

基于深度学习及改进模糊KMeans的寻常型银屑病智能诊断方法

石丽平¹, 杜笑青², 李静¹, 刘丽娟¹, 张国强¹

1. 河北医科大学第一医院皮肤科, 河北 石家庄 050000; 2. 中国人民解放军联勤保障部队第九八〇医院皮肤科, 河北 石家庄 050000

【摘要】为了解决寻常型银屑病在样本分布不平衡的数据中可能会导致的深度学习模型诊断效果下降等问题,通过结合改进模糊KMeans聚类算法对高聚类复杂度数据的处理能力以及Visual Geometry Group 13(VGG13)深度卷积神经网络模型的预测能力,提出一种基于改进模糊KMeans聚类算法的VGG13深度卷积神经网络(VGG13-KMeans)模型,并将其应用于寻常型银屑病的诊断任务中。实验结果表明,相较于VGG13以及ResNet18两种方法,本文方法更适用于对银屑病特征的识别。

【关键词】寻常型银屑病;改进模糊KMeans聚类算法;VGG13;深度卷积神经网络模型

【中图分类号】R318;R758.63

【文献标志码】A

【文章编号】1005-202X(2024)02-0253-05

Intelligent diagnosis of psoriasis vulgaris based on deep learning and improved fuzzy KMeans

SHI Liping¹, DU Xiaoqing², LI Jing¹, LIU Lijuan¹, ZHANG Guoqiang¹

1. Department of Dermatology, the First Hospital of Hebei Medical University, Shijiazhuang 050000, China; 2. Department of Dermatology, the 980th Hospital of PLA Joint Logistic Support Force, Shijiazhuang 050000, China

Abstract: In order to address issues such as the decline in diagnostic performance of deep learning models due to imbalanced data distribution in psoriasis vulgaris, a VGG13-based deep convolutional neural network model is proposed by integrating the processing capability of the improved fuzzy KMeans clustering algorithm for highly clustered complex data and the predictive capability of VGG13 deep convolutional neural network model. The model is applied to the diagnosis of psoriasis vulgaris, and the experimental results indicate that compared with VGG13 and resNet18, the proposed approach based on deep learning and improved fuzzy KMeans is more suitable for identifying psoriasis features.

Keywords: psoriasis vulgaris; improved fuzzy KMeans clustering algorithm; Visual Geometry Group 13; deep convolutional neural network

前言

作为一种常见的慢性自身免疫性炎症性皮肤病,寻常型银屑病的主要症状是皮肤上出现红色、厚重、覆盖着银白色鳞屑的斑块^[1]。这些斑块通常出现在头皮、膝盖、肘部、背部和骨盆等部位。该种疾病具体病因尚不清楚,可能与遗传因素和免疫系统异常有关^[2]。免疫系统的异常活跃导致皮肤细胞的生长加速,使它们在数天内成熟,导致了皮肤细胞的过

度堆积,形成斑块和鳞屑。由于寻常型银屑病患者常常面临皮肤痒、疼痛、紧绷和灼热感等不适症状;同时一些寻常型银屑病的患者可能出现关节炎症状,称为银屑病性关节炎^[3],这会导致关节肿痛、活动受限和关节功能损害。因此,及时准确地对寻常型银屑病进行诊断有助于指导治疗、监测病情、提供心理支持,并为患者及其家族提供更好的管理和预防策略,对于患者的健康和生活质量改善至关重要。然而,寻常型银屑病的症状和体征在不同患者之间可能存在很大的变异性。有些患者可能只有轻微的斑块,而有些患者可能全身都受到影响。这种变异性可能导致在一些轻度病例中被忽视或误诊。同时,寻常型银屑病的症状与其他一些皮肤病如湿疹、药物反应等相似,这可能导致在初步诊断阶段存在

【收稿日期】2023-07-11

【基金项目】河北省医学科学研究课题计划(20190470, 20231294)

【作者简介】石丽平,硕士研究生,主治医师,研究方向:银屑病、玫瑰痤疮、黄褐斑等,E-mail: shily890@126.com

困难。此外,目前并没有特定的实验室检查手段或生物标记物能明确诊断寻常型银屑病。诊断主要依赖医生对患者的观察和临床经验,这在一定程度上可能存在主观性。再者,在某些情况下,确诊寻常型银屑病可能需要较长的时间和成本。这涉及到患者进行多次就诊、进行必要的检查和排除其他病因等,对患者来说可能是一种负担。而最后,在早期和轻度病例中,寻常型银屑病的症状可能不明显或容易被忽视,导致延误诊断和治疗的时机,使疾病进一步恶化。

传统的寻常型银屑病诊断方法主要包括病史询问^[4]、临床分析^[5]、生物检测^[6]等人工诊断方法,面临着效率低、易混淆、成本过高等一系列问题,在临床上有着很大的局限性。因此,为了解决上述问题,近年来国内外的学者们研究并提出了多种基于人工智能的高效寻常型银屑病诊断方法。在国外,Raj等^[7]提出一种改进的U-Net模型并将其应用于银屑病的诊断任务中,通过自动分割与测量患者皮肤照片中的病变图块,该模型可以准确识别并估计银屑病及其病变程度。Roslan等^[8]提出一种基于卷积神经网络的银屑病诊断方法,经过实验验证,该方法对多种常见银屑病有着令人满意的诊断准确率。Raj等^[9]提出一种基于深度学习的全自动单阶段框架,有助于根据银屑病患者不同身体区域的彩色图像客观诊断银屑病及其病变程度,同时借助迁移学习与图像分割模型,该方法取得了较为优秀的平均诊断准确率,能够简单有效地辅助医生进行诊断。Arunkumar等^[10]使用一种基于可分离卷积的MobileNet深度学习模型以保持原有诊断精度的基础上加快模型的推理速度,经过与多个深度学习模型的对比试验,该方法能够在较低的资源开销的基础上有着与其他模型相似的诊断效果,这有助于医生使用手持设备辅助银屑病的诊断。在国内,基于深度学习方法的银屑病诊断研究较少,主要有谢斌等^[11]提出的基于卷积神经网络的皮肤疾病鉴别方法以及胡晟等^[12]提出的基于U型结构深度学习模型的皮肤病分割与识别方法等。然而,由于银屑病的数据可能在不同数据集和不同医疗机构之间存在差异,包括不同人群、不同照明条件等,可能会导致数据集分布产生偏差,进而影响深度学习模型的诊断性能以及泛化能力^[13]。为解决该问题,在数据的预处理阶段使用聚类算法可能有助于帮助模型使用更加平衡的数据标签分布提升模型在推理过程中的性能以及泛化能力^[14]。Ito^[15]使用一种改进的KMeans算法对偏差数据进行聚类以降低数据分布的偏差。Yuan等^[16]提出一种改进KMeans聚类算法分析变压器故障的影响因素,并取得较好的效果。

根据以上国内外不同学者的研究成果,为了解决寻常型银屑病样本分布不平衡可能会导致的深度学习模型诊断效果下降等问题,本文提出一种基于改进模糊KMeans聚类算法的VGG13 (Visual Geometry Group 13)深度卷积神经网络模型,并将其应用于寻常型银屑病的诊断任务中,经过实验验证,聚类后的深度学习模型的准确率显著高于未聚类的深度学习模型。

1 方法

1.1 VGG13 深度卷积神经网络

VGGNet作为一种被广泛使用的深度卷积神经网络模型由牛津大学计算机视觉组研发^[17]。相较于传统的VGGNet所使用的浅层卷积网络结构,为了提升模型的推理效果,本文所使用的VGG13模型由10个卷积层以及3个全连接层组成,模型的结构如图1所示。本文所使用的模型的卷积结构由5个VGG块所组成,在同一个VGG块中的每个卷积层结构相同。卷积核的大小均为 3×3 ,其步长为1,使得每层矩阵的输入与输出维度均相同,有利于同一VGG块内多个卷积层的堆叠与复用。在每一个VGG块的结尾,使用大小为 2×2 的过滤器以2为步长进行最大减半池化操作,其目的为精简模型参数的同时提高模型的泛化能力。在最后的3个全连接层结构中,本文模型使用Dropout将神经元以0.5的概率随机失活来优化模型的过拟合程度。最后,使用SoftMax层将全连接层的输出转化为对银屑病诊断结果的概率密度。

1.2 改进模糊KMeans聚类算法

本文所提出的改进模糊KMeans算法在Krishna等^[18]所提出的通用KMeans聚类算法的基础上,提出一种改进的数据聚点的隶属度度量方法以及相应的聚点计算方法,其目的是为了更有效地处理相较于传统KMeans难以处理的高聚类复杂度数据^[19]。

高聚类复杂度数据是一种具有高度重叠聚类、聚类形状与数据密度不规则、噪点影响较大的复杂数据集^[20]。传统的KMeans聚类算法由于其二进制的聚点隶属度计算方式,使得数据点只能服从单一聚点,进而导致难以处理数据点和多个聚点关系近似的情况。为了解决上述问题,本文将传统的KMeans算法所使用的二进制聚点关系改进为能够反映数据点与多个聚点潜在隶属关系的浮点数隶属度矩阵。而为了根据改进的浮点数隶属度矩阵Affi计算聚点位置,本文提出一种改进的聚点计算公式,其公式如下:

$$\text{Cetr}_j = \frac{\sum_{i=1}^N (\text{Affi}_{ij}^m \cdot x_i)}{\sum_{i=1}^N \text{Affi}_{ij}^m} \quad (1)$$

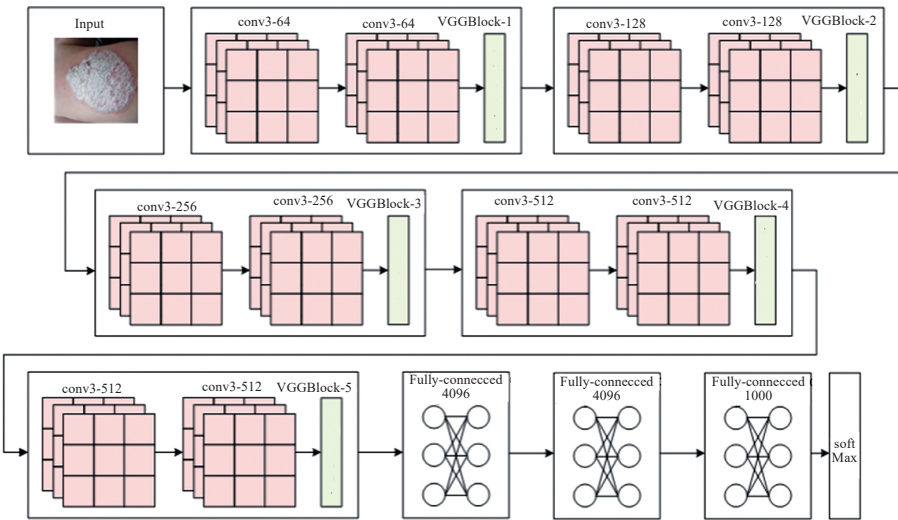


图1 VGG13结构图
Figure 1 VGG 13 structure

其中, $Cetr_j$ 是第 j 类的聚类中心, 矩阵元素 $Affi_{ij}$ 代表大小为 N 的数据集中第 i 个数据点对于第 j 类的隶属度值, x_i 表示第 i 个数据点, 模糊系数 m 作为外部参数, 表示模糊程度。

根据式(1), 通过 Lagrange 乘法, 本文提出一种聚点隶属度计算函数, 其公式如下:

$$Affi_{ij} = [\sum_{k=1}^{Cetr} (\frac{Dist_{ij}}{Dist_{kj}})^{\frac{2}{(m-1)}}]^{-1}$$

(2)

其中, 元素 $Dist_{ij}$ 代表数据点 i 与聚类中心 $Cetr_j$ 之间的欧式距离, 其公式如下:

$$Dist_{ij} = \|x_i - Cetr_j\|^2$$

(3)

改进模糊 KMeans 聚类算法的具体步骤即为反复迭代式(1)和式(2), 直到 $\max \{ \|Cetr - Cetr_{prev}\|^2 \} < \varepsilon$ 成立, 即每一个聚类中心的位置收敛。

1.3 基于深度学习及改进模糊KMeans聚类算法的寻常型银屑病智能诊断方法

本文所提出的基于深度学习及改进模糊 KMeans 的寻常型银屑病智能诊断方法流程如图 2 所示。本文方法首先需要对数据进行清洗、剪裁、变换以及归一化等预处理操作, 以尽可能地提升数据质量。然后, 将处理后的图像数据进行聚类, 获得聚类标签后将其与真实标签融合以建立数据分布与真实数据类别的映射关系。之后, 将融合后的图像使用 VGG13 深度卷积神经网络进行模型训练并调优以获取效果较优的识别模型。最后, 将待诊断的银屑病图像输入模型进行诊断, 得到其类别, 即完成完整的模型训练与诊断流程。

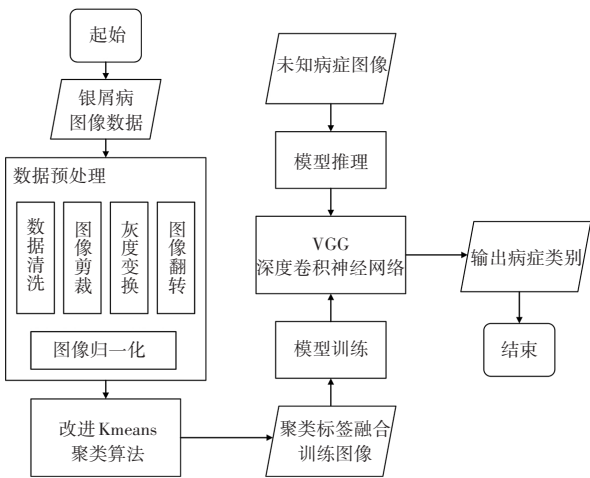


图2 本文方法流程图
Figure 2 Flowchart of the proposed method

2 结果

2.1 环境与参数

本文的代码是在 Windows 10 22H2 操作系统上构建的, 使用 Python 3 作为编程语言。具体而言, 笔者采用了 Python 的 Scikit-Learn 和 PyTorch 框架, 构建了 VGG13 神经网络结构, 并通过 CUDA 11.8 加速模型训练。

2.2 数据来源

实验使用的数据来源于河北医科大学第一医院皮肤科银屑病患者图像数据集。图像数据集共包含 89 张图像, 且所有图像均被缩放为 224×224 像素。部分患者图像数据如图 3 所示。



图3 部分患者图像数据
Figure 3 Some pictures

2.3 模型比较

为了验证基于深度学习及改进模糊 KMeans 聚类算法(VGG13-KMeans)模型的寻常型银屑病智能诊断方法的有效性和可行性,本文使用 VGG13 以及 ResNet18 与该方法进行比较。3 种方法得到的训练过程中的损失曲线如图 4 所示;3 种方法得到的准确率的对比结果如图 5 所示。

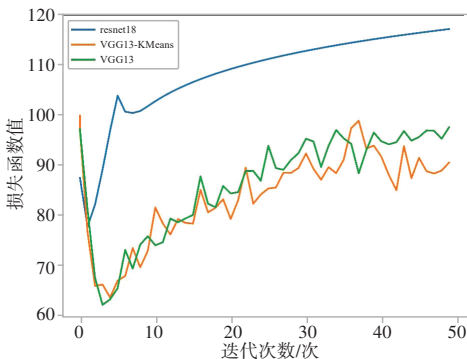


图4 不同识别方法损失值的对比结果
Figure 4 Comparison of the loss values among different methods

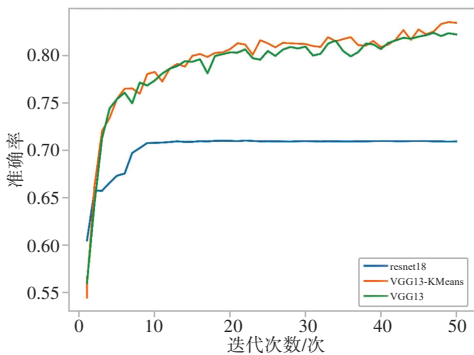


图5 不同识别方法准确率的对比结果
Figure 5 Comparison of the accuracy among different methods

通过图 4 的观察,随着迭代次数的增加,3 种方法得到的损失值逐渐减小,并在达到最低点后产生了一定程度的反弹。这说明随着训练的进行,3 种模型都出现了或多或少的过拟合现象。这一现象的出现可以解释为在训练过程中所使用的 L1 范数损失函数的影响。在高维数据处理中,基于 L_1 范数损失函数存在解的稀疏性,因此在损失函数值达到极小值后,模型会出现一定程度的反弹。相对于 VGG13 以及 ResNet18 方法,VGG13-KMeans 得到的损失更小,这表明其在模型训练中取得了更好的性能。由图 5 可知,当迭代次数增加时,3 种方法的准确率都在提升,但 ResNet18 方法得到的曲线很快趋于平稳。在训练准确率方面,VGG13-KMeans 与 VGG13 两种方法要远优于 ResNet18 方法;且 VGG13-KMeans 方法得到的曲线在 VGG13 方法得到的曲线之上,表明 VGG13-KMeans 方法优于 VGG13 方法。

对于识别任务,笔者比较了 3 种方法的准确率:采用 VGG13-KMeans 方法获得了 0.834 6 的准确率,VGG13 方法达到了 0.822 3 的准确率,而基于 resnet18 方法的准确率为 0.709 5。实验结果表明,与其他方法相比,VGG13-KMeans 方法在数据集中对银屑病特征的识别准确率最高,其次为 VGG13 方法,最后为 ResNet18 方法。综上所述,相较于 VGG13 以及 ResNet18 两种方法,本文提出的 VGG13-KMeans 方法更适用于对银屑病特征的识别。

3 结 论

本文针对寻常型银屑病智能诊断问题,提出一种基于深度学习及改进模糊 KMeans 聚类算法的模型,该模型在通用 KMeans 聚类算法的基础上,提出一种改进的数据聚点的隶属度度量方式以及相应的聚点计算方式,可以更有效地处理高聚类复杂度数据,同时结合深度卷积神经网络模型 VGG13 实现有效识别。将本文方法应用于银屑病患者图像数据集,实验结果表明,相较于 VGG13 以及 ResNet18 两种方法,本文方法更适用于对银屑病特征的识别。

【参考文献】

[1] 刘爱民,张步鑫,赵巍,等.寻常型银屑病病因病机探讨[J].北京中医药大学学报,2021,44(3): 266-271.
Liu AM, Zhang BX, Zhao W, et al. Exploring the etiology and pathogenesis of plaque psoriasis[J]. Journal of Beijing University of Traditional Chinese Medicine, 2021, 44(3): 266-271.

[2] Gisondi P, Bellinato F, Girolomoni G, et al. Pathogenesis of chronic plaque psoriasis and its intersection with cardio-metabolic comorbidities[J]. Front Pharmacol, 2020, 11: 117.

[3] 张恒,杨斌,李小花,等.高频超声对银屑病性关节炎诊断价值的研究进展[J].中华诊断学电子杂志,2022,10(1): 67-70.
Zhang H, Yang B, Li XH, et al. Research progress on the diagnostic

- value of high-frequency ultrasound in psoriatic arthritis[J]. Journal of Chinese Electronic Journal of Diagnosis, 2022, 10(1): 67-70.
- [4] 谢秀丽, 卢传坚, 姚丹霓. 基于德尔菲法的寻常型银屑病常见证候诊断指标研究[J]. 中医杂志, 2015, 56(16): 1409-1413.
- Xie XL, Lu CJ, Yao DN. Research on diagnostic index of common symptoms in psoriasis vulgaris based on delphi method[J]. Journal of Traditional Chinese Medicine, 2015, 56(16): 1409-1413.
- [5] 曹珊, 王娜, 颜潇潇, 等. 10例寻常型银屑病合并大疱性类天疱疮临床分析[J]. 中国麻风皮肤病杂志, 2019, 35(11): 643-646.
- Cao S, Wang N, Yan XX, et al. Clinical analysis of 10 patients with psoriasis vulgaris complicated with bullous pemphigoid[J]. Chinese Leprosy Skin Disease Journal, 2019, 35(11): 643-646.
- [6] 肖勤, 黄巍, 汪亚利, 等. 寻常型银屑病患者血清IL-4, IL-8, TNF- α 水平和T细胞亚群结构与预后转归的关系[J]. 河北医药, 2019, 41(22): 3376-3379.
- Xiao Q, Huang W, Wang YL, et al. Relationship between serum IL-4, IL-8, TNF- α , T lymphocyte subset and prognosis of patients with psoriasis vulgaris[J]. Hebei Medicine, 2019, 41(22): 3376-3379.
- [7] Raj R, Londhe ND, Sonawane RS. Deep learning based multi-segmentation for automatic estimation of psoriasis area score[C]// 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 2021: 1137-1142.
- [8] Roslan RB, Razly IN, Sabri N, et al. Evaluation of psoriasis skin disease classification using convolutional neural network[J]. IAES International Journal of Artificial Intelligence, 2020, 9(2): 349.
- [9] Raj R, Londhe ND, Sonawane R. PsLSNetV2: end to end deep learning system for measurement of area score of psoriasis regions in color images[J]. Biomed Signal Process Control, 2023, 79: 104138.
- [10] Arunkumar TR, Jayanna HS. A novel light-weight approach for the classification of different types of psoriasis disease using depth wise separable convolution neural networks[J]. Indian Journal of Science and Technology, 2022, 15(13): 561-569.
- [11] 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(9): 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]. Journal of Central South University (Medical Edition), 2019, 44(9): 113-120.
- [12] 胡晟, 田生伟, 禹龙. TUNet: 基于U型结构深度学习模型皮肤病图像分割研究[J]. 微电子学与计算机, 2022, 39(10): 71-79.
- Hu S, Tian SW, Yu L. TUNet: research on skin disease image segmentation based on U-shaped deep learning model [J]. Microelectronics & Computer, 2022, 39(10): 71-79.
- [13] Kim B, Kim H, Kim K, et al. Learning not to learn: training deep neural networks with biased data[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 9012-9020.
- [14] Petegrosso R, Li Z, Kuang R. Machine learning and statistical methods for clustering single-cell RNA-sequencing data[J]. Brief bioinform, 2020, 21(4): 1209-1223.
- [15] Ito A. Successive binary partition K-means algorithm-a clustering algorithm with less sample bias[J]. IEICE Tech Rep, 2022, 121(362): 37-42.
- [16] Yuan P, Wan T, Wu X, et al. Method of analyzing transformer DC magnetic bias based on big data clustering and relevance analysis[J]. IOP Conf Ser Earth Environ Sci, 2019, 310(3): 032027.
- [17] Tammina S. Transfer learning using VGG-16 with deep convolutional neural network for classifying images[J]. International Journal of Scientific and Research Publications (IJSRP), 2019, 9(10): 143-150.
- [18] Krishna K, Murty MN. Genetic K-means algorithm[J]. IEEE Trans Syst Man Cy B, 1999, 29(3): 433-439.
- [19] Askari S. Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: review and development[J]. Expert Syst Appl, 2021, 165: 113856.
- [20] Yuan C, Yang H. Research on K-value selection method of K-means clustering algorithm[J]. J-Multidiscip Sci, 2019, 2(2): 226-235.

(编辑:薛泽玲)