✧专论

Intelligent auxiliary model based on Swin-Transformer for diagnosing fetal ocular malformations

TAO Xiongjie[1], DI Zhenwei[1], LIANG Bocheng[2], OUYANG Shuyuan[3], GUO Hui[1], HE Jie[1], TONG Rui[4], CHEN Jiaxi[1], XIE Di[1], ZHAO Yingli[1], QIN Ni[1], LI Shengli[2*]

(1. Guangxi Key Laboratory of Machine Vision and Intelligent Control, Wuzhou University, Wuzhou 543002, China; 2. Department of Ultrasound, 3. Medical Genetics Center, Shenzhen Maternity and Child Healthcare Hospital, Women and Children's Medical Center, Southern Medical University, Shenzhen 518028, China; 4. Department of Ultrasound, Kunming MCH Hospital, Kunming 650032, China)

[**Abstract**] **Objective** To observe the value of intelligent auxiliary model based on Swin-Transformer for diagnosing fetal eye malformations. **Methods** A total of 1 282 ophthalmic ultrasound images of fetuses with ocular malformations confirmed by prenatal screening and 526 ophthalmic ultrasound images of normal fetuses were retrospectively collected and divided into training set, validation set and test set in a ratio of 8∶1∶1. An intelligent auxiliary diagnosis model was built based on Swin-Transformer, and its efficiency for diagnosing fetal eye malformations was compared with that of MobileNet-V2, ResNet-50, VGG-16 and Vision-Transformer models. **Results** Intelligent auxiliary diagnosis model based on Swin-Transformer achieved a sensitivity of 88.31%, specificity of 97.37%, area under the receiver operating characteristic (ROC) curve of 0.990, precision of 87.31% and F1 score of 87.71% for diagnosing fetal ocular malformations in test set, all superior to those of the other 4 mainstream models. Swin-Transformer model exhibited a high degree of focus in all anomaly diagnosis heat maps, with clear clustering shown by confusion matrix analysis. ROC curve showed that Swin-Transformer model had the best performance in simultaneously diagnosing each anomaly, its t-SNE feature distribution clustering boundary was clear, and the performance was stable. **Conclusion** Intelligent auxiliary model based on Swin-Transformer exhibited high accuracy and stability for diagnosing fetal eye malformations, promising to provide crucial technical support for the auxiliary diagnosis of fetal ocular malformations.

[**Keywords**] abnormalities; fetus; ultrasonography, prenatal; Swin-Transformer; intelligent auxiliary diagnosis

DOI:10.13929/j.issn.1003-3289.2025.12.005

基于 Swin-Transformer 智能辅助模型用于诊断胎儿眼部畸形

陶雄杰[1],邸臻炜[1],梁博诚[2],欧阳淑媛[3],郭　慧[1],贺　杰[1],仝　蕊[4],陈家希[1],解　迪[1],赵英丽[1],覃　妮[1],李胜利[2*]

(1.梧州学院广西机器视觉与智能控制重点实验室,广西 梧州　543002;2.南方医科大学妇女儿童医学中心深圳市妇幼保健院超声科,3.医学遗传中心,广东 深圳　518028;4.昆明市妇幼保健院超声医学科,云南 昆明　650032)

[**摘　要**] **目的**　观察基于 Swin-Transformer 的智能辅助模型用于诊断胎儿眼部畸形的价值。**方法**　回顾性收集经产

[**基金项目**]国家自然科学基金(62162054)、广西自然科学基金(2025GXNSFAA069497)、梧州市科技计划(202302036)、梧州学院校级科研青年项目(2024QN001)。

[**第一作者**]陶雄杰(1995—),男,广西贵港人,在读硕士,工程师。研究方向:人工智能医学影像。E-mail:384576861@qq.com

[**通信作者**]李胜利,南方医科大学妇女儿童医学中心深圳市妇幼保健院超声科,518028。E-mail:lishengli63@vip.126.com

[**收稿日期**]2025-07-26　[**修回日期**]2025-11-28

前筛查确诊眼部畸形胎儿的 1 282 幅及 526 幅正常胎儿眼部声像图，按 8∶1∶1 比例划分训练集、验证集及测试集。基于 Swin-Transformer 构建智能辅助诊断模型，并与 4 种主流模型 MobileNet-V2、ResNet-50、VGG-16 及 Vision-Transformer 比较其效能。**结果**　基于 Swin-Transformer 智能辅助模型诊断测试集胎儿眼部畸形的敏感度为 88.31%、特异度为 97.37%，受试者工作特征(ROC)曲线的曲线下面积为 0.990、精确率为 87.31%、F1 分数为 87.71%，均优于 4 种主流模型。Swin-Transformer 模型在诊断所有畸形的热力图中均呈高度聚焦，混淆矩阵分析显示聚集明显，ROC 曲线显示其同时诊断各畸形的效能最佳，t-SNE 特征分布聚类边界更清晰且性能稳定。**结论**　基于 Swin-Transformer 智能辅助模型用于产前诊断胎儿眼部畸形具有较高准确性与稳定性，有望为辅助诊断胎儿眼部畸形提供关键技术支撑。

［**关键词**］　畸形；胎儿；超声检查，产前；Swin-Transformer；智能辅助诊断

［**中图分类号**］　R714.53；R445.1　［**文献标识码**］　A　　［**文章编号**］　1003-3289(2025)12-1960-06

胎儿眼部畸形指胚眼生长发育异常或停滞所致外观畸形、组织器官缺陷或功能障碍[1]，主要包括眼附件(eye accessories，EA)、眼距(eye distance，ED)、眼眶(eye socket，ES)及眼球(eye ball，EB)畸形(图 1)，系较少见的体表异常，多与染色体异常或遗传疾病紧密关联[2-3]；严重眼部畸形如独眼和无眼可能影响视力或致盲[4]。产前超声现已成为筛查胎儿疾病的首选方法[5-6]，但具有主观因素依赖性[7]；亟需发展客观、稳定的智能辅助诊断方法，将人工智能(artificial intelligence，AI)技术用于医学影像以识别异常、预估风险及实现全流程辅助，进而提高诊断效率与质量。

1　资料与方法

1.1　研究对象

回顾性收集 2011 年 9 月—2024 年 6 月深圳市妇幼保健院经产前超声诊断眼部畸形胎儿的 1 282 幅胎儿眼部声像图，包括 EA 畸形 526 幅、ED 畸形 126 幅、ES 畸形 145 幅、EB 畸形 485 幅，以及 526 幅正常胎儿眼部声像图。按 8∶1∶1 比例划分训练集、测试集及验证集，见表 1。纳入标准：依据《超声产前筛查指南》[8]，正常对照孕周 11～32 周，EA 畸形 27～32周、ED 畸形 12～28 周、ES 畸形 13～24 周、EB 畸形 13～32 周；超声图像完整清晰显示病变，满足横切面或纵切面评估要求。由具有 10 年以上工作经验的产前超声医师针对所有图像进行诊断并存档，由 20 年以上工作经验的超声医师进行复核；剔除存在伪影、切面不完整或因胎儿体位及运动致关键结构显示不清等的图像。本研究获孕妇及其法定代理人知情同意，并经医院医学伦理委员会批准(LLYJ2024-202-089)。

1.2　方法

1.2.1　构建辅助诊断模型　将所有超声图像统一调整为像素 224×224，数据脱敏后进行归一化处理。所涉胎儿眼部畸形种类包括 EA 畸形(包含鼻泪管囊肿)、ED 畸形(包含眼距过宽、眼距过窄)、ES 畸形(包含眼眶囊肿、

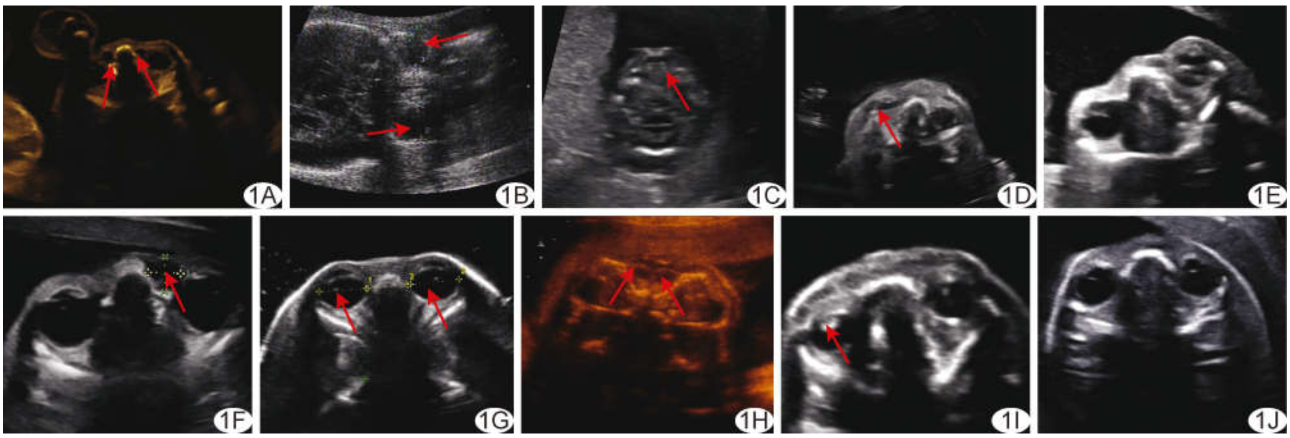

图 1　声像图示眼部畸形与正常眼部　A. EA 畸形(27 周)；B. ED 畸形(24 周)；C. ES 畸形(13 周)；D. EB 畸形(25 周)；E. 正常眼部(29 周)；F. ED 畸形(28 周)；G. ES 畸形(31 周)；H. EA 畸形(25 周)；I. EB 畸形(24 周)；J. 正常眼部(30 周)　(红箭示病灶)

表 1　各集别内眼部畸形及正常眼部声像图分布(幅)

集别	EA 畸形	ED 畸形	ES 畸形	EB 畸形	正常	合计
训练集	420	100	116	456	420	1 512
验证集	52	12	14	14	52	144
测试集	54	14	15	15	54	152
合计	526	126	145	485	526	1 808

四眼畸形)及 EB 畸形(包含无眼、小眼、白内障、无晶状体、眼球出血)。基于 PyTorch 框架、采用 Swin-Transformer[9]构建胎儿眼部畸形智能辅助诊断模型。将输入图像划分为 4×4 非重叠像素 patch,经线性嵌入后送入分层 Transformer Block,通过全局平均池化与分类头输出多集别概率分布。将所获模型与具有代表性的卷积神经网络(convolutional neural network, CNN)模型 MobileNet-V2[10]、ResNet-50[11]、VGG-16[12]及 Transformer 类的代表模型 Vision-Transformer[13]进行对比。

1.2.2 模型训练环境 采用 Linux 服务器、Intel Xeon Platinum 8375C 处理器、378GB 内存及 4 块 NVIDIA 3090 GPU。以自适应梯度优化算法(adaptive moment estimation, Adam)结合余弦退火学习率调度策略提升模型诊断能力与泛化性能。

1.3 统计学分析 采用 Python 3.8.20 软件,通过集成 PyTorch、Scikit-learn 及 Matplotlib 等核心库构建模型性能评估指标,包含敏感度(sensitivity, Sen)、特异度(specificity, Spe)、受试者工作特征(receiver operating characteristic, ROC)曲线的曲线下面积(area under the curve, AUC)、精确率(precision, Pre)及 F1 分数(F1 Score)。基于 ROC 曲线、t 分布随机邻域嵌入算法(t-distributed stochastic neighbor embedding, t-SNE)可视化、混淆矩阵及梯度加权类激活映射(gradient-weighted class activation mapping, Grad-CAM)热力图对模型性能进行多维度分析。

2 结果

2.1 模型效能 Swin-Transformer 模型诊断测试集胎儿眼部畸形的 Sen 为 88.31%,Spe 为 97.37%,AUC 为 0.990,Pre 为 87.31%,F1 Score 为 87.71%,均属最佳,见表 2。

表 2 各模型诊断胎儿眼部畸形效能

模型	Sen (%)	Spe (%)	AUC	Pre (%)	F1 Score (%)
MobileNet-V2	85.59	97.18	0.983	83.84	84.62
ResNet-50	85.44	97.13	0.984	84.84	85.06
VGG-16	86.22	97.09	0.980	82.91	84.29
Vision-Transformer	85.67	97.22	0.984	83.44	84.42
Swin-Transformer	88.31	97.37	0.990	87.31	87.71

2.2 Grad-CAM 热力图 热力图中的红色高亮区域对应模型关注的关键解剖结构(图 2)。Swin-Transformer 模型在所有畸形的诊断热力图均呈高度聚焦,其关注区域边界清晰且背景噪声低,能较准确地提取临床诊断特征,展现出较好的诊断能力。

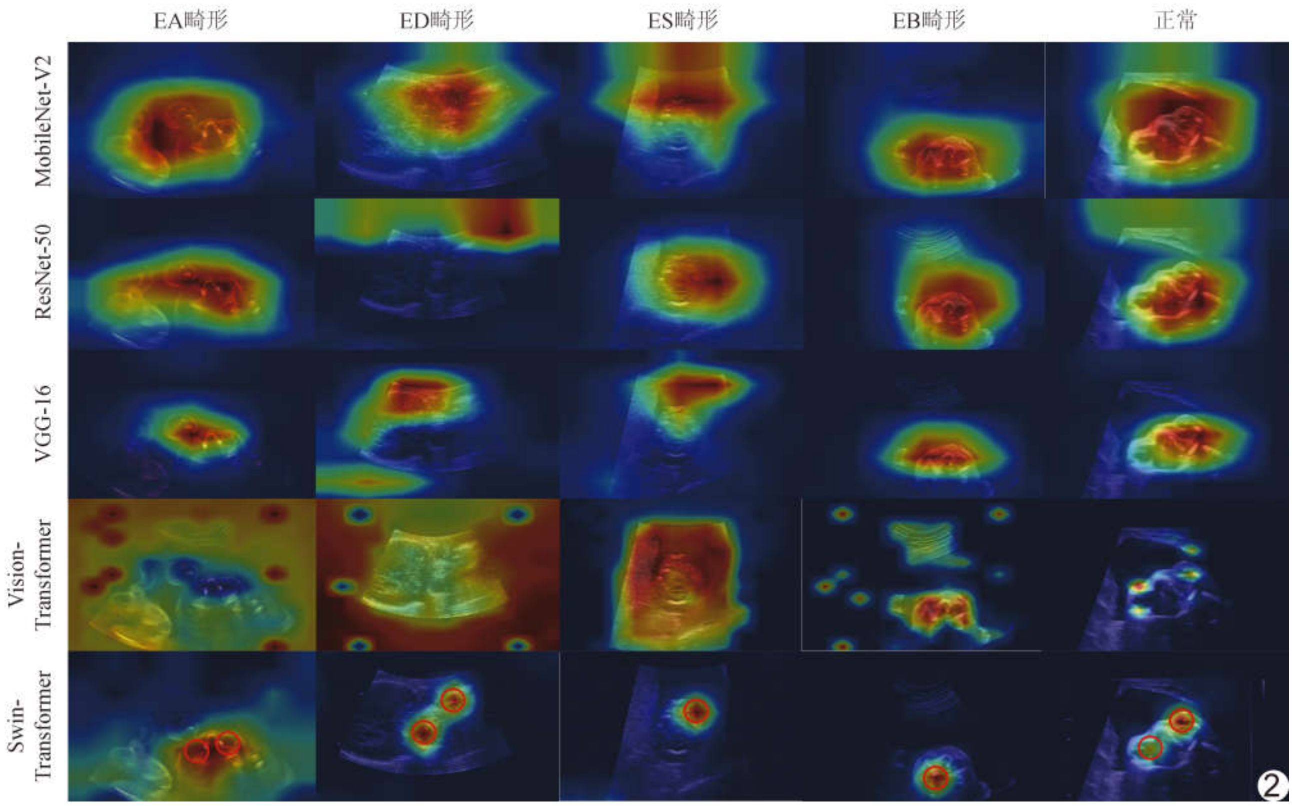

图 2 各模型诊断胎儿眼部畸形与正常结构的 Grad-CAM 热力图(红圈为 Swin-Transformer 模型识别的眼部畸形)

2.3　混淆矩阵分析　混淆矩阵对角线上的数值颜色越深，代表正确分类数越多、聚集越明显，诊断畸形越精准、漏诊误诊风险越低。Swin-Transformer 模型诊断测试集 EA、ED、ES、EB 畸形与正常的精准度为分别为 48、12、51、12 及 39；除 EA 畸形外，其诊断胎儿眼部各种畸形的精准度均高于其他模型。见图 3。

2.4　ROC 曲线分析　各模型均能准确识别正常结构（AUC 均为 1.000）。Swin-Transformer 模型诊断 EA、ED、ES 及 EB 畸形的 AUC 分别为 0.993、0.995、0.981 及 0.986，且同时诊断各畸形的效能最佳。见图 4 和表 3。

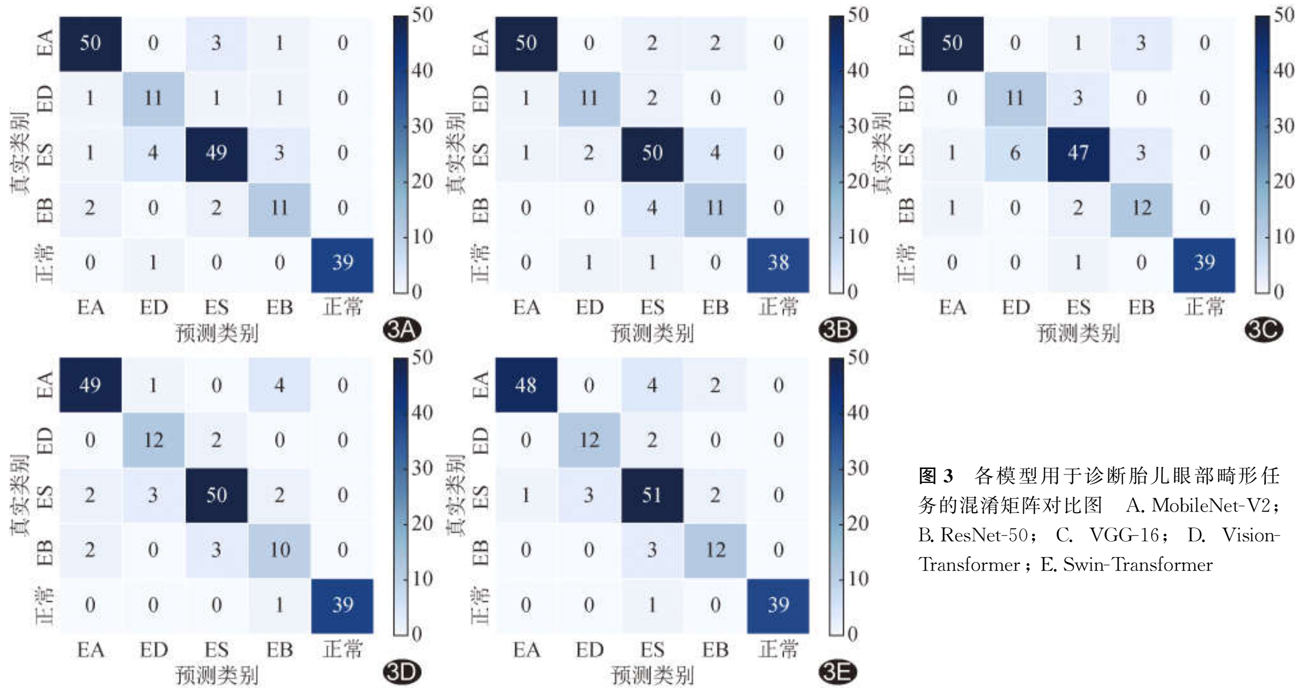

图 3　各模型用于诊断胎儿眼部畸形任务的混淆矩阵对比图　A. MobileNet-V2；B. ResNet-50；C. VGG-16；D. Vision-Transformer；E. Swin-Transformer

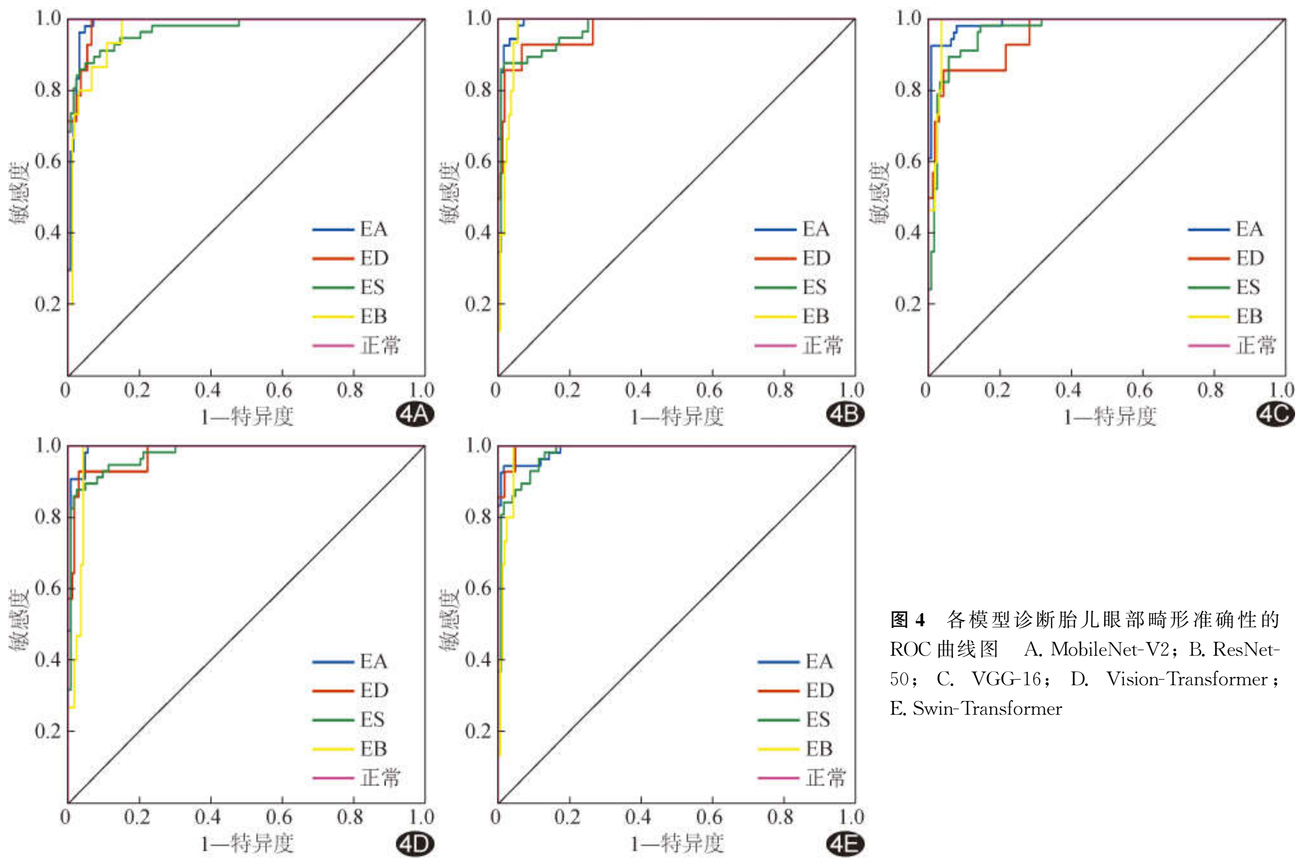

图 4　各模型诊断胎儿眼部畸形准确性的 ROC 曲线图　A. MobileNet-V2；B. ResNet-50；C. VGG-16；D. Vision-Transformer；E. Swin-Transformer

表 3　各模型诊断胎儿眼部畸形的 AUC

模型	EA 畸形	ED 畸形	ES 畸形	EB 畸形
MobileNet-V2	0.987	0.987	0.972	0.969
ResNet-50	0.993	0.972	0.973	0.979
VGG-16	0.990	0.956	0.968	0.985
Vision-Transformer	0.992	0.977	0.976	0.975
Swin-Transformer	0.993	0.995	0.981	0.986

2.5　t-SNE 可视化分析　t-SNE 以不同颜色代表不同畸形或正常结构。不同集别散点图在三维特征空间中呈聚类趋势，表明模型具备类型判断能力(图 5)。Swin-Transformer 模型 t-SNE 特征分布聚类边界更清晰，即其在智能辅助诊断胎儿眼部畸形任务中的性能稳定。

3　讨论

AI 技术已在超声医学领域取得显著进展，其应用包括识别标准切面[14]、控制质量[15]、测量参数[16-17]及辅助诊断疾病[18]。凭借其分层结构和滑动窗口注意力机制，Swin-Transformer 能针对图像局部与全局特征依赖关系实现高效建模，用于诊断乳腺癌[19]及食管癌[20]等表现优异。胎儿眼部结构细微、形态多变，且需关联解剖上下文信息以辅助诊断；而 Swin-Transformer 处理此类任务具有潜在优势。

本研究基于 Swin-Transformer 构建智能辅助诊断胎儿眼部畸形模型，并以 1 282 幅胎儿眼部声像图证实了其性能优越。该模型在测试集的 Sen 88.31% 高于次优模型 VGG-16 的 86.22%，Pre(87.31%)及 AUC (0.990)亦均高于较次优模型[分别为 ResNet-50(84.84%)及 Vision-Transformer(0.984)]，各项指标数据较对比模型平均提升 2.09%～4.26%，展现出良好的辅助诊断能力。从临床角度分析，该模型对于无眼和小眼等严重畸形的准确识别有助于产前遗传学评估[4]；而对于鼻泪管囊肿等良性病变，该模型有助于避免过度诊断与不必要的临床随访并进一步完善超声评估体系。

现有诊断胎儿眼部畸形的临床研究多针对小样本进行探讨[1]，缺乏 AI 辅助。本研究所获基于 Swin-Transformer 智能辅助模型用于诊断胎儿眼部畸形展现出较强稳定性与鲁棒性，整体辅助诊断性能更优，有助于提升诊断效率。本研究的主要不足：①样本量较小，结果可能存在选择偏倚；②病例多样性欠缺，导致模型在跨地域、跨设备场景下的泛化效能受到制约。未来将联合多中心扩大数据集多样性，进一步优化模型，以适应跨地域、跨设备等更多场景。

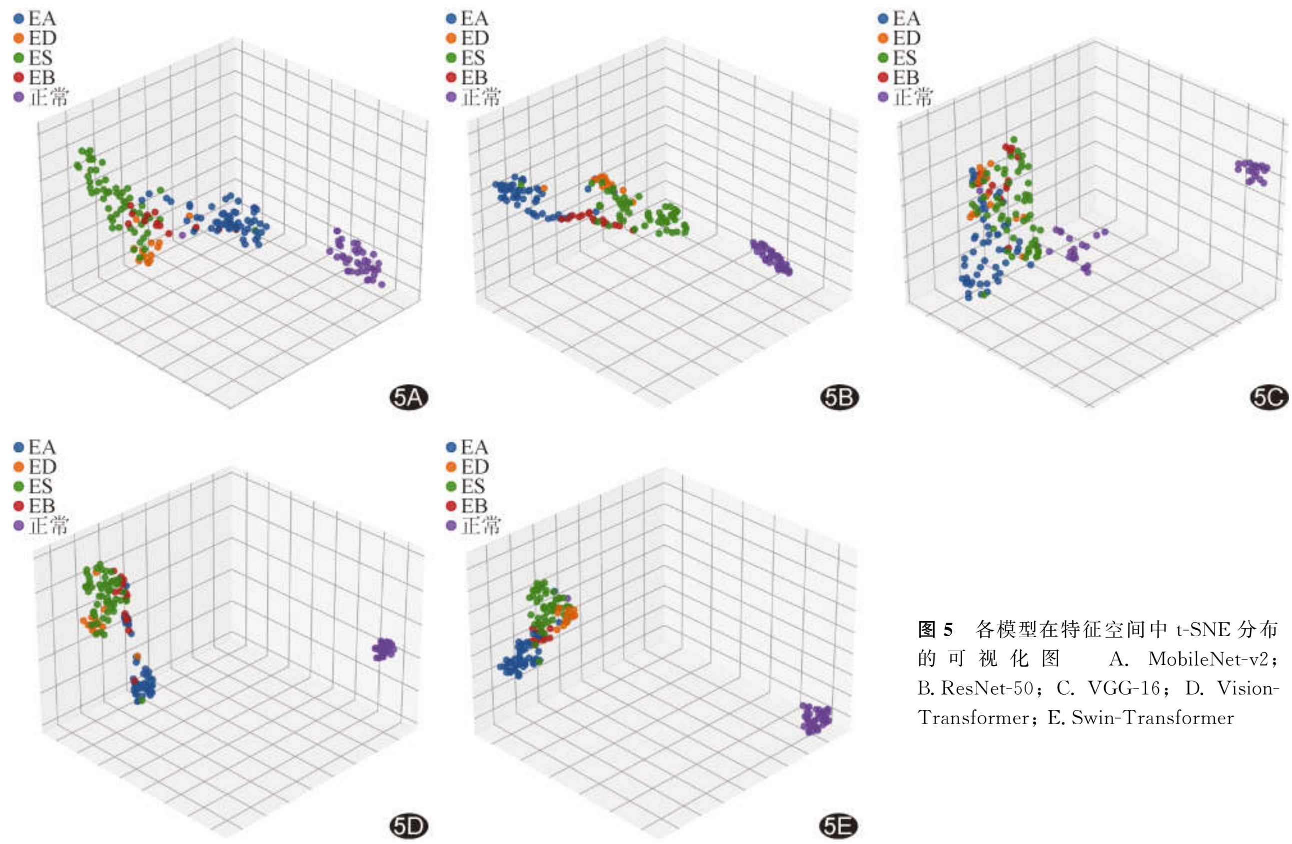

图 5　各模型在特征空间中 t-SNE 分布的可视化图　A. MobileNet-v2；B. ResNet-50；C. VGG-16；D. Vision-Transformer；E. Swin-Transformer

利益冲突：全体作者声明无利益冲突。

作者贡献：陶雄杰撰写和修改文章；邱臻炜、梁博诚、欧阳淑媛研究设计；郭慧、贺杰和仝蕊数据分析；陈家希和解迪研究实施；赵英丽和覃妮查阅文献；李胜利指导、审阅文章。

［参考文献］

[1] 钟燕秋，张艳红，钟晓红，等．产前超声在诊断胎儿眼部异常中的价值[J]．中国医疗器械信息，2023，29(2)：15-17．

[2] JIMENEZ C J，BOCKENHOFF P，TASCON P L，et al. Sonographic diagnosis of fetal eye anomalies and their association with syndromal diseases：A retrospective multicenter analysis of 264 cases［J］. Acta Obstet Gynecol Scand，2025，104（5）：850-859.

[3] DUBUCS C，CAILLET A，FREMONT F，et al. Prevalence of congenital ocular anomalies in 15 countries of Europe：Results from the Medikeye study［J］. Birth Defects Research，2024，116 (11)：e2414.

[4] 黄筱桐，罗红．产前超声诊断胎儿小眼畸形的价值[J]．实用妇产科杂志，2020，36(2)：154-156．

[5] 曾晴，李胜利．产前超声评估胎儿颅脑：可见与不可见异常[J]．中国医学影像技术，2024，40(11)：1636-1641．

[6] 秦越，李胜利．产前超声观察外侧裂评估胎儿脑皮质发育研究进展[J]．中国医学影像技术，2024，40(11)：1633-1635．

[7] ZHANG J，LI Z，LIN H，et al. Deep learning assisted diagnosis system：Improving the diagnostic accuracy of distal radius fractures[J]. Front Med (Lausanne)，2023，10：1224489.

[8] 中华医学会超声医学分会妇产超声学组，国家卫生健康委妇幼司全国产前诊断专家组医学影像组．超声产前筛查指南[J]．中华超声影像学杂志，2022，31(1)：1-12．

[9] LIU Z，LIN Y，CAO Y，et al. Swin transformer：Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021：10012-10022.

[10] SANDLER M，HOWARD A，ZHU M，et al. MobileNetV2：Inverted residuals and linear bottlenecks［C］//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018：4510-4520.

[11] HE K，ZHANG X，REN S，et al. Deep residual learning for image recognition［C］//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016：770-778.

[12] DOAOVITSKIY A. An image is worth 16 × 16 words：Transformers for image recognition at scale［J］. Arxiv preprint arxiv：2010.11929，2020.

[13] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition［J］. Arxiv preprint arxiv：1409.1556，2014.

[14] PU B，LI K，LI S，et al. Automatic fetal ultrasound standard plane recognition based on deep learning and IIoT［J］. IEEE Transactions on Industrial Informatics，2021，17（11）：7771-7780.

[15] PU B，ZHU N，LI K，et al. Fetal cardiac cycle detection in multiresource echocardiograms using hybrid classification framework[J]. Future Gener Comp Sy，2021，115：825-836.

[16] LIANG B，PENG F，LUO D，et al. Automatic segmentation of 15 critical anatomical labels and measurements of cardiac axis and cardiothoracic ratio in fetal four chambers using nnU-NetV2[J]. BMC Med Inform Decis Mak，2024，24 (1)：128.

[17] LU Y，LI K，PU B，et al. A YOLOX-based deep instance segmentation neural network for cardiac anatomical structures in fetal ultrasound images［J］. IEEE/ACM Transactions on Computational Biology and Bioinformatics，2022.

[18] 梁博诚，罗丹丹，罗彩群，等．产前超声诊断先天性心脏病胎儿的基因特征[J]．中国医学影像技术，2024，40(2)：251-255．

[19] MO Y，HAN C，LIU Y，et al. Hover-trans：Anatomy-aware hover-transformer for ROI-free breast cancer diagnosis in ultrasound images[J]. IEEE Trans Med Imaging，2023，42(6)：1696-1706.

[20] 王润媛，陈星材，吴蔚，等．基于增强 CT 图像和 Swin Transformer 网络的食管癌 T 分期智能诊断模型的构建与评估[J]．陆军军医大学学报，2023，45(16)：1770-1778．