊断早期 MF 的受试者工作特征曲线及与皮肤科医师诊断结果比较  
圆点为 13 名皮肤科医师的诊断灵敏度和特异度(存在部分医师诊断数据一致,圆点重合的情况),▲为其均值。当圆点位于曲线右下方时,说明 CNN 二分类模型的诊断效能优于皮肤科医师平均水平  
MF、CNN; 同图 2; AUC: 同表 2

3 讨论

MF 是最常见的 CTCL,早期 MF 易被误诊为慢性湿疹、银屑病、玫瑰糠疹、脂溢性皮炎、扁平苔藓等炎症性皮肤病,通过影像学检查进行评估可提高 CTCL 检出率,并避免不必要的活检操作,具有重要的临床应用价值。本研究基于皮肤镜图像,探究 CNN

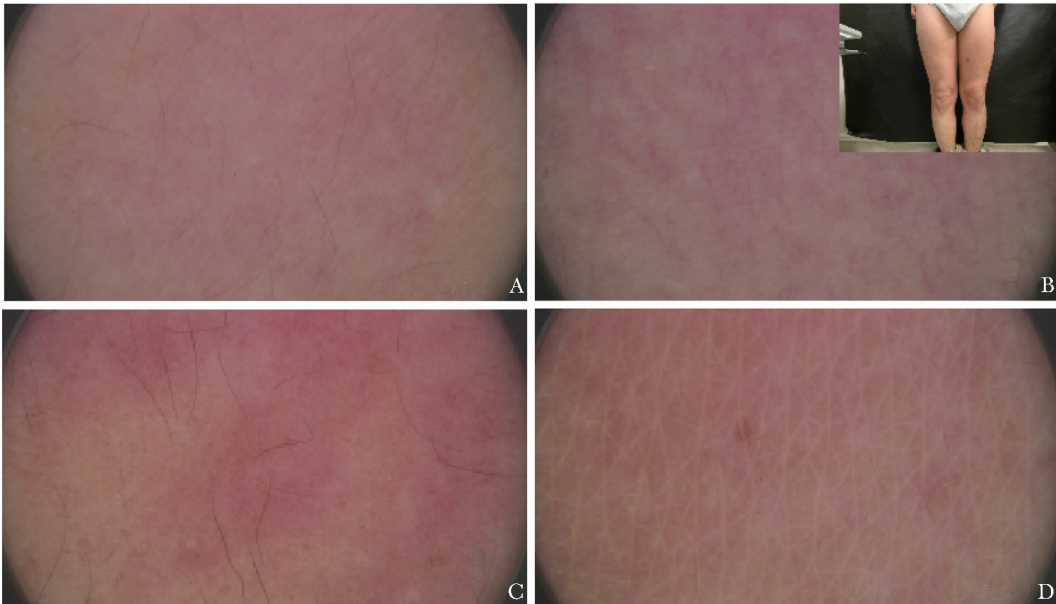


图 4 CNN 二分类模型误诊的 1 例早期 MF 病例的皮肤镜图像及对应的临床图像  
MF、CNN; 同图 2



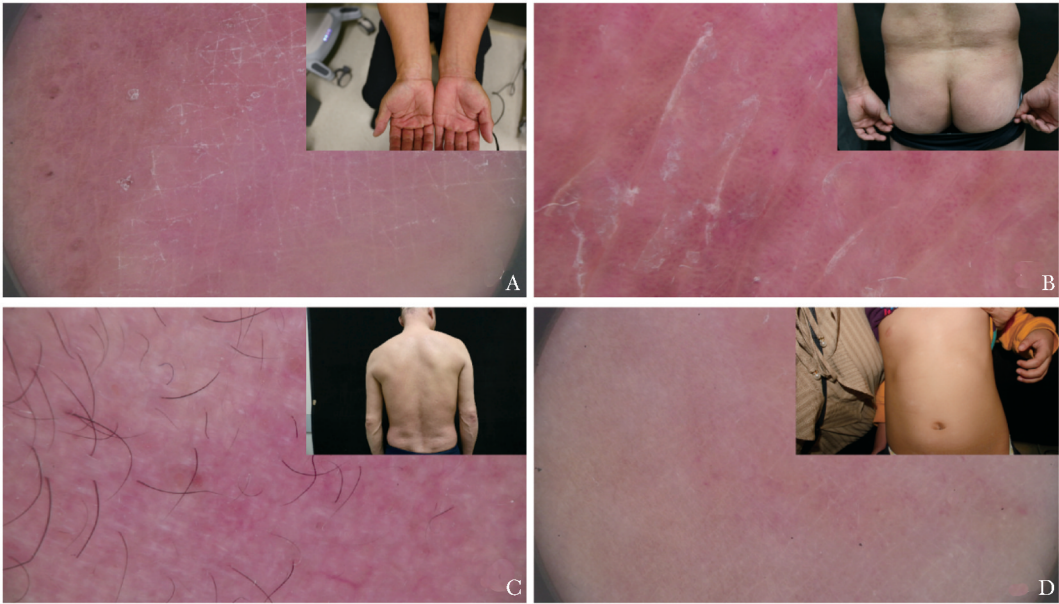


图 5 皮肤科医生诊断准确率较低的早期 MF 病例皮肤镜图像及对应的临床图像  
A. 皮肤科医师的诊断准确度为 0；B~D. 皮肤科医师的诊断准确度均为 69.23%  
MF：同图 2

二分类模型对早期 MF 与炎症性皮肤病的鉴别诊断能力，结果显示，测试集中皮肤科医师鉴别诊断早期 MF 与炎症性皮肤病的灵敏度和特异度分别为 70.19% 和 94.74%，Kappa 值为 0.677。按图像分类时，CNN 二分类模型对早期 MF 与炎症性皮肤病鉴别诊断的 AUC 为 0.87，灵敏度和特异度分别为 75.02% 和 82.02%，Kappa 值为 0.563。按病例分类时，CNN 二分类模型对早期 MF 与炎症性皮肤病鉴别诊断的 AUC 为 0.97，灵敏度和特异度分别为 87.50% 和 93.85%，Kappa 值为 0.920，提示该模型对早期 MF 具有较好的识别能力，尤其按病例分类时，灵敏度、特异度均较满意，且不同网络结构模型之间，诊断结果的一致性较高。进一步 ROC 曲线分析显示，按病例分类时，EfficientNet-B0 网络结构 CNN 二分类模型诊断早期 MF 的 AUC 为 0.99，灵敏度和特异度分别为 88.9% 和 100%，亦提示 CNN 二分类模型对早期 MF 的分类效果较好，且优于皮肤科医师的平均水平。

3.1 皮肤镜图像的人工诊断

皮肤镜作为一种无创的影像学检查手段，具有便捷、实时、患者接受度高等优点。Lallas 等<sup>[10]</sup> 和 Ghahramani 等<sup>[11]</sup> 分别通过回顾性研究发现早期 MF 最常见的皮肤镜特征为短线状血管、橘黄色斑片状区域和“精子样”血管，其中“精子样”血管为早期 MF 的特征性血管结构。本课题组于 2019 年回顾性分

析了北京协和医院皮肤科诊治的 31 例早期 MF 患者的皮肤镜特征，并与慢性湿疹和银屑病患者患者的皮肤镜表现进行比较，结果显示出现线状血管（灵敏度：90.3%，特异度：92.9%）、“精子样”血管（灵敏度：74.2%，特异度：100%）以及橘黄色斑片状区域（灵敏度：90.3%，特异度：91.4%）等皮肤镜表现高度提示皮损为早期 MF<sup>[12]</sup>。Bilgic 等<sup>[13]</sup> 对 MF、斑块状银屑病、扁平苔藓、玫瑰糠疹和结节性痒疹 5 种皮肤病的皮肤镜图像表现进行总结后发现，玫瑰糠疹皮肤镜下可见黄色背景，其他皮损背景以亮红色和暗红色为主；斑块状银屑病、玫瑰糠疹和结节性痒疹主要表现为点状血管，MF 和扁平苔藓中以点状和线状血管为主；在血管分布上，斑块状银屑病呈规则分布，扁平苔藓呈周边分布，玫瑰糠疹、MF 和结节性痒疹呈斑片状分布；在鳞屑颜色上，斑块状银屑病、玫瑰糠疹、MF 以白色鳞屑为主，扁平苔藓、结节性痒疹以黄白色鳞屑为主；在鳞屑分布上，除玫瑰糠疹的鳞屑呈外周“领圈样”分布外，其余疾病的鳞屑分布均呈斑片状，提示皮肤镜可用于评估并鉴别诊断 MF 与炎症性皮肤病。但准确识别皮肤病图像对医生的皮肤镜判读能力有较高要求，一般需经系统培训和长期临床实践，且人工判读皮肤镜图像存在主观差异性。

3.2 皮肤镜图像智能分析

随着人工智能（artificial intelligence，AI）深度

学习技术的不断发展,借助皮肤镜图像计算机辅助诊断(computer aided diagnosis, CAD)系统可客观地对多种皮肤科疾病进行诊断,且诊断结果具有较高的重复性<sup>[14]</sup>。Schindewolf 等<sup>[15]</sup>将 CAD 系统应用于皮肤恶性肿瘤的协助诊断中,发现恶性黑色素瘤的诊断准确度由 75%提高至 92%。但传统皮肤镜图像分类模型常使用人工设计的特征,不具有高层语义信息,无法准确描述图像。CNN 能利用卷积运算提取图像的高层语义特征,近年来被应用于皮肤镜图像的诊断分类中。谢斌等<sup>[16]</sup>利用 CNN 算法成功构建皮肤疾病诊断模型,该模型对基底细胞癌和色素痣分类的正确率为 93.5%,证实 CNN 模型在皮肤疾病的诊断中具有巨大潜力。Serener 等<sup>[17]</sup>基于 CNN 算法构建常见恶性色素性皮肤病图像分类模型,可实现基底细胞癌、鳞状细胞癌和角质形成细胞肿瘤的分类。

本研究对训练集数据进行 5 折交叉验证以比较不同网络结构的 CNN 二分类模型对早期 MF 与炎症性皮肤病疾病的分类性能。结果显示 6 种网络结构的 CNN 二分类模型均取得了较好的分类准确度、灵敏度和特异度,尤以 EfficientNet-B0 网络结构的综合能力最佳,其基于复合系数,通过平衡网络宽度、深度优化网络结构,在皮肤镜图像分类任务中表现出良好的能力。

本研究结果显示,按病例分类时, CNN 二分类模型在测试集中对 MF 与炎症性皮肤病鉴别诊断的 AUC 为 0.97,灵敏度和特异度分别为 87.50%和 93.85%, Kappa 值为 0.920。而皮肤科医师鉴别诊断的灵敏度和特异度分别为 70.19%和 94.74%, Kappa 值为 0.677。提示 CNN 二分类模型在早期 MF 的分类中具有较高的识别能力,且在不同的网络结构模型间达到了高度一致性。进一步对基于 EfficientNet-B0 网络结构的 CNN 二分类模型进行 ROC 曲线分析,结果表明该模型诊断早期 MF 的 AUC 高达 0.99,且灵敏度和特异度优于 13 名皮肤科医师的均值,亦说明 CNN 二分类模型对早期 MF 的识别能力优于皮肤科医生整体诊断水平。

在对单幅图像识别能力的比较中, CNN 二分类模型的灵敏度为 75.02%,略高于皮肤科医师(70.19%),但无统计学差异,提示 CNN 二分类模型根据单幅皮肤镜图像识别早期 MF 的灵敏度已达临床医生水平。但其特异度、准确度、Kappa 值、阳性预测值、阴性预测值均低于皮肤科医师。分析原因:单幅皮肤镜图像所含信息较少,而皮肤科医师可综合皮肤镜图像和临床图像的特征,获得更多的诊断信息与

依据。数据量是影响 CNN 学习能力的关键因素,由于 MF 较罕见,本研究纳入的患者较少,以致 CNN 二分类模型的分类能力降低,即便如此, CNN 二分类模型在单幅图像早期 MF 的识别中 AUC 亦达 0.87,提示其对早期 MF 具有较高的分类能力。

### 3.3 分类错误的图像分析

在对错误分类病例的皮肤镜图像分析中发现, CNN 二分类模型对其中 1 例病例 4 幅图像中的 3 幅图像分类错误,可能与该患者的 MF 皮损数目较少、局限性分布、血管结构不明显,与慢性湿疹、银屑病等炎症性皮肤病表现极为相似有关。而 13 名皮肤科医师对 6 例早期 MF 患者的诊断准确度较低,其中 1 例病例均误诊,考虑与该患者皮损呈局限性分布,皮肤镜图像为非浸润模式拍摄、血管结构显示不清晰有关。对于该病例,6 种网络结构 CNN 二分类模型的诊断准确度均为 100%。皮肤科医师诊断准确度较低的其余 5 例病例中, CNN 二分类模型的诊断准确度整体亦高于皮肤科医师,提示多点取材、同一病例提供多张皮肤镜图像可提高 AI 的诊断准确度,达到与临床医生相当甚至更高的诊断水平。

本研究局限性:(1) MF 为罕见病,发病率较低,本研究收集的病例数较少,深度学习模型从原始数据中获取的信息较少,造成模型过拟合现象。此外,本研究参与皮肤镜图像诊断的医师均为经北京协和医院培训、有 1 年以上皮肤镜判读经验的皮肤科医师,其诊断能力不能代表全国医师的水平。(2) 为符合临床实践情况,皮肤科医师在进行早期 MF 分类鉴别时,为其提供了与皮损对应的临床图像以辅助诊断,而 CNN 二分类模型仅提供皮肤镜图像,两者之间存在操作差异。(3) 本研究测试集病例仍来自本研究中心,仅对模型的性能进行内部验证。因此, CNN 二分类模型对早期 MF 的识别能力仍需多中心、大样本量数据研究加以证实,并进行外部数据验证。

综上,基于深度学习的 CNN 辅助诊断模型具有学习耗时短、受环境因素干扰少、可重复性高等优点,可广泛应用于多种学科疾病的诊断。已有研究表明,应用医生结合 AI 辅助诊断模型可提高疾病的诊断准确度<sup>[18]</sup>。在临床实践中,除皮肤影像检查外,病史、病程、组织病理、实验室检查、治疗反应等信息均可可为医生作出诊疗决策提供重要信息。尽管本研究中 CNN 二分类模型对早期 MF 已取得较好的识别、分类能力,但其仅依赖于皮肤镜图像信息的诊断方法与临床诊断过程并不完全相符。未来,本研究团队将尝试在网络结构中纳入不同模态的临床信息,并构建



基于多模态的 CNN 模型,以提高对不同皮肤病的分类准确度,使 AI 诊断模型更接近临床实际并具有充分合理的依据,从而进一步用于临床疾病的辅助诊断,提高诊断准确度。

**作者贡献:** 刘兆睿和张漪澜负责研究实施和论文撰写; 刘洁和谢凤英负责研究设计和论文审校。

**利益冲突:** 无

参 考 文 献

[1] 刘洁, 邹先彪. 实用皮肤镜学 [M]. 北京: 人民卫生出版社, 2021: 1-249.

[2] Olsen E, Vonderheid E, Pimpinelli N, et al. Revisions to the staging and classification of mycosis fungoides and Sezary syndrome: a proposal of the International Society for Cutaneous Lymphomas (ISCL) and the cutaneous lymphoma task force of the European Organization of Research and Treatment of Cancer (EORTC) [J]. Blood, 2007, 110: 1713-1722.

[3] Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks [C]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.

[4] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv, 2014: 1409. 1556V6.

[5] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[6] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.

[7] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[8] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning, 2019: 6105-6114.

[9] Hunt RJ. Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability [J]. J Dent Res, 1986, 65: 128-130.

[10] Lallas A, Apalla Z, Lefaki I, et al. Dermoscopy of early stage mycosis fungoides [J]. J Eur Acad Dermatol Venereol, 2013, 27: 617-621.

[11] Ghahramani GK, Goetz KE, Liu V. Dermoscopic characterization of cutaneous lymphomas: a pilot survey [J]. Int J Dermatol, 2018, 57: 339-343.

[12] Xu C, Liu J, Wang T, et al. Dermoscopic patterns of early-stage mycosis fungoides in a Chinese population [J]. Clin Exp Dermatol, 2019, 44: 169-175.

[13] Bilgic SA, Cicek D, Demir B. Dermoscopy in differential diagnosis of inflammatory dermatoses and mycosis fungoides [J]. Int J Dermatol, 2020, 59: 843-850.

[14] 谢凤英, 刘洁, 崔勇, 等. 皮肤镜图像计算机辅助诊断技术 [J]. 中国医学文摘 (皮肤科学), 2016, 33: 45-50.

Xie FY, Liu J, Cui Y, et al. Computer aided diagnosis of dermoscopic images [J]. Zhongguo Yixue Wenzhai (Pifu Kexue), 2016, 33: 45-50.

[15] Schindewolf T, Stolz W, Albert R, et al. Classification of melanocytic lesions with color and texture analysis using digital image processing [J]. Anal Quant Cytol Histol, 1993, 1: 1-11.

[16] 谢斌, 何小宇, 黄伟红, 等. 基于卷积神经网络的基底细胞癌和色素痣的临床图像鉴别 [J]. 中南大学学报 (医学版), 2019, 44: 113-120.

Xie B, He XY, Huang WH, et al. Clinical image identification of basal cell carcinoma and pigmented nevi based on convolutional neural network [J]. Zhongnan Daxue Xuebao (Yixueban), 2019, 44: 113-120.

[17] Serener A, Serte S. Keratinocyte carcinoma detection via convolutional neural networks [C]. 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2019: 1-5.

[18] Pangti R, Mathur J, Chouhan V, et al. A machine learning - based, decision support, mobile phone application for diagnosis of common dermatological diseases [J]. J Eur Acad Dermatol Venereol, 2021, 35: 536-545.

(收稿: 2021-06-26 录用: 2021-08-02)  
(本文编辑: 董 哲)