Modern Engineering and Application 现代工程与应用, 2026, 4(2), 56-67
Published Online August 2026.
https://ufascience.com/mea/index

卷积神经网络 DenseNet 在阅读障碍手写图像分类中的应用

谢穗芷

东北大学理学院，沈阳 110819，辽宁，中国

摘要： 阅读障碍是一种与词汇层面阅读困难相关的特定学习障碍（SLD）。由于语音加工与正字法加工存在共同缺陷，该障碍常在儿童手写中表现为间距不规则、字母大小不一致。早期识别至关重要，但传统诊断流程耗时较长，不适用于大规模筛查。本研究旨在利用 DenseNet121 卷积神经网络（CNN）模型开展段落级手写分析，为资源受限的教育场景提供一种低成本的阅读障碍筛查工具。本文对 100 幅英文手写图像进行预处理并标准化为 200 个样本，其中 70%的数据采用 4 折交叉验证进行评估，其余 30%用于测试。该模型取得 90%的测试准确率与 92.86%的训练准确率，显著优于随机森林基线模型（训练准确率 83.57%、测试准确率 63.33%），其统计显著性经 McNemar 检验确认。本研究的主要贡献在于证明：采用段落级分析的轻量化单一架构 DenseNet121，能够在计算资源需求显著更低、流水线更为简化的前提下，取得与依赖复杂混合模型和字符级分析的既有研究相当的性能。上述发现表明，DenseNet121 可为资源有限的教育环境中的阅读障碍初步筛查提供稳健而低成本的解决方案。

关键词： 卷积神经网络；深度学习；阅读障碍；K 折交叉验证

Convolutional Neural Network DenseNet in Classifying Dyslexic Handwriting Images

Suizhi Xie

College of Sciences, Northeastern University, Shenyang 110819, Liaoning, China

Abstract: Dyslexia is a specific learning disability (SLD) associated with word-level reading difficulties and often manifests in childhood handwriting through irregular spacing and inconsistent letter sizing, due to shared phonological and orthographic processing. Early identification is critical; however, traditional diagnostic procedures are time-consuming and unsuitable for large-scale screening. This study aimed to develop a handwriting analysis at the paragraph-level using a DenseNet121 convolutional neural network (CNN) model as a low-cost dyslexia screening tool for resource-constrained educational settings. One hundred English handwriting images were preprocessed and standardized into two hundred samples, with 70% of the dataset evaluated using 4-fold cross-validation and the remaining 30% used for testing. The model achieved 90% test accuracy and 92.86% training accuracy, significantly outperforming a random forest baseline that reached 83.57% train accuracy and 63.33% test accuracy, with statistical significance confirmed by McNemar's test. The main contribution of this study is the demonstration that a lightweight,

文章引用：谢穗芷. 卷积神经网络 DenseNet 在阅读障碍手写图像分类中的应用[J]. 现代工程与应用, 2026, 4(2): 56-67.
DOI: https://doi.org/10.61784/mea2021

single-architecture DenseNet121 using paragraph-level analysis can achieve competitive performance compared to prior studies that relied on more complex hybrid models and character-level analysis, while requiring substantially lower computational resources and a simplified pipeline. These findings indicate that DenseNet121 provides a robust and low-cost solution for preliminary dyslexia screening in resource-limited educational environments.

Keywords: Convolutional neural network; Deep learning; Dyslexia; K-fold cross validation

一、引言

阅读障碍作为一种特定学习障碍（SLD），其主要特征是难以准确或流畅地识别词汇、难以解码不熟悉的单词[1-3]，其根源在于语音加工与正字法加工方面的缺陷，而这两种加工能力对读写习得至关重要[4]。这些困难独立于认知能力而出现，且在有效的课堂教学下仍常持续存在[2-3]。发展性阅读障碍在全球小学生中的发生率约为 7%—10%[5-6]，其中相当数量的个案仍未被诊断[6]。除读写困难外，阅读障碍儿童往往表现出消极的自我认知与较差的学业表现[7]；不过，早期识别与针对性的教育干预能够同时改善其自我概念与学业结果[2-7]。传统的阅读障碍识别依赖全面的心理教育评估，而由于学界对阅读障碍标志物长期存在争议，各评估方案差异较大[3]，致使这类评估成本高昂、难以用于大规模筛查。近期的共识强调其核心特征——词汇层面阅读的准确性与流畅性困难[3]，并主张采用以阅读与拼写表现为依据、优先进行早期评估的筛查模式[2]。

为支持可及的早期识别，近期研究探索了在多种模态下用机器学习（ML）与深度学习（DL）方法自动检测阅读障碍[8]，如脑电图（EEG）、眼动追踪与磁共振成像（MRI）。SESHADRI 等[9]在 EEG 数据上使用 K 近邻（KNN）取得 96.7%的平均准确率，但其研究关注的是认知与注意指标，而非阅读或读写相关技能。眼动追踪方面，SVARICEK 等在有限数据集上以注视点图像可视化结合 ResNet18 取得 86.65%的准确率；VAITHEESHWARI 等则报告称，采用基于虚拟现实的眼动特征并结合卷积神经网络（CNN）、深度神经网络（DNN）与基于 Transformer 的双向编码器表示（BERT）的融合模型，准确率达 98%。多模态研究亦有开展：ALKHURAYYIF 与 SAIT 将多种深度学习模型（MobileNetV3、EfficientNetB7 与双向长短期记忆网络 Bi-LSTM）应用于 MRI、功能磁共振（fMRI）与 EEG 数据，准确率分别为 98.6%、98.9%与 98.8%。这些模态虽然效果出色，却需要专门的设备与专业知识，难以用于大规模、低成本的教育筛查[10]。此外，近期研究指出，尽管阅读障碍个体存在神经生物学差异，以读写困难的行为学指标而非神经影像来识别阅读障碍更为有效可靠[11]。

鉴于阅读与书写高度相关，且研究证实阅读障碍儿童即使在简单任务中也表现出较差的书写清晰度与较慢的书写速度，两种技能都反映出语音加工与正字法加工方面的共同缺陷[12]。因此，手写分析已成为一种实用的行为筛查手段。呈现空间特征（如间距不一致、字母书写不规则、整体组织混乱）的手写样本，可以在教育场景中无需昂贵设备便捷采集。已有若干深度学习研究探索了这一模态：PATIL 等在 IAM 数据集与一所小学的数据集上，采用 CNN-Bi-LSTM 混合架构进行整体手写分析，取得 95.6%的准确率；ALQAHTANI 等采用 CNN-SVM 混合模型对三类手写字符（正常、镜像、修正）进行分类，准确率达 99.33%；DysDiTect 采用 CNN-LSTM 混合模型，在中文字符数据集上取得 83.2%的准确率；ZAIBI 与 BEZINE 在“Handyg23”阿拉伯语段落 / 文本数据集上，用梯度提升（GB）与随机森林（RF）模型均取得 99%的准确率。此外，SAIT 与 ALKHURAYYIF 的跨模态研究将手写字符数据与 MRI、EEG 数据相结合，采用基于 Transformer 的混合模型取得 99.1%—99.2%的准确率。

尽管取得上述进展，仍存在明显的研究空白。第一，基于手写的研究中很少开展段落级分析，多数前沿方法依赖字符级或序列级分析，需要复杂的混合架构，且限制了模型学习整体书写运动模式（而非孤立字形）的能力。第二，字母文字体系下的段落级阅读障碍手写数据集十分稀缺；采用中文、阿拉伯文等非字母文字的研

究，由于书写运动需求不同，未必能推广到字母文字体系。第三，许多既有数据集规模小且不公开，使复现困难，因而需要在小数据条件下仍能保持稳定性能的模型。第四，混合架构虽然准确率高，但其复杂性阻碍了在硬件条件有限的普通学校中实际部署。

为弥补上述空白，本研究的目标是构建一种计算高效的单一架构阅读障碍筛查模型：采用基于 CNN 的 DenseNet121，通过段落级手写分析完成分类，从而在资源受限的教育环境中实现低成本、可扩展的初步筛查。选择 DenseNet121 是因其对过拟合具有稳健性，在小数据集上尤为如此。本研究使用开源的英文段落级手写样本数据集，以确保可复现性以及对字母文字体系的实际适用性。本文的主要贡献如下：（1）提出配套预处理流水线的单一架构 DenseNet121 阅读障碍分类模型，证明轻量化 CNN 可在计算资源需求显著更低的条件下，取得与既有混合或集成模型相当的准确率；（2）在字母文字手写上实施段落级分析以捕捉整体空间特征——不同于以往多数聚焦字符级或词级切分的研究，该方法既缓解了数据集稀缺问题，也免去了耗时的切分流水线；（3）在小规模字母文字手写数据集上验证了 DenseNet121 的稳健性，回应了医学与教育深度学习应用中常因缺乏大规模数据而面临的关键挑战。

为使行文清晰连贯，本文其余部分安排如下：第二部分介绍材料与主要方法；第三部分给出结果与讨论；第四部分总结全文并指出未来工作方向。

二、方法

图 1 给出了本研究采用的总体研究流程，自数据获取直至最终的性能评估。流程始于数据集获取（第（一）节）以及旨在产出干净、标准化的类扫描件手写图像的结构化预处理流水线（第（二）节）。DenseNet121 的模型架构与配置详见第（三）节；第（四）节涵盖完整的训练与验证流水线，包括数据集划分、4 折交叉验证（CV）的设定、模型编译以及训练与验证过程。模型测试、性能指标评价以及统计与对比分析在第（五）节说明。

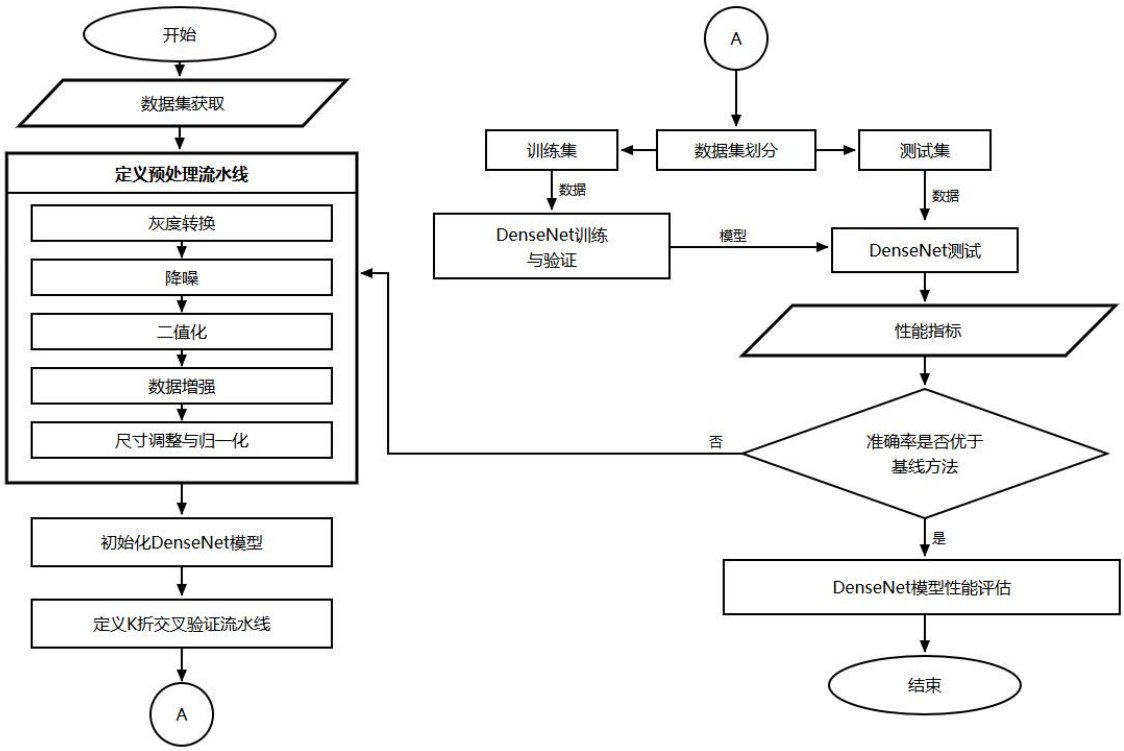

Figure 1 Research Workflow

图 1 研究流程

（一）数据集获取

数据集取自用户 dlsathvik04 的开源 GitHub 仓库“Dyslexia_Detection”，包含 100 幅英文手写图像，阅读障碍类与非阅读障碍类各占一半。图 2 给出了非阅读障碍（左）与阅读障碍（右）手写样本，每份样本由一至两个段落构成。该数据集虽缺少受控的受试者元数据（如年龄、年级），但因其类别均衡、公开可得且适用于整体手写二分类任务而被选用。

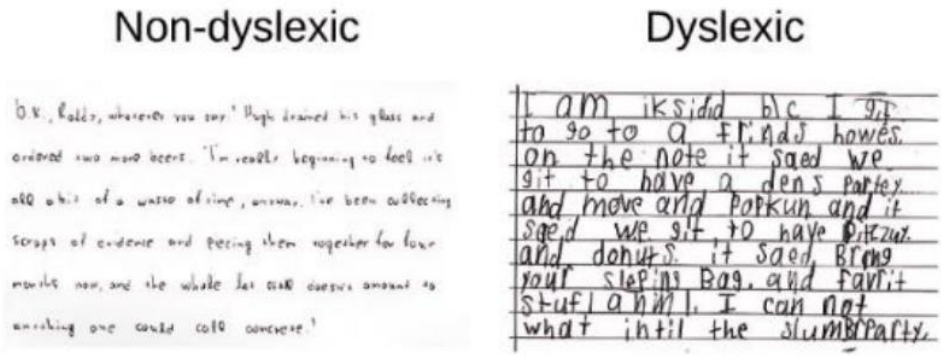

Figure 2 Dataset Samples

图 2 数据集样本

（二）数据集预处理

为把原始数据转化为干净、标准化的输入，本研究采用了适用于图像数据的预处理技术，包括灰度转换、降噪、二值化、数据增强与标准化（尺寸调整与归一化）。本研究按灰度转换、降噪、二值化的特定顺序处理，以获得更干净的类扫描件图像。灰度转换降低了色彩复杂度并简化了计算；使用中值滤波降噪可去除灰度转换过程中可能引入或被放大的噪声；采用自适应阈值的二值化则依据局部强度分布把灰度图像转换为干净的黑白图像。在上述步骤之后，再施加随机旋转与亮度调整等增强方法，为数据集另外增加 100 幅图像（图像总数为 200 幅），以扩充训练数据并提升泛化能力。最后，将图像统一调整为 200×200 像素、3 通道，并作归一化处理，使像素值落在合理区间内。

（三）模型架构定义

CNN 作为一种深度学习算法，通过卷积与池化运算，利用多个隐藏层检测数据中的复杂模式。卷积运算取代传统的矩阵乘法来提取空间特征；在多数深度学习框架中，卷积以互相关方式实现，不对卷积核作翻转，其定义见式（1）。

$$S(i,j) = (I * K)(i,j) = \Sigma_m \Sigma_n I(i+m, j+n) \cdot K(m,n) \tag{1}$$

式中，S 为特征图，I 为输入，K 为卷积核，(m, n)为卷积核位置，(i, j)为输出像素位置。池化运算则通过汇总局部特征来降低空间维度：最大池化返回每个区块中的最大值，平均池化返回区块内元素的均值。本研究采用平均池化，其定义为：

$$f\_ave(X) = (1/N) \Sigma_{i=1}^{N} x_i \tag{2}$$

式中，X 为特征区块，x_i为区块中的第 i 个元素，N 为元素总数。

DenseNet 作为一种 CNN 模型，在稠密块内采用稠密连接：每一层都接收其之前所有层的特征图，并把自

身输出传递给之后的所有层。该结构改善了信息流动、促进了特征复用，并缓解了梯度消失问题。稠密连接定义为：

$$x\_l = H\_l([x_0, x_1, \cdots, x\_\{l-1\}]) \tag{3}$$

式中，x_l 为第 l 层的特征图，H_l(·)为复合函数，[x0, x1, ⋯ , x_{l−1}]为其前所有特征图的拼接。

每个复合函数 H_l 由批归一化（BN）、修正线性单元（ReLU）激活与卷积构成，其排列顺序为 BN → ReLU → 1×1 卷积（瓶颈层）→ ReLU → 3×3 卷积。BN 通过归一化激活值稳定训练；ReLU 引入非线性以学习复杂模式；瓶颈层用于减少参数量，最后的卷积则提取空间特征。BN 与 ReLU 的核心公式依次见式（4）与式（5）。

$$H' = (H - \mu) / \sigma \tag{4}$$

$$g(z) = \max\{0, z\} \tag{5}$$

式中，H 为一个小批量的激活值，μ为每个单元的平均激活值，σ为每个单元的标准差；g(z)表示作为恒等函数形式的 ReLU，z 为输入。

稠密块之间的过渡层包含 BN、1×1 卷积与 2×2 平均池化，用以降低空间维度并控制模型复杂度。DenseNet 通常采用全局平均池化后接 Softmax 层，而本研究针对阅读障碍与非阅读障碍手写的二分类任务，采用阈值为 0.5 的 Sigmoid 激活。Sigmoid 函数把任意实数 z 映射到区间[0,1]，其定义见式（6）。

$$\sigma(z) = 1 / (1 + e^{-z}) \tag{6}$$

基于上述优势，本文选用预训练的 DenseNet121 主干网络，因其效率高且参数量相对较少，并在 Google Colab 中以 TensorFlow/Keras 实现。该模型在数据量与算力受限的条件下针对二分类任务作了适配，具体见伪代码 1。

（四）训练与验证流水线

200 幅手写图像的数据集按 70∶30 划分，其中 70%（140 幅）用于训练与验证，30%（60 幅）用于测试。选择该留出比例是为了减轻测试集过小带来的乐观偏差，并确保有足够数量的未见样本。70%的部分采用 4 折交叉验证进行评估，划分时打乱顺序并固定随机种子（42）以保证可复现性。每一折约使用 105 个训练样本与 35 个验证样本。

DenseNet121 模型以 AdamW 优化器（学习率=1×10^{-4}，权重衰减=1×10^{-5}）与二元交叉熵损失编译。每折训练 15 轮，批大小为 16，并使用模型检查点依据验证准确率保存表现最佳的权重。该 4 折交叉验证设置可缓解过拟合与偏差，为小数据集提供更好的泛化能力。完整的训练过程见伪代码 1。

伪代码 1　DenseNet121 结合 4 折交叉验证的训练算法

输入：预处理后的训练数据集 D（140 幅图像，200×200×3），类别标签 y，折数 K=4，轮数 E=15，
批大小 B=16，学习率 ρ=1e-4，权重衰减 λ=1e-5，Dropout 比率（0.2、0.4），
L2 正则化参数 α，随机种子 s=42。

输出：各折最优模型 {M1,…,M4}，各折验证指标 {Acc, Prec, Rec, F1}，训练历史 H。

1: 载入以 ImageNet 预训练权重初始化的 DenseNet121 主干，并去除顶层

2: 冻结 DenseNet121 中的全部基础层

3: 构建分类头：

添加 GlobalAveragePooling2D 层

添加 Dropout 层（0.2）

添加 Dense 层（16，ReLU，L2 = α）

添加 Dropout 层（0.4）

添加 Batch Normalization 层

添加 Dense 层（1 个单元，Sigmoid）

4: 初始化 4 折交叉验证（shuffle=True, random_state=s）

5: 将 D 划分为 4 折 $\{F_1,\ldots,F_4\}$

对每一折 i=1 到 4 执行：

6: 将 F_i 划分为训练集 D_train（约 105 幅）与验证集 D_val（约 35 幅）

7: 以 DenseNet121 主干与分类头初始化模型 M_i

8: 用 AdamW 优化器（ρ, λ）与二元交叉熵损失编译 M_i

9: 依据验证准确率定义模型检查点

对每一轮 e=1 到 E 重复：

在 D_train 上训练 M_i（批大小 = B）

在 D_val 上评估 M_i

若验证准确率提高则保存最优权重

直至达到轮数 E

10: 载入 M_i 已保存的最优权重

11: 在 D_val 上评估 M_i 并记录指标（Acc, Prec, Rec, F1）

12: 保存最优模型 M_i 与训练历史

循环结束

13: 汇总全部 K 折的验证指标

14: 返回 $\{M_1, M_2, M_3, M_4\}$、验证指标与训练历史

（五）模型测试与评价

从 4 折交叉验证中选出的最佳模型在包含 60 幅图像的测试集上进行评估。主要评价工具为混淆矩阵，它可反映误分类的类型与模型的总体性能。混淆矩阵中使用四个数值，见表 1。由这些数值可导出四项评价指标：（1）准确率衡量正确预测数占预测总数的比例；（2）精确率衡量预测为阳性的样本中真阳性所占的比例；（3）召回率衡量实际阳性样本中真阳性所占的比例；（4）F1 值为精确率与召回率的调和平均。各指标见式（7）至式（10）。

Table 1 Confusion Matrix Values for Binary Classification

表 1 二分类混淆矩阵的取值

实际类别	预测类别：阳性	预测类别：阴性
阳性	真阳性（TP）	假阴性（FN）
阴性	假阳性（FP）	真阴性（TN）

此外，本文还纳入其他测度以增强研究的说服力。受试者工作特征曲线下面积（AUC-ROC）用于评价模型在所有决策阈值下的分类性能，提供与阈值无关的判别能力度量。由 Sigmoid 激活函数（式（6））导出的置信度分数被用于评估单个分类结果的确定性。为给出总体准确率的合理取值范围，本文采用 Wilson 得分法计算测试准确率的 95%置信区间（CI）。交叉验证的稳定性通过计算 4 折准确率的均值与标准差、并以变异系数（CV% = 标准差/均值×100）来评估性能的一致性。

本文还在α=0.05 水平上采用 McNemar 检验开展统计评价，在同一数据集上比较 DenseNet 与基线模型，其原假设为两模型的错误率相等。

b 为 DenseNet 正确而 RF 错误的样本数，c 为 RF 正确而 DenseNet 错误的样本数。本研究采用的基线模型

为随机森林（RF），其参数为 100 棵树、最大深度 10、叶节点最小样本数 2，并采用与 DenseNet 完全相同的 4 折交叉验证设置以保证比较公平。选择 RF 作为基线是因其在分类任务中稳健且计算高效。在训练 RF 之前，先将预处理后的数据展平为一维向量，并通过主成分分析（PCA）降至 100 个成分。若 DenseNet 的准确率未能超过 RF 基线，则重新评估预处理方法与超参数。

1. 与前沿模型的对比分析

为评估所提方法的竞争力，本研究依据分类性能、模型复杂度、计算效率、数据需求以及实际可部署性（硬件要求、实现复杂度）等标准，对近期基于手写的阅读障碍检测研究进行了比较。比较对象为 2020—2025 年间发表的研究，具体包括 PATIL 等、ALQAHTANI 等、DysDiTect 以及 ZAIBI 与 BEZINE。这些研究代表了目标相近但方法学选择不同的当前前沿方案。

三、结果与讨论

（一）模型性能结果

1. DenseNet 模型在训练与测试中的表现

图 3 给出了 DenseNet 模型在 4 折上的训练历史。各折均表现出一致的收敛趋势：训练与验证准确率（蓝线）持续上升，训练与验证损失（红线）总体下降。各折的训练准确率稳步提高，范围约为 80%至 90%以上，验证准确率则始终超过 88%。不过，相对较高的训练 / 验证损失（>50%）以及各折之间的差异，反映了在小数据集（200 幅图像）上训练深层网络的困难。DenseNet121 稳健的架构可能使模型过于自信，因而需要较强的正则化来约束特征学习能力、防止过拟合，从而造成训练历史曲线出现小幅波动、不够平滑。

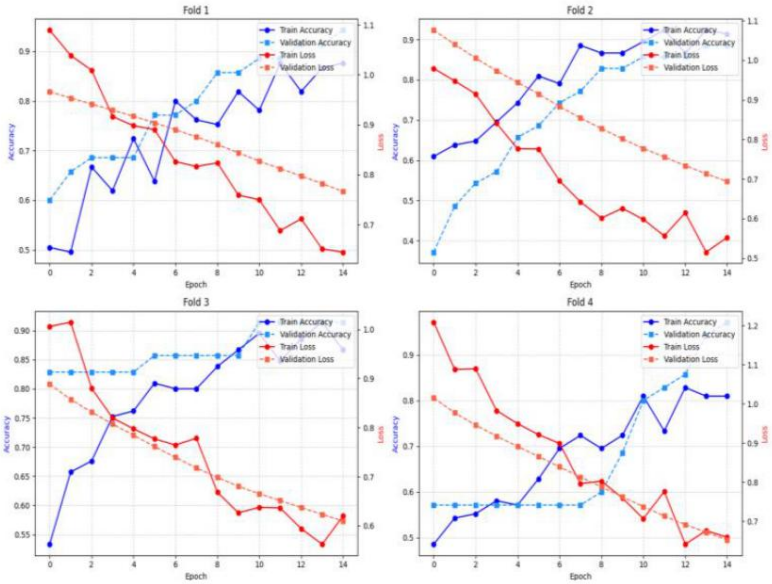

Figure 3 Training and Validation History on Each Fold

图 3 各折的训练与验证历史

DenseNet 模型取得 90.0%的测试准确率，其余指标见表 2。95%置信区间表明，尽管测试集相对较小，性能估计依然可靠；15.49%的区间宽度反映了教育类数据集中样本量有限所固有的不确定性。变异系数 CV%=3.44%（<5%）表明稳定性良好，训练准确率均值为 92.86%±3.19%，证实了模型在不同数据划分下的稳健表现。AUC-ROC 为 98.44%，说明模型在所有分类阈值下都具有很强的类别判别能力。预测置信度分数分析显示：阅

读障碍样本的置信度范围较宽（8%—49%）且分数持续偏低，反映出模型在识别阅读障碍空间模式时确定性较高；非阅读障碍样本的范围较窄但存在重叠（42%—64%），部分预测值低于 50%的分类阈值。这种非对称的置信度分布解释了模型偏保守的分类行为及其完美的精确率——模型能够自信地识别阅读障碍模式，但对某些可能具有非典型空间特征的非阅读障碍样本存在不确定性。

Table 2 Overall DenseNet Metrics Performances

表 2 DenseNet 的总体指标表现

指标	取值
测试准确率	90%
测试精确率	100%
测试召回率	80%
测试 F1 值	88.89%
训练准确率	92.86%
训练精确率	89.01%
训练召回率	98.33%
训练 F1 值	93.13%
95%置信区间	[79.85%, 95.34%]
AUC-ROC	98.44%
交叉验证变异系数	3.44%
置信度分数（阅读障碍）	8%—49%
置信度分数（非阅读障碍）	42%—64%

2. 与基线模型的性能比较

表 3 给出了所提 DenseNet 与基线 RF 模型的各项指标表现，其中“差值”列表示两模型之间的指标差异。在训练阶段，DenseNet 在全部指标上均表现更优：准确率高出 9.29%，召回率高出 19.28%，F1 值高出 10.40%，精确率略高 0.32%。DenseNet 还表现出更好的稳定性，各指标的标准差均低于基线 RF（标准差：3.19%对 8.42%），说明其学习过程更为一致。在测试阶段，性能差距显著扩大：DenseNet 的准确率高出 26.67%、精确率高出 33.33%、召回率高出 26.67%、F1 值高出 29.63%，进一步表明深度学习方法不仅学习效果更好，对未见数据的泛化能力也更强。

Table 3 Metrics Performances Comparison of DenseNet and Baseline Model

表 3 DenseNet 与基线模型的指标性能比较

指标	DenseNet（训练）	RF（训练）	DenseNet 标准差	RF 标准差	差值（训练）	DenseNet（测试）	RF（测试）	差值（测试）
准确率	92.86%	83.57%	3.19%	8.42%	+9.29%	90%	63.33%	+26.67%
精确率	89.01%	88.69%	8.01%	7.93%	+0.32%	100%	66.67%	+33.33%
召回率	98.33%	79.05%	2.89%	14.24%	+19.28%	80%	53.33%	+26.67%
F1 值	93.13%	82.73%	3.49%	8.97%	+10.40%	88.89%	59.26%	+29.63%

为确认 DenseNet 模型确实优于基线 RF 模型，本文施加了 McNemar 统计检验（见表 4）。检验证实差异具有统计显著性（χ^2=11.25，p=0.000796）：DenseNet 正确分类了 18 个 RF 误判的样本，而 RF 仅纠正了 2 个 DenseNet 误判的样本（比例为 9∶1）。这一有利于 DenseNet 的 9∶1 不一致比例，说明其相对于传统机器学习方法具有实质性且统计显著的优势。p 值 0.000796 表明，随机偶然观察到这一性能差异的概率不足 0.08%，为该改进的真实性提供了有力证据。

Table 4 McNemar's Test Summary (DenseNet121 vs RF)

表 4 McNemar 检验结果汇总（DenseNet121 与 RF 对比）

统计量	取值
两模型均正确	36
仅 DenseNet 正确（b）	18
仅 RF 正确（c）	2
两模型均错误	4
χ^2（卡方值）	11.25
p 值	0.000796
α（显著性水平）	0.05

（二）讨论

1. 消融分析

为验证各组成部分的贡献，本文通过系统性地移除所提流水线中的关键要素开展了消融实验。

（1）预处理流水线的影响：类扫描件图像转换、数据增强与标准化对于获得稳健性能至关重要。移除这些步骤（仅保留基本的尺寸调整与归一化）后，训练准确率由 92.86%降至 75.57%（±14.67%），测试准确率由 90%降至 83.33%；更为关键的是，训练稳定性劣化了 360%（标准差由 3.19%升至 14.67%），说明各折之间学习过程不一致。测试精确率由 100%降至 77.78%，引入的假阳性会削弱筛查的可靠性。原始手写图像含有大量噪声、光照不一致与背景伪影，会干扰空间特征提取；预处理流水线的各项操作（灰度转换、降噪与二值化）把输入统一为类扫描件格式，使 DenseNet 能够聚焦于空间组织模式而非图像质量的干扰因素。6.67%的准确率提升与 360%的稳定性改善，足以证明预处理开销对于部署一致性是必要的。

（2）架构选择：DenseNet121 的稠密连接模式能够保留空间特征，这对于捕捉阅读障碍手写所特有的细微紊乱模式尤为重要。其 706 万参数在表达能力与计算效率（208.57 ms/幅）之间取得了平衡，不同于需要多级顺序处理的混合模型（见表 5）。

（3）数据增强：在仅有 100 幅原始图像的条件下，数据增强对泛化至关重要。增强扩大了有效训练样本量，帮助模型学习到对旋转不变的空间特征；不过基础样本量偏小仍是一项局限，这体现在 15.49%的置信区间宽度上。

2. 与前沿方法的比较

表 5 给出了所提方法与近期基于手写的方法之间的对比分析。所提方法在取得有竞争力性能的同时，在可部署性与资源效率方面具有明显的实用优势。与以往研究采用的混合或多级架构（如 PATIL 等的 CNN-BiLSTM 模型以及 DysDiTect 中的 CNN-位置编码-LSTM-注意力网络）不同，所提模型采用参数量为 706 万的单一 DenseNet121 架构。尽管配置更为简单，本文模型的准确率仍比 DysDiTect 高 6.8%，并与 PATIL 等保持相当水

平，同时通过 4 折交叉验证展现出更好的泛化稳定性（CV%=3.44%）。虽然 ALQAHTANI 等取得了最高准确率（CNN-SVM，99.33%），但其字符级方法的实现要复杂得多；同样，ZAIBI 与 BEZINE 用集成模型报告了 99%的准确率，却需要多个模型流水线与专门的数据采集工具。

Table 5 Comparing State-of-the-Art Models

表 5 与前沿模型的比较

相关研究	分类性能	模型复杂度	计算效率	所用数据集	实际可部署性（硬件要求与实现复杂度）
本文方法	准确率 90%，精确率 100%，召回率 80%，F1 值 88.89%，AUC 98.44%	单一架构 DenseNet121，总参数量 7 058 017	208.57 ms/幅	200 幅英文段落图像（140 幅用于 4 折交叉验证训练，60 幅用于测试），类别均衡 50∶50	硬件：12 GB 内存、Core i3 2.00 GHz CPU、128 MB 显存、500 GB 存储；实现：单一模型，配套预处理流水线，特征提取与分类在一个模型内直接完成
PATIL 等	准确率 95.6%，精确率 94.38%，召回率 91.51%，F1 值 92.61%	混合模型（CNN-BiLSTM＋CTC 损失），总参数量 8 743 247	未说明	IAM 数据集：1539 页、657 名书写者，字符级分析	硬件：未说明；实现：混合模型，需预处理与词切分，由 CNN+BiLSTM 完成特征提取与序列分析
ALQAHTANI 等	CNN 98.59%，CNN-RF 98.44%，CNN-SVM 99.33%（仅报告准确率）	单一与混合模型（CNN、CNN-SVM、CNN-RF），总参数量未说明	未说明	176 673 幅英文字符图像（按 70/15/15 划分）	硬件：未说明；实现：多个模型，需预处理与切分，CNN＋分类器
DysDiTect	仅手写：准确率 83.2%，敏感度 79.2%，特异度 86.4%，AUC 91.2%；加入年级信息后：准确率 85%，敏感度 83.3%，AUC 89.7%	混合 CNN-位置编码-LSTM-注意力模型，总参数量未说明	最多 50 轮，采用早停	100 000 个中文字符，1064 名儿童（483 名发展性阅读障碍、581 名典型发展），词级	硬件：未说明；实现：复杂混合模型，需字符切分、序列特征与迁移学习
ZAIBI 与 BEZINE	最佳 99%（GB、RF），AdaBoost 97%，SVM-RBF 94%（仅报告准确率）	集成模型（GB、RF、AdaBoost、SVM），总参数量未说明	未说明	120 份阿拉伯语样本（7—12 岁），12 项手写任务	硬件：Wacom 数位板＋MovAlyzer 软件；实现：多个集成模型，12 项任务规程

一项关键的区别性因素在于所提方法采用段落级分析，而字符级或词级方法则需要大得多的数据集，如

ALQAHTANI 等、DysDiTect 与 PATIL 等。在教育工作者采集自然书写样本的筛查场景中，段落级分析具有更好的生态效度，且无需耗时的词切分。完美的精确率（100%）对初步筛查尤为可贵：它确保模型标记出潜在阅读障碍时能够避免假阳性，从而不会在教育场景中造成不必要的焦虑与资源错配。

所提方法在实际可部署性方面表现优越：在配置一般的硬件（Intel Core i3 CPU、12 GB 内存、128 MB 显存）上，单幅图像推理时间仅为 208.57 ms。相比之下，既有研究要么未报告计算指标，要么需要资源开销更大的架构。单一模型设计免除了混合流水线所需的特征提取、序列处理与集成分类等分离环节。这种简洁的实现方式使模型可直接经云平台（Google Colab）部署：教育工作者上传手写照片、运行自动化预处理流水线，即可立即获得初步筛查结果，无需专门硬件或技术专长，从而为资源受限教育环境中的实际阅读障碍筛查提供了一种可行且可及的方案。

3. 局限性

若干技术局限源自实验设计与数据集条件。在稳健的深度学习模型上使用小数据集（200 幅图像），迫使本文采用较强的正则化（L2、Dropout、批归一化）以防止过拟合，这同时也限制了模型的特征学习能力，表现为训练 / 验证损失偏高（>50%）、学习曲线不够平滑，以及尽管取得 90%测试准确率但仍可能存在的欠拟合。15.49%的置信区间宽度反映了样本量的限制，说明性能估计本身带有固有的不确定性。此外，手写格式的一些差异（如段落长短不一、印刷体或连笔体）引入了不受控的变异，可能影响分类的一致性：部分学生只写了简短的单段样本，另一些则写了篇幅较长的多段文本，造成样本间空间复杂度的异质性。模型偏保守的分类倾向导致出现 6 例假阴性（非阅读障碍学生被误判为可能存在阅读障碍），这些个案很可能具有非典型空间特征，如书写本身较不规整、疲劳效应或书写过于匆忙。仅凭段落级空间分析无法捕捉词级拼写错误、字母镜像或语音模式，而这些信息在临床阅读障碍评估中可提供互补的诊断价值。因此，该模型只应被视为研究用的初步辅助工具，而非教育场景中可独立使用的诊断仪器。

尽管完整的预处理、训练与测试流水线已整合进单个 Google Colab 笔记本，使模型在技术上可在真实教育场景中部署，但当前实现对不具备信息技术背景的教师、学校工作人员或家长而言仍不够友好：运行该笔记本仍需具备 Python、Google Colab 与文件管理的基本知识。在缺少图形界面或自动化输入 / 输出流程的情况下，非技术用户可能难以上传手写样本、解读输出或排查运行错误。因此，即便模型可执行且可复现，它尚不能直接用于实际筛查，仍需额外的设计工作（如网页界面或移动应用）以支持其在基础教育场景中的真正落地。

四、结论

本研究提出了一种基于 DenseNet121、通过段落级手写分析进行阅读障碍初步筛查的方法，在数据集较小的条件下取得 92.86%的训练准确率与 90%的测试准确率。该模型相对于传统机器学习 RF 基线具有统计显著的优势（McNemar 检验：χ^2=11.25，p<0.001），并与更复杂的前沿方法性能相当，同时保持了计算开销低（208.57 ms/幅）、单一架构工作流、易于部署（硬件要求不高、书写样本易于采集）等实用优势，因而适用于资源受限的教育场景。鉴于手写变异会出现在多种特定学习障碍中，且可能受多种非诊断性因素影响，本工具被定位为早期筛查支持系统而非确定性的诊断仪器：完美的精确率确保被标记的个案值得进一步评估，而 80%的召回率则意味着在存在其他阅读障碍指征时，筛查结果为阴性也不应排除进一步评估。不过，小数据集（200 幅图像）与自由书写格式带来的局限仍有待未来工作解决。尽管存在上述约束，所提方法证明了基于深度学习的阅读障碍筛查作为一种实用、高效且可及的初步评估工具在教育场景中的可行性。

未来工作应沿三条相互关联的方向推进。第一，标准化的书写规程可减少混杂变异，并支持将模型输出与

专业诊断评估直接比对的临床验证研究。第二，扩充数据集、获得规模更大且更均衡的样本，可增强泛化能力，并支持探索注意力机制或基于 Transformer 的模型等更先进的架构，以便把词级与字符级分析中更细致的序列级书写运动模式整合进段落级表征。第三，应把模型转化为可及的筛查工具，如面向教师的友好界面或用于家庭早期筛查的移动应用，以促进教育工作者与家长的实际采用。

参考文献

[1] 丁汉. 工业机器人精密装配控制技术. 机械工程学报, 2022, 58(02): 1-14.

[2] 周济. 智能制造系统架构与实施路径. 中国机械工程, 2022, 33(03): 253-262.

[3] 张建锋. 云计算分布式调度优化算法. 计算机学报, 2022, 45(03): 567-584.

[4] 聂建国. 装配式混凝土结构抗震性能研究. 土木工程学报, 2022, 55(02): 1-12.

[5] 汤广福. 柔性直流输电装备核心技术. 中国电机工程学报, 2022, 42(05): 1561-1576.

[6] 彭苏萍. 深部煤炭开采围岩控制技术. 岩石力学与工程学报, 2022, 41(02): 225-238.

[7] 李培根. 数字孪生车间构建关键技术. 计算机集成制造系统, 2022, 28(03): 645-658.

[8] Lyon G R, Shaywitz S E, Shaywitz B A. A definition of dyslexia. Annals of Dyslexia, 2003, 53(1): 1-14.

[9] Miciak J, Fletcher J M. The critical role of instructional response for identifying dyslexia and other learning disabilities. Journal of Learning Disabilities, 2020, 53(5): 343-353.

[10] Siegel L S, Hurford D P, Metsala J L, et al. Thoughts on the definition of dyslexia. Annals of Dyslexia, 2025.

[11] Berninger V W, Nielsen K H, Abbott R D, et al. Writing problems in developmental dyslexia: under-recognized and under-treated. Journal of School Psychology, 2008, 46(1): 1-21.

[12] Yang L. Prevalence of developmental dyslexia in primary school children: a systematic review and meta-analysis. Brain Sciences, 2022, 12(2): 240.