Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Medical Engineering—Article

贝叶斯推理和动态神经反馈促进先天性心脏病智能诊断的临床应用

谭伟敏[a,#]，曹银银[b,#]，马晓静[b,#]，茹港徽[a,#]，李吉春[a]，张璟[b]，高燕[b]，杨佳伦[b]，黄国英[b,*]，颜波[a,*]，李健[b,*]

[a] *Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai 200433, China*
[b] *Cardiovascular Center & Clinical Laboratory Center, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai 201102, China*

ARTICLE INFO

Article history:
Received 26 March 2022
Revised 1 September 2022
Accepted 10 October 2022
Available online 16 January 2023

关键词
先天性心脏病
人工智能
深度学习
模型不确定性

摘要

先天性心脏病（CHD）是婴幼儿死亡的主要原因。基于人工智能的先天性心脏病诊断网络（CHDNet）是一种基于超声心动图视频的二分类模型，用于判别超声心动图视频是否包含心脏缺陷。现有的CHDNet模型表现出与医学专家相当甚至更好的判别性能，但它们在训练集之外的样本上的不可靠性已成为模型部署的关键瓶颈。而这是当前大多数基于AI诊断方法的共性问题。为了克服这一挑战，本文提出了两种基本机制——贝叶斯推理和动态神经反馈——分别用于衡量和提高人工智能诊断的可靠性。贝叶斯推理允许神经网络模型输出CHD判别的可靠性而不仅仅是单一的判别结果，而动态神经反馈是一个计算神经反馈单元，允许神经网络将知识从输出层反馈给浅层，使神经网络有选择地激活相关神经元。为了评估这两种机制的有效性，我们在包含三种常见CHD缺陷的4151个超声心动图视频上训练了CHDNet，并在1037个超声心动图视频的内部测试集和从其他心血管成像设备新收集的692个外部视频集上对其进行了测试。每个超声心动图视频对应于一位患者和一次就诊。我们在多种代表性神经网络架构上展示了贝叶斯推理获得的可靠性如何解释和量化神经网络内部和外部测试集之间的性能显著差异，以及尽管输入被噪声破坏或使用外部测试集时，设计的反馈单元如何帮助神经网络保持高精度和可靠性。

1. 引言

先天性心脏病（congenital herat disease, CHD）发病率约占所有活产新生儿的9/1000。在CHD所有亚型中，室间隔缺损（ventricular septal defect, VSD）、房间隔缺损（atrial septal defect, ASD）和动脉导管未闭（patent ductus arteriosus, PDA）是三种最常见类型[1–2]。ASD是指胚胎时期房间隔的发育异常，导致左、右心房的间隔出现缺损。根据缺损位置不同，ASD可进一步分为继发孔型ASD、原发孔型ASD、静脉窦型ASD或其他类型的ASD，其中继发孔型ASD是最常见的类型。VSD是指左、右心室之间存在异常交通，是最常见的先天性心脏畸形，占所有先天性心脏病的50%。根据解剖位置不同，VSD可进一步分为膜周型、嵴下型、嵴内型、干下型及肌部缺损。其中膜周型VSD最常见，约占全部VSD的60%。动脉导管是指主动脉弓降部与肺动脉之间相连的管道，在胎儿时期，胎儿循环系统依赖其存在，但在出生后应自然闭合，如未能闭合，则在主动脉与肺动脉之间存留一个通

* Corresponding authors.
E-mail addresses: gyhuang@shmu.edu.cn (G. Huang), byan@fudan.edu.cn(B. Yan), lijianjulia@fudan.edu.cn (J. Li).
# These authors contributed equally to this work.

英文原文：Engineering, 2023, 23(4): 90–102
引用本文：Weimin Tan, Yinyin Cao, Xiaojing Ma, Ganghui Ru, Jichun Li, Jing Zhang, Yan Gao, Jialun Yang, Guoying Huang, Bo Yan, Jian Li. Bayesian Inference and Dynamic Neural Feedback Promote the Clinical Application of Intelligent Congenital Heart Disease Diagnosis. *Engineering*, https://doi.org/10.1016/j.eng.2022.10.015

道，即PDA。足月儿中孤立性PDA的发病率约为活产新生儿的1/2000，占所有CHD的5%~10%。

早发现、早诊断是降低CHD自然死亡率和改善预后的关键因素。经胸超声心动图是非侵入性的成像方式，被认为是检测和诊断CHD的首选方法[3–5]。超声心动图是制定外科手术或介入治疗方案及评价疗效的主要依据。近年来基于深度学习技术的心脏超声诊断方法逐渐被开发出来[6–10]，并在内部测试集上取得了良好的性能，甚至超过了专家水平。然而，当它们被应用于没有受过训练的病例时，诊断结果的准确性可能会受到影响。解决此类情况的可行方法是测量模型诊断结果的可靠性。

可靠性是诊断模型在给出正确诊断结果的前提下，将分布内（IND）和分布外（OOD）测试样本识别为低不确定性（高置信度）或高不确定性（低置信度）的能力。测量诊断结果的可靠性和追求高诊断准确性是同样重要的目标。在临床实践中，医生是否愿意使用基于机器学习的诊断方法取决于他们是否相信这些方法能够准确可靠地诊断CHD。这是一个适用于任何诊断方法的有效关注点。因此，CHD诊断方法的可靠性测量越来越重要。近年来，在各种假设条件下构建具有覆盖保证的预测集的研究不断涌现[11–13]。当构建预测的数据分布与生成预测模型的数据分布相匹配时，这些方法中的大多数都提供了理论上的覆盖保证。Conformal prediction是其中最著名的方法之一，它可以保证高概率地覆盖新的观测值[12,14]。作为另一个方向的推广，该方法[11]提供了风险控制预测集，它们具有较低的预测风险和较高的数据随机性概率。

与保形预测不同，贝叶斯概率论为模型不确定性的推理提供了基于数学的工具，但这些工具通常需要大量的计算。最近报道的贝叶斯神经网络[15]、贝叶斯近似[16]和变分推理[17–18]都是贝叶斯推理方法的例子，它们可以测量模型输出的可靠性，从而为是否应该采用该模型的输出提供客观评估。在本研究中，我们使用贝叶斯推理的近似方法，该方法只需要常用的dropout技术来评估模型的不确定性，并且具有快速和易于实现的优点。直观地说，内部测试集上的诊断可靠性应该高于外部测试集上的诊断可靠性，因此可靠性有助于识别训练集之外的病例。

测量诊断结果的可靠性很重要，但还不够，因为提高外部测试集的可靠性和鲁棒性更有意义，并且需要基于深度学习的诊断方法。人类视觉系统经过长期的自然进化，具有极高的可靠性和鲁棒性，抗干扰能力极强。神经反馈是人类视觉皮层中复杂的基本机制，它可以选择性地激活相关神经元，抑制不相关的干扰噪声或模式，这在处理带有干扰物或杂乱背景的输入图像时非常有用[19–21]。最近，一种反馈机制被探索并应用于各种视觉任务[22–29]。在认知理论中，连接皮层视觉区域的反馈连接可以将反应信号从高阶区域传递到低阶区域[22–23]。这启发了学者[24–25]设计包含反馈模块的高级深度架构。这些体系结构中的反馈机制将网络深层的信息传递回前一层，指导底层编码信息的提取，实现了自顶向下的方法。反馈网络方法[30]是与本研究最相关的工作，因为它将具有语义信息的深度特征传输到输入图像的中间表示中，以实现深度网络中的反馈。然而，它只传递深层信息，不更新浅层表示。因此，迫切需要一种能够广泛应用于深度神经网络的动态神经反馈单元。

在本研究中，我们提出了一个计算动态神经反馈单元，它可以将知识从输出层反馈回浅层，允许诊断模型在前馈推理期间改变浅层的特征。对于三种常见的先天性心脏缺陷（VSD、PDA和ASD），我们在各种具有代表性的深度体系结构上证明了即使输入受到噪声的严重损坏或来自外部测试集，反馈单元仍可以显著提高体系在区分正常心脏和常见CHD方面的可靠性、鲁棒性和准确性。考虑到所提出的反馈单元的高可转移性，该单元有可能在可靠性、鲁棒性和准确性方面改进其他诊断模型。

2. 方法

2.1. 训练和测试数据收集

基于深度学习的诊断模型[9–10,31–34]通常需要足够的数据和准确的标注来进行模型训练。缺乏具有准确标注的心脏缺陷类型的大规模临床超声心动图数据阻碍了CHD诊断的研究进展。为此，在伦理审查通过的前提下，我们共收集了2015年1月1日到2021年6月30日来自复旦大学附属儿科医院的5880例超声心动图数据，包括1213例ASD、1078例VSD、970例PDA和2619例健康对照儿童（图1；附录A中的表S1和表S2）。每位纳入对象均包括超声心动图视频和静态图像，用来训练CHD诊断网络（CHDNets）。我们使用Philips iE33心脏超声仪器，探头频率范围为3~8 MHz，或1~5 MHz。二维（2D）成像结合彩色多普勒血流图像可显示缺损的位置、大小和血流方向。根据三类CHD的解剖特征，对房间隔、室间隔和降主动脉与左肺动脉连接处进行观察来确定心脏是否存在未闭合情况。本文采集了三种标准二维视图和彩色多普勒血流图（双模型），包括VSD和PDA患者的胸骨旁大动脉短轴切面（PSSAX）和观察ASD存在与否的剑突下两房心切面（SXLAX）。所有患者的诊断结果均由至少两名资深超声心动图医师或术中最终诊断证实。本研究获得了复

旦大学附属儿科医院伦理委员会的批准（批准号：258），研究遵循《赫尔辛基宣言》。本研究获得了患儿父母或监护人的知情同意，并严格确保患者信息的安全和保密。

2.2. 数据标注和质量控制

我们下载了DICOM格式的超声心动图数据，每个超声心动视频中的关键帧均由经验丰富的超声心动图医师手动选取。数据标注的过程如下：每次超声心动图成像均通过三级评估系统进行评估。一级评估由一名接受过质量控制培训的学士或更高学历的医学生进行，二级评估由两名初级超声心动图医师进行，三级评估由两名具有10年以上临床经验的资深超声心动图医师进行评估。三级评估系统确保每张心脏超声图像都有正确的诊断标签和心脏缺损位置。完成数据标注后，随机选取采集数据中的100个对象，由第三位具有20多年临床经验的超声心动图医师进行检查，以尽量减少人为错误对计算建模过程的影响。最后随机抽取881例ASD、772例VSD、688例PDA进行模型训练，训练还包括它们对应的健康对照组，分别有584例、644例和582例图像。

2.3. 基于关键帧的超声心动图视频诊断

经验丰富的心脏超声医师遵循的诊断步骤为：从超声心动图视频中选择心脏缺损最清晰视图的关键帧，然后根据选定的关键帧做出诊断决定[图2（a）]。诊断决定包括患者是否健康、患有何种先天性心脏缺陷以及心脏缺陷的位置。这个诊断过程启发我们设计了一种基于关键帧的CHD诊断模型[图2（b），左]。在诊断模型中，首先使用训练好的分类模型判断视频是否包含心脏缺陷（可以将视频帧合并为Batch形式进行并行计算），然后通过特征距离比较选择潜在的包含心脏缺陷的视频帧。最后，使用Faster-RCNN模型检测所选视频帧上心脏缺陷的位置。诊断结果如图2（c）~（e）所示。

心脏超声医生的诊断过程非常合理，因为从模型方面来看，Faster-RCNN在网络参数和计算复杂度方面通常比分类模型大很多。此外，从超声心动图视频来看，包含先天性心脏缺损的超声心动图视频中通常只有少量视频帧有心脏缺损以供心脏超声医生分析。因此，从模型和数据两方面考虑，先识别再检测心脏缺陷的方法是基于超声心动图视频的先心病诊断的一个很好的选择，因为它在诊断准

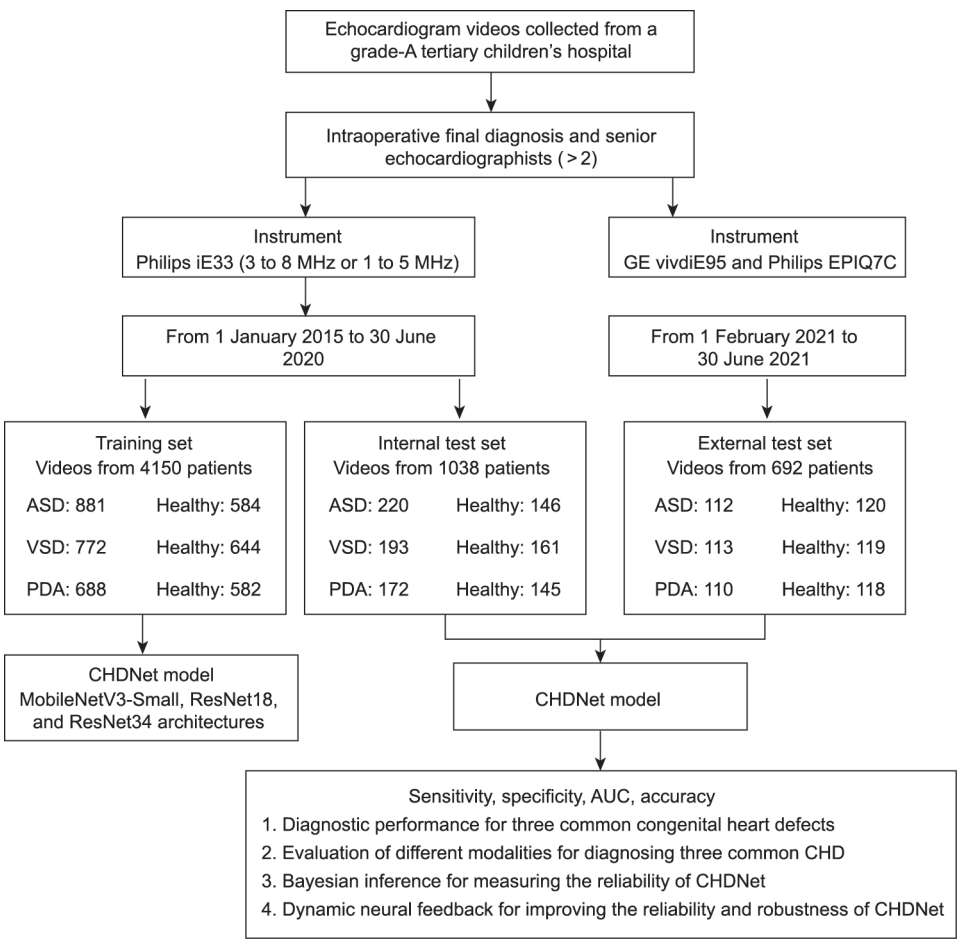


图1. CHDNet训练和评估数据采集流程图。

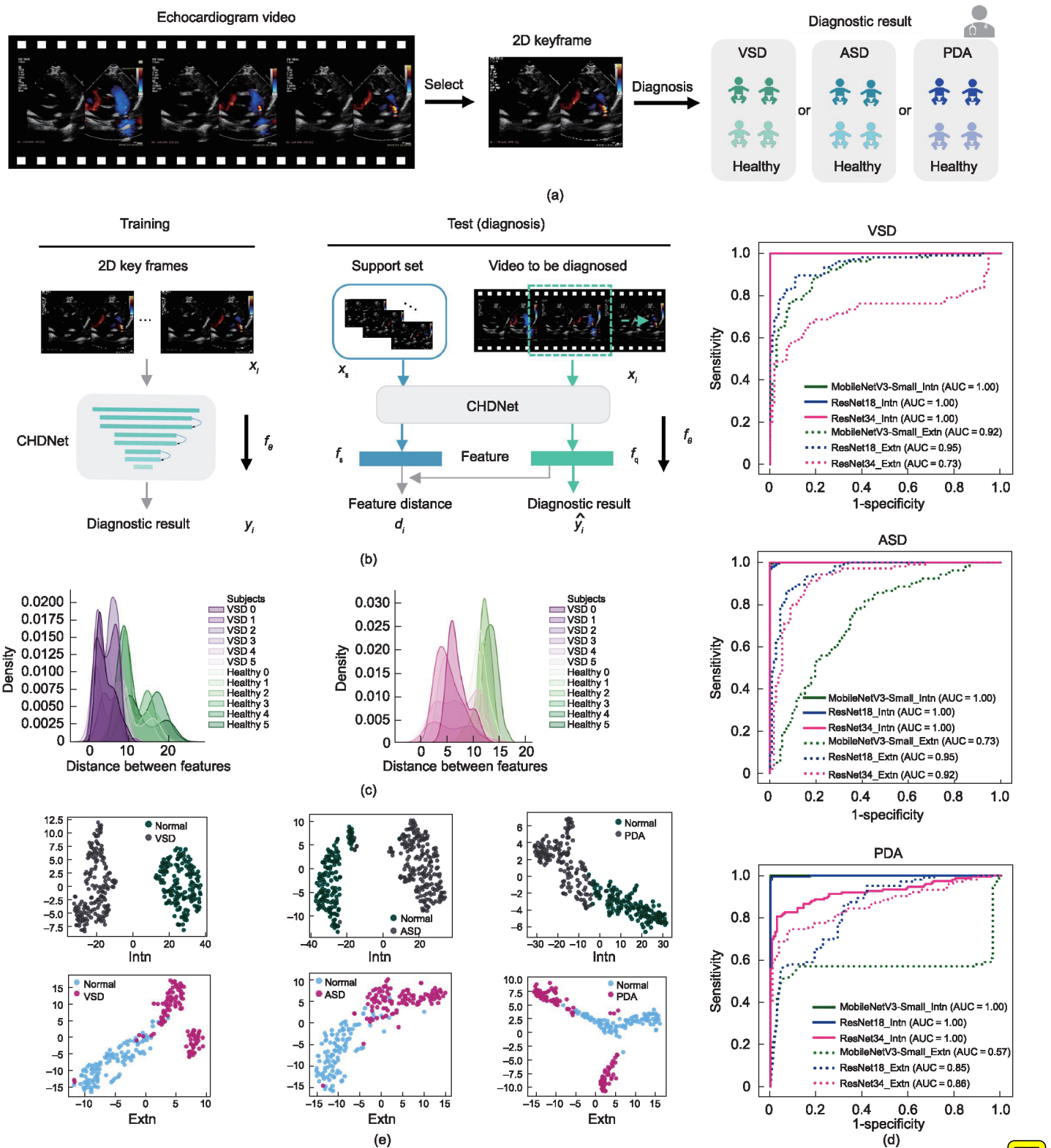


图2. 根据二维关键帧诊断三种常见的婴儿先天性心脏缺陷。（a）经验丰富的心脏超声医师的诊断过程：从超声心动视频中选择有明显心脏缺陷的二维关键帧进行诊断。（b）提出的心脏病诊断流程概览。二维关键帧对$(x_i; y_i)$和基于术中最终诊断的相应注释用于训练CHDNet，以根据x_i预测y_i。经过训练的CHDNet $f_\theta(\cdot)$可用于诊断之前未见过的超声心动视频x_i，得出$(d_i; \hat{y}_i)$的特征距离和诊断结果，其中，x_s表示包含心脏缺陷异常病例的支持集。f_s是支持集x_s的元表示，f_q是输入视频帧的特征表示。（c）支持集与待诊断VSD和ASD的超声心动图视频之间的特征距离统计。（d）用于诊断三种常见先天性心脏缺陷的三种代表性神经网络架构的接收者操作特征（ROC）曲线。（e）CHDNet（ResNet18）在VSD内部和外部测试集上的分类特征可视化。Intn：内部；Extn：外部。

确性和效率之间进行了较好权衡。

2.4. CHDNet训练与评估

基于上述分析和获取的超声心动图数据，我们选取了三种具有代表性的神经网络架构来具体实现CHDNet（CHD诊断网络）：①MobileNetV3-Small [35]，这是ImageNet数据库上的典型分类网络，计算资源开销低；②ResNet18 [36]，ImageNet数据库上的分类网络，具有众所周知的深度残差连接；③ResNet34 [36]。经过训练后，CHDNet模型被应用于训练过程中未见过的内部和外部测试集以验证其性能[图2（b），右]。

对于诊断三种常见先天性心脏病（ASD、VSD、PDA）的实验，采用MobileNetV3-Small、ResNet18和ResNet34在训练集中的超声心动图视频的二维关键帧上进行训练。训练集由经验丰富的超声心动图医师从超声心动图视频中选取的关键帧（包含代表性的正常和异常帧）组合而成。对于所有其他实验，采用ResNet18来实现CHDNet。所有实验都是使用Pytorch在Python中实现的。训练和评估的源代码可在https://drive.google.com/file/d/17plnqZVyBGADlYRXMulZRu4te-zkbDBu/view?usp=sharing获取。所有CHDNet模型都采用相同的训练设置，即网络超参数、训练数据和数据预处理。通过为网络参数设置不同的初始种子，对每个CHDNet模型进行三次训练。在模型训练过程中，我们将一小部分训练集划分为验证集，在验证集上具有最高准确性的模型被用作最终模型。

2.5. 基于训练好的CHDNet自动选择关键帧

超声心动图视频通常包含大量帧（>50帧）。选择能够清楚显示心脏缺陷的关键帧并将其推荐给心脏超声医师进行进一步诊断非常重要，可以节省大量人力和时间。为此，我们首先通过对训练集中所有异常病例的分类特征进行平均，得到每个心脏缺陷的支持原型（support prototype）特征$f_s^d, d\in\{\mathrm{VSD, PDA, ASD}\}$[图1（b），右]，表述如下:

$$f_s^d=\frac{1}{S}\sum_{k=1}^{S}f_\theta\left(x_s(k)\right), d\in\{\mathrm{VSD, PDA, ASD}\} \quad (1)$$

式中，$x_s=\{x_s(k)|k=1,2,...,S\}$表示包含$S$个心脏缺陷异常超声心动图的支持集（support set）。$f_\theta(\cdot)$表示经过训练的CHDNet模型，参数为θ。支持原型特征f_s^d是每个心脏缺陷的元表示。

然后，我们计算待诊断超声心动图视频每一帧的支持原型特征与分类特征之间的欧氏距离。最后，那些与支持原型特征距离最小的帧被选为关键帧。

$$L_{\mathrm{keyframe}}(x_i)=\underset{x_i}{\mathrm{argmin}}\left\|f_s^d-f_\theta(x_i)\right\|_2,\ i=\{1,2,\cdots,T\} \quad (2)$$

式中，T表示超声心动图视频中的帧数。通常，选择3~4帧作为关键帧即可。

2.6. 基于关键帧的心脏缺陷检测

仅仅筛选出当前超声心动图视频中的帧是否包含心脏缺陷是不够的，告诉心脏超声医生心脏缺损的具体位置和大小更有意义。因此，两位经验丰富的心脏超声医师被邀请使用边界框（附录A中的表S3和表S4）在超声心动图中准确标注每个心脏缺损的位置和大小，然后用于训练Faster-RCNN模型[37]——用于心脏缺陷检测的经典检测网络。最后，将经过训练的Faster-RCNN模型应用于选取的关键帧而不是超声心动图视频中的所有帧进行诊断，从而显著减少诊断时间。

检测网络的处理流程如下。选择ResNet-50 [36]作为特征提取器。之后，基于ResNet-50第1层到第4层的学习表示，使用特征金字塔网络[38]计算多尺度特征表示，从而处理超声心动图中不同大小的心脏缺损问题。最后，使用基于多尺度表示的区域推荐网络来生成感兴趣区域，以确定每个预定义的锚点是否有心脏缺陷，以及优化这些锚点的大小和位置。在我们的研究中，我们将预定义的锚点设置为具有三个纵横比{1:2, 1:1, 2:1}和五个尺度$\{16^2,32^2,64^2,128^2,256^2\}$。该设置充分考虑了超声心动图中心脏缺损的实际大小和轮廓。在第二阶段，我们使用感兴趣区域池化操作来提取推荐区域的特征。类别预测器用于预测区域类型并进一步优化每个边界框的位置和大小。最后，我们使用非最大抑制（NMS）[39]算法去除置信度低于0.3的冗余预测边界框。

2.7. 可靠性度量

传统的分类网络使用交叉熵函数进行训练，可以描述如下：输入图像和输出标签的对$\left\{\left(x_i^d,y_i^d\right)\right\}_{i=1}^{N}$，$d\in\{\mathrm{VSD, PDA, ASD}\}$用于通过最小化交叉熵函数$L_{\mathrm{ce}}(\theta)$来训练CHDNet。

$$L_{\mathrm{ce}}(\theta)=-\frac{1}{N}\sum_{i=1}^{N}y_i\log\left(p\left(x_i\right)\right)+\left(1-y_i\right)\log\left(1-p\left(x_i\right)\right) \quad (3)$$

式中，N是训练样本的总数。$p\left(x_i\right)=\mathrm{Softmax}\left(\hat{y}_i=1|x_i\right)=\frac{\exp\left(f_\theta^1\left(x_i\right)\right)}{\sum_{j=0}^{1}\exp\left(f_\theta^j\left(x_i\right)\right)}$为样本$x_i$预测为异常的概率。$\hat{y}_i$表示诊断结果。

贝叶斯神经网络被用于为深度网络提供概率解释，但它们的前馈推理受到计算开销过大的影响[40–42]。为了克服这个问题，Gal和Ghahramani [16]提出使用dropout技术来近似贝叶斯推理。他们的研究工作表明，模型预测的后验分布可以通过使用dropout方法的Monte Carlo sampling获得。我们这里沿用他们的方法，让CHDNet具备输出结果不确定性的能力，即近似贝叶斯推理的能力。我们选择$L_{\mathrm{uce}}(\theta,\sigma)$作为损失函数，而不是$L_{\mathrm{ce}}(\theta)$，以优化模型参数$\theta$和附加方差$\sigma$。

$$\begin{aligned}L_{\mathrm{uce}}(\theta,\sigma)&=\frac{1}{N}\sum_{i=1}^{N}-\log p\left(y_i=1\middle|f_\theta(x_i),\sigma\right)\\&=\frac{1}{N}\sum_{i=1}^{N}-\log \mathrm{Softmax}\left(y_i=1\middle|f_\theta(x_i),\sigma\right)\\&=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma^2}L_{\mathrm{ce}}(\theta)+\log\frac{\sum_{j=0}^{1}\exp\left(\frac{1}{{\sigma_i}^2}f_\theta^j(x_i)\right)}{\left(\sum_{j=0}^{1}\exp\left(f_\theta^j(x_i)\right)\right)^{\frac{1}{\sigma^2}}}\\&=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma^2}L_{\mathrm{ce}}(\theta)+\log\sigma\end{aligned}\tag{4}$$

式中，一种假设$\left(\sum_{j=0}^{1}\exp\left(f_\theta^j(x_i)\right)\right)^{\frac{1}{\sigma^2}}\approx\frac{1}{\sigma}\sum_{j=0}^{1}\exp\left(\frac{1}{{\sigma_i}^2}f_\theta^j(x_i)\right)$用于简化上式最终转换[16]。

在网络优化中，由于数值不稳定，直接优化σ是困难的。相反，我们决定优化$\log\sigma$。然后，公式（4）改为：

$$L_{\mathrm{uce}}(\theta,z)=e^{z}L_e(\theta)+z\tag{5}$$

其中

$$z=\log\sigma\tag{6}$$

式中，方差σ用于衡量偶然不确定性。

认知不确定性可以通过计算多个模型结果的方差来衡量。由于dropout的优势，需要通过为模型初始化设置不同的初始种子，在同一个数据集上独立训练M个诊断模型。因此，我们可以通过在测试期间保持dropout打开来对多个不同的网络进行采样。

$$\hat{y}_i=E\left[P_j\left(y=1\middle|f_\theta(x_i,\sigma_i)\right)\right],j=1,2,\cdots,M\tag{7}$$

式中，E表示数学期望；$\hat{y}_i$表示模型输出的期望；P_j是第j个模型的预测概率。

$$u_i=\frac{1}{M}\sum_{j=1}^{M}\left(P_j\left(y=c\middle|f_\theta(x_i),\sigma_i\right)-\hat{y}_i\right)^2\tag{8}$$

式中，u_i表示模型输出的认知不确定性。认知不确定性u_i是评估训练后CHDNet知识盲点的有效客观指标，其特征表现如图3所示。

2.8. 用于提高模型鲁棒性的动态反馈单元

动态反馈单元的主要思想是利用分类层的特征-类别变换矩阵$W_{\mathrm{clf}}\in R^{C_{\mathrm{out}}\times C}$获得注意力图，然后利用注意力图更新分类模型中的中间层特征F_{in}^{i-1}，从而获得更鲁棒的分类特征。整个反馈过程如算法1所示。

算法1 动态反馈单元

Input:	Input $F_{\mathrm{in}}^{i-1}\in R^{H\times W\times C_{\mathrm{in}}}$ and output $F_{\mathrm{out}}^{i-1}\in R^{H\times W\times C_{\mathrm{out}}}$ features of a middle module (contains several convolutional layers) in a classification model, matrix $W_{\mathrm{clf}}\in R^{C_{\mathrm{out}}\times C}$ in classification layer, max iterations T. C_{in} and C_{out} denote feature channels. C denotes the number of categories.
Output:	Feedback feature F_{fb}^{i-1}
Target:	Updating the input feature F_{in}^{i-1}
1	**for** i = 1 to T **do**
2	$F_{\mathrm{fb}}^{\mathrm{cls}}=\mathrm{Softmax}\left(F_{\mathrm{out}}^{i-1}.W_{\mathrm{clf}}\right)$
3	$F_{\mathrm{fb}}^{\mathrm{cls}}\leftarrow F_{\mathrm{fb}}^{\mathrm{cls}}[\ldots,k]$% Slice, where k indexes the abnormal channel and use it as the attention map
4	$F_{\mathrm{fb}}^{\mathrm{fuse}}=\left(F_{\mathrm{fb}}^{\mathrm{cls}}\otimes F_{\mathrm{in}}^{i-1}\right)\oplus F_{\mathrm{in}}^{i-1}$
5	$F_{\mathrm{fb}}^{i-1}=Conv_{1\times1}\left(F_{\mathrm{fb}}^{\mathrm{fuse}}\right)$
6	$F_{\mathrm{in}}^{i-1}\leftarrow F_{\mathrm{fb}}^{i-1}$
7	**end**

动态神经反馈单元的架构如图4（a）所示。对于深度神经网络中的浅层，输入和输出分别表示为F_{in}^{i-1}和F_{out}^{i-1}。反馈单元以分类层开始：

$$F_{\mathrm{fb}}^{\mathrm{cls}}=\mathrm{Softmax}\left(F_{\mathrm{out}}^{i-1}\cdot W_{\mathrm{clf}}\right)\tag{9}$$

式中，$F_{\mathrm{fb}}^{\mathrm{cls}}$是分类层的输出，它表示浅层特征$F_{\mathrm{out}}^{i-1}$对分类结果的贡献。

然后，$F_{\mathrm{fb}}^{\mathrm{cls}}$和$F_{\mathrm{in}}^{i-1}$通过两个串行操作进行整合：逐点乘法$\otimes$和逐通道级联$\oplus$。

$$F_{\mathrm{fb}}^{\mathrm{fuse}}=\left(F_{\mathrm{fb}}^{\mathrm{cls}}\otimes F_{\mathrm{in}}^{i-1}\right)\oplus F_{\mathrm{in}}^{i-1}\tag{10}$$

式中，$F_{\mathrm{fb}}^{\mathrm{fuse}}$是融合结果，其特征通道数是$F_{\mathrm{in}}^{i-1}$的两倍。为了让$F_{\mathrm{fb}}^{\mathrm{fuse}}$在不对网络做任何修改的情况下传入原网络，我们使用一个核大小为1×1（Conv1×1）的卷积层来压缩$F_{\mathrm{fb}}^{\mathrm{fuse}}$的通道，获得与$F_{\mathrm{in}}^{i-1}$相同数量的通道。

$$F_{\mathrm{fb}}^{i-1}=\mathrm{Conv}_{1\times1}\left(F_{\mathrm{fb}}^{\mathrm{fuse}}\right)\tag{11}$$

式中，F_{fb}^{i-1}是反馈单元的输出，可以看作是F_{in}^{i-1}的反馈改进后的“干净特征”。从而F_{in}^{i-1}中不相关的背景噪声和模式被抑制，而在F_{fb}^{i-1}中不存在。

从式（9）至式（11），我们发现反馈单元的关键思想是充分利用分类层中的变换矩阵，因为该矩阵可以将网络在输入上提取的高维特征向量投影到类别空间。这个投影过程是一个降维过程，保留与分类相关的特征信

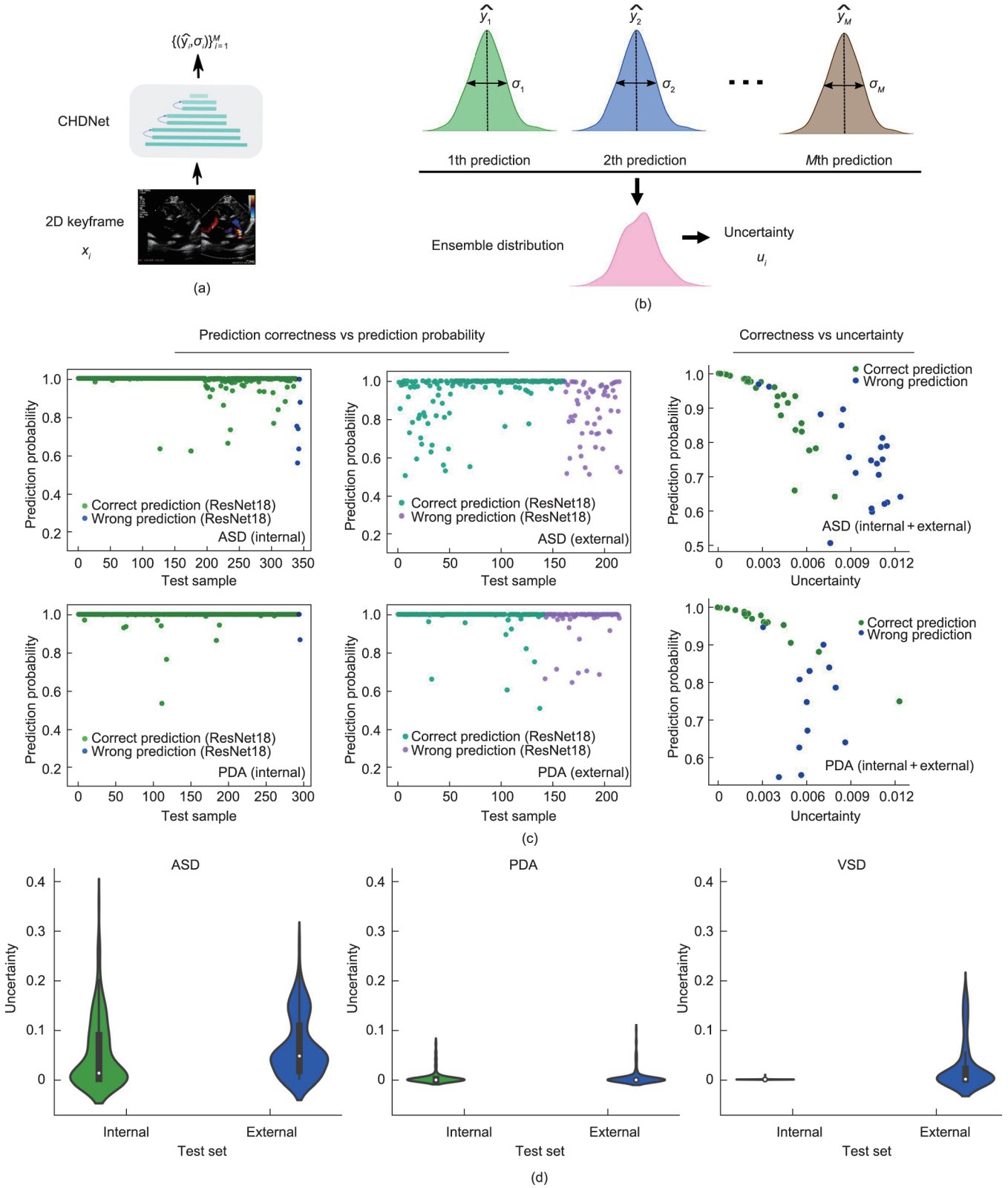


图3. 采用贝叶斯推理的CHDNet的可靠性。（a）CHDNet (ResNet18)的输出被修改为一个向量$(\hat{y}_i, \sigma_i)$，i={1, 2, ⋯, M}，其中，σ是每个输出的方差；（b）对M个分布进行组合，得到输入x_i的认识不确定性u_i；（c）预测正确率与预测概率（左），以及预测正确率与不确定性（右）；（d）从左到右，ASD、PDA和VSD内部和外部测试集的不确定性统计。

息，忽略与分类无关的特征信息。因此，我们提出应用这个变换矩阵来更新网络中间层的特征，从而帮助网络更好地过滤掉与分类无关的特征，即获得反馈改进后的“干净特征”。

2.9. 评价指标

在这项工作中，采用了常用的分类评估指标，包括特异性、敏感性、准确性、F1分数和曲线下面积（AUC）。为了评估检测模型的性能，我们计算了每种心脏缺陷中真阳性（D_{tp}）、假阳性（D_{fp}）和假阴性（D_{fn}）的数量。需要注意的是，我们根据预测框与医生标注的边界框的重合程度来判断一次检测是否准确。基于这些统计数据，我们使用以下公式计算了两个评估指标，即召回率和精确率：

$$D_{recall} = \frac{D_{tp}}{D_{tp} + D_{fn}} \tag{12}$$

$$D_{precision} = \frac{D_{tp}}{D_{tp} + D_{fp}} \tag{13}$$

式中，D_{recall}和$D_{precision}$用于评估检测性能。

2.10. 噪声鲁棒性评估

为了测试CHDNet模型的稳健性，我们将方差σ_{noise}为0.01~0.05的高斯噪声添加到输入超声心动图中。噪声输入定义如下：

$$x_{noise} = x_i + N\left(0, \sigma_{noise}^2\right) \tag{14}$$

式中，x_i表示输入超声心动图；$N\left(0, \sigma_{noise}^2\right)$表示以0为均值、$\sigma_{noise}$为方差的高斯分布。

2.11. 数据可用性

由于数据集整体较大，我们只上传了部分数据到GoogleDrive上。尽管如此，这部分数据仍然可以帮助我们了解CHDNet模型是如何开发和验证的。数据可在https://drive. google. com/file/d/16ZuZw6JuIqKHUTvYDt2z_0xgq8Ua7e1J/view?usp=sharing获得。

2.12. 代码可用性

用于诊断三种常见先天性心脏缺陷的CHDNet (ResNet18)的Pytorch代码及用于检测心脏缺陷位置和大小的Faster-RCNN可在https://drive. google. com/file/d/17pln-qZVyBGADlYRXMulZRu4te-zkbDBu/view?usp=sharing获得。

3. 结果

3.1. 用于诊断三种常见CHD的典型神经网络结构的评估

图2显示了三种CHDNet模型在三种常见先天性心脏缺陷（VSD、ASD和PDA）的内、外测试集上的诊断性能。受试者工作特征（ROC）曲线显示，这三种模型在内部测试集上一致达到较高的诊断性能，甚至在VSD和ASD上达到100%的AUC [图2（d）]。在外部测试集上，模型的性能显著下降。例如，CHDNet (ResNet18)的AUC在PDA的内部测试集中从100%下降到外部测试集中的57.8%（附录A中的表S5至表S7）。这些结果表明，来自不同采集设备的数据会部分导致CHDNet模型失败，这是大多数基于机器学习的诊断方法的常见问题。图2（e）为使用*t*-SNE算法降维后学习到的心脏缺陷分类特征的可视化结果。很明显，来自内部测试集的案例分离得更好，而来自外部测试集的案例分离得更差。基于另外两种架构（MobileNetV3-Small和ResNet34）的CHDNets结果如附录A中的图S1所示，这一观察结果与上述诊断结果一致。

我们接下来的问题是，在二维关键帧上训练的CHDNet是否可应用于基于超声心动图视频的CHD诊断。为此，我们将训练集中某一先天性心脏缺陷的所有异常病例均作为支持集。因此，ASD、VSD和PDA分别对应于一个支持集。支持集的平均分类特征被认为是支持原型特征，是CHD异常超声心动图的元表示。超声心动图视频的支持原型特征与帧特征之间的特征距离可以帮助我们选择最可能出现异常的帧。图2（c）为从VSD和ASD外部测试集中随机抽取的12个案例（对照组6例，病例组6例）的特征距离统计。VSD和ASD的支持原型特征更接近患者特征，而远离健康个体特征。通过选择异常关键帧，我们可以进一步使用检测模型检测心脏缺陷的空间位置和大小[附录A中的图S2（a）]。ASD内部集的检测精度和召回率分别为0.955和0.907，外部集的检测精度和召回率分别为0.962和0.752 [图S2（b）]。模型预测框与医生的注释高度吻合[附录A中的图S2（c）和图S3]。在另外两种类型的心脏缺陷（VSD和PDA）中也观察到类似的结果。

3.2. 采用贝叶斯推理测量CHD诊断网络可靠性

我们展示了CHDNet模型对三种常见先天性心脏缺陷内、外部测试集的诊断性能，以及模型与不同超声心动图模式的关系。然而，对于任何一种诊断方法，都必须解决诊断结果可靠性的测量问题，因为心脏超声医师往往不愿意使用可靠性未知的诊断模型。为此，在蒙特卡罗采样[16]之后，我们将贝叶斯推理嵌入到CHDNet模型中，并获得了诊断的不确定性[图3（a）和（b）]。我们首先绘制了CHD病例的预测精度与预测概率（Softmax后）之间的关系[图3（c），左]。对于错误的预测情况，模型仍然以相对较高的概率给出预测结果，即对自己的预测具有较高的置信度。因此，预测概率不能很好地衡量诊断的可靠性。

接下来，我们询问模型输出的方差σ与检验样本的预测正确性之间是否存在密切关系（附录A中的图S4）。直

观上，方差小表示对CHD诊断的置信度高，方差大表示置信度低。我们观察到，方差σ不能帮助我们直观地理解哪种情况的预测是可靠的或不可靠的。因此，我们可视化了案例的不确定性[图3（c），右]。预测错误的情况的不确定性明显高于预测正确的情况。这一发现表明，不确定性可以作为反映预测结果置信度的候选度量。图3（d）显示了VSD、ASD和PDA内部和外部测试集的不确定性。外部测试集的不确定性高于内部测试集的不确定性，这间接表明外部测试集的病例诊断比内部测试集的病例诊断更困难。

3.3. 动态神经反馈提高CHD诊断网络的可靠性和鲁棒性

贝叶斯推理可以获得CHDNet模型的不确定性并作为衡量其可靠性的指标，但这是不够的。提高CHDNet模型的可靠性和鲁棒性也至关重要，对于任何基于机器学习的诊断方法都是如此。CHDNet模型通常在具有特定参数设置的特定成像设备上进行训练。将经过训练的CHDNet应用于从不同设备或不同参数设置获得的未知情况时，其诊断性能可能会受到很大影响，表现出较低的可靠性和鲁棒性。然而，我们的人类视觉系统表现出极高的可靠性和鲁棒性，其抗干扰能力是非常强大的。受人类视觉皮层神经反馈机制的启发，我们提出了一种计算动态神经反馈单元，可将分类层（可以区分不同类别）的深层知识反馈回浅层，以消除与诊断任务无关的噪声或模式[图4（a）]。该反馈单元可应用于现有的深层神经网络和深层网络的不同层次。

为了测试所提出的反馈单元选择性激活相关神经元的能力，测试过程中在输入图像中加入不同程度的高斯噪声（σ 噪声= 0.01~0.05）。随着噪声的增加，除了PDA外部测试集无反馈外，其余内部和外部测试集的诊断准确度均逐渐降低[图4（b）]，表明即使是认真训练的CHDNet仍然显示出低鲁棒性。这种观察结果可以用网络特征来解释[图5（a）]。反馈单元帮助CHDNet模型抑制与CHD诊断目标无关的背景噪声，从而获得清晰的输出结果。在将CHDNet模型嵌入反馈单元后，它在三种CHD诊断的准确性、召回率和AUC方面均表现出更好的鲁棒性和更高的诊断性能[图4（b）和图5（b）、（c）]。

3.4. 与另一种模型进行比较

将我们的反馈单元与最近流行的一种方法——LassoNet [43]——进行了比较，后者可以在外部数据集上或加入噪声后提高模型预测的准确性。Lemhadri等[43]提出了LassoNet方法，通过对损失函数进行约束修改，使神经网络具有特征稀疏性。这种方法可以帮助网络选择特征子集，从而形成稀疏网络。我们使用已发布的Office代码，并对其稍加修改以使其适应我们的任务。我们的方法与LasoNet的比较如表1 [43]所示。使用解码器网络和基于树的分类器的LasoNet在内部数据集上的性能略低，但比在外部数据集上没有反馈单元的Resnet18模型的性能好。总的来说，我们发现带有反馈单元的ResNet18在内部和外部测试集上均取得了更好的性能。

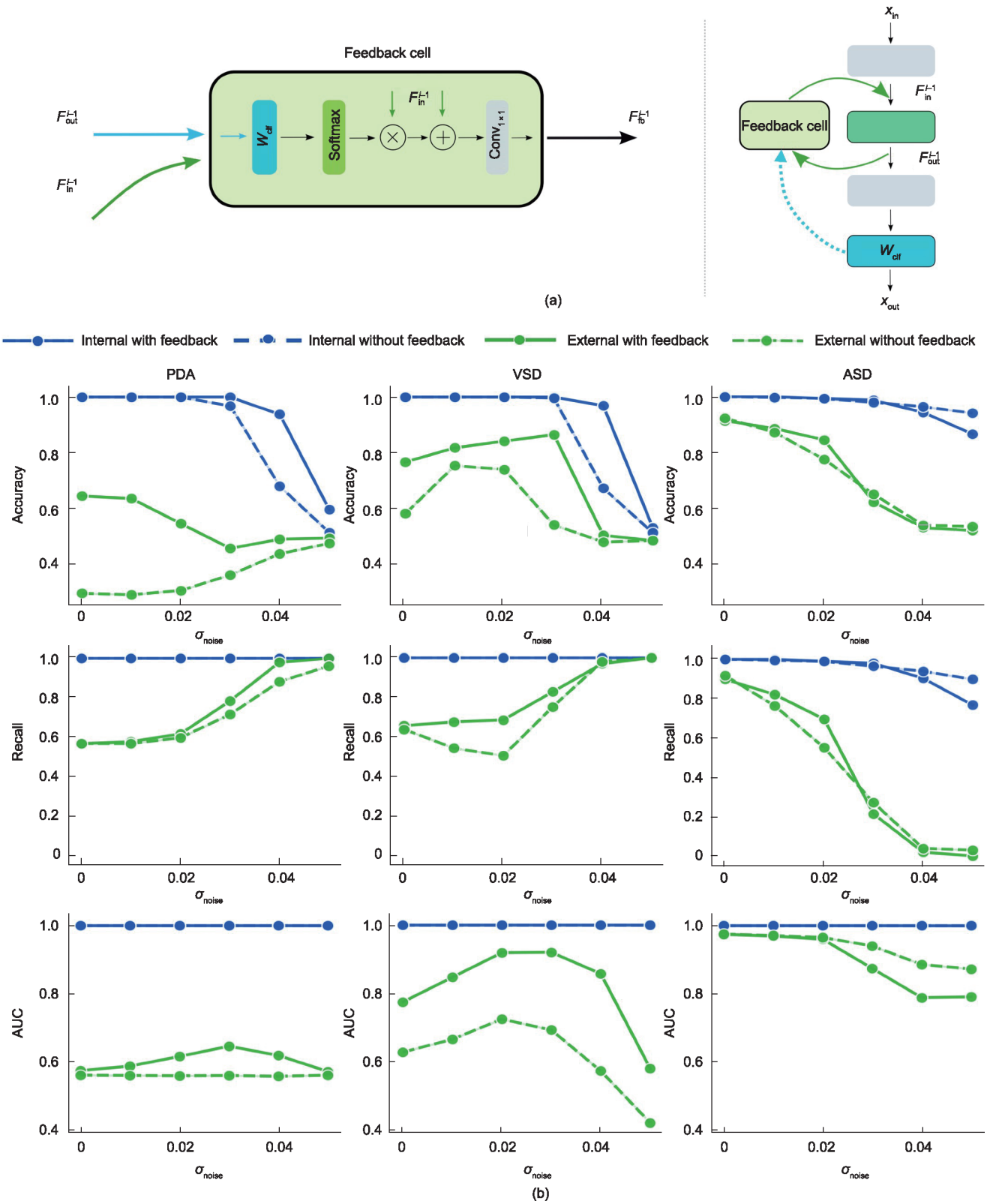
3.5. 与人类专家比较

超声心动图是心血管疾病筛查及诊疗评估中最重要且性价比最高的一项检查。目前我国大多数医院的心脏超声检测都是由专业的心脏超声医生来完成和出具报告，但不同级别医院的心脏超声医生水平差距很大。另外，心血管医生、新生儿及重症监护病房的医生也经常会遇到心脏超声医师不能及时到达时，他们需对患儿进行初步的心脏超声成像评估的情况。尤其是危重患儿，CHD的存在与否及类型会影响进一步药物和液体治疗的选择。因此，快速、准确地解读超声心动图像对这些医生来说非常重要。在临床工作中，我们经常遇到一些医生可以进行心脏超声成像，但却不能很好的给出诊断。

我们邀请了23名医生，包括8名心超医生和15名临床医生（来自不同医院的心血管专科医生、新生儿科医生和重症监护室医生）分别评估来自内、外部测试集的心超图像（50∶115），医生仅需给出诊断，模型则在给出诊断的同时进行缺损定位及初步测量。最后对比分析各自评估的准确性。附录A中的表S8显示了比较结果。在内部测试集的判读中，我们的模型具有绝对的优势。在外部测试集的判读中，与我院经验丰富的超声心动图专家和心内科

表1 本方法与LassoNet的分类精度比较[43]

Method	PDA		VSD		ASD	
	Internal test set	External test set	Internal test set	External test set	Internal test set	External test set
ResNet18	1.000	0.321	1.000	0.583	0.997	0.921
ResNet18 + feedback cell	**1.000**	**0.638**	**1.000**	**0.782**	**1.000**	**0.945**
LassoNet (decoder networks) [43]	0.972	0.548	0.981	0.674	0.974	0.933
LassoNet (tree-based classifiers) [43]	0.978	0.565	0.983	0.685	0.976	0.935



(b)

图4. 动态神经反馈机制。(a) 反馈单元示意图，它更新分类模型中的中间层特征F_{in}^{l-1}，通过使用分类层中的特征-类别转换矩阵W_{clf}获得更稳健的分类特征。(b) 从上到下是在不同噪声水平（σ_{noise} = 0.01~0.05）的干扰下，有反馈单元和无反馈单元的CHDNet的准确率、召回率和AUC。从左到右分别是PDA、VSD和ASD内部和外部测试集的诊断性能。

医生相比，我们的模型在准确性上不占优势，但与医疗水平相对逊色的超声科医生和心血管医生以及我院新生儿科和重症监护室的医生相比，我们模型的准确率明显高于他们，尤其是针对ASD和VSD的预测模型。

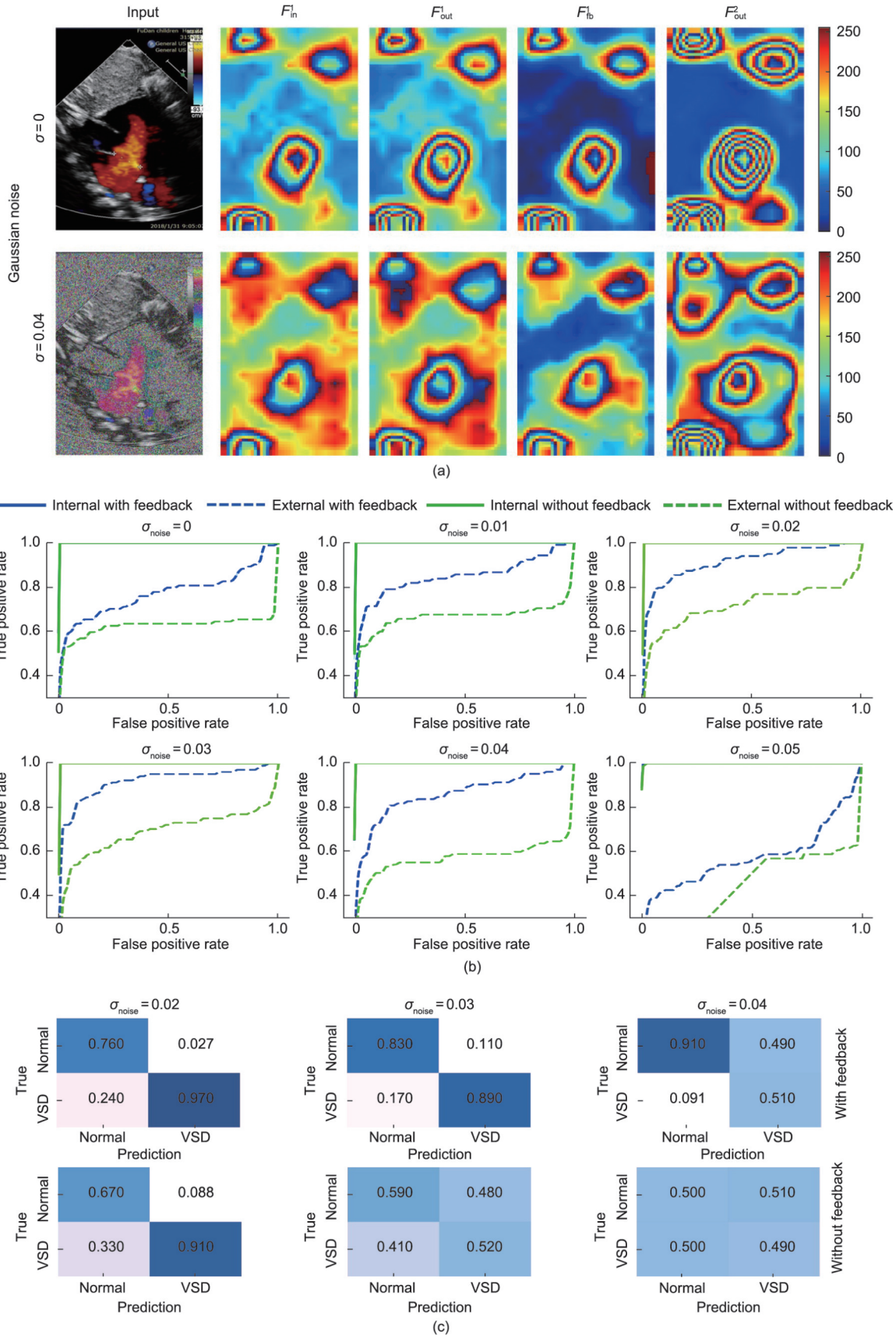


图5. 反馈单元的特征。（a）模型中加入反馈单元前后的网络特征对比。反馈单元抑制了与诊断无关的空间区域。（b）σ_{noise} = 0.01~0.05 时的 ROC 曲线。带反馈单元和不带反馈单元的 CHDNet (ResNet18)模型在 VSD 内部和外部测试集上的比较。（c）在高斯噪声 σ_{noise} = 0.02~0.04 条件下，有反馈单元和无反馈单元的 CHDNet 模型的混淆矩阵。

我们的模型受到了心脏病专家和临床医生的认可，有助于医疗水平相对落后的超声科医生提高工作效率及业务能力，同时有助于超声成像领域之外的医务人员更好地开展临床工作。

4. 讨论

对于三种常见的先天性心脏缺陷，我们基于获得的超声心动图数据，展示了用三种代表性神经网络架构（MobileNetV3-Small [35]、ResNet18 [36]和ResNet34 [36]）实现的CHDNet的诊断性能。我们观察到内部和外部测试集的诊断性能有显著差异。此外，我们进行了消融研究，讨论了不同超声心动图模式对CHDNet性能的影响，结果表明彩色血流模式优于灰阶模式，两者结合是最佳选择。这些结果表明，现有的先进深度模型具有足够的回归能力来实现CHD的诊断。然而，开发CHD诊断模型最紧迫的任务是解决临床相关性，而不仅仅是追求内部测试的高诊断性能。使我们的工作与众不同的是，我们通过嵌入贝叶斯推理使CHDNet输出其诊断可靠性，这使得心脏超声医师能够识别结果可能不正确的诊断病例。我们已经证明，CHDNet输出分布的方差不能准确地衡量诊断的可靠性，而CHDNet输出集合分布的不确定性可以很好地衡量诊断的可靠性。

在人类视觉皮层中，反馈的连接比前馈的连接更容易控制[44–45]。这一事实与目前流行的深层神经网络相反，深层神经网络在评估过程中基本上只包含前馈推理，其连接参数即使在输入变化时也保持不变。在现有的机器学习研究中，很少有人研究反馈推理的强大功能。在这项工作中，我们提出了一个计算动态神经反馈细胞，以提高CHD诊断的可靠性、鲁棒性和准确性。这种反馈单元可以很容易地嵌入现有的深层神经网络中，并在不增加大量计算复杂度的情况下对其进行显著改进。当输入超声心动图受到噪声严重干扰时，有反馈单元的心脏超声网络的表现明显优于没有反馈单元的心脏超声网络，这从三种常见CHD的评估结果可见一斑。

为了进一步了解反馈单元的能力，我们在CHDNet模型中观察了反馈单元前后神经层的特征。可视化特征映射表明，反馈单元赋予CHDNet模型选择性地激活相关神经元和抑制非相关干扰噪声或模式的能力，以完成CHD诊断任务。值得注意的是，反馈细胞与递归神经网络（RNN）、门控循环单元（GRU）或长期短期记忆网络（LSTM）有着显著的不同。这些循环单元使用前一刻的信息，而反馈单元使用分类层的变换矩阵来更新中间层的特征。RNN、GRU和LSTM的目的是利用序列中词的相关性，而反馈单元的目的是去除与分类目标无关的背景噪声。

然而，结合贝叶斯推理和动态神经反馈的CHDNet模型不能直接应用于临床诊断。大量的工作仍然是必要的，包括数据公平、道德合规、临床试验等。贝叶斯推理使CHDNet能够输出诊断的可靠性，该诊断已通过对三种常见先天性心脏缺陷获得的大规模真实世界内部和外部测试集进行了评估。然而，在临床实践之前，仍然需要对外部测试集进行全面而广泛的评估，包括不同种族、不同年龄、不同心脏超声成像设备。

通过抗噪声实验，我们在非增强和增强的CHDNet模型上评估了所提出的动态神经反馈机制的有效性。这种评价对临床实践是不够的，因为临床超声心动图成像不仅包括噪声，还包括模糊、抖动、患者内部和患者之间的变化等。因此，进一步的临床试验评价是必要的。

综上所述，我们的研究结果表明，CHDNet结合贝叶斯推理和动态神经反馈对三种常见先天性心脏缺陷的诊断具有更好的准确性、更高的可靠性和更强的鲁棒性。本文介绍的两种技术可以很容易地嵌入现有的基于深度神经网络的诊断方法中，从而提高其性能和可靠性。我们预测，当前超声心动图数据丰富度的爆炸式增长，以及CHDNet通过神经反馈动态适应各种先心病案例的能力，将释放CHDNet的巨大临床潜力，并使这种学习方法在临床中普及。

致谢

本研究得到复旦大学附属儿科医院大数据与人工智能项目（2020DSJ07）、国家自然科学基金（U2001209; 61902076; 81670281）、上海市卫生健康委员会科研项目（20204Y0100）及上海市自然科学基金项目（21ZR1406600）的资助。

作者贡献

颜波、李健、黄国英、谭伟敏、曹银银、马晓静、茹港徽负责整个研究的构思设计；谭伟敏、曹银银、马晓静、茹港徽实现了该算法；曹银银、张璟、高燕、杨佳伦、马晓静、李健收集并制作数据标签；谭伟敏、李吉春、茹港徽提出基于深度学习的CHD诊断；颜波、谭伟敏、茹港徽设计验证实验；谭伟敏、茹港徽对网络进行训练并进行验证实验；颜波、李健、

黄国英负责项目监督，所有作者均参与研究并撰写论文。

Compliance with ethics guidelines

Weimin Tan, Yinyin Cao, Xiaojing Ma, Ganghui Ru, Jichun Li, Jing Zhang, Yan Gao, Jialun Yang, Guoying Huang, Bo Yan, and Jian Li declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2022.10.015.

References

[1] Erikssen G, Liestøl K, Seem E, Birkeland S, Saatvedt KJ, Hoel TN, et al. Achievements in congenital heart defect surgery: a prospective, 40-year study of 7038 patients. Circulation 2015;131(4):337–46, Discussion 346.

[2] Luo H, Qin G, Wang L, Ye Z, Pan Y, Huang L, et al. Outcomes of infant cardiac surgery for congenital heart disease concomitant with persistent pneumonia: a retrospective cohort study. J Cardiothorac Vasc Anesth 2019;33(2):428–32.

[3] Liu S, Wang Y, Yang X, Lei B, Liu L, Li SX, et al. Deep learning in medical ultrasound analysis: a review. Engineering. 2019;5(2):261–75.

[4] Rong G, Mendez A, Bou Assi E, Zhao B, Sawan M. Artificial intelligence in healthcare: review and prediction case studies. Engineering. 2020;6(3):291–301.

[5] Sedghi S, Huang B. Real-time sssessment and diagnosis of process operating performance. Engineering 2017;3(2):214–9.

[6] O'Neill S. Handheld ultrasound advances diagnosis. Engineering 2021;7(11):1505–7.

[7] Cui Z, Yang B, Li RK. Application of biomaterials in cardiac repair and regeneration. Engineering 2016;2(1):141–8.

[8] Li C, Pisignano D, Zhao Y, Xue J. Advances in medical applications of additive manufacturing. Engineering 2020;6(11):1222–31.

[9] Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, et al. Video-based AI for beat-to-beat assessment of cardiac function. Nature 2020;580(7802):252–6.

[10] Arnaout R, Curran L, Zhao Y, Levine JC, Chinn E, Moon-Grady AJ. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. Nat Med 2021;27(5):882–91.

[11] Bates S, Angelopoulos A, Lei L, Malik J, Jordan M. Distribution-free, risk-controlling prediction sets. J Assoc Comput Mach 2021;68(6):1–34.

[12] Sadinle M, Lei J, Wasserman L. Least ambiguous set-valued classifiers with bounded error levels. J Am Stat Assoc 2019;114(525):223–34.

[13] Affenit RN, Barns ER, Furst JD, Rasin A, Raicu DS. Building confidence and credibility into CAD with belief decision trees. In: Proceedings of the Medical Imaging 2017: Computer-Aided Diagnosis; 2017 Mar 3, Orlando, FL, USA.

[14] Scheffe H, Tukey JW. Non-parametric estimation. I. validation of order statistics. Ann Math Stat 1945;16(2):187–92.

[15] McClure P, Kriegeskorte N. Robustly representing uncertainty in deep neural networks through sampling. In: Proceedings of the Second Workshop on Bayesian Deep Learning (NIPS 2017); LongBeach, CA, USA.

[16] Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning; 2016 Jun 16 – 24; New York, NY, USA. JMLR: W&CP; 2016. p.1050–9.

[17] Vishnevskiy V, Walheim J, Kozerke S. Deep variational network for rapid 4D flow MRI reconstruction. Nat Mach Intell 2020;2(4):228–35.

[18] Piray P, Daw ND. A model for learning based on the joint estimation of stochasticity and volatility. Nat Commun 2021;12(1):6587.

[19] Lazar A, Lewis C, Fries P, Singer W, Nikolic D. Visual exposure enhances stimulus encoding and persistence in primary cortex. Proc Natl Acad Sci USA 2021;118(43):e2105276118.

[20] Chariker L, Shapley R, Hawken M, Young LS. A theory of direction selectivity for macaque primary visual cortex. Proc Natl Acad Sci USA 2021;118(32):e2105062118.

[21] Ferro D, van Kempen J, Boyd M, Panzeri S, Thiele A. Directed information exchange between cortical layers in macaque V1 and V4 and its modulation by selective attention. Proc Natl Acad Sci USA 2021;118(12):e2022097118.

[22] Gilbert CD, Sigman M. Brain states: top-down influences in sensory processing. Neuron 2007;54(5):677–96.

[23] Hupé JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 1998;394(6695):784–7.

[24] Stollenga MF, Masci J, Gomez F, Schmidhuber J. Deep networks with internal selective attention through feedback connections. In: Proceedings of the 27th International Conference on Neural Information Processing Systems; 2014 Dec 8–13; Montreal, Canada. Cambridge, MA: MIT Press; 2014. p. 3545–53.

[25] Ozawa T, Ycu EA, Kumar A, Yeh LF, Ahmed T, Koivumaa J, et al. A feedback neural circuit for calibrating aversive memory strength. Nat Neurosci 2017;20(1):90–7.

[26] Williams MA, Baker CI, Op de Beeck HP, Shim WM, Dang S, Triantafyllou C, et al. Feedback of visual object information to foveal retinotopic cortex. Nat Neurosci 2008;11(12):1439–45.

[27] Cao C, Huang Y, Yang Y, Wang L, Wang Z, Tan T. Feedback convolutional neural network for visual localization and segmentation. IEEE Trans Pattern Anal Mach Intell 2019;41(7):1627–40.

[28] Cao C, Liu X, Yang Y, Yu Y, Wang J, Wang Z, et al. Look and think twice: capturing top-down visual attention with feedback convolutional neural networks. In: Proceedings of the 2015 IEEE International Conference on Computer Vision; 2015 Dec 7 – 13; Santiago, Chile. Washington, DC: IEEE Computer Society; 2015. p. 2956–64.

[29] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18 – 23; Salt Lake City, UT, USA. Washington, DC: IEEE; 2018. p. 1664–73.

[30] Gal Y, Islam R, Ghahramani Z. Deep Bayesian active learning with image data. In: Proceedings of the 34th International Conference on Machine Learning; 2017 Aug 6–11; Sydney, NSW, Australia. JMLR.org; 2017. p. 1183–92.

[31] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115–8.

[32] Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 2018;24(10):1559–67.

[33] Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019;25(6):954–61.

[34] Arnaout R. Toward a clearer picture of health. Nat Med 2019;25(1):12.

[35] Howard A, Sandler M, Chen B, Wang W, Chen L, Tan M, et al. Searching for mobileNetV3. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision; 2019 Oct 27 – Nov 2; Seoul, Republic of Korea. Washington, DC: IEEE; 2019.

[36] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2016 Jun 27 – 30; Las Vegas, NV, USA. Washington, DC: IEEE; 2016. p. 770–8.

[37] Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017;39(6):1137–49.

[38] Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition; 2017 Jul 21 – 26; Honolulu, HI, USA. Washington, DC: IEEE; 2017. p.: 936–44.

[39] Liu Z, Hu J, Weng L, Yang Y. Rotated region based CNN for ship detection. In: Proceedings of the 2017 IEEE International Conference on Image Processing; 2017 Sep 17–20; Beijing, China. Washington, DC: IEEE; 2017.p. 900–4.

[40] Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. Nature 2021;599(7883):91–5.

[41] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach

Intell 2017;39(12):2481–95.
[42] Krygier MC, LaBonte T, Martinez C, Norris C, Sharma K, Collins LN, et al. Quantifying the unknown impact of segmentation uncertainty on image-based simulations. Nat Commun 2021;12(1):5414.
[43] Lemhadri I, Ruan F, Abraham L, Tibshirani R. LassoNet: neural networks with feature sparsity. J Mach Learn Res 2021;22:1–29.
[44] Suway SB, Schwartz AB. Activity in primary motor cortex related to visual feedback. Cell Rep 2019;29(12):3872–84.e4.
[45] Marques T, Nguyen J, Fioreze G, Petreanu L. The functional organization of cortical feedback inputs to primary visual cortex. Nat Neurosci 2018;21(5):757–64.