
苏州医工所在脑疾病辅助诊断研究方面取得进展

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/20180.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

目前，机器学习方法已被用于脑疾病的辅助诊断，在帮助临床医生及时诊断和干预脑疾病方面，具有重要的临床意义。然而，脑疾病临床数据在样本分布和特征表达方面还存在问题，造成脑疾病辅助诊断模型在实际应用时面临挑战。主要问题包括：样本分布不平衡，主要指不同类别的样本量差异很大，临床数据集上，脑疾病患者组与正常对照组的样本集在数量上有较大的差异，致使样本集只包含少量的脑疾病患者影像数据，但有较多的正常对照影像数据（图1）；二是存在特征表达能力弱的问题，患者和正常组的功能连接特征的细微差异在原始特征空间中不能很好表示，致使样本在原始特征空间中很难线性分开（图2）。上述两方面问题会对模型鲁棒性、泛化性造成影响，并且限制脑疾病辅助诊断精度的提高。因此，在实际应用时需要优化样本分布和特征表示来帮助模型提升分类性能，从而提高脑疾病辅助诊断的准确性、鲁棒性。中国科学院苏州生物医学工程技术研究所戴亚康课题组多年来在神经影像分析领域开展深入研究，在脑网络分析方法、辅助诊断模型构建及临床应用方面取得了一系列研究进展。针对上述问题，课题组研究人员等提出一种基于课程学习和特征表示增强的级联分类模型用于脑疾病辅助诊断方法（SPL-EKM-cmcRVFL+），通过自步学习（self-paced learning）的模式，根据样本难易程度决定样本的加权系数，从易到难地训练分类模型，从而达到平衡学习的目的，同时通过核映射（EKM）的方式，对提取的功能连接特征进行增强，使样本在特征空间中具有更好的分布，从而提升脑疾病分类性能。

SPL-EKM-cmcRVFL+是一种基于级联机器学习框架的脑疾病分类方法（图3）。该框架可以概括为脑网络构建、特征表达增强和级联分类模型构建。脑网络构建模块是从结构磁共振（structural magnetic resonance imaging, sMRI）或功能磁共振（functional magnetic resonance imaging, fMRI）图像中提取结构/功能连接特征，使用t检验和最大相关最小冗余方法进行多重特征选择，从脑网络中选择与疾病相关的结构/功能连接特征。特征表达增强模块是采用经验核映射将原始特征映射到高维特征空间，以增强特征表示，从而使样本在特征空间中具有更好的分布特性。级联分类模型构建模块是将增强的特征输入多列随机向量函数链接网络（RVFL）分类器，其预测标签（用作PI）与增强特征一并送入下一层多列RVFL分类器（RVFL+），形成级联分类框架，通过自步学习（SPL）确定样本权重和其对分类器的贡献，形成最优的分类器模型。在训练过程中，通过调整样本分布，该研究使用自定步长学习来提高模型的鲁棒性、泛化性。在测试阶段，将测试样本输入训练得到的最优模型，可得到脑疾病辅助诊断的结果。相关成果发表在Biomedical Signal Processing and Control上。

针对自闭症谱系障碍数据集（ABIDE）和帕金森病数据集（PPMI），所提出的方法具有优异的分类性能，表明增强的特征表达和优化的样本分布有助于提高脑部疾病辅助诊断效果。与经典的机器学习方法（ELM、SVM和RVFL）相比，该方法在ASD分类中的分类精度分别提高了6.31%-8.36%，AUC提高了0.18%-0.35%。对于PD数据集，分类精度得到了更显著的提高，ACC提高了9.0

5%-12.44%。与单独使用RVFL分类器相比，所提出的优化级联分类模型在ASD和PD分类任务中也具有较好表现。结果表明，所提出的方法可以降低特征聚类分布以及样本分布对模型鲁棒性的影响。该方法在临床帕金森病辅助诊断中也表现出优异的性能，相关成果发表在Neuroscience Letters上。

同时，研究人员还将该方法应用在临床实际高血压脑损伤的辅助诊断中。2017年至2021年，研究人员共收集了700余例不同程度高血压患者的磁共振影像数据，高血压诊断标准符合2017 ACC/AHA (2017 American College of Cardiology/ American Heart Association)，新指南建议将血压分为四类：正常 (Grade 1)、升高 (Grade 2)、高血压1级 (Grade 3) 和2级 (Grade 4)。研究结果表明，特征表达增强的机器学习模型在高血压脑损伤的辅助诊断中也能取得较好的分类结果。可见，该方法在脑疾病辅助诊断临床应用中的潜在价值或可帮助判定高血压等代谢性疾病对脑神经损伤的程度。相关成果发表在BioMedical Engineering OnLine上。

综上所述，该研究提出的样本优化和特征表示增强的级联分类模型，有助于提升脑疾病辅助诊断性能。与现有方法相比，所提出的方法在不同脑疾病分类任务中性能均得到显著提升。同时，在临床实际中，该方法可应用于自闭症谱系障碍、帕金森病、认知功能障碍、高血压脑损伤等脑疾病的辅助诊断。相关研究工作得到国家自然科学基金、江苏省重点研发计划等项目的支持。

论文链接：[1](#)、[2](#)、[3](#)

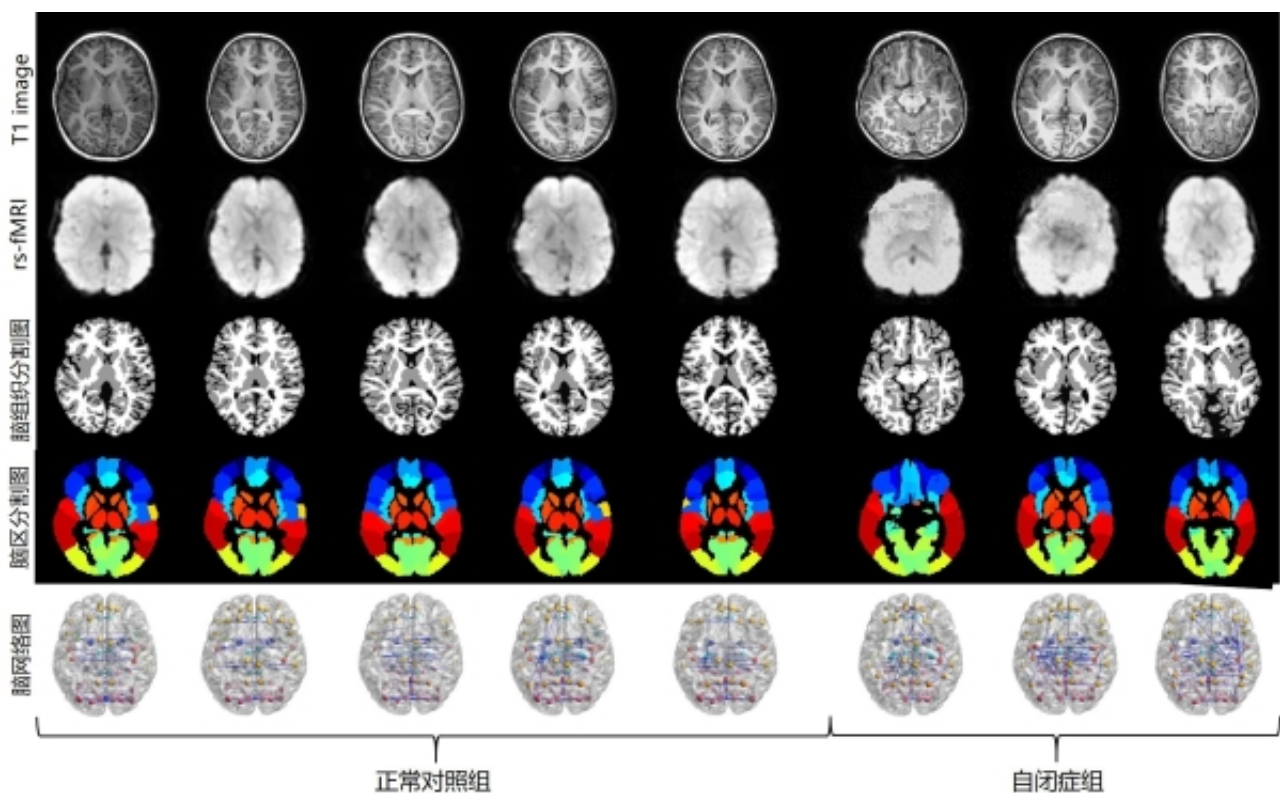


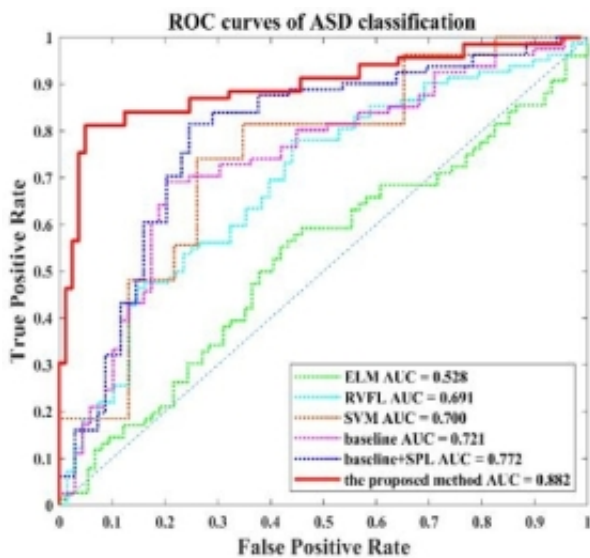
图1 自闭症患者组和正常对照组的脑影像可视化图，由图可知存在样本分布不平衡的问题

图2 ABIDE数据集和PPMI数据集的样本分布t-SNE可视化结果。左侧(a)图和(c)图显示优化样本分布之前患者组样本(橙色)和正常对照组样本(蓝色)的分布重叠在一起,很难分开;右侧(b)图和(d)图显示优化特征表示之后两组样本的分布更容易区分开。

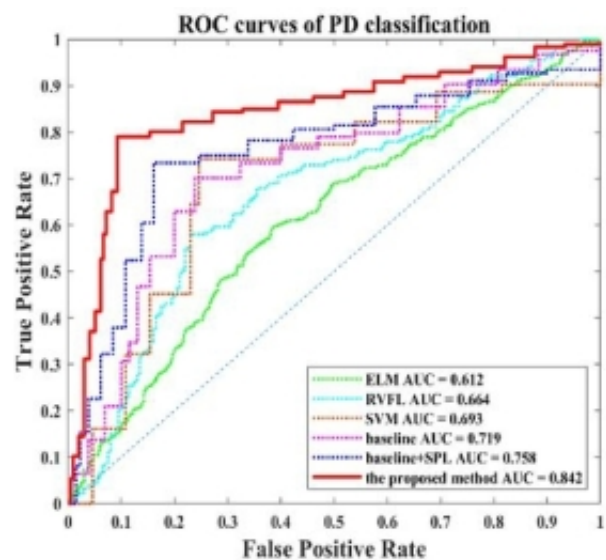
图3 基于课程学习和特征表示增强的脑疾病分类方法框图

Algorithm	ABIDE dataset				PPMI dataset			
	ACC	SEN	SPC	Precision	ACC	SEN	SPC	Precision
ELM	81.67	83.01	83.97	81.03	72.87	72.43	75.29	73.11
SVM	82.46	84.28	88.00	87.38	74.80	75.47	77.06	75.88
RVFL	82.35	85.03	80.00	80.33	76.26	76.22	76.23	75.32
cmcRVFL+	83.48	84.29	83.11	82.09	78.26	75.02	81.27	80.45
SPL-cmcRVFL+	85.56	85.33	86.22	85.22	80.42	77.89	82.82	81.34
SPL-EKM-cmcRVFL+	90.03	83.44	95.56	93.86	85.31	80.09	90.26	89.22

表1 SPL-EKM-cmcRVFL+方法在ABIDE和PPMI数据集上的对比试验结果

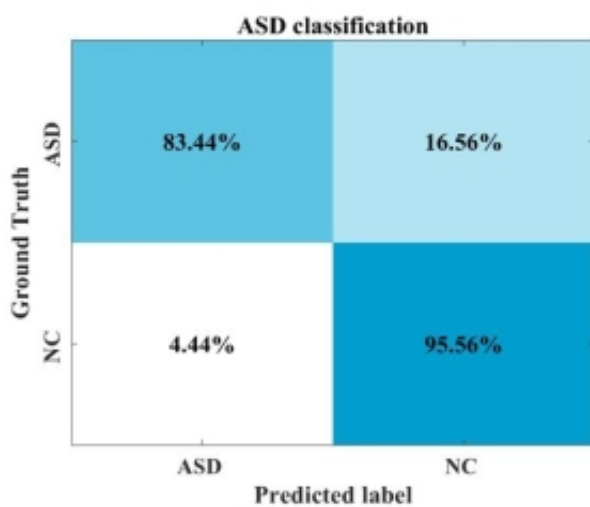


(a)

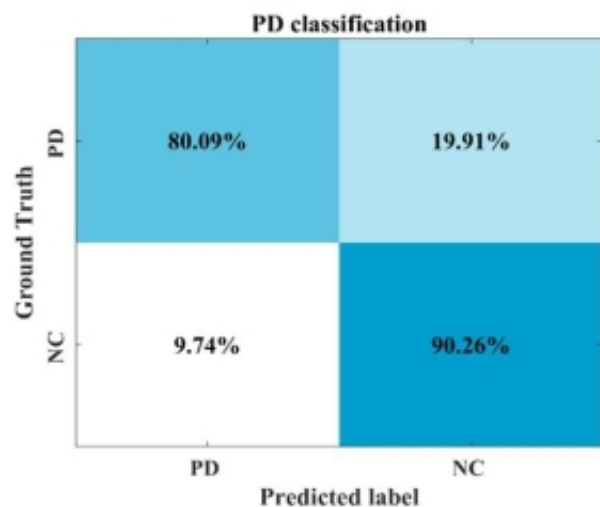


(b)

图4 两个数据集上的ROC曲线的可视化结果。(a)所提出的方法在ABIDE数据集上分类任务中的AUC最高 (AUC = 0.882) ; (b)所提出的方法在PPMI数据集上分类任务中的AUC最高 (AUC = 0.8422) 。



(a)



(b)

图5 所提出的方法在两个数据集上的混淆矩阵可视化结果。(a) 在ABIDE数据集上, 真阳性率为83.44%, 真阴性率为95.56%; (b) 在PPMI数据集, 真阳性率为80.09%, 真阴性率为90.26%。

	GMV	WMV	CSFV	Thickness	Area
Grade 1 and Grade 2					
SVM	60.90 ± 7.21	58.21 ± 5.56	58.90 ± 9.67	54.09 ± 8.96	54.81 ± 8.52
KELM	70.47 ± 6.11	66.40 ± 4.11	67.75 ± 4.95	68.49 ± 4.32	70.49 ± 3.58
KELM	74.34 ± 5.40	69.85 ± 4.57	73.89 ± 5.52	73.32 ± 9.42	69.85 ± 6.63
EKM-KELM	76.73 ± 4.39	73.20 ± 5.13	76.63 ± 6.04	70.52 ± 4.84	75.98 ± 2.18
Grade 1 and Grade 3					
SVM	78.13 ± 6.41	66.47 ± 5.27	61.11 ± 10.89	67.70 ± 8.81	69.27 ± 9.69
KELM	82.24 ± 7.19	72.70 ± 7.42	69.87 ± 4.88	77.99 ± 7.15	74.77 ± 11.24
KELM	89.05 ± 4.40	80.29 ± 7.28	77.46 ± 4.74	78.70 ± 5.97	83.67 ± 8.10
EKM-KELM	93.19 ± 4.01	83.70 ± 6.97	80.87 ± 5.97	80.05 ± 5.56	83.69 ± 8.50
Grade 1 and Grade 4					
SVM	87.65 ± 3.93	72.63 ± 5.72	76.61 ± 5.04	78.61 ± 8.03	71.92 ± 3.56
KELM	88.98 ± 6.20	80.82 ± 7.91	83.48 ± 3.37	80.75 ± 5.52	84.20 ± 5.87
KELM	92.43 ± 3.00	82.25 ± 5.42	86.22 ± 3.78	86.91 ± 5.43	84.22 ± 3.92
EKM-KELM	95.15 ± 3.98	82.93 ± 4.56	88.24 ± 5.50	86.91 ± 5.43	84.27 ± 3.14

表2 特征表达增强的机器学习模型在不同程度高血压脑损伤辅助诊断中的分类性能

图6 不同程度高血压脑损伤组之间，最具辨别力的脑网络特征可视化结果。其中，红色线条表示同侧脑功能的关联程度，灰色线表示对侧脑功能的关联程度，线条的粗细反映其在模型训练中被选择的频次，例如，越粗的线表示被选择的频率越高。

研究团队单位：苏州生物医学工程技术研究所

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发