

# 基于深度学习的桡骨远端骨折自动分型研究

杨锋<sup>1</sup>, 丛日坤<sup>1</sup>, 王卫国<sup>2</sup>, 丁波<sup>1\*</sup>

<sup>1</sup> 山东中医药大学网络信息中心, 山东 济南 250355;

<sup>2</sup> 山东中医药大学第一临床医学院, 山东 济南 250355

**摘要** 为解决桡骨远端骨折内部病灶区域骨碎块多且不规则, 致使医生漏诊及误诊率高的问题, 利用课题组前期收集的临床桡骨远端骨折病例设计了一种监督式桡骨远端骨折自动分型的深度学习诊断模型。实验中还引入迁移学习思想, 提高了诊断模型的训练效率。最后采用交叉验证的方法对模型进行评估, 结果表明, 本文提出诊断模型的结果优于传统机器学习及经典深度学习分类模型, 分类准确率达到 84.2%, 较经典深度学习模型提升了 4% 左右, 且网络结构简单, 运算速度快, 具有一定鲁棒性和较强的泛化能力。

**关键词** 图像处理; 桡骨远端骨折; 预处理; 深度学习; 迁移学习; 过拟合

中图分类号 O436

文献标志码 A

doi: 10.3788/LOP202158.1210020

## Research on Automatic Classification of Distal Radius Fractures Based on Deep Learning

Yang Feng<sup>1</sup>, Cong Rikun<sup>1</sup>, Wang Weiguo<sup>2</sup>, Ding Bo<sup>1\*</sup>

<sup>1</sup> Network Information Center of Shandong University of Traditional Chinese Medicine, Jinan, Shandong 250355, China;

<sup>2</sup> The First Clinical College of Shandong University of Traditional Chinese Medicine, Jinan, Shandong 250355, China

**Abstract** In order to solve the problem that there are many and irregular bone fragments in the focal area of the distal radius fracture, which causes the doctor's missed diagnosis and high rate of misdiagnosis, this paper uses the clinical cases of distal radius fracture collected by the research group to propose a supervised automatic distal radius fracture deep learning model. The experiment also introduces the concept of migration learning, which improves the training efficiency of the diagnostic model. Finally, the experiment uses a cross-validation method to evaluate the model. The results show that the classification results of the proposed diagnostic model are better than traditional machine learning and classic deep learning classification models. The classification accuracy rate reaches 84.2%, which is 4% higher than the classic deep learning model. The network structure is simple, the calculation speed is fast, with certain robustness and strong generalization ability.

**Key words** image processing; distal radius fracture; pretreatment; deep learning; transfer learning; overfitting

**OCIS codes** 100.2000; 110.1650; 100.4996; 100.2960; 100.4999

## 1 引言

桡骨远端骨折(DRF)<sup>[1]</sup>主要分为关节内和关节外骨折,是指桡骨远端关节面近端 2~3 cm 以内的骨折,多为跌倒时手掌撑地或直接暴力导致。大量临床数据显示,DRF 严重影响腕关节的生理功能,

如不能根据患者骨折的具体情况,准确合理地选择治疗方案,将严重影响骨折部位的愈合及腕关节功能的恢复。DRF 的 AO 分型系统<sup>[2]</sup>是目前公认且常用的分型方法,实现 DRF 的正确分型对后续手术入路、固定方式的选择以及患者的预后效果评价有着很大的指导作用。但因腕部骨质结构复杂,骨碎

收稿日期: 2020-08-11; 修回日期: 2020-09-06; 录用日期: 2020-10-29

基金项目: 山东省社科规划研究项目(19CSHJ13)、教育部高等教育司产学研合作协同育人项目(201902205004)

\* E-mail: dingbo@126.com

块多且形状不一,存在重叠、旋转、嵌插等诸多复杂的骨折形式,对骨科医生判断 DRF 具体类型、骨碎程度及稳定性带来一定难度,极易造成误诊,从而影响后续骨折手术及复位方式的选择。

为提高 DRF 的诊断效率,解决当前临床漏诊及误诊率高的难题,设计自动检测骨折类型的诊断模型具有重要实用价值。当前,计算机辅助诊断技术(CAD)<sup>[3]</sup>在骨折方面的研究已取得重要进展,主要分为传统机器学习的诊断方法以及深度学习诊断方法。如 Yap 等<sup>[4]</sup>针对胫骨骨折图像特点提出了一种自动检测胫骨骨折的诊断模型。首先通过使用 sobel 算子定位胫骨骨折图像边界,然后利用灰度共生矩阵(GLCM)<sup>[5]</sup>算法提取骨折图像的局部纹理特征,后续利用传统机器学习分类器如 k 近邻(kNN)<sup>[6]</sup>、支持向量机(SVM)<sup>[7]</sup>、极限学习机(ELM)<sup>[8]</sup>完成胫骨骨折图像的自动分类,并达到了 78.2% 的骨折分类准确率。此外,深度学习在骨折诊断方面取得了一系列成功应用。Olczak 等<sup>[9]</sup>使用在非医学图像上预先训练好的卷积神经网络(CNN),并利用迁移学习将其用于普通腕部 X 射线图像上以实现骨折的自动检测。Szegedy 等<sup>[10]</sup>使用了初始 Inception V3 模型,训练数据采用放射科医生手动标记 1389 张肱骨 X 射线图像以重新训练 InceptionV3 网络顶部的初始层从而实现二分类问题。这证明了 CNN 模型如果已经在非医学图像上进行了预训练,可以成功地迁移应用于骨折诊断领域,且诊断的准确率超过了机器学习的诊断方法。尽管上述工作在骨折诊断过程中取得了不错的诊断效果,但过拟合的现象依然存在。为此, Kim 等<sup>[11]</sup>使用经过预训练的 ResNet-152 模型将其应用于桡骨骨折 X 射线图像。为了避免出现过拟合现象,作者首先对实验数据进行扩增。然后使用 Inception V3 网络迭代训练了 2 万次。除此之外,黄刚等<sup>[12]</sup>通过定位腕部骨折的目标区域来减少过拟合。作者首先使用目标检测算法从腕部的正位 X 射线图像中标记出感兴趣的区域(ROI)<sup>[13]</sup>,然后对 Inception V4 模型进行迭代训练,形成腕部骨折诊断模型,最后通过对比实验证明该模型的可行性,结果显示该诊断模型具有更高的骨折识别准确率。

尽管上述研究在骨折诊断领域取得了不同程度的进展,但仍存在一些问题。DRF 最常用的检查手段是 X 射线正侧位摄片,该方法对骨折线的具体走向、断端成角及位移具有直观呈现,但因其是二维成像,所以对关节面的塌陷、分离以及桡关节损伤程

度的情况显示欠佳。因此,研究一种高效的算法提取 DRF 图像的“病灶特征”并实现 DRF 的多分类诊断具有重要意义。CNN 的卷积层可以提取 DRF 图像的深层特征,且可通过适当增大网络层数及调节参数来降低特征冗余与特征表达能力。当前深度学习<sup>[14]</sup>模型凭借其在辅助脑、乳腺、肺等器官的临床诊断方面取得的重要进展成为医学图像处理的重要方法,但受限于数据集大小及类型,以及二维成像对 DRF 图像特征提取的限制等问题,当前对骨折诊断的研究还局限于二分类,在辅助骨折疾病的多分类诊断研究方面基本处于空白。本文利用课题组前期收集并由专业骨伤科医生手工标注的 DRF 临床图像数据,设计基于 DRF 自动分型的监督式深度卷积神经网络(DRF-Net),利用卷积层提取图像纹理、形状等传统特征和表达 DRF 深层语义的深度特征,然后将其进行特征融合形成全面 DRF 图像的特征表示。CNN 的训练通常需要大规模数据集,本文采用的 DRF 数据集太小容易造成过拟合,为解决上述问题,本文引入迁移学习的思想,提高了 DRF-Net 模型训练效率。

## 2 实验数据及预处理

### 2.1 实验数据来源

本文实验数据采用山东中医药大学附属医院和山东省中医院于 2014 年 3 月至 2020 年 3 月从临床采集的 DRF 图像数据,以及课题组前期积累的大量 DRF 临床诊疗病例和临床带教搜集的临床病例信息所建立的骨折图像样本库总共 1189 例样本。在 DRF 数据集中,每个 DRF 图像的具体类别均由有着 10 年以上工作经验的专业骨伤科专家医生手动标注。分型标准如下<sup>[15]</sup>:A 型,即关节外骨折,主要包括桡骨关节外侧的粉碎和嵌插型骨折,共 489 例;B 型,部分关节内骨折,分为桡骨远端矢状面、背侧缘及掌侧缘骨折,共 344 例;C 型,即复杂的完全关节内骨折,分为关节内粉碎性、无干骺端粉碎骨折,共 282 例。数据进行批量脱敏处理,包括去除涉及患者基本信息的姓名、身份识别号字段以及图像中的患者姓名拼音。图 1 给出 DRF 的 3 种具体骨折类型示例。

### 2.2 图像预处理

实验采用的 DRF 分类数据集均为 X 射线图像,X 射线图像具有亮度低且对比度差的问题,因此必须对 DRF 图像进行预处理。限制对比度自适应直方图均衡化算法(CLAHE)<sup>[16]</sup>能够提高 DRF 图

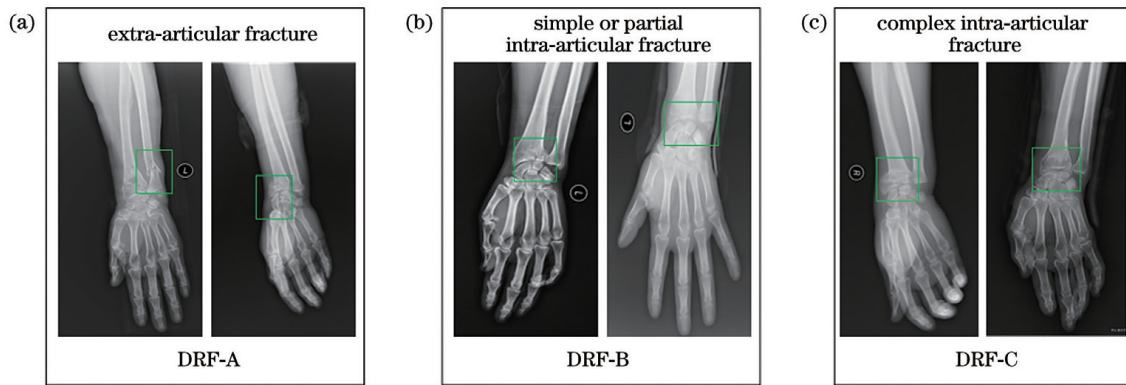


图 1 DRF 的 3 种具体骨折类型示例。(a) DRF-A; (b) DRF-B; (c) DRF-C

Fig. 1 Examples of 3 specific fracture types of DRF. (a) DRF-A; (b) DRF-B; (c) DRF-C

像亮度,增加图像对比度,还能抑制噪声,突出 DRF 图像的内部细节,从而更好地显示 DRF 图像的病灶信息。

CLAHE 算法的实现过程如下。

1) 图像分块,将单张 DRF 图像分为  $M \times N$  个单独的图像块。

2) 以块为单位,计算单个分割后图像块的灰度直方图,然后平均图像块的像素值到各个灰度级。灰度级中平均像素的数目可表示为

$$\bar{N}_a = \frac{N_{xp} - N_{yp}}{N_g}, \quad (1)$$

式中: $N_g$  为图像块中灰度级的数目; $N_{xp}$  和  $N_{yp}$  分别为图像块沿  $x$  轴和  $y$  轴方向的像素数目。

3) 截取灰度直方图中大于  $N_L$  的值, $N_L$  计算公式为

$$N_L = S\bar{N}_a, \quad (2)$$

式中: $S$  为截取系数。计算截取的像素数目以及像素总数平均到单个灰度级的像素数目。

4) 直方图均衡化。对经过步骤 3 得到图像块进行直方图均衡化处理,并利用变换函数求出经均衡化处理后的灰度值。图 2 表示 DRF 图像经过 CLAHE 算法处理的前后对比图。

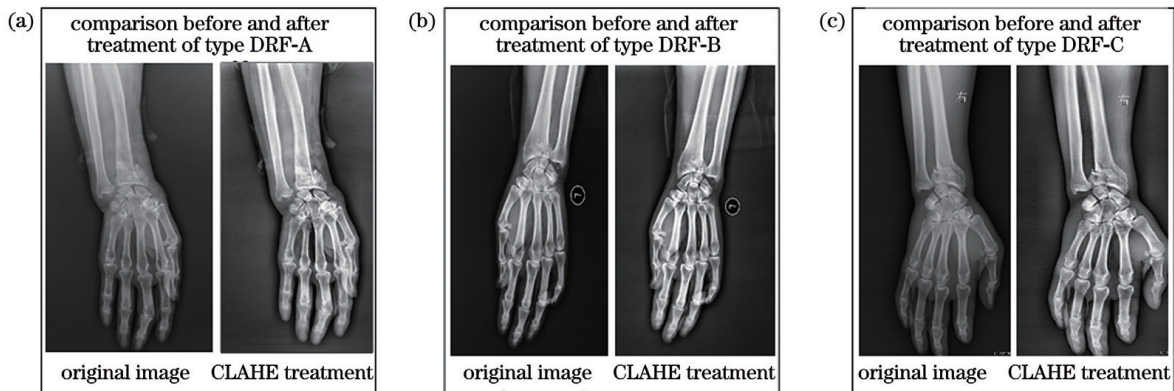


图 2 DRF 图像经过 CLAHE 算法处理的前后对比图。(a) DRF-A; (b) DRF-B; (c) DRF-C

Fig. 2 DRF images before and after processing by CLAHE algorithm. (a) DRF-A; (b) DRF-B; (c) DRF-C

### 3 方法介绍

#### 3.1 DRF-Net 模型

本文在文献[17]的基础上设计了一个 DRF 图像自动分型的模式识别模型,因其网络层数少、结构简单,所以具有诊断高效的优点。DRF-Net 多分类识别模型使得网络在提取 DRF 图像特征的同时,能够自动前向传播损失值,从而导致模型参数经多次迭代后调整到最优。

DRF-Net 模型结构共包括 3 层卷积层、3 层池化层、全连接层和输出层各 1 层,总共 8 层网络结构,DRF-Net 诊断模型的基本框架如图 3 所示。其中该诊断模型的优势在于其全连接层和输出层。针对 DRF 图像病灶目标区域小而集中,骨碎块多且不规则等复杂的骨折特点,本实验通过将 DRF-Net 模型提取的“分布式特征”利用全连接的方式映射到样本标记空间,从而更有效地利用 DRF-Net 模型提取的特征,提高 DRF 多分类准确率。



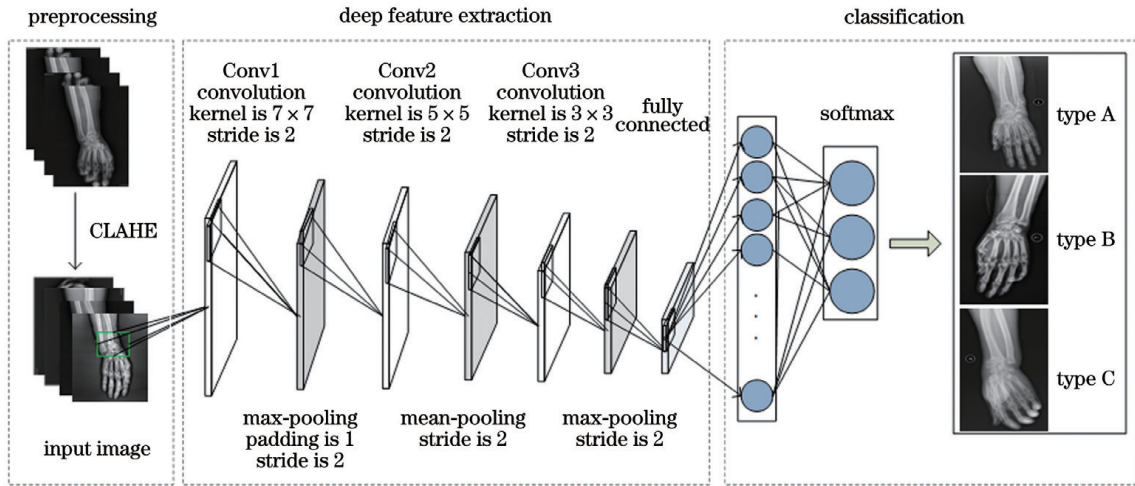


图 3 DRF-Net 模型主要结构

Fig. 3 Main structure of DRF-Net model

DRF-Net 模型主要结构及功能如下。

1) 输入层。在 DRF-Net 诊断模型中输入经预处理后的 DRF 图像。

2) 卷积层。其主要用来进行特征提取,利用固定维度的卷积核来提取图像的局部感知信息。其中,浅层的卷积层能够提取 DRF 图像的形状、边缘、纹理等低级特征,卷积层的更深层提取的是 DRF 图像的深层抽象特征。其中卷积核是指输入卷积的权重,又称为滤波器,即卷积核通过与上一层的特征图作卷积,得到的元素进行累加后再加上偏置量,将最终结果传递给非线性激活函数,DRF-Net 模型选择的非线性激活函数为 ReLU 函数<sup>[18]</sup>。同时,为降低 DRF-Net 的运算复杂度,本文设置同一层中某些神经元之间实现连接权重共享。卷积核的大小分别为  $7 \times 7, 5 \times 5, 3 \times 3$ ,步长设置为 2;并由高斯分布初始化,标准差设置为 0.01。卷积层中特征图的数目分别为 16、32 和 64。

具体计算过程为

$$x_{x_j}^l = f \left[ \left( \sum_{i \in M^{l-1}} x_i^{l-1} * K_{ij}^l \right) + b_j^l \right], \quad (3)$$

式中: $l$  为卷积层数总数; $K_{ij}^l$  为连接第  $l$  层与第  $l-1$  层的特征图  $j$  和  $i$  的卷积核; $M^{l-1}$  为第  $l-1$  层输入的特征图; $*$  为卷积符号; $b_j^l$  为第  $l$  层中特征图  $j$  的偏置量; $f(\cdot)$  为非线性激活函数。

3) 池化层,也称降采样层。主要作用是通过减少网络参数数目来降低计算量,从而实现特征图的降维。其在网络参数数量减少的同时还保证了标量不变,能够在一定程度上抑制过拟合。DRF-Net 诊断模型的 3 个池化层分别采用 max-pooling、mean-pooling 和 max-pooling,步长为 2。池化层与其前

端卷积层的特征图个数一致,特征图个数依次为 16、32、64。

4) 全连接层。用于整合获得的所有特征图,并通过 softmax 分类器获得 DRF 图像具体类别的预测结果。共包含 256 个神经元,将每一个单元都与上一层的每一个单元进行全连接,最后得到 256 维特征向量。

5) 输出层。输出层由欧氏径向基函数单元组成。计算公式为

$$j_i = \sum_j (x_i - w_{ij})^2, \quad (4)$$

式中: $x_i$  为最后一层卷积层的输入; $w_{ij}$  代表权重。

## 3.2 算法训练过程

### 3.2.1 DRF-Net 参数训练

DRF-Net 模型的参数训练过程选用从头训练和迁移学习相结合的策略。从头训练指利用网络的随机初始化参数从头训练 DRF-Net 模型;迁移学习则是指利用其他经典网络已经训练好的权重参数,使 DRF-Net 模型预先学到类似形状、边缘等最基本特征,从而提高 DRF-Net 的分类性能。当前最常用的 CNN 模型训练方法是对原始网络参数随机初始化,但本文采用的 DRF 数据集规模太小,使得深度学习在 DRF 数据集上容易出现过拟合。为解决上述问题,提高 DRF-Net 模型分类性能,本研究借鉴文献 [19] 所采用的方法,在本文提出的 DRF-Net 模型中引入迁移学习<sup>[20]</sup>的概念。

迁移学习原理如下:存在带有任务  $T_s = \{Y_s, f_s(\cdot)\}$  的源域  $D_s = \{X_s, P_s(X_s)\}$  和带有任务  $T_t = \{Y_t, f_t(\cdot)\}$  的目标域  $D_t = \{X_t, P_t(X_t)\}$ ,其中,  $D_s \neq D_t$  和/或  $T_s \neq T_t$  作为域和任务之间的条件。

$D_s$  和  $T_s$  获得“知识”并帮助  $D_t$  中预测函数  $f_t(\cdot)$  的学习。其中一个域由两部分组成,分别为边缘概率分布和特征空间,模型训练其他数据集的任务和域为  $T_s$  和  $D_s$ ,  $T_t$  和  $D_t$  则是训练本文 DRF 数据集的任务和域。本实验首先利用 ImageNet<sup>[21]</sup> 数据集上对本文提出的 DRF-Net 模型进行预训练,随后将预训练过程中获得的网络参数迁移到 DRF 数据集,多次训练后获得最终的参数。

### 3.2.2 DRF-Net 的激活函数

DRF-Net 使用了 Relu 激活函数,其中,Relu 函数表达式为

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}, \quad (5)$$

式中: $x$  为输入数据; $f(x)$  为对应输入  $x$  的输出。

使用 Relu 可以增加模型的非线性,由(5)式可以得出,Relu 可以让部分神经元的输出为 0,这就使得网络更加稀疏,并且减少了参数的相互依存关系,能够在一定程度上缓解过拟合现象。

最后一层选择 softmax 分类器来概率量化 DRF-Net 模型提取的“分布式特征”。其中训练集由  $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$  表示,与其相对应的标签为  $y^{(i)} \in \{1, 2, \dots, k\}$ ,由于本文是 3 分类模型,因此设定  $k=3$ 。 $p(y=i|x)$  表示输入的  $x$  与其相对应具体类别的概率,softmax 函数计算公式为

$$P(j) = \frac{\exp(\theta_j^T x)}{\sum_{i=1}^k \exp(\theta_i^T x)}, \quad (6)$$

式中: $\theta$  为 Softmax 的参数; $(\theta_i^T, x)$  为最后一层向输出层的输入。训练优化过程其实是逼近最佳  $\theta^T$  的过程。

则 Softmax 的损失函数记为

$$J(\theta) = -\frac{1}{m} \left\{ \sum_{i=1}^m \sum_{j=1}^k \{y^{(i)} = j\} \cdot \log [P(j)] \right\}. \quad (7)$$

(7)式中  $1\{y^{(i)} = j\}$  的取值规则为: $1\{值为真的表达式\} = 1, 1\{值为假的表达式\} = 0$ 。

## 4 评价标准

为证明本文提出的 DRF-Net 诊断模型的可行性,避免单一准确率来衡量算法优劣的片面性,现从以下 4 个方面对本文提出的算法进行综合评价。

准确率

$$A_{Acc} = (T_{TP} + T_{TN}) / (T_{TP} + T_{TN} + F_{FP} + F_{FN}), \quad (8)$$

查准率

$$P = T_{TP} / (T_{TP} + F_{FP}), \quad (9)$$

查全率

$$R = T_{TP} / (T_{TP} + F_{FN}), \quad (10)$$

式中: $T_{TP}$  为真阳性; $T_{TN}$  为真阴性; $F_{FP}$  表示假阳性; $F_{FN}$  表示假阴性。

F-score 计算公式为

$$F_{F-score} = (1 + \beta^2) \frac{P \times R}{\beta^2 \times P + R}, \quad (11)$$

式中: $\beta$  用来调节查准率和查全率两部分的比重,当  $\beta=1$  时,查准率和查全率具有相同权重,都很重要,称为  $F_1$ -score。

$$F_1 = \frac{2 \times P \times R}{P + R}. \quad (12)$$

此外,本文还采用受试者操作特性曲线下面积 (AUC) 来评估本文提出的实验模型。

## 5 实验结果与分析

### 5.1 实验环境

本文所有实验均在 Windows 10, MATLAB R2019a, Inter (R) Core (TM) i9-9820X CPU 和 128 GB RAM, NVIDIA GeForce RTX 2080 Ti GPU 中运行。同时为防止 DRF 数据集中的特殊个例对本文实验结果产生影响,本文采用五折交叉验证来训练 DRF-Net 模型。将 DRF 数据集平均分为 5 份,每次进行实验时取其中 1 份用作测试,1 份用作验证,其余 3 份用作训练,且确保用于构建训练集的图像不再用于测试和验证,最终实验结果取平均值作为整个模型的分分类准确率,从而保证了整个实验过程的严谨性。

### 5.2 对比实验

为验证本文提出的 DRF-Net 模型的有效性,将本文所提出的算法与传统机器学习算法做比较。当前机器学习的诊断方法主要分为以下两步:传统特征提取和分类。在传统特征提取算法的选择上,采取当前常用的纹理特征提取算法,分别为:Gabor 特征<sup>[22]</sup>、局部二值模式 (LBP)<sup>[23]</sup>、GLCM、方向梯度直方图特征 (HOG)<sup>[24]</sup> 及灰度特征<sup>[25]</sup>。表 1 列举了上述 5 种传统特征的具体维度。

在分类器的选择上,选取经典机器学习分类算法:SVM、ELM、kNN 以及随机森林算法 (RF)<sup>[26]</sup>,来评估上述单一传统特征的分分类效果。

表 1 5 种传统特征提取及其具体维数

Table 1 Five traditional feature extraction algorithms and their specific dimensions

Feature extraction algorithm	Feature vector dimension
Gabor feature	48
LBP feature	59
GLCM feature	8
HOG feature	324
Gray feature	6

### 5.3 实验结果

#### 5.3.1 DRF-Net 训练过程

经过多次实验,最终 DRF-Net 模型训练时的参数设置如下:选择优化器 RMSProp<sup>[27]</sup>,学习率设为 0.001,最大迭代次数为 3500 次,动量设置为 0.9,选取随机梯度下降 (SGD) 作为模型的优化函数。DRF-Net 模型的具体训练过程如图 4 所示。本文模型最终分类准确率为 84.2%,损失值为 0.63。

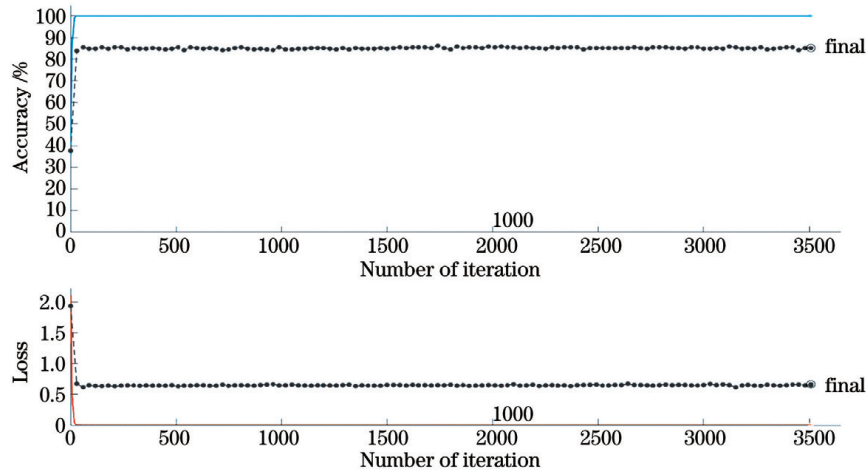


图 4 DRF-Net 的训练过程。(a)准确率与迭代次数示意图;(b)损失值与迭代次数示意图

Fig. 4 DRF-Net training process. (a) Schematic of accuracy rate and iteration number; (b) schematic of loss value and iteration number

#### 5.3.2 DRF-Net 与传统机器学习算法分类性能对比

为验证 DRF-Net 的分类性能,本文的对比实验选择当前 4 种经典的传统机器学习分类算法进行实验。并将准确率、 $F_1$  作为其多分类性能的评价指标,由实验结果可以得出以下结论:本文提出的 DRF-Net 诊断模型的准确率达到 84.2%,其结果显著优于 4 种经典的机器学习分类算法。另外,

对于同一类型的分类算法,5 种传统特征描述子的表现大体一致,差距在 3% 以内,在后续的分类结果中稳定且相近。且经过纵向对比,Gabor 特征描述子的表达能力最为突出,虽然 Gabor 特征多用于对象识别,但其针对图像细节的分类结果要优于其他 4 种纹理特征。表 2 为不同传统分类方法的分类性能比较。

表 2 不同传统分类方法的分类性能统计表

Table 2 Classification performance statistics table of different traditional classification methods

Feature extraction algorithm	Classifier	Validation set $A_{Acc}$ %	Test set $A_{Acc}$ %	$F_1$	AUC
Gabor feature	SVM	$67.8 \pm 1.3$	$68.4 \pm 2.4$	0.737	0.664
LBP feature		$58.7 \pm 2.7$	$57.7 \pm 1.1$	0.621	0.552
GLCM feature		$69.4 \pm 2.2$	$72.5 \pm 1.3$	0.731	0.683
HOG feature		$65.1 \pm 2.6$	$69.4 \pm 2.1$	0.714	0.626
Gray feature		$60.1 \pm 3.2$	$62.2 \pm 3.4$	0.628	0.542
Gabor feature	ELM	$68.5 \pm 2.7$	$72.1 \pm 1.2$	0.708	0.711
LBP feature		$68.1 \pm 3.1$	$70.8 \pm 1.2$	0.721	0.674
GLCM feature		$62.5 \pm 1.4$	$67.1 \pm 2.7$	0.692	0.670
HOG feature		$67.2 \pm 1.3$	$65.4 \pm 1.4$	0.734	0.702
Gray feature		$54.2 \pm 1.9$	$52.9 \pm 2.4$	0.533	0.556

续表

Feature extraction algorithm	Classifier	Validation set $A_{Acc}$ %	Test set $A_{Acc}$ %	$F_1$	AUC
Gabor feature	RF	72.1±2.3	70.1±1.2	0.724	0.679
LBP feature		65.5±1.2	64.2±1.9	0.679	0.604
GLCM feature		68.7±1.9	69.4±2.1	0.739	0.626
HOG feature		63.9±1.6	67.8±1.4	0.704	0.725
Gray feature		57.6±2.4	59.3±2.8	0.604	0.571
Gabor feature	kNN	67.3±1.5	68.8±3.2	0.697	0.592
LBP feature		71.4±2.8	65.9±4.1	0.721	0.661
GLCM feature		63.5±2.1	59.4±4.1	0.693	0.685
HOG feature		65.9±2.6	58.8±4.4	0.701	0.729
Gray feature		51.6±3.6	52.3±2.2	0.671	0.507
DRF-Net	Softmax	78.3±3.7	79.1±2.4	0.815	0.792
DRF-Net+TL		<b>84.2±2.3</b>	<b>82.1±0.9</b>	<b>0.834</b>	<b>0.826</b>

(Note: TL is transfer learning)

### 5.3.3 DRF-Net 同其他经典深度学习算法分类性能对比

本文还利用 AlexNet、GoogleNet<sup>[28]</sup>、ResNet50 等经典的 CNN 网络对 DRF 数据集进行分类,其分类性能的对比结果如表 3 所示。实验中,本文算法 DRF-Net 的分类准确率达到 84.2%,普遍优于上述经典深度学习分类算法。且由实验的分类对比结果我们能够得出以下结论,拥有过多层数的 CNN 并未在 DRF 数据集上表现出其层级优势,恰恰相反,因其层数较多,网络结构复杂,如 GoogLeNet,

使得分类模型出现了一定的过拟合现象,因此并非 CNN 的层数越高分类效果越好,最终本实验选用较为简单的网络结构。此外,由实验结果可以得出 DRF-Net 验证集和测试集的分类结果接近一致,这证明 DRF-Net 诊断模型的泛化能力较强,能够在一定程度上避免过拟合现象。另外,采用迁移学习+微调的分类准确率高于一开始训练参数的 4.9%,说明直接利用其他网络预训练好的网络参数能够提高模型的诊断效率和准确率。

表 3 DRF-Net 同其他深度学习模型分类性能对比

Table 3 Comparison of classification performance between DRF-Net and other deep learning models

Deep learning model	Validation set $A_{Acc}$ %	Test set $A_{Acc}$ %	$F_1$	AUC
AlexNet	76.7±1.8	75.6±2.2	0.732	0.774
GoogleNet	84.5±1.6	80.2±2.1	0.793	0.812
ResNet50	83.5±3.7	80.2±2.8	0.822	0.806
DRF-Net	79.3±2.1	76.9±1.7	0.815	0.792
<b>DRF-Net+TL</b>	<b>84.2±2.3</b>	<b>83.1±0.9</b>	<b>0.834</b>	<b>0.826</b>

## 6 讨 论

本文针对桡骨远端骨折临床误诊率高的问题,提出一种基于深度学习的自动检测 DRF 的多分类诊断模型。实验首先对临床采集的 DRF 图像数据进行预处理,利用 CLAHE 算法增强图像对比度,从而突出病灶区域,减少噪声干扰。在图像特征提

取方面,跳过传统机器学习方法,不对图像进行手动的特征选择(如每张图像进行形状、纹理等特征提取),而是利用深度学习的卷积层逐层提取图像的深层语义特征。并引入迁移学习的概念,提高了 DRF-Net 模型的性能。在训练 DRF-Net 模型的过程中,存在多种因素影响模型的训练效果。如:一次训练过程中选取的样本数量 batch size 值,训



练过程中样本的轮次数目 epochs 值,以及优化器的种类、学习率的选择,模型中每一层的隐藏单元数等超参数。实验过程中需要不断调整模型结构及超参数,从而使得 DRF-Net 模型采用最优的参数值。

由于本实验采用的是小规模样本数据集,因此在模型评估方面可能具有随机性。如果将数据集中较难区分的样本集中分到训练集或测试集中,则得到模型的分类结果可能差别较大。这使得模型分类结果不可靠,为解决上述问题,本实验采用交叉验证的方法来提高分类模型的可信度。最后,为了验证本文提出的 DRF-Net 模型的有效性,本文采取传统机器学习分类模型及当前常用的深度学习分类模型在 DRF 数据集上的分类结果作对比,通过比较实验结果,得出 DRF-Net 的分类准确率和  $F_1$  均高于上述分类模型,从而进一步证明了本文提出诊断方法的有效性。

## 7 结 论

为提高 DRF 多分类准确率,本文提出了一种基于深度学习的 DRF 图像的多分类诊断模型 DRF-Net。且在本文提出 DRF-Net 诊断模型的基础上,利用迁移学习提升了算法的分类性能。通过对多方面对本文提出的方法进行分析比较,得出本文提出的 DRF-Net 模型分类效果显著优于传统机器学习方法,相比于其他深度学习模型具有层数少,结构简单,分类速度快且分类性能优越的特点,可为临床辅助骨折诊断提供有效的技术支撑。在后续的研究工作中,将针对 DRF 图像的多分类工作,进一步扩展并融合 DRF 图像的特征,并从优化诊断模型算法的角度出发,进一步提高 DRF 图像的分类识别算法的诊断性能,从而为临床提供更加有效的辅助诊断技术。

## 参 考 文 献

- [1] Jo Y H, Lee B G, Kim H S, et al. Incidence and seasonal variation of distal radius fractures in Korea: a population-based study[J]. *Journal of Korean Medical Science*, 2018, 33(7): 78-84.
- [2] Shen W, Zhou M, Yang F, et al. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification[J]. *Pattern Recognition*, 2017, 61: 663-673.
- [3] Han Z Y, Wei B Z, Leung S, et al. Automated pathogenesis-based diagnosis of lumbar neural foraminal stenosis via deepmultiscale multitask

learning[J]. *Neuroinformatics*, 2018, 16(3/4): 325-337.

- [4] Yap D, Chen Y, Leow W K, et al. Detecting femur fractures by texture analysis of trabeculae[C]// *Proceedings of the 17th International Conference on Pattern Recognition*, 2004. ICPR 2004, August 26, 2004, Cambridge, UK. New York: IEEE Press, 2004: 730-733.
- [5] Haralick R M, Shanmugam K, Dinstein I. Textural features for image classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, SMC-3(6): 610-621.
- [6] Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. *IEEE Transactions on Evolutionary Computation*, 2002, 6(2): 182-197.
- [7] Cortes C, Vapnik V. Support-vector networks[J]. *Machine Learning*, 1995, 20(3): 273-297.
- [8] Whitley D, Starkweather T, Bogart C. Genetic algorithms and neural networks: optimizing connections and connectivity[J]. *Parallel Computing*, 1990, 14(3): 347-361.
- [9] Olczak J, Fahlberg N, Maki A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs[J]. *Acta Orthopaedica*, 2017, 88(6): 581-586.
- [10] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]// *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2818-2826.
- [11] Kim D H, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks [J]. *Clinical Radiology*, 2018, 73(5): 439-445.
- [12] Huang G, Liu X L. Automatic extraction and classification of road markings based on deep learning [J]. *Chinese Journal of Lasers*, 2019, 46(8): 0804002.  
黄刚, 刘先林. 基于深度学习的道路标线自动提取与分类方法[J]. *中国激光*, 2019, 46(8): 0804002.
- [13] Huang J L, Pan J, Zhao H S. A control study of MSCT and X-ray in the differentiation of different types of distal radius fractures[J]. *Chinese Journal of CT and MRI*, 2016, 14(3): 120-123.  
黄吉利, 潘捷, 赵鸿声. MSCT 对不同分型桡骨远端骨折的鉴别及与 X 线的对照研究[J]. *中国 CT 和 MRI 杂志*, 2016, 14(3): 120-123.
- [14] Gan K F, Xu D L, Lin Y M, et al. Artificial intelligence detection of distal radius fractures: a



- comparison between the convolutional neural network and professional assessments[J]. *Acta Orthopaedica*, 2019, 90(4): 394-400.
- [15] Pan S J, Yang Q. A survey on transfer learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345-1359.
- [16] Yang F, Wei G H, Cao H, et al. Research progression content-based medical image retrieval[J]. *Laser & Optoelectronics Progress*, 2020, 57(6): 060003.  
杨锋, 魏国辉, 曹慧, 等. 基于内容的医学图像检索研究进展[J]. *激光与光电子学进展*, 2020, 57(6): 060003.
- [17] Hong Y F, Wei B Z, Liu C, et al. Deep learning based automatic multi-classification algorithm for intervertebral foraminal stenosis[J]. *CAAI Transactions on Intelligent Systems*, 2019, 14(4): 708-715.  
洪雁飞, 魏本征, 刘川, 等. 基于深度学习的椎间孔狭窄自动多分级研究[J]. *智能系统学报*, 2019, 14(4): 708-715.
- [18] Chu J H, Wu Z R, Lü W, et al. Breast cancer diagnosis system based on transfer learning and deep convolutional neural networks[J]. *Laser & Optoelectronics Progress*, 2018, 55(8): 081001.  
褚晶辉, 吴泽蕤, 吕卫, 等. 基于迁移学习和深度卷积神经网络的乳腺肿瘤诊断系统[J]. *激光与光电子学进展*, 2018, 55(8): 081001.
- [19] Hong H C, Zheng L X, Pan S W. Fast computational technique for gray-levelco-occurrence matrix based on graphics process unit in biomedical engineering applications[J]. *Journal of Medical Imaging and Health Informatics*, 2018, 8(2): 309-312.
- [20] Liu X L, Hou F, Qin H, et al. Multi-view multi-scale CNNs for lung nodule type classification from CT images[J]. *Pattern Recognition*, 2018, 77: 262-275.
- [21] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database [C] // 2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE Press, 2009: 248-255.
- [22] Wong P K, Gao X H, Wong K I, et al. Online extreme learning machine based modeling and optimization for point-by-point engine calibration[J]. *Neurocomputing*, 2018, 277: 187-197.
- [23] Shen D, Wu G, Suk H I. Deep learning in medical image analysis [J]. *Annual Review of Biomedical Engineering*, 2017, 19: 221-248.
- [24] Zhou F Y, Jin L P, Dong J. Review of convolutional neural network[J]. *Chinese Journal of Computers*, 2017, 40(6): 1229-1251.  
周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. *计算机学报*, 2017, 40(6): 1229-1251.
- [25] Jiang M L, Zhang S T, Li H S, et al. Computer-aided diagnosis of mammographic masses using scalable image retrieval [J] *IEEE Transactions on Biomedical Engineering*, 2015, 62(2): 783-792.
- [26] Tsochatzidis L, Zagoris K, Arikidis N, et al. Computer-aided diagnosis of mammographic masses based on a supervised content-based image retrieval approach[J]. *Pattern Recognition*, 2017, 71: 106-117.
- [27] Huang P, Park S, Yan R K, et al. Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study [J]. *Radiology*, 2018, 286(1): 286-295.
- [28] Tomita N, Cheung Y Y, Hassanpour S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans [J]. *Computers in Biology and Medicine*, 2018, 98: 8-15.