DOI:10.19651/j.cnki.emt.2518754

基于跨模态特征融合的自闭症筛查研究*

黄嘉瑒 赵英亮 韩星程
(中北大学信息与通信工程学院 太原 030051)

摘 要:由于自闭症儿童早期在视觉注意方面存在异常,为早期干预提供了重要的区分标准。针对自闭症研究中对于模态间的语义对齐、动态交互等关注不足,本研究提出一种融合显著性图与眼动轨迹数据特征的多模态模型,为自闭症的诊断提供一种客观的实现方法。该方法构建了一个双流网络架构:采用 U-Net 特征提取器处理显著性图,利用时序卷积网络对眼动轨迹进行时序建模,为了实现两种不同模态数据间的动态加权融合,引入跨模态注意力机制。并在时序建模的过程中,同时进行眼动轨迹预测,额外将预测误差作为区分特征引入分类过程中,来提升模型的分类性能。通过对比实验验证,所提出的模型在自闭症早期筛查任务中取得了 98.89%的准确率。

关键词:自闭症;显著性图;眼动轨迹;跨模态注意力机制;分类
中图分类号:TN911.7 **文献标识码**:A **国家标准学科分类代码**:510.40

Research on autism screening based on cross-modal feature fusion

Huang Jiayang Zhao Yingliang Han Xingcheng
(School of Information and Communications Engineering, North University of China, Taiyuan 030051, China)

Abstract: Because children with autism show abnormalities in visual attention in their early years, it provides an important distinguishing criterion for early intervention. In view of the insufficient attention paid to semantic alignment and dynamic interaction between modalities in autism research, this study proposes a multimodal model that integrates saliency maps and eye movement trajectory data features, providing an objective implementation method for the diagnosis of autism. This method constructs a dual-stream network architecture: the U-Net feature extractor is used to process the saliency map, and the temporal convolutional network is utilized to conduct temporal modeling of the eye movement trajectory. To achieve dynamic weighted fusion between two different modal data, a cross-modal attention mechanism is introduced. During the process of time series modeling, eye movement trajectory prediction is carried out simultaneously. Additionally, the prediction error is introduced as a distinguishing feature into the classification process to enhance the classification performance of the model. Through comparative experiments, it was verified that the proposed model achieved an accuracy rate of 98.89% in the early screening task of autism.

Keywords: autism; saliency map; eye-tracking; cross-modal attention mechanism; classification

0 引 言

自闭症谱系障碍(autism spectrum disorder, ASD)是一种复杂的神经发育障碍,通常隐匿存在于婴幼儿阶段,较难在早期发现,主要表现为社会交流障碍、感知觉障碍与兴趣受限与刻板行为[1]。由于自闭症患病的复杂性,在每个个体上的表现都不尽相同,为后续诊断和治疗带来了很大的困难。到目前为止,并没有完全能治愈自闭症的方法,但通过一系列的干预与治疗,可以明显地改善其发病的症状[2]。随着时代的发展,自闭症的诊断也呈现出多样化的形式。通过评估行为或量表形式的传统自闭症的诊断方法具有较大的主观性、耗时较长,且难以在较早时期发现[3],因此结合科学技术的诊断方法就应运而生。由于自闭症儿童在婴幼儿期就已经表现出明显的视觉偏差和社交注意力异常[4],眼动追踪技术能够在儿童时期检测到这些变化,为早期诊断提供了可能[5]。眼动追踪技术提供了一种非侵入式的研究方法,这种研究方法降低了测试者的不适感,增加了数据的可用性[6]。

利用人工智能技术的自闭症诊断方法逐渐流行并趋于成熟。加州大学与肯塔基大学[7]基于儿童在注视图像时的

收稿日期:2025-05-07
* 基金项目:国家自然科学基金(62203405)、山西省应用基础研究计划基金(202303021212206)、山西省重点研发计划(202202110401015)项目资助

扫视路径数据提出了两种分类方法：一种是合成眼动方法，另一种是基于图像的方法。Praveena 等[8]采用卷积神经网络(convolution neural network，CNN)分别对 ASD 儿童与正常发育(typically developing，TD)儿童的注视点图提取空间特征从而实现分类。Mumenin 等[9]开发了一种轻量级的内卷神经网络(involutional neural network，INN)架构来分别从扫描路径、热图与注视图图像中诊断 ASD。以上是使用单一模态数据来进行研究分类的，具有一定的局限性。为了提供更丰富的分类信息，使用多模态融合的分析方法，但利用眼动数据进行多模态融合的研究较少。迈阿密大学[10]提出了一种联合网络模型，根据眼动数据的注视点坐标从显著性图中提取出一个图像块，利用 CNN 提取视觉特征，再将视觉特征与注视时长进行拼接，最终使用长短期记忆网络(long short-term memory，LSTM)进行分类。张小帅[11]提出一种加入图像目标信息的融合网络，将根据眼动数据生成眼动数据图、显著性预测模型得到的显著性图与通过 YOLOv8(you only look once version 8)得到的目标检测图三者叠加为一张融合图像，再输入残差卷积神经网络(residual neural network，ResNet)进行分类。Benabderrahmane 等[12]研究出一种新型机器学习分类模型，利用 CNN 提取眼动路径可视化后的图像特征，门控循环单元(gated recurrent unit，GRU)来处理眼动距离序列与持续时间序列，通过人工神经网络(artificial neural network，ANN)将两种模态特征融合后再进行分类。Colonnese 等[13]提出一种融合多模态眼动数据的网络框架，利用 CNN 分别处理原始图像与注视点图，利用 LSTM 网络处理时间序列，将三者经过线性变换后融合起来进行分类，最后引入训练数据影响(training data influence，TracIn)方法，对训练样本进行数据归因分析，提升模型的性能。针对以上的研究方法，了解到关于现有的自闭症研究方法中存在一定的局限性，没有考虑到图像模态与眼动数据模态之间的对齐关系，忽略了不同模态间深层语义的动态交互机制，只是进行简单的静态特征拼接操作。

由此本文提出了一种多模态深度学习模型，通过使用跨模态注意力机制来实现显著性图与眼动数据的动态加权融合。此外，利用时序卷积网络(temporal convolutional network，TCN)在进行时序建模过程中实现多任务学习，同时进行轨迹预测，将眼动轨迹预测误差作为分类辅助特征，进一步提升了模型的分类性能。通过多模态数据的融合与深度学习技术的应用，建立了一种客观、准确的自闭症早期诊断方法。

1　模型与原理

1.1　模型框架

自闭症早期筛查模型算法主要包含两个分支：显著性图与眼动轨迹数据处理分支。通过复现已有的显著性预测模型对本研究使用数据集中的图像生成显著性图，然后利用 U-Net 特征提取网络对显著性图进行处理；利用 TCN 对眼动数据进行特征提取并同时进行眼动轨迹预测。通过跨模态注意力机制实现显著性图特征与眼动轨迹特征的选择性交互，再与计算的预测误差特征结合起来送入分类器中，自闭症筛查模型架构如图 1 所示。

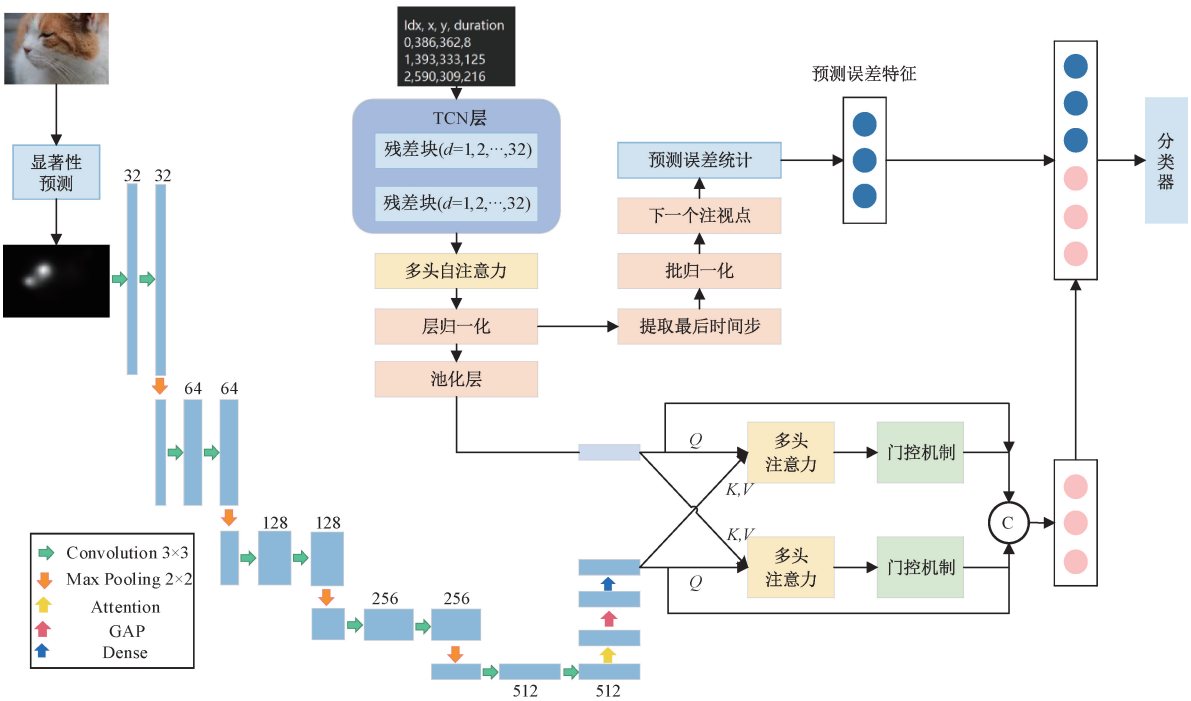


图 1　基于眼动特征的自闭症筛查模型架构

Fig. 1　Architecture of an autism screening model based on eye movement feature

1.2 显著性预测

显著性图体现了图像中不同像素点对人眼注视的吸引程度[14]。由于 ASD 儿童与 TD 儿童对事物的关注点不同,在视觉显著性预测过程中,对图像感兴趣的区域也不尽相同,因此对正常人视觉显著性预测的结果可以作为区分自闭症儿童的依据。

本文选择使用 Marcella 等[15]提出的显著性预测模型得到数据集中图像对应的显著性预测图。该模型引入了注意力机制,并且进行高斯先验学习,使模型更好地模拟人类视觉注意的中心偏差特点,从而提升显著性预测的性能。由于训练模型使用的数据集仅有 300 张图像,在训练时无法进行充分学习,所以优先使用 SALICON[16]数据集进行训练,该数据集为典型的显著性预测数据集,可以让模型学习到更丰富的视觉特征,然后再进行微调来提高模型在该数据集上的泛化能力。显著性图预测网络模型如图 2 所示。

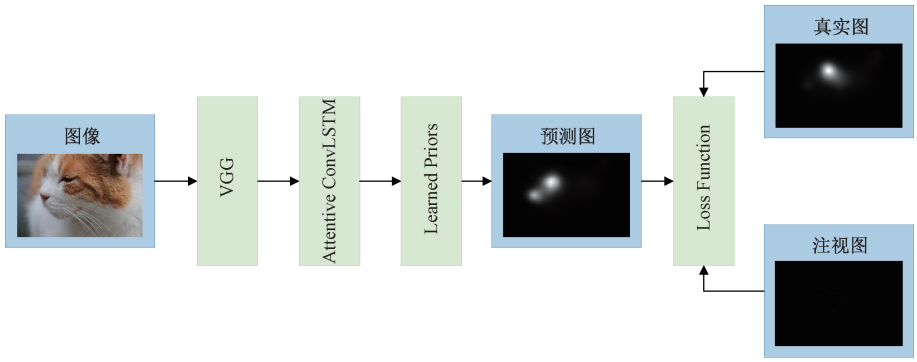


图 2 显著性图预测网络模型

Fig. 2 Significance map prediction network model

1.3 自闭症分类

针对自闭症筛查系统的实现,使用到的数据包括显著性图片与眼动数据,分别对这两种不同类型的数据进行分析,因此构建一种多模态融合模型,从而可以更加准确地进行自闭症儿童与正常儿童的区分。多模态模型采用双流深度学习网络结构,分别对显著性图与眼动数据进行分析,再通过跨模态注意力机制将提取到的两种模态的特征融合起来。此外,模型在对眼动数据提取特征时,还要进行 ASD 儿童与 TD 儿童的眼动轨迹预测,进而利用预测误差来进行辅助诊断,将误差特征与通过注意力机制处理后的特征融合起来,再将其输送至分类器中,实现自闭症的有效分类。

1)显著性图特征提取

在双流多模态模型中关于显著性图处理的分支中,将显著性预测模块生成的显著性图作为该部分的输入,为了可以充分地提取图像中的信息,采用改进的 U-Net 架构特征提取网络来对图像进行处理。U-Net 网络具有强大的多尺度特征提取能力与空间信息保留能力,因此比较适合处理具有模糊边界以及空间对比明显的特点的灰度显著性图像。该模型中的 U-Net 网络仅采用了传统网络中关于编码器的内容,专门用来进行特征提取,通过 4 个编码层渐进式下采样来提取多尺度的空间信息。由于 ASD 个体的视觉注视差异可能会体现在显著性图的全局信息中,而位于编码器最后的桥接层中拥有很大的感受野,可以帮助获取这些全局信息,并且在桥接层中引入空间注意力机制,这可以帮助模型识别显著性图中与 ASD 诊断最相关的视觉注意信息。随后将提取的特征通过全局平均池化操作来压缩空间维度,并通过全连接层生成最终的显著性图特征表示,参与多模态融合。

2)眼动数据处理

眼动数据是由注视点坐标与持续时间构成的,具有明显的时间顺序与局部信息等特征,针对这些特征,在双流模型中采用 TCN 对眼动数据进行时序建模,同时通过多任务学习机制来进行眼动轨迹的预测。对于实际状况下的眼动轨迹来说,是一个相对动态的过程,但眼动数据只是反映了静态的结果,因此决定构建一个眼动轨迹预测模型,来实现动态的分析过程,同时计算眼动轨迹的预测误差。由于 TD 个体的视觉轨迹通常是有规律的,并且具有更强的时序性,因此其眼动轨迹更容易预测,计算出的预测误差就会相对较小;而 ASD 个体在视觉探索的过程中通常表现出较强烈地随机性、重复性与跳跃性,因此较难预测其视觉轨迹,计算出的预测误差相对来说就会大一些。因此可以将预测误差作为一个额外特征来进行自闭症个体的识别。

TCN 网络常常被用来处理序列数据,相较于传统的 LSTM 网络与 GRU,TCN 具备更强大的局部时间建模能力。模型中 TCN 网络通过扩张因果卷积实现了对眼动数据的短期动态模式与长期依赖关系的建模:

$$y(t) = \sum_{i=0}^{k-1} f(i) \cdot x(t - d \cdot i) \tag{1}$$

其中,$y(t)$ 表示当前输出,$x(t)$ 表示输入序列,$f(i)$ 表示卷积核权重,d 表示扩张率,k 表示卷积核大小。

因果性的实现通过在前向传播时对输入序列的左侧进行零填充来对时序数据建模进行约束,确保某一时刻的

输出仅依赖当前时刻以及之前时刻的输入序列，防止未来信息的泄露；扩张性的实现通过进行指数级感受野扩展，使用扩张率为(1，2，4，8，16，32)的多尺度卷积结构，低扩张率的结构来捕捉眨眼、快速注视等短时特征，高扩张率的结构来建模持续注视、注意力转移等长期模式。本文使用的 TCN 残差块包括扩张因果卷积、层归一化、Relu 激活函数以及 SpatialDropout，图 3 展示了残差块的结构。

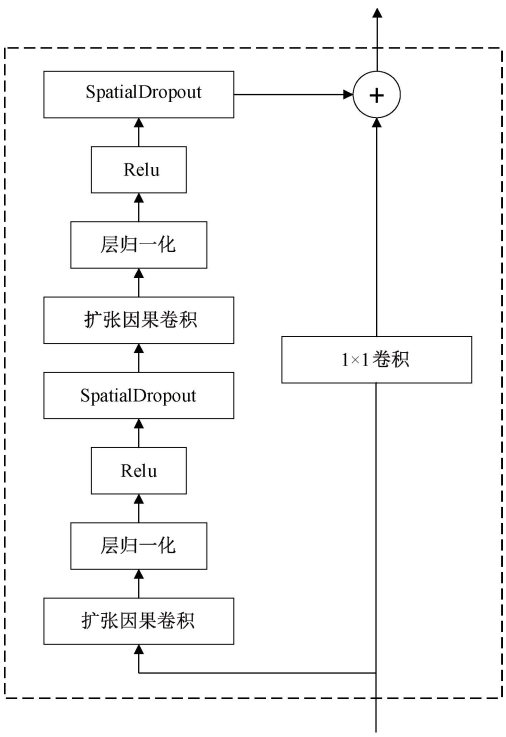


图 3　残差块结构

Fig. 3　Residual block structure

此外在 TCN 中引入了多头自注意力机制，将 TCN 层输出再进行全局注意力建模，强化了时间序列中的长距离依赖建模能力，有效克服了 TCN 原生感受野增长受限的缺陷，提升了对序列结构例如注视模式中重复、回视等行为的捕捉能力。

为实现眼动轨迹的预测，模型使用了一个基于多任务学习策略的 TCN 模块。眼动数据通过 TCN 层使特征提取任务与轨迹预测任务共享底层特征，让网络可以学习到对两个任务都有用的眼动特征表示。在输出结果时，特征提取任务使用全局平均池化与最大池化获取整个序列的信息，再通过全连接层得到最终特征表示；眼动轨迹预测任务使用 TCN 输出的最后一个时间步，通过全连接层预测下一个注视点坐标。在训练眼动轨迹预测阶段时，模型采用滑动窗口的方式，充分利用眼动序列中的时间信息，扩充了训练数据，同时保留了眼动轨迹中的动态变化信息，使模型可以学习局部的注视行为。使用均方误差(mean square error，MSE)作为损失函数，用于评估真实值与预测值之间的差异：

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

其中，n 为样本总数，y_i 为真实值，$\hat{y}_i$ 为预测值。

在完成训练后，对每条眼动序列进行逐步预测，并计算预测位置与真实注视点的欧氏距离，进而统计每条样本的平均误差、标准差、最大误差和最小误差，这些预测误差特征可以反映观测者眼动轨迹预测的难易程度。

3)跨模态注意力机制

为了有效地获取显著性图与眼动数据的信息，本文使用跨模态注意力机制[17]对这两种特征进行选择性交互与特征融合。该机制首先将两种模态的特征投影到一个共同的特征空间中：

$$\boldsymbol{X} = \boldsymbol{W}\boldsymbol{X}' + b \tag{3}$$

其中，$\boldsymbol{X}'$ 表示显著性图特征与眼动轨迹特征，$\boldsymbol{W}$ 表示投影权重矩阵，$\boldsymbol{X}$ 表示映射到共同特征空间的特征，b 为偏置项。

该注意力机制包括两个注意力层，第一个注意力层为显著性图特征与眼动数据特征进行交互，显著性图特征作为查询(query)，眼动数据的特征作为键(key)和值(value)，得到注意力加权特征：

$$\boldsymbol{A}_{s\to e} = \mathrm{softmax}\left(\frac{(\boldsymbol{X}_s\boldsymbol{W}_s^Q)(\boldsymbol{X}_e\boldsymbol{W}_e^K)^{\mathrm{T}}}{\sqrt{d_k}}\right)(\boldsymbol{X}_e\boldsymbol{W}_e^V) \tag{4}$$

第二个注意力层为二者进行反向交互，眼动数据特征作为查询(query)，显著性图的特征作为键(key)和值(value)，得到眼动特征到显著性图特征的注意力加权特征表示。

同时在该注意力机制中引入了自适应门控机制[18]，用于调节两种模态之间的信息流，使模型可以根据特定的输入动态调整每种模态的重要性。门控机制先通过 sigmoid 函数计算特征重要性：

$$\boldsymbol{G} = \sigma(\boldsymbol{W}_g\boldsymbol{A} + b_g) \tag{5}$$

分别得到显著性图与眼动数据的门控值，然后，原始特征与门控值进行主元素相乘得到显著性图与眼动数据门控后的特征：

$$\boldsymbol{X}^{gated} = \boldsymbol{X}' \odot \boldsymbol{G} \tag{6}$$

通过这种方式，眼动数据和显著性图之间的交互关系会得到加强，这两种特征都可以选择性地关注另一种模态中的相关特征，同时保留自己的原始特征。最后将所有的特征串联起来，输出一个表示眼动轨迹和显著性图相互作用后的最终特征向量。

4)分类器

对于分类器，本模型选择使用多层感知机(multilayer perceptron，MLP)与 XGBoost(extreme gradient boosting)两种分类器进行分类。MLP 是一个由全连接层构成的前馈神经网络，包括输入层、隐藏层以及输出层。XGBoost 是一种基于梯度提升决策树(gradient boosting decision tree，GBDT)的增强学习框架，通过集成多个弱分类器来构建一个强分类器[19]。

2 实验

2.1 数据集

本文使用由 Saliency4ASD 挑战大赛组织者提供的数据集[20],该数据集旨在探索 ASD 儿童与 TD 儿童在视觉注意方面上的差异性。该数据集包含了 300 张不同类型的图像、相对应的显著性图真值标定以及通过眼动仪获取的被测试者观察图像时的眼动数据。300 张图像分别被 14 名正常发育的儿童和 14 名被诊断为自闭症的儿童观察。这些图像包含了多种类型,包括人物、动物以及风景图等等,从各个方面模拟了日常生活中的视觉刺激,从而能够更全面地捕捉两类儿童在注视行为上的差异。该数据集所选用的测试者均为 5～12 岁的儿童,为自闭症的早期研究提供了重要支持。

2.2 评估指标

对于分类模型地评估,本文选取了 4 项经常被用于二分类问题的评估指标:准确率(accuracy)、精确率(precision)、召回率(recall)以及 F1 分数($f1score$)[21]。

准确率为正确预测的样本数占总样本数的比例:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

其中,TP 表示真实类别为正类,并且被正确预测为正类;TN 表示真实类别为负类,并且被正确预测为负类;FP 表示真实类别为负类,但被错误预测为正类;FN 表示真实类别为正类,但被错误预测为负类。

精确率表示在被预测为正类的样本中,真实为正类的的比例:

$$precision = \frac{TP}{TP + FP} \tag{8}$$

召回率表示在真实为正类的样本中,被正确预测为正类的比例:

$$recall = \frac{TP}{TP + FN} \tag{9}$$

F1 分数表示精确率与召回率的调和平均值:

$$f1score = 2\frac{precision \cdot recall}{precision + recall} \tag{10}$$

3 实验结果与分析

3.1 实验结果

在分类模型的算法中,同样为了可以更好地评估模型的泛化能力,按照 7∶1.5∶1.5 的比例把实验用的数据集分成训练集、验证集与测试集。在该分类算法中,训练周期被设置为 100 轮,学习率被定为 0.001。为了高效控制训练过程,还引入了回调机制,用早停法来监控验证集上的损失。

为了验证该模型的有效性,进行单一模态的实验,获得仅使用眼动数据实现分类的结果,并且在 Saliency4ASD 数据集上比较本文的两种分类器模型与其他文献提出的多模态 ASD 分类方法的结果,其性能对比如表 1 所示。

表 1 不同模型的分类性能对比

Table 1 Comparison of classification performance of different models

模型	准确率	精确率	召回率	F1 分数
Single mode	0.744 4	0.823 5	0.622 2	0.708 9
SP-ASDNet[10]	0.579 0	0.592 1	0.562 6	0.569 7
ResNet-SVM[11]	0.938 0	0.948 3	0.927 0	0.940 0
GBAC[13]	0.943 5	0.931 0	0.932 0	0.931 5
XGBoost	0.955 6	0.992 0	0.918 5	0.953 8
MLP	0.988 9	0.992 5	0.985 2	0.988 8

针对眼动轨迹预测任务的实现,本文利用了两种预测方法来进行实现,分别是 LSTM 与 TCN。为了了解两种不同预测眼动轨迹的方法对最后分类结果的影响,使用之前提出的 4 个指标来进行定量分析,两种方法的分类结果如表 2 所示。从表 2 的数据中可以得出在使用 TCN 网络进行多任务学习预测眼动轨迹时,可以达到更高的分类准确率。

表 2 三种预测眼动轨迹方法的分类结果对比

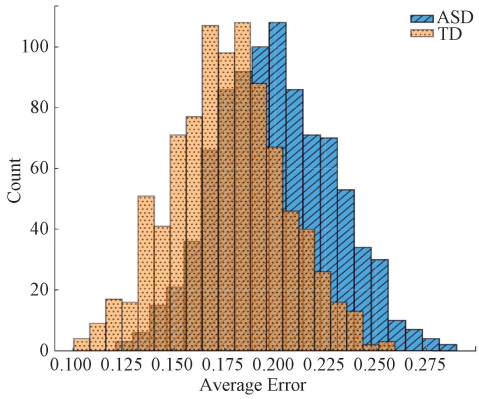
Table 2 A comparison of classification results of three methods for predicting eye movement trajectories

模型	准确率	精确率	召回率	F1 分数
无预测任务	0.963 0	0.992 1	0.933 3	0.961 8
LSTM	0.977 8	0.992 4	0.963 0	0.977 4
TCN	0.988 9	0.992 5	0.985 2	0.988 8

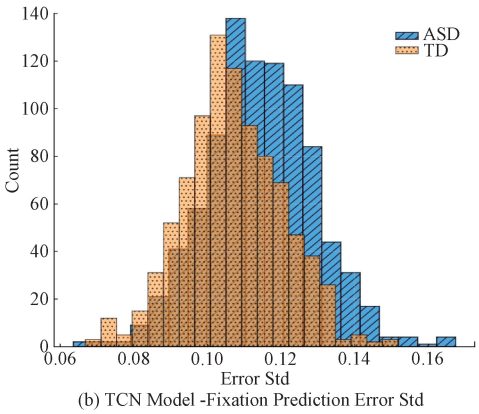
在对眼动轨迹进行预测时,得到了眼动轨迹预测的平均预测误差值,ASD 组的平均预测误差约为 0.20,而 TD 组的平均预测误差约为 0.18,由于在处理眼动数据时对其进行了标准化处理,因此 0.02 的差距可以表示它们之间存在的差异。ASD 组与 TD 组的预测误差分布可视化图如图 4 所示。

顶部图展示了平均预测误差的分布,ASD 组(点状)显示出较高的平均预测误差,其分布集中在 0.21～0.23;而 TD 组(斜线状)的误差分布集中在 0.17～0.19,明显小于 ASD 组的平均预测误差。底部图展示了误差标准差的分布,其中 ASD 组标准差以 0.12 为中心,表示预测误差的可变性较高;而 TD 组以 0.10 为中心,表现出较低的可变性。以上误差分布表明了 ASD 个体的眼球运动具有不规律性,眼动轨迹比较难预测一些,TD 个体则表现出更规律的眼球运动模式以及眼动轨迹的可预测性。

为了更好地分析模型性能,本文通过使用混淆矩阵,直观地展示了分类模型预测的结果与真实标签之间的对应关系,反映了正确分类与错误分类的样本数量,模型在验证集上的混淆矩阵可视化图如图 5 所示。验证集包含 270 个样本,其中 ASD 样本与 TD 样本各 135 个,由图可



(a) TCN Model -Fixation Prediction Average Error



(b) TCN Model -Fixation Prediction Error Std

图 4　ASD 与 TD 预测误差分布图

Fig. 4　The distribution of prediction errors for ASD and TD

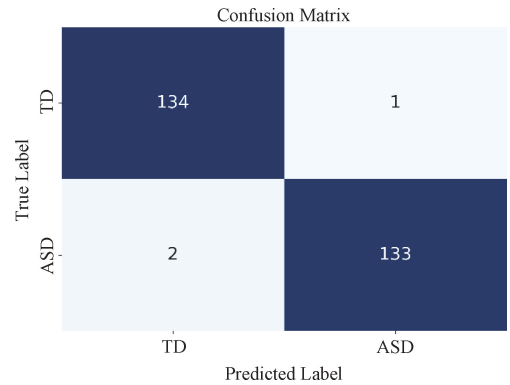
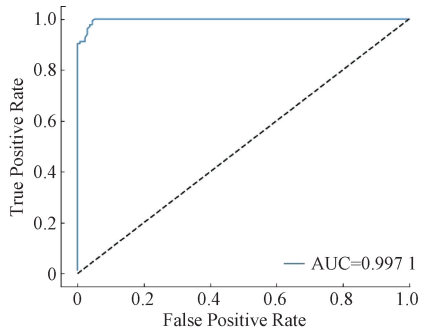


图 5　混淆矩阵

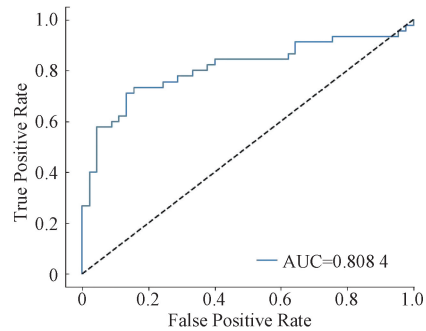
Fig. 5　Confusion matrix

知，有 134 个 TD 样本与 133 个 ASD 样本被正确分类，而被错误分类的样本数分别为 1 和 2。图 5 中左上角表示 TN，左下角为 FN，右上角为 FP，右下角为 TP，从图 5 中可以得出对角线 TN 与 TP 的数值较高，说明模型的性能较好。

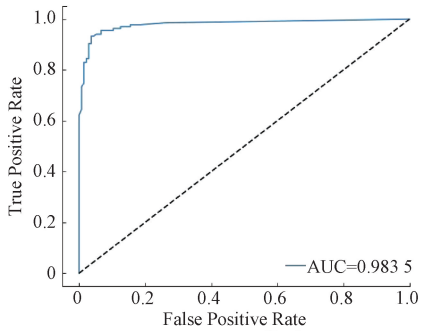
ROC 曲线被用来体现模型在不同分类阈值下区分正负样本的能力，其横轴代表假正率，纵轴代表真正率[22]。ROC 曲线下方的面积表示 AUC 的值，是评估模型性能的一种量化指标，其取值范围介于 0～1。AUC 值越趋近于 1，意味着模型的分类表现越优异。ROC 曲线图如图 6 所示。图 6(a)表示本文中的多模态模型的 ROC 曲线，多模态模型的 AUC 值为 0.997 1，非常接近于 1，表明该模型能



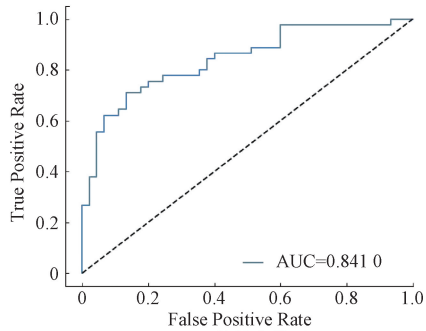
(a) 本文模型
(a) The model of this article



(b) CNN-LSTM模型
(b) CNN-LSTM model



(c) GBAC模型
(c) GBAC model



(d) 单模态模型
(d) Single-mode model

图 6　不同模型的 ROC 曲线对比图

Fig. 6　ROC curve comparison chart of different models

以极高的敏感度区分 ASD 和 TD 个体;图 6(b)与(c)分别表示模型 SP-ASDNet 与模型 GBAC 的 ROC 曲线图;图 6(d)表示单一模态下的 ROC 曲线图。以上均说明本文多模态模型存在较好的分类性能。

3.2 实验分析

由于神经网络在提取特征时,是一个非常复杂的过程,为了理解模型在做出分类决策时关注的显著性图与眼动数据包含的信息,分别对两种特征进行可解释性分析。通过对显著性图进行遮挡敏感度分析,来帮助模型了解显著性图在分类时最依赖哪些区域的信息。生成的敏感度分析图包含 3 个部分:原始显著性图、遮挡敏感度与叠加图,如图 7 所示。

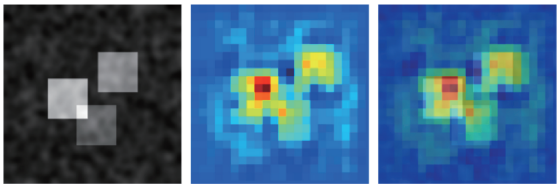


图 7 敏感性分析图

Fig. 7 Sensitivity analysis chart

左侧的原始显著性图像中心区域有几个较亮的块状区域,这些代表了被测试者视觉注意的焦点区域;中间为遮挡敏感度图,展示了模型对显著性图不同区域的敏感程度,图中红色与黄色的位置表示对模型的预测结果影响较大,而蓝色区域表示影响较小;而右侧叠加图将原始显著性图与敏感度热图叠加在一起,直观地反映出模型关注的关键区域与原始显著性图的关系。

对眼动数据进行可视化眼动轨迹模式,来了解 ASD 儿童与 TD 儿童在观察同一张图像时的眼动模式,眼动轨迹可视化图如图 8 所示。由图 8 可知,ASD 个体的眼动轨迹显示了较分散的视觉注意模式,视线集中停留在中间区域,但明显有向四周扩散的趋势;而 TD 个体的眼动轨迹却显示出了更系统的视觉探索,视线停留点大多在图像的偏左下区域,呈现出比较集中的范围。关于注视时长的分布可视化如图 9 所示,由图 9 可知,在 ASD 个体的注视时长分布图中,早期注视点有较高的停留时间,但整体的分布并不均匀;TD 个体的注视时长分布中存在多个注视点有较长的停留时间,时长分布更加均匀。以上这些差异均反映出了两类个体在视觉注意方面的不同之处,TD 个体表现出更集中、更规律的视觉探索模式,而 ASD 个体却表现出更分散、更不规则的视觉注视行为,这些均与自闭症患者的临床观察特征相符合。

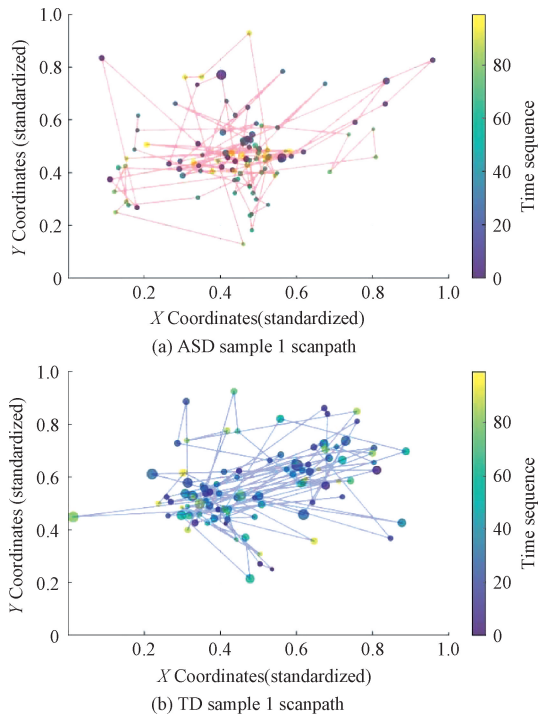


图 8 眼动轨迹可视化图

Fig. 8 Eye movement trajectory visualization

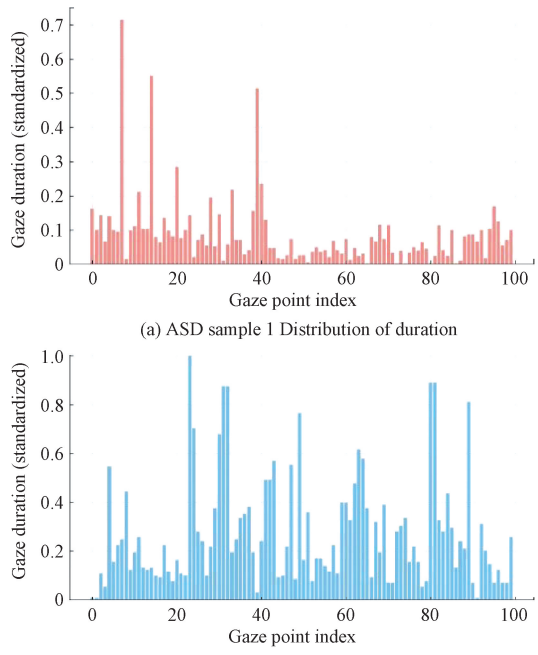


图 9 注视时长分布可视化图

Fig. 9 Visualization of the distribution of gaze duration

4 结 论

针对传统自闭症诊断方法的主观性与局限性,提出一种新的模型来实现对自闭症的早期筛查。由于自闭症儿童在视觉注意方面与正常发育儿童存在显著差异,该模型采用将显著性图与眼动数据结合起来,通过分析观察者的眼动特征来实现自闭症儿童的分类。同时将眼动轨迹预测误差特征作为辅助诊断 ASD 的额外特征,来提升模型的分类性能。该模型的分类准确率达到了 98.89%,AUC 值达到

了 0.998 5，经过与其他分类模型进行对比，各项指标均有所提升，表明该多模态分类模型具有较好的分类性能。

参考文献

[1] 陈柯颖，柴森潮，宫之羽，等. 自闭症儿童中医体质与临床表现相关性的研究[J]. 中国中医药现代远程教育，2024，22(3)：81-84.
CHEN K Y，CHAI S CH，GONG ZH Y，et al. Study on the correlation between traditional chinese medicine constitution and clinical manifestations in autistic childrenclinical manifestations in autistic children[J]. Chinese Medicine Modern Distance Education of China，2024，22(3)：81-84.

[2] 王亚民，潘礼正，闵云霄，等. 基于时-频-空域特征表征方法的自闭症儿童诊断[J]. 电子测量技术，2024，47(23)：76-83.
WANG Y M，PAN L ZH，MIN Y X，et al. Diagnosis of autistic children based on temporal-spectral-spatial feature representation[J]. Electronic Measurement Technology，2024，47(23)：76-83.

[3] 曾泳添，陈日玲，农雪艳，等. 数字疗法在自闭症筛查到干预的临床研究进展与挑战[J]. 中国全科医学，2025，28(14)：1702-1708.
ZENG Y T，CHEN R L，NONG X Y，et al. Clinical research progress and challenges of digital therapeutics from screening to intervention in autism spectrum disorder[J]. Chinese General Practice，2025，28(14)：1702-1708.

[4] CONGIU S，CONGIU S，DONEDDU G，et al. Attention toward social and non-social stimuli in preschool children with autism spectrum disorder：A paired preference eye-tracking study[J]. International Journal of Environmental Research and Public Health，2024，21(4)：421.

[5] RAHMAN M K，MOHAN S M C. Investigation of eye-tracking scan path as a biomarker for autism screening using machine learning algorithms[J]. Diagnostics，2022，12(2)：518.

[6] HOU W，JIANG Y，YANG Y，et al. Evaluating the validity of eye-tracking tasks and stimuli in detecting high-risk infants later diagnosed with autism：A meta-analysis[J]. Clinical psychology review，2024，112：102466.

[7] WU C，SIDRAH L，CHEUNG S，et al. Predicting autism diagnosis using image with fixations and synthetic saccade patterns.[J]. IEEE International Conference on Multimedia and Expo workshops，2019：647-650，DOI：10.1109/ICMEW.2019.00125.

[8] PRAVEENA K N，MAHALAKSHMI R. Classification of autism spectrum disorder and typically developed children for eye gaze image dataset using convolutional neural network[J]. International Journal of Advanced Computer Science and Applications（IJACSA），2022，13（3），DOI：10.14569/IJACSA.2022.0130345.

[9] MUMENIN N，YOUSUF A M，NASHIRY A M，et al. ASDNet：A robust involution-based architecture for diagnosis of autism spectrum disorder utilising eye-tracking technology[J]. IET Computer Vision，2024，18(5)：666-681.

[10] TAO Y，SHYU M L. SP-ASDNet：CNN-LSTM based ASD classification model using observer scanpaths[C]. 2019 IEEE International conference on multimedia & expo workshops（ICMEW）. IEEE，2019：641-646.

[11] 张小帅. 基于眼动数据和视觉信息的自闭症筛查算法研究[D]. 武汉：长江大学，2024.
ZHANG X SH. Research on autism screening algorithms based on eye tracking data and visual information[D]. Wuhan：Yangtze University，2024.

[12] BENABDERRAHMANE B，GHARZOULI M，BENLECHEB A. A novel multi-modal model to assist the diagnosis of autism spectrum disorder using eye-tracking data[J]. Health Information Science and Systems，2024，12(1)：40.

[13] COLONNESE F，DI LUZIO F，ROSATO A，et al. Enhancing autism detection through gaze analysis using eye tracking sensors and data attribution with distillation in deep neural networks[J]. Sensors，2024，24(23)：7792.

[14] 李可新，何丽，刘哲凝，等. 基于跨模态特征融合的RGB-D显著性目标检测[J]. 国外电子测量技术，2024，43(6)：59-67.
LI K X，HE L，LIU ZH N，et al. RGB-D salient object detection based on cross-modal feature fusion[J]. Foreign Electronic Measurement Technology，2024，43(6)：59-67.

[15] MARCELLA C，LORENZO B，GIUSEPPE S，et al. Predicting human eye fixations via an lstm-based saliency attentive model[J]. IEEE Transactions on Image Processing，2018，27(10)：5142-5154.

[16] JIANG M，HUANG S，DUAN J，et al. Salicon：Saliency in context[C]. IEEE conference on computer vision and pattern recognition，2015：1072-1080.

[17] 邢致恺，何怡刚，姚其新. 基于多模态信息融合的变压器在线故障诊断方法[J]. 电子测量与仪器学报，2024，38(9)：95-103.

XING ZH K, HE Y G, YAO Q X. Transformer online fault diagnosis method based on multi-modal information fusion[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(9): 95-103.

[18] LI X, GWAN C, ZHAO S, et al. Multimodal temperature prediction for lithium-ion battery thermal runaway using multi-scale gated fusion and bidirectional cross-attention mechanisms[J]. Journal of Energy Storage, 2025, 116:116098.

[19] 杨勇,胡东,代浩,等. 基于实域粗糙集和 NRBO-XGBoost 的变压器故障诊断[J]. 电子测量技术, 2025, 48(5): 30-39.
YANG Y, HU D, DAI H, et al. Research on transformer fault diagnosis method based on real domain rough set and NRBO-XGBoost[J]. Electronic Measurement Technology, 2025, 48(5): 30-39.

[20] DUAN H, ZHAI G, MIN X, et al. A dataset of eye movements for the children with autism spectrum disorder[C]. 10th ACM Multimedia Systems Conference, 2019: 255-260.

[21] 沈胤宏,郑秀娟,张畅. 基于融合特征与优化随机森林的眼动模式识别[J]. 电子测量技术, 2023, 46(15): 10-17.
SHEN Y H, ZHENG X J, ZHANG CH. Eye movement pattern recognition based on fused features and optimized random forest[J]. Electronic Measurement Technology, 2023, 46(15): 10-17.

[22] FAWCETT T. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2005, 27(8):861-874.

作者简介

赵英亮(通信作者),博士研究生,副教授,主要研究方向为工业 CT 图像三维重建技术研究、多维信号目标跟踪等。
E-mail:zhaoyl18@nuc.edu.cn

黄嘉瑒,硕士研究生,主要研究方向为图像与眼动数据的处理。
E-mail:h15735362672@163.com