DOI:10.19651/j.cnki.emt.2520019

基于 Mamba 自注意力和多模态融合的 DVT 风险预测*

肖连禹[1,2] 陆向艳[1]

(1.广西大学计算机与电子信息学院 南宁 530004;2.中国人民解放军陆军军医大学第一附属医院 重庆 400037)

摘 要:深静脉血栓可能导致肺栓塞等严重并发症危及患者生命安全,早期进行 DVT 风险预测具有重要临床意义。针对当前 DVT 风险预测存在仅对单一结构化文本数据或图像数据进行预测,未有结合这两种模态数据进行预测的方法及深度学习最新模型应用于 DVT 风险预测方法较少这两个挑战问题。本研究将 Mamba 状态空间模型与多模态融合结合,首次提出一种基于 Mamba 自注意力机制和多模态融合的 DVT 风险预测方法。所提方法以患者超声影像和病史、实验室检验指标等结构化文本数据作为多模态输入数据,首先构建双通道特征编码框架,该框架利用 ViT 编码捕获医学超声图像特征,DNN 编码获取结构化临床数据特征;然后设计基于 Mamba 自注意力和多模态特征融合框架,该框架首先拼接图像和结构化文本特征得到联合特征,采用原始 Mamba 训练联合特征得到多模态融合特征,然后设计 Mamba 自注意力、前馈网络和 CNN 实现多模态数据全局和局部、高层和底层特征提取和融合,实现多角度保留原始多模态特征;最后进行多层次 MLP 特征降维获得 DVT 预测结果。在临床数据集上与其他 13 种组合模型进行对比实验,结果表明该模型效果最佳,AUC 达到 0.912,较结构化数据单模态方法平均提高了 11.97%,$F1$ 分数平均提高了 13%。与传统图像数据单模态对比模型 AUC 平均提高 14.7%,准确率和 $F1$ 分数均提高 20%以上。在多模态对比模型中,该模型与表现较优的 ResNet 与 Transformer 融合的模型(AUC=0.871)相比,其准确率、精确率、召回率、$F1$ 分数均约提高了 6%,与同结构的 Transformer 混合模型相比,AUC 与其余 4 项性能评估指标以及模型推理速度都提高 20%以上。结果表明本研究中的模型为 DVT 的早期预防和预测提供了有力支持,具有良好的应用前景和临床价值。

关键词:Mamba;自注意力;多模态;深静脉血栓 DVT;预测;多层级融合

中图分类号:TN919.5;TP391.4 **文献标识码**:A **国家标准学科分类代码**:510.40

DVT risk prediction via Mamba self-attention and multimodal fusion

Xiao Lianyu[1,2] Lu Xiangyan[1]

(1. School of Computer and Electronic Information, Guangxi University, Nanning 530004, China;
2. The First Affiliated Hospital of Army Military Medical University, Chongqing 400037, China)

Abstract: Deep vein thrombosis can potentially give rise to severe complications like pulmonary embolism, which poses a threat to the life safety of patients. Therefore, early prediction of DVT risk holds significant clinical implications. However, current DVT risk prediction methods mainly focus on only predict using either single-text or single-image data, and there are few studies which integrate these two types of modal data for DVT risk prediction. To address these challenges, this study combines the Mamba state space model with multimodal fusion and proposes a novel DVT risk prediction method based on Mamba self-attention and multimodal fusion for the first time. This method takes the patient's ultrasound images and structured text data such as medical history and laboratory test indicators as multimodal input data. Firstly, a dual-channel feature encoding framework is constructed, which uses ViT to capture the features of ultrasound images and DNN to obtain the features of structured clinical data. Then, this paper proposes a multimodal feature fusion framework based on Mamba self-attention. This framework first concatenates the image and text features to obtain the joint features, and then uses the original Mamba to train the joint features to obtain the multimodal fusion features. Subsequently, Mamba self-attention, feedforward network, and CNN are designed to extract and fuse global and local, high-level and low-level features of multimodal data, thereby preserving the original multimodal features from multiple perspectives. Finally, multi-level MLP is used for feature dimension reduction to obtain the DVT prediction results. Comparative experiments were conducted on a clinical dataset with 13 other combined models. The results show that this model outperforms the others, with an AUC of 0.912, an average improvement of 11.97% compared to the single structured data model, and an average improvement of 13% in $F1$ score. Compared with the traditional single image data model, the AUC is improved by an average of 14.7%, and the accuracy and $F1$ score are both increased by more than 20%. Among the multimodal comparison models, this model outperforms the ResNet and Transformer fusion model (AUC=0.871) in terms of accuracy, precision, recall, and $F1$ score by approximately 6%. Compared with the same-structured Transformer hybrid model, the AUC and the other four performance evaluation indicators as well as the model inference speed are all improved by more than 20%. The results indicate that the model proposed in this study provides strong support for the early prevention and prediction of DVT and has good application prospects and clinical value.

Keywords: Mamba; self-attention; multiple model; deep vein thrombosis; prediction; multi-level fusion

0 引 言

深静脉血栓(deep vein thrombosis,DVT)是静脉血管中出现异常血凝块使静脉血液回流异常,引发深静脉功能障碍的血管疾病(多发生于下肢深静脉)[1],DVT 可能导致肺栓塞等严重并发症危及患者生命安全[2-4],因此早期进行

收稿日期:2025-10-13

*基金项目:广西软科学研究计划(桂科 AB17205002)、广西多源信息挖掘与安全重点实验室开放基金(MIMS20-06)、重庆市社会科学规划博士和培育项目(2025PY27)资助

DVT 风险预测，预防其发生具有重要临床意义。临床上 DVT 风险预测主要采用 Caprini 评分、Padua 评分量表等传统评分表或改进量表进行评估[5-8]。传统评分模型[9]是静态评分，依赖专家经验，在实验室指标波动、多因素交互影响等情况下，难以满足临床精准预测需求。随着人工智能的发展，有不少研究者开始采用机器学习进行 DVT 风险预测。Ryan 等[10]应用梯度增强型机器学习算法预测患者在发病前 12 和 24 h 窗口内发生 DVT 的风险。孙玉萍等[11]通过常规超声检查或造影检测 UEDVT，利用最小绝对收缩和选择算子(least absolute shrinkage and selection operator regressio，LASSO)回归算法建立 Seeley 量表模型和机器学习模型，比较机器学习与现有预测工具的预测效果。张建楠等[12]构建下肢骨科手术后 DVT 形成风险的预测模型，通过单因素和多因素逐步回归分析法从 31 个变量中最终筛选出 5 个变量构建预测模型。Abbasi 等[13]利用决策树(decision tree，DT)等多种模型预测接受体外膜肺氧合治疗的患者发生出血和血栓形成的风险，结果表明决策树模型预测发生出血风险的准确率很高。Wang 等[14]在膝关节置换术后 DVT 风险预测的研究中，采用极限梯度提升(extreme gradient boosting，XGBoost)、随机森林(random forest，RF)、逻辑回归(Logistic Regression，LR)等多种机器学习算法和集成模型进行风险预测，结果表明运用多个模型的集成模型效果最优。Wang 等[15]主要是使用 XGBoost 模型预测患者后路腰椎融合术后的 VTE 风险，结果表明明显优于传统评估方式。徐青等[16]构建预测全髋关节置换术(THA)患者下肢深静脉血栓风险的机器学习模型，运用包括 K-最近邻(K-nearest neighbors，KNN)在内的 6 种机器学习算法开发预测模型，并用多种指标评估性能。

相较于传统评估方法，基于机器学习的 DVT 风险预测方法更能全面捕捉变量间的复杂交互作用，使预测的精准性得到很大的提升，但由于机器学习方法依赖于所选取的特征、处理的数据量较小、图像处理能力不强等原因导致预测精度仍有待提升。近年来，深度学习技术[17]使图像处理[18-19]能力得到极大提升，研究者转向利用深度学习方法对患者超声和磁共振等影像数据进行 DVT 风险预测以进一步提升精度。Seo 等[20]应用基于卷积神经网络(convolutional neural network，CNN)的模型来检测计算机断层扫描图像中的静脉血栓形成。段丽芬等[21]利用磁共振黑血血栓成像，基于残差网络(residual network，ResNet)、视觉 Transformer(vision transformer，ViT)等构建 3 个下肢 DVT 分期预测模型，计算准确率、曲线下面积评估其预测性能。Nakayama 等[22]对深度学习模型 ResNet101 进行了微调，用于对腘静脉超声图像进行 DVT 风险预测。Xue 等[23]采用 RF、深度神经网络(deep neural network，DNN)等模型联合手术前后数据预测手术并发症。

综上所述，基于机器学习和深度学习的 DVT 预测方法使预测精度得到很大的提高，但当前研究存在以下两个问题。一是仅对单一结构化文本数据或图像数据进行预测，未有结合这两种模态数据进行预测的方法；二是深度学习最新模型应用于 DVT 风险预测方法较少，多运用于图像识别分割[24-25]或其他常见疾病的风险预测，例如 Mitra 等[26]使用深度卷积神经网络(very deep convolutional networks for large-scale image recognition，VGG)进行脑肿瘤 MRI 分类，Wang 等[27]提出了一种基于 ViT 的多类病变检测模型在心血管图像数据集上展现较优的病变定位效果。为了应对这些挑战，进一步提高预测精度，Gu 等[28-30]提出了一种基于 Mamba 自注意力和多模态融合的 DVT 风险预测模型(M_MATT)。该方法首先采用 ViT[31]和 DNN 双分支结构提取图像与结构化文本特征，增强特征提取能力，提升疾病预测模型的鲁棒性和可解释性；接着采用改进后的 Mamba 进行多模态特征融合和多层次多尺度信息全局捕获；最后利用 CNN 增强局部再提取。本文主要贡献为：

1)模型架构创新。提出了一种基于 Mamba 自注意力和多模态融合的 DVT 风险预测模型(M_MATT)，首次应用该架构模型进行 DVT 静脉血栓风险预测。采用 Mamba 替代 Transformer[32]，与 ViT，DNN，CNN 分阶段作用于模型，实现线性复杂度与全局和局部特征提取，将底层特征与高层特征结合，融合了多尺度特征，大大增强模型的表达能力，预测效果和推理速度优于表现较好的同结构 Transformer 的混合模型。

2)多模态特征融合。设计 ViT＋DNN 双分支异构网络提取图像与结构化文本特征，有效解决临床数据的跨模态特征提取难题，显著提升多源数据融合效能。

3)针对临床 DVT 数据收集的难点，本研究不仅收集了临床患者基础信息、病史、检验指标，还包含超声图像，实现了多模态数据的收集。

1　方法模型

1.1　框架结构

基于 Mamba 自注意力和多模态融合的 DVT 风险预测模型(M_MATT)分为 3 个阶段(如图 1 所示)：ViT＋DNN 多模态特征编码、Mamba 自注意力和多模态特征融合和 DVT 预测输出。

ViT＋DNN 多模态特征编码是 M_MATT 模型的第 1 阶段，作用是对 DVT 患者数据进行多模态特征编码。本文选择患者 DVT 超声图像和结构化文本数据作为模型输入，由于 Mamba 的输入是序列数据，因此先提取高维图像和结构化数据后再转为序列数据。为了充分获取图像全局特征，采用 ViT 进行图像特征编码，ViT 能够解决局部感受野导致的远距离依赖建模困难问题，且具备良好的全局建模能力。患者结构化文本包括基础数据、实验室指标和

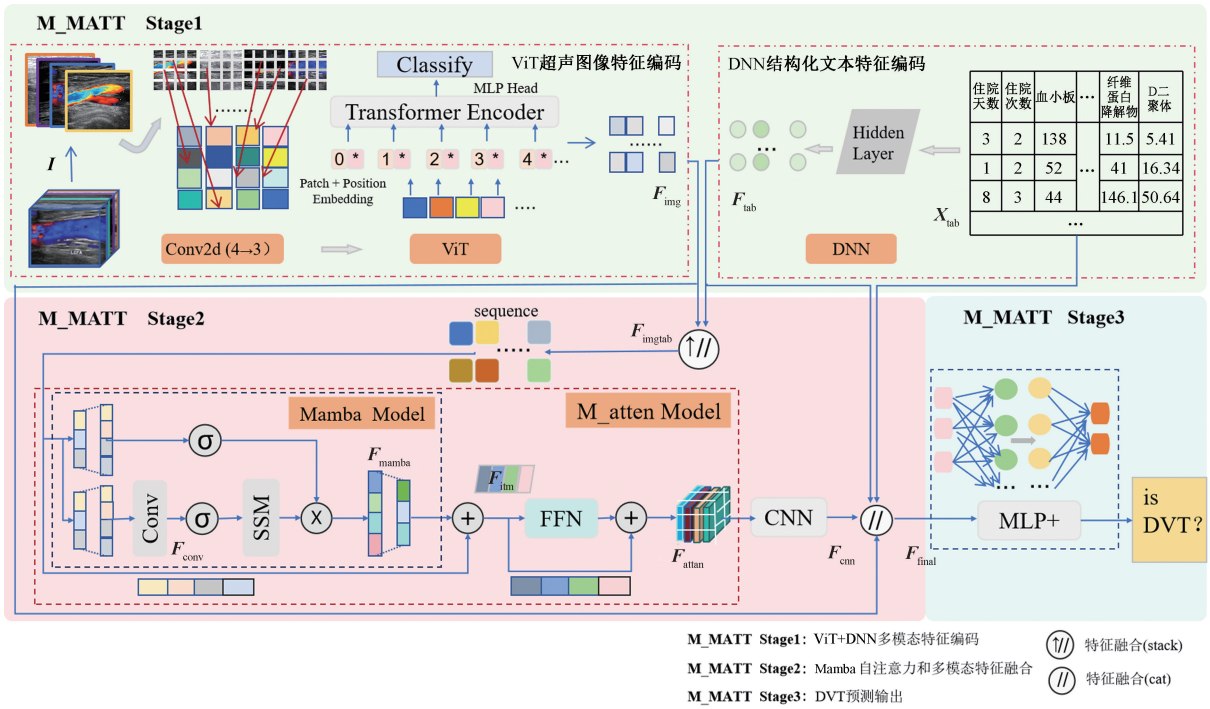


图 1 基于 Mamba 自注意力和多模态融合的模型(M_MATT)架构

Fig. 1 The architecture of Mamba-based self-attention and multimodal fusion model (M_MATT)

病史,采用 DNN 对结构化文本数据进行特征编码,DNN 通过多层非线性变换和端到端训练能实现复杂数据的建模,提升文本数据的语义建模能力。

基于 Mamba 自注意力和多模态特征融合是 M_MATT 模型的第 2 阶段,作用是实现患者超声图像特征和结构化文本特征的多模态融合。该阶段分为 3 个过程,第 1 个过程是图像和文本多模态特征对齐和初级融合;第 2 个过程是基于 Mamba 自注意力机制的全局融合特征捕获和多模态特征融合;第 3 个过程是基于 CNN 的局部细节特征捕获和多层级融合特征整合。

预测输出是 M_MATT 模型的第 3 阶段,该阶段的作用是输出预测结果。使用两次全连接层和激活函数构成的多层感知机处理特征融合层的输出结果,通过特征降维与抽象,输出最终分类和预测结果。

1.2 ViT+DNN 多模态特征编码

1)ViT 超声图像特征编码

ViT 在全局建模上具有不错的表现,ViT 通过自注意力机制直接建模图像所有区域的关系,无需逐层卷积即可捕获全局信息。相比于传统的卷积操作,ViT 能够更直接地捕捉图像中的长距离依赖和全局信息,有助于发现 DVT 在血管中的分布模式、形态特征等关键信息,因此在处理超声图像编码时选用 ViT 模型,ViT 特征编码网络流程包括通道变换和 ViT 处理两个步骤(如图 2 所示)。第 1 步通道变换,本研究每个患者选择了 4 张同一次检查的下肢超声图片进行预测分析,它们的采集位置和采集时间存在细微差异,所以样本图像数据为 4 通道,但标准 ViT 只接收 3 通道 RGB 图像,因此在 ViT 编码前需将 4 通道医学超声图像转成 3 通道自然图像,然后再提取多图融合特征和跨图关联信息。第 2 步是 ViT 处理,包括图像分块、Transformer 编码器、分类头设计。ViT 输入图像的大小为 224×224,将图像分块处理,分块大小为 16×16,将每个图像块映射为一个向量,形成序列化的图像表示,然后通过多头自注意力机制计算图像块之间的相互关系,生成全局的图像特征。

ViT 通道变换首先将超声图像裁剪填充为 224×224 大小,然后将向模型输入图像 $\boldsymbol{I} \in \mathbf{R}^{H\times W\times 4}$,随后采用 3×3 可学习卷积层:

$$\boldsymbol{I}' = Conv2D_{4\to 3}(\boldsymbol{I}) \in \mathbf{R}^{H\times W\times 3} \tag{1}$$

其中,$\mathbf{R}$ 表示图像特征,H 和 W 分别表示图像的高宽,$\boldsymbol{I}'$为映射特征图像切片。卷积核参数通过反向传播学习,实现 4 通道医学影像到 3 通道自然图像域的映射,如图 3 所示。ViT 处理将图像分割成多个图像块 $N = \frac{H}{16} \times \frac{W}{16}$ 个 16×16 的图块进行分块嵌入,嵌入映射公式为:

$$\begin{cases} \boldsymbol{P} = unfold(\boldsymbol{I}', \text{kernel} = 16) \in \mathbf{R}^{N\times 768} \\ \boldsymbol{X}_p = \boldsymbol{P} \cdot \boldsymbol{E}_p + \boldsymbol{b}_p \\ \boldsymbol{E}_P \in \mathbf{R}^{768\times d}, \boldsymbol{b}_p \in \mathbf{R}^d \end{cases} \tag{2}$$

其中,每个块展平后的原始维度为 16×16×3=768,$\boldsymbol{P}$ 表示分割的图块,d 为隐式特征维度,$\boldsymbol{X}_p$ 表示分块嵌入后

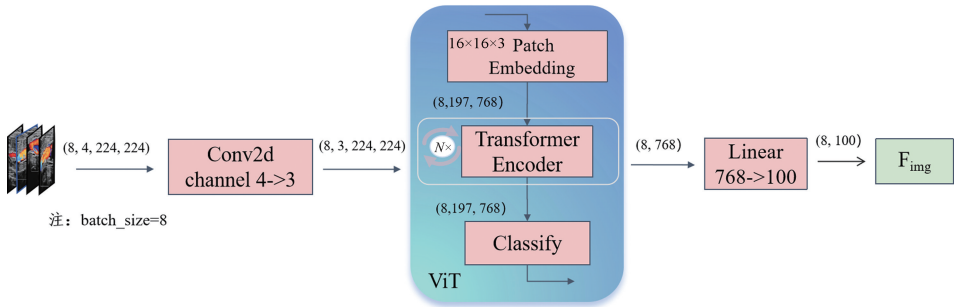


图 2　ViT 超声图像特征编码网络

Fig. 2　Ultrasound image feature encoding based ViT

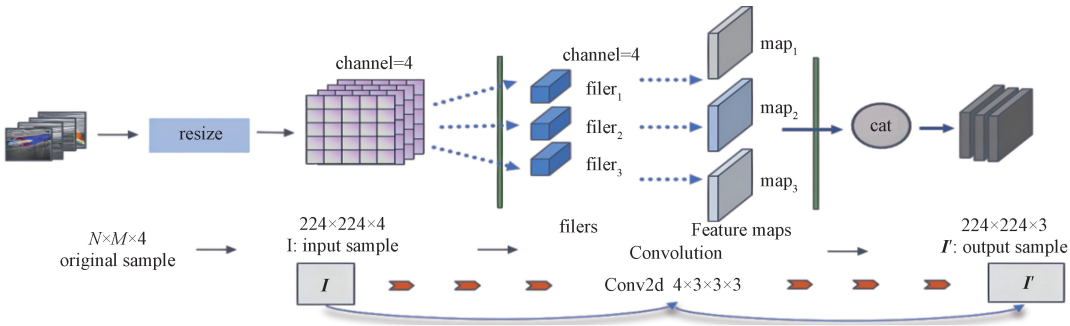


图 3　卷积 4 通道转 3 通道

Fig. 3　4-channel to 3-channel

的图像块。图像分块嵌入后与位置编码相加，保留图像块的空间位置信息。通过 12 层 Transformer 编码器堆叠处理后，输出图像特征 $\boldsymbol{F}_{img}$：

$$\boldsymbol{F}_{img}=\boldsymbol{Z}_L^0\in\mathbf{R}_d \tag{3}$$

其中，$\boldsymbol{Z}_L^0$ 为第 L 层类别标记。

2）DNN 结构化文本特征编码

DNN 适用于处理结构化的文本数据。本研究采用 DNN 处理结构化数据，通过多层神经元的非线性变换，学习数据中的复杂映射关系，提取出与 DVT 风险相关的潜在特征，经过 2 层 MLP 和激活函数映射后输出结构化文本特征 $\boldsymbol{F}_{tab}$：

$$\begin{cases}\boldsymbol{h}_1=Relu(\boldsymbol{W}_1\boldsymbol{X}_{tab}+b_1)\\ \boldsymbol{h}_2=\boldsymbol{W}_2\boldsymbol{h}_1+b_1\\ \boldsymbol{F}_{tab}=\boldsymbol{h}_2\end{cases} \tag{4}$$

其中，$\boldsymbol{X}_{tab}$ 表示结构化数据，$\boldsymbol{W}$ 表示权重矩阵，b 表示偏置，$\boldsymbol{h}$ 表示特征，DNN 特征编码网络流程如图 4 所示。

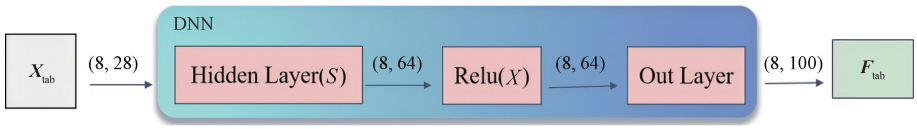


图 4　DNN 结构化文本特征编码网络

Fig. 4　DNN structured text feature encoder

1.3　基于 Mamba 自注意力机制的多模态特征融合

处理多模态数据需整合不同模态的信息，既要保证全局特征又要兼顾局部信息不被丢失，不仅需要提升模型对复杂场景的理解和决策能力，还需保证模型的鲁棒性和特征对齐。本研究在多模态特征融合阶段细分为 3 阶段处理：图像、结构化文本特征编码初级融合、基于 Mamba 自注意力机制的多模态特征融合、基于 CNN 的多层级多模态特征融合，以保证模型在具有线性复杂度的同时能够保证特征表达、全局特征提取和稳定训练，避免深层网络信息损失，高效的完成 DVT 疾病的分类预测。

1）图像、文本特征编码初级融合

为整合不同模态数据的优势，和后续多特征融合、跨模态多层次交互提供结构化输入以及为后续风险预测提供更全面、准确的特征基础，同时保留模态独立性，便于分别处理或联合分析。本研究将 ViT 编码图像特征与 DNN 编码结构化文本特征进行初级融合，生成多模态联合特征。用 stack 函数将图像特征和结构化特征无损拼接，保留跨模态特征，输出多模态联合特征 $\boldsymbol{F}_{imgtab}$：

$$\boldsymbol{F}_{imgtab}=\uparrow[\boldsymbol{F}_{img}//\boldsymbol{F}_{tab}] \tag{5}$$

其中，图像、结构化文本特征的无损融合能够保留多

模态的原始特征、对齐多尺度模态维度,为后续模态动态交互输入做准备。

2)基于 Mamba 自注意力机制的多模态特征融合

本研究模型引入 Mamba 注意力层,实现高效建模序列依赖关系,在长距离和短距离序列、全局特征和多模态交互上做到了互相补充。模型经过 Mamba 和前馈网络的训练,在高效捕获全局特征和增强多模态特征间动态交互能力的同时也降低了信息损失。因模型层次结构较深,因此在模型中增加了残差连接、Dropout 和层归一化操作,残差连接能够避免梯度消失,Dropout 能防止过拟合,层归一化可以加速收敛。基于 Mamba 自注意力机制的多模态特征融合包括 3 个步骤,如图 5 所示。

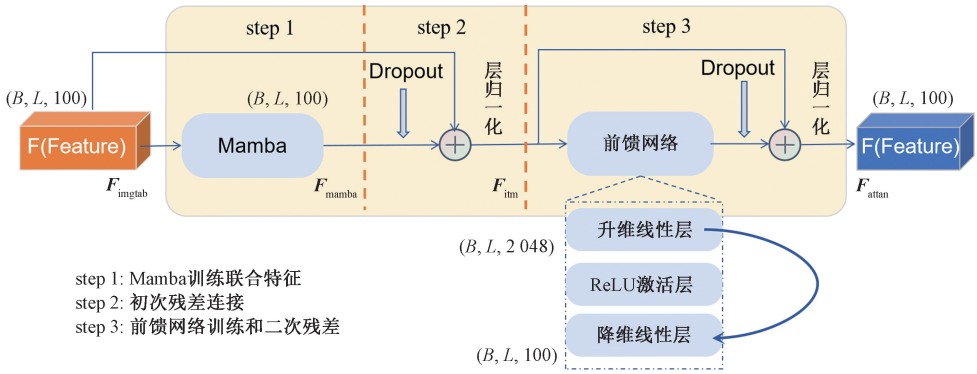


图 5 基于 Mamba 自注意力机制的多模态特征融合网络

Fig. 5 A multimodal feature fusion based on the Mamba self-attention network

第 1 步是采用原始 Mamba 训练联合特征得到多模态融合特征。方法是在 SSM 前施加深度可分离卷积,深度卷积能够有效捕捉序列的短距离依赖,从而避免纯 SSM 对局部信息的忽略,其输出卷积局部增强特征 $\boldsymbol{F}_{conv}$ 与 SMM 状态向量相结合,实现在综合不同模态局部特征的同时并增强位置敏感性。利用可学习门控系数 λ 动态特征选择,自动分配图像与结构化数据的注意力权重,自适应调节 SMM 处理后内容与多模态联合特征 $\boldsymbol{F}_{imgtab}$ 的融合权重,最终输出多模态融合特征 $\boldsymbol{F}_{mamba}$:

$$\boldsymbol{F}_{mamba} = LayerNorm(\boldsymbol{F}_{imgtab} + \lambda SSM(\boldsymbol{F}_{conv})) \quad (6)$$

第 2 步是进行残差连接。将 Mamba 模块前后输入输出特征 $\boldsymbol{F}_{imgtab}$ 和 $\boldsymbol{F}_{mamba}$ 进行残差连接,增强多模态全局信息的捕获,避免多尺度特征信息被 Mamba 序列化操作破坏和梯度消失,输出 $\boldsymbol{F}_{itm}$:

$$\boldsymbol{F}_{itm} = LayerNorm(\boldsymbol{F}_{imgtab} + Dropout(\boldsymbol{F}_{mamba})) \quad (7)$$

第 3 步是建立前馈网络并进行残差连接。为防止高频细节信息在深层网络衰减丢失、加速深层网络训练,建立前馈网络(feedforward neural network,FFN),包含线性层和 *ReLU* 激活函数。采用先升高维度以获取高阶交互特征,再通过激活函数和降维度层进行非线性变换,达到增强模型的非线性表达能力和增强深层模型的特征表达的目的。最后再一次进行残差连接,将 $\boldsymbol{F}_{itm}$ 特征和前馈网络的输出进行跨层连接,保护特征传递信息流,以免非线性变化导致梯度消失,最终 Mamba 自注意力模块输出为 $\boldsymbol{F}_{attan}$:

$$\boldsymbol{F}_{attan} = LayerNorm(\boldsymbol{F}_{itm} + Dropout(FFN(\boldsymbol{F}_{itm}))) \quad (8)$$

3)基于 CNN 的多层级多模态特征融合

上阶段基于 Mamba 自注意力的多模态融合,主要完成了全局特征的提取和多模态信息的融合,但是 Mamba 结构自身在局部特征提取和模态权重稳定性上存在局限性。针对模型的不足,本研究设计 CNN 对 Mamba 自注意力模块输出结果进行局部特征再提取,增强模型对局部信息的捕获和感知能力,弥补 Mamba 短程依赖建模的不足;设计多层级多模态特征拼接融合,保留各层级特征影响,增强模型的稳定性,整合全局信息。

CNN 作为经典的深度学习架构,在图像处理领域有着广泛的应用,多层卷积层的堆叠能够逐步提取图像的高级语义特征,在本文中,CNN 在 Mamba 自注意力网络之后处理数据,进一步增强了模型对特征的提取能力,与 Mamba 网络提取的全局特征相互补充,提高对 DVT 早期迹象的识别准确率。

首先 CNN 对 Mamba 自注意力机制输出的融合特征 $\boldsymbol{F}_{attan}$ 进行空间细化,由多个 3×3 卷积层和池化层堆叠,输出混合细化特征 $\boldsymbol{F}_{cnn}$。然后通过 cat 函数融合原始结构化数据 $\boldsymbol{X}_{tab}$ 与中间表示(图像特征 $\boldsymbol{F}_{img}$、文本特征 $\boldsymbol{F}_{tab}$、混合细化特征 $\boldsymbol{F}_{cnn}$),进行特征拼接,保留多层次信息,整合全局不同层级特征,保障模型训练的稳定性,最后输出多层级多模态融合特征 $\boldsymbol{F}_{final}$:

$$\boldsymbol{F}_{final} = Flatten(\boldsymbol{F}_{cnn})//\boldsymbol{X}_{tab}//\boldsymbol{F}_{img}//\boldsymbol{F}_{tab} \quad (9)$$

1.4 DVT 预测结果输出

模型的决策输出层由多层感知机(MLP)实现,经过两层隐藏层和一层输出层,将模型维度从高维度进行压缩和逐步降维。隐藏层通过全连接层和 ReLU 激活函数进行非线性变换(428→128→64 维),最后一层直接输出 2 维 logits(64→2 维),如图 6 所示。输出二分类预测结果 y',$\boldsymbol{W}$ 表示权重矩阵,b 表示偏置。

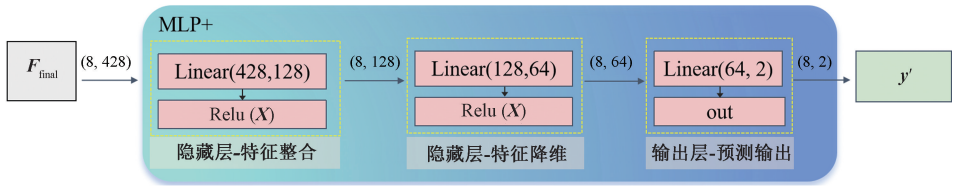


图 6　DVT 预测输出网络

Fig. 6　DVT prediction output

$$y' = \boldsymbol{W}_3 Rule(\boldsymbol{W}_2 Rule(\boldsymbol{W}_1 \boldsymbol{F}_{final} + b_1) + b_2) + b_3 \quad (10)$$

2　实验与结果分析

本文实验全部采用 python 语言编程，基于 kaggle 云平台实现，实验运行环境为 Ubuntu 22.04.4，CPU 为 Intel Xeon，GPU 型号为 T4×2，CUDA 版本为 12.6，Python 版本为 3.11。

2.1　实验数据

DVT 多模态数据缺乏公开的数据源，并且临床 DVT 数据收集难度大。本研究的实验数据来源于真实的临床数据，共收集了 402 例多模态数据样本，其中包括患有 DVT 的确诊病例和无血栓的阴性病例。影像学数据收集了病例的下肢静脉超声图像，示例如图 7 所示，图像分辨率经过统一处理，尺寸调整为适合模型输入的大小。

结构化文本数据则包含年龄、性别、病史（如是否患有心脏病、糖尿病等）、手术史等共 24 个特征，临床检验指标包括 D-二聚体、纤维蛋白原、血小板计数等共 5 项指标[33]。确诊和非确诊数据集的部分风险因素分布情况如表 1 所示，除性别、住院次数、血小板计数、活化部分凝血活酶时间外，其余变量与结局事件之间差异存在统计学意义（$P<0.05$）。在数据收集过程中，严格遵循伦理规范和隐私保护原则，对患者的个人信息进行脱敏处理，确保数据的合法性和安全性。数据集按照 8∶2划分为训练集和测试集，用于模型的训练、参数调整和性能评估。

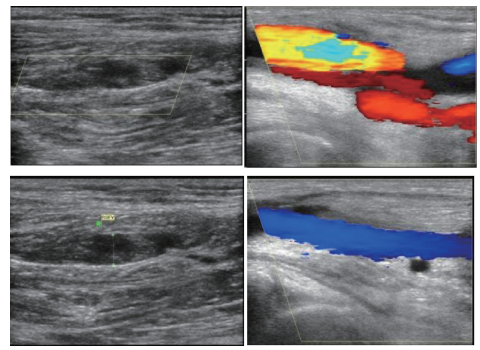


图 7　某样本超声图示例

Fig. 7　An example of ultrasound images

表 1　部分风险因素分布情况

Table 1　Distribution of some risk factors

指标因素	确诊（$n=200$）	非确诊（$n=202$）	P 值
年龄/岁			
18～40	16	38	
40～60	49	76	<0.001
60～80	110	63	
>80	25	10	
纤维蛋白（原）降解产物（$\bar{X}\pm s$，mg/L）	17.84±34.56	3.77±4.48	<0.001
D 二聚体水平（$\bar{X}\pm s$，mg/L）	7.50±14.13	0.98±2.02	<0.001
凝血酶原时间（$\bar{X}\pm s$，s）	11.95±2.96	11.21±1.34	<0.001
活化部分凝血活酶时间（$\bar{X}\pm s$，s）	29.01±19.60	27.89±3.24	0.418
急性感染或风湿免疫性疾病			
1-是	67	26	<0.001
0-否	133	176	
恶性肿瘤类疾病			
1-是	58	18	<0.001
0-否	142	184	
高血压			
1-是	87	65	0.019
0-否	113	137	
下肢肿胀			
1-是	43	15	<0.001
0-否	159	187	

针对收集到原始数据,首先进行数据清洗和数据处理。将收集到的检验数据表现形式进行统一,将性别等二分类属性进行数字标准化标识,将一些缺失值采用平均值补齐,对于必要信息缺失或者有明显异常的数据剔除。

2.2 评估指标

为了全面、客观地评估所提模型的性能,本文采用了多种评估指标,包括准确率(Accuracy)、召回率(Recall)、精确率(Precision)、$F1$ 分数(F1-Scorce)以及 AUC-ROC 曲线。这些指标从不同角度全面地反映了模型在 DVT 风险预测任务中的表现,有助于对模型进行准确的评价和比较。准确率反映模型的整体预测能力;召回率关注模型对 DVT 确诊患者的识别能力,避免漏诊;精确率体现模型预测的可靠性;$F1$ 分数是精确率和召回率的调和平均数,平衡了模型的性能。AUC-ROC 曲线则评估模型对正负样本的区分能力,AUC 值越接近 1,表示模型的性能越好。计算公式为:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{14}$$

其中,TP 表示正确预测的正样本,TN 表示正确预测的负样本,FP 表示错误预测的正样本,FN 错误预测的负样本。

2.3 对比实验

本文研究内容是基于 Mamba 自注意力机制和多模态融合的预测模型,为了验证所提方法的有效性和优越性,将其与主流的方法进行了对比实验,包括仅使用单一影像模态的预测模型、结构化数据的非影像模态模型,以及采用主流深度学习算法融合的多模态预测模型。结构化模态预测模型对比实验中,本文借鉴文献[13]中的 DT、文献[14]中采用的 RF 和 LR、文献[16]中采用的 KNN 以及文献[15]中的 XGBoost 进行模型训练。单一影像模态预测模型对比实验中,参考文献[20]的 CNN、文献[21]的 ResNet、文献[26]的 VGG 进行模型构建。由于多模态 DVT 风险预测方法本文首次提出,目前没有对比文献,本研究组合了当前可选的其他方法进行多模态对比实验,分别是 CNN + DNN, ResNet + DNN, ResNet + DNN + Transformer, ResNet + DNN + Transformerr + CNN, ViT+DNN+Transformer+CNN。

使用分类报告的综合指标评估各类模型,对比实验结果如表 2 所示。经过多次实验结果显示本研究的模型 M_MATT 具有最优的 AUC(0.912),明显高于其他模型的预测效能。

表 2 不同模型对比实验结果

Table 2 Comparison experiment of different models

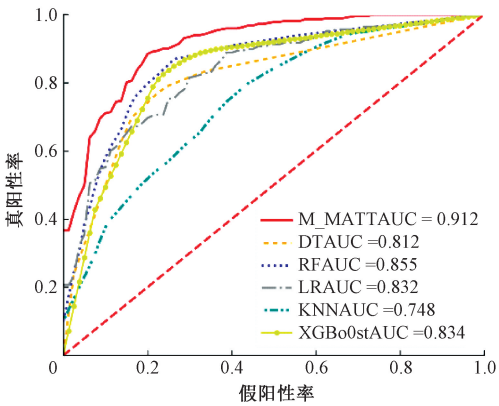
模型分类	模型	AUROC (±<2%)	Accuracy (±<1%)	Precision (±<1%)	Recall (±<1%)	F1 Score (±<1%)
结构化数据模型	DT[13]	0.812	0.808	0.813	0.808	0.806
	RF[14]	0.855	0.810	0.811	0.808	0.808
	LR[14]	0.832	0.746	0.751	0.747	0.743
	KNN[16]	0.748	0.657	0.658	0.656	0.653
	XGBoost[15]	0.834	0.816	0.820	0.812	0.812
图像数据模型	CNN[20]	0.724	0.621	0.676	0.627	0.594
	ResNet[21]	0.849	0.751	0.759	0.752	0.748
	VGG[26]	0.823	0.740	0.742	0.737	0.736
大众数据模型	CNN+DNN	0.702	0.664	0.664	0.660	0.645
	ResNet+DNN	0.798	0.749	0.751	0.749	0.743
	ResNet+DNN+Transformer	0.849	0.792	0.791	0.785	0.786
	ResNet+DNN+Transformer+CNN	0.872	0.810	0.808	0.808	0.807
	ViT+DNN+Transformer+CNN	0.729	0.664	0.681	0.671	0.656
	M_MATT(本文)	**0.912**	**0.861**	**0.863**	**0.859**	**0.858**

本研究模型在准确率(0.861)、精确率(0.863)、召回率(0.859)、$F1$ 分数(0.858)上均显著优于其他对比方法。与仅使用非影像模态的结构化数据模型相比,本研究模型的 AUC 平均提高了 12%,ROC 曲线对比如图 8(a)所示,最低提高了 6.7%,模型的 $F1$ 分数平均提升了 13%,说明影像数据在 DVT 风险预测中具有重要的价值,而本研究

模型有效地利用了这部分信息。与仅使用超声图像的模型相比，所提模型的准确率平均提高了 23%，ROC 曲线也表现最优如图 8(b)，这是因为本模型整合了多种模态的数据，能够更全面地捕捉 DVT 相关的特征信息。根据表 2 和图 8(c)，与实验中其他融合方法相比，所提模型的融合策略在各指标上也有明显优势，与表现较好的 ResNet＋DNN＋Transformer＋CNN 模型相比，M_MATT 准确率、精确率、召回率、F1 分数均提升了 6%左右，AUC 上升了 4.6%；M _ MATT 模型与同结构的 ViT ＋ DNN ＋ Transformer＋CNN 模型相比，AUC 及其他 4 项指标均提高了 20%以上，推理速度提升 20%，从 121 ms 提升至 96 ms，体现了其在特征提取和融合方面的高效性。

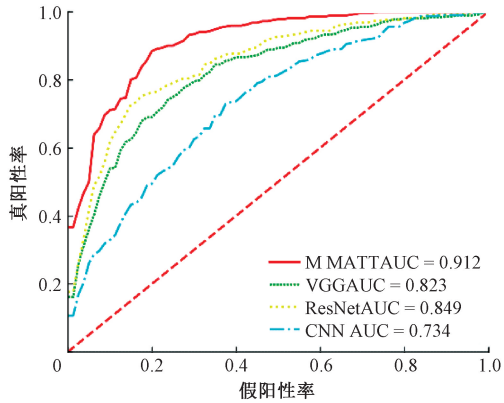
2.4　消融实验

为了进一步分析模型中各个组件对预测性能的贡献，本文进行了消融实验。分别去掉模型中的层级融合模块、CNN 模块，Mamba attention 模块、ViT、DNN 等关键组件。重新训练和测试模型，观察各指标的变化情况，如表 3 所示。



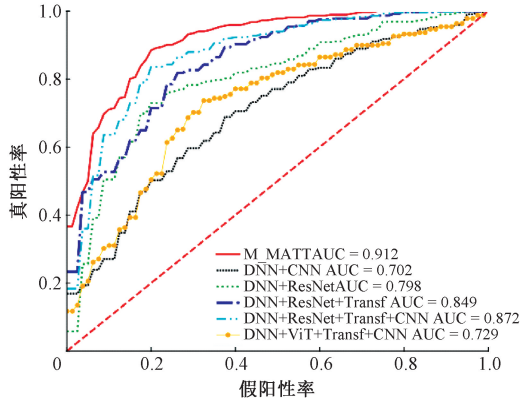
(a) 单一结构化数据模型ROC曲线

(a) Comparison for ROC curves with structured models



(b) 单一图像数据模型ROC曲线

(b) Comparison for ROC curves with pic models



(c) 多模态数据模型ROC曲线

(c) Comparison for ROC curves with multi models

图 8　各模型 ROC 曲线对比

Fig. 8　ROC curves of different models

表 3　消融实现各模型对比

Table 3　Comparison of ablation experiment results

DNN	ViT	MambaAtten	CNN	Multi-level Fusion	Auc (±<1%)	Accuracy (±<2%)	Precision (±<2%)	Recall (±<2%)	F1 Score (±<2%)
√	×	×	×	×	0.780	0.712	0.724	0.722	0.711
√	√	×	×	×	0.801	0.738	0.742	0.738	0.734
√	√	√	×	×	0.854	0.788	0.790	0.790	0.786
√	√	√	√	×	0.839	0.753	0.758	0.757	0.752
√	√	√	×	√	0.872	0.794	0.802	0.794	0.789
√	√	√	√	√	**0.912**	**0.861**	**0.863**	**0.859**	**0.858**

实验结果如表 3 所示，所有指标随模块的移除整体呈阶梯式下降，AUC 从 0.912 下降到 0.780，实验中各模型 AUCROC 曲线如图 9 所示。当模型层级融合(Multi-level Fusion)模块缺失时，AUC 从 0.912 下降至 0.839，核心模块 Mamba 自注意力网络缺失时，AUC 下降至 0.801，分别下降 8%、12%，模型的特征提取识别能力显著下降，多模态特征之间的交互和融合效果变差，评估指标(准确率、召回率等)均下降了 10%以上，说明该两个模块对模型长序列建模和跨模态多层级特征融合具有关键作用。当只去掉 CNN 模块时，模型能保持较好的基础性能(AUC＝0.872)，但对多模态的细节特征提取能力受限，与目标模型相比，AUC 下降 8%，其余 4 项指标均降低 7%左右，说

明局部特征对全局建模具有补充作用;而缺少 ViT + DNN 融合部分时,模型缺少对影像模态数据的处理,引发严重的性能衰减,AUC 骤跌 14.5%,从 0.912 到 0.780,且在所有消融配置中表现最差,说明其对医学影像特征解析具有重要作用,证明视觉特征提取的必要性。

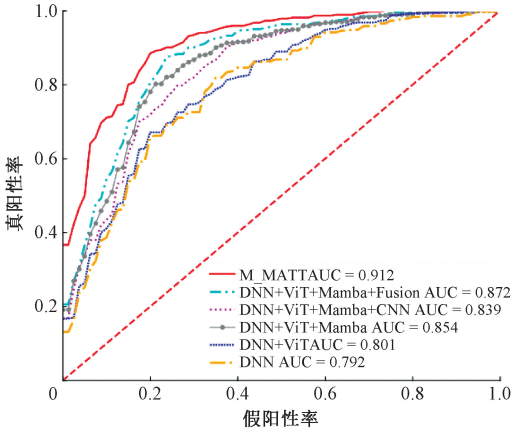


图 9 消融实验 ROC 曲线

Fig. 9 ROC curve of the ablation experiment

这说明模型中的每个组件都发挥着重要作用,相互配合共同提升了模型的预测性能,验证了所提模型架构的合理性和有效性。目标模型具有最高的 AUC 值,说明最终的混合模型具有最优的预测性能。

3 结 论

本研究提出了一种基于 Mamba 多模态融合和自注意力机制的 DVT 风险预测方法,通过整合影像学数据、临床结构化检验指标数据等多种模态信息,利用 Mamba 自注意力网络、ViT、DNN、CNN 等技术构建了一个跨层级融合和多模态融合的风险预测模型。实验结果充分证明了该方法在预测准确率、召回率等关键指标上的优越性,能够为临床医生提供更早期、更可靠的 DVT 风险评估,在辅助制定预防和治疗方案上,具有重要的临床应用价值。同时该模型提出的 Mamba 和多模态融合结合的模型创新也为人工智能在医疗领域的应用提供了新的思路。在未来的研究中,将进一步优化模型架构,探索更多的数据模态(例如基因信息、电生理数据等)的应用,以进一步提高预测分类性能,为 DVT 的预防和治疗提供更有力的技术支持。

参考文献

[1] 陈思宇, 但敏, 姜永生. 肿瘤相关静脉血栓栓塞风险评估模型的研究进展[J]. 华中科技大学学报(医学版), 2024,53(3):409-413.
CHEN S Y, DAN M, JIANG Y SH. Research progressof risk assessment models for cancer-associated venous thromboembolism [J]. Acta Med Univ Sci Technol Huazhong,2024,53(3):409-413.

[2] 中华医学会呼吸病学分会肺栓塞与肺血管病学组,中国医师协会呼吸医师分会肺栓塞与肺血管病工作委员会,全国肺栓塞与肺血管病防治协作组. 肺血栓栓塞症诊治与预防指南[J]. 中华医学杂志, 2018, 98(14):1060-1087.
Group of Pulmonary Embolism and Pulmonary Vascular Disease, Chinese Thoracic Society, Chinese Medical Association; Working Committee of Pulmonary Embolism and Pulmonary Vascular Disease, Chinese Respiratory Physician Association; National Collaborative Group for Prevention and Treatment of Pulmonary Embolismand Pulmonary Vascular Disease. Guidelines for the Diagnosis, Treatment and prevention of pulmonary thromboembolism [J]. Chinese Medical Journal, 2018, 98(14):1060-1087.

[3] BRUNING G, WOITALLA-BRUNING J, QUEISSER A C, et al. Diagnosis and treatment of postthrombotic syndrome[J]. Hamostaseologie,2020, 40(2):214-220.

[4] GHORBANI A, GREATHOUSE J, BAKHSHAEI S, et al. Persistent peril: Recurrent deep vein thrombosis and pulmonary embolism in a patient with protein S deficiency despite optimal anticoagulation therapies [J]. Cureus,2024,16(5):60517.

[5] FENG Y, ZHENG R J, FU Y, et al. Assessing the thrombosis risk of peripherally inserted central catheters incancer patients using Caprini risk assessment model: A prospective cohort study[J]. Supportive Care Cancer,2021,29(9):5047-5055.

[6] WANG P, WANG Y, YUAN ZH Y, et al. Venous thromboembolism risk assessment of surgical patients in Southwest China using real-world data: Establishment and evaluation of an improved venous thromboembolism risk model[J]. BMC Medical Informatics Decision Making, 2022,22(1):59.

[7] DING Y, YAO L J, TAN T, et al. Risk assessment for postoperative venous thromboembolism using the modified Caprini risk assessment model in lung cancer[J]. Journal of Thoracic Disease, 2023,15(6):3386-3396.

[8] 盛昌, 贺爱兰, 万凌燕, 等. 普通外科恶性肿瘤患者术后下肢深静脉血栓形成预测模型的构建[J]. 中国普通外科杂志,2023, 32(6):850-858.
SHENG CH, HE AI L, WAN L Y, et al. Development of a prediction model for postoperative lower extremit deep venous thrombosis in patients with malignant tumors undergoing general surgery[J]. China Journal of General Surgery, 2023, 32(6):

850-858.

[9] 吴倩，许静，王铁皓，等. 住院成年患者静脉血栓栓塞风险评估模型的研究进展[J]. 血管与腔内血管外科杂志，2025,11(1):54-60,66.
WU Q, XU J, WANG T H, et al. Research progress in risk assessment models for venous thromboembolism in hospitalized adult patients has been made[J]. Journal of Vascular and Endovascular Surgery, 2025, 11(1):54-60,66.

[10] RYAN L, MATARASO S, SIEFKAS A, et al. A machine learning approach to predict deep venous thrombosis among hospitalized patients[J]. Clinical and Applied Thrombosis/Hemostasis, 2021, 27, DOI:10.1177/1076029621991185.

[11] 孙玉萍，宋岗，张建美，等. 基于机器学习的 PICC 相关性上肢深静脉血栓形成预测[J]. 循证护理，2021,7(15):2071-2075.
SUN Y P, SONG G, ZHANG J M, et al. Prediction of PICC-related upper extremity deep vein thrombosis based on machine learning[J]. Chinese Evidence-Based Nursing, 2021,7(15):2071-2075.

[12] 张建楠，李荣华，周红梅，等. 下肢骨科手术后深静脉血栓形成风险的预测模型构建与验证[J]. 实用临床医药杂志,2023,27(23):73-78.
ZHANG J N, LI R H, ZHOU H M, et al. Establishment of a predictive model for the risk of deep vein thrombosis after orthopedic surgery in the lower extremities and its verification[J]. Journal of Clinical Medicine in Practice,2023,27(23):73-78.

[13] ABBASI A, KARASU Y, LI C, et al. Machine learning to predict hemorrhage and thrombosis during extracorporeal membrane oxygenation[J]. Critical Care, 2020,24(1):689.

[14] WANG X G, XI H X, GENG X, et al. Artificial intelligence-based prediction of lower extremity deep vein thrombosis risk after Knee/Hip arthroplasty[J]. Clinical and Applied Thrombosis/Hemostasis, 2023, 29: 255476124.

[15] WANG K Y, IKWUEZUNMA I, PUVANESARAJAH V, et al. Using predictive modeling and supervised machine learning to identify patients at risk for venous thromboembolism following posterior lumbar fusion[J]. Global Spine Journal,2023,13(4):1097-1103.

[16] 徐青，余冰，周佩敏，等. 基于机器学习与 SHAP 的全髋关节置换术患者下肢深静脉血栓可解释性预测模型构建研究[J]. 中国医院统计，2024,31(1):11-18,24.
XU Q, YU B, ZHOU P M, et al. Construction of interpretable prediction model for lower limb deep vein thrombosis in patients undergoing total hip arthroplasty based on machine learning and SHAP[J]. Chinese Journal of Hospital Statistics, 2024,31(1):11-18,24.

[17] 王琦，张涛，徐超炜，等. 多尺度注意力融合与视觉 Transformer 方法优化的电阻抗层析成像深度学习方法[J]. 仪器仪表学报,2024,45(7):52-63.
WANG Q, ZHANG T, XU CH Y, et al. Optimized learning method for electrical impedance tomography with multi-scale attention fusion and vision Transformer[J]. Chinese Journal of Scientific Instrument, 2024,45(7):52-63.

[18] 陈广秋，尹文卿，温奇璋，等. 基于双重注意力机制生成对抗网络的偏振图像融合[J]. 电子测量与仪器学报,2024,38(4):140-150.
CHEN G Q, YING W Q, WEN Q ZH, et al. Polarization image fusion based on dual attention mechanism for generating adversarial networks[J]. Journal of Electronic Measurement and Instrumentation, 2024,38(4):140-150.

[19] SIDDIQUI N A, QADRI M T, AKHTE M O, et al. High-precision brain tumor segmentation using a progressive layered U-Net(PLU-Net)with multi-scale data augmentation and attention mechanisms on multimodal magnetic resonance imaging[J]. Instrumentation, 2025, 12(1):77-92.

[20] SEO J W, KIM Y J, KIM K G. Deep vein thrombosis detection based on deep learning for CT images[C]. 2021 International Conference on Information and Communication Technology Convergence (ICTC), 2021:682-684.

[21] 段丽芬，叶裕丰，陈秋梅，等. 深度学习和磁共振黑血血栓成像可用于下肢深静脉血栓分期预测[J]. 分子影像学杂志,2024,47(12):1335-1340.
DUAN L F, YE Y F, CHEN Q M, et al. Deep learning and black-blood magnetic resonance thrombus imaging can be used for predicting the staging of deep vein thrombosis in the lower limbs[J]. Journal of Molecular Imaging,2024,47(12):1335-1340.

[22] NAKAYAMA Y, SATO M, OKAMOTO M, et al. Deep learning-based classification of adequate sonographic images for self-diagnosing deep vein thrombosis[J]. PLOS ONE, 2023,18(3):0282747.

[23] XUE B, LI DW, LU CH Y, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications[J]. JAMA Network

Open, 2021,4(3): 212240.

[24] 成荣,朱文忠,王文. UltraLight CrackNet:基于 VMamba 的轻量化裂缝分割网络[J]. 电子测量技术, 2025,48(22):224-234.
CHENG R, ZHU W ZH, WANG W. UltraLight CrackNet: A VMamba-based lightweight network for crack segmentation[J]. Electronic Measurement Technology, 2025, 48(22):224-234.

[25] 林哲,潘慧琳,陈丹. 融合改进 YOLO 和语义分割的遮挡目标抓取方法[J]. 电子测量与仪器学报, 2024, 38(12):190-201.
LIN ZH, PAN H L, CHEN D. Grasp method for occlusion method by fusing improved YOLO with semantic segmentation[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(12): 190-201.

[26] MITRA A, SRIDAR K, RATHNA S, et al. Optimizing brain tumor mri classification using modified Vgg16 model[C]. 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), 2024: 1-7.

[27] WANG Z X, SHAO Y F, SUN J Y, et al. Vision transformer based multi-class lesion detection in IVOCT [C]. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2023: 327-336.

[28] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces [J]. ArXiv preprint arXiv:2312.00752, 2024.

[29] FANG ZH J, ZHU SH H, CHEN Y F, et al. GFE-Mamba: Mamba-based AD multi-modal progression assessment via generative feature extraction from MCI[J]. ArXiv preprint arXiv:2407.15719, 2025.

[30] 周景,董晖,刘心,等. Bi-EMamba:基于 Mamba 的高压电缆接地电流预测模型[J]. 电子测量技术, 2025, 48(22):66-77.
ZHOU J, DONG H, LIU X, et al. Bi-EMamba: A Mamba-based prediction model for ground currents in high-voltage cable[J]. Electronic Measurement Technology, 2025, 48(22):66-77.

[31] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical vision transformer using shifted windows[J]. 2021 IEEE/CVF International Conference on Computer Vision(ICCV), 2021:10012-10022.

[32] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. ArXiv preprint arXiv: 1706.03762, 2017.

[33] 刘华茜,王海东,聂丽,等. 肺癌患者术后深静脉血栓形成风险预测模型的构建[J]. 陆军军医大学学报, 2024,46(17):1994-2001.
LIU H X, WANG H D, NIE L, et al. Construction of a risk prediction model for postoperative deep vein thrombosis in lung cancer patients [J]. Journal of Army Medical University, 2024,46(17):1994-2001.

作者简介

肖连禹,软件设计师,硕士研究生,主要研究方向为医疗智能、医院信息化。

E-mail:932000660@qq.com

陆向艳(通信作者),副教授,硕士,主要研究方向为人工智能、大数据分析。

E-mail:xiangyanlu@126.com