

基于多模态交叉注意力机制融合的 1型糖尿病血糖浓度预测方法

王梦星¹, 薛付忠^{1,2,3}, 杨帆^{1,2,3}

(1. 山东大学齐鲁医学院公共卫生学院医学数据学系, 山东 济南 250012;

2. 国家健康医疗大数据研究院, 山东 济南 250003; 3. 山东大学齐鲁医院, 山东 济南 250012)

摘要: **目的** 通过融合扫描式葡萄糖监测(flash glucose monitoring, FGM)数据与结构化电子健康记录(electronic health records, EHR)中的多模态信息,构建1型糖尿病(type 1 diabetes mellitus, T1DM)患者血糖浓度预测模型,以解决传统单一模态模型在捕捉复杂血糖波动规律时的局限性,为临床制定个性化控糖策略及早期风险预警提供数据支持。**方法** 基于T1DiabetesGranada数据集,整合FGM数据、生化检测指标、人口统计学信息及诊断编码等模态特征,构建多模态时序预测模型XCLA-Net。该模型采用一维卷积神经网络提取血糖序列的局部动态特征,结合长短时记忆网络建模长期时间依赖关系,通过交叉注意力机制实现多模态语义对齐,并引入自归一化神经网络增强融合特征的稳定性。**结果** XCLA-Net在RMSE、MAE、MAPE等评估指标上显著优于多种对比模型。在未来1h和3h预测任务中,MAPE分别为19.64%和37.81%,表明模型在不同时间尺度下具备较强的预测准确性;克拉克误差网格分析表明,模型预测结果高度集中于A区,具备良好的临床一致性。消融实验验证了交叉注意力机制、一维卷积神经网络与长短时记忆网络对提升模型预测性能的关键作用。**结论** 本研究提出的XCLA-Net通过多模态数据融合与时序建模,显著提升了血糖预测的精度与稳定性,具备良好的临床解释性和实用价值,为T1DM患者的个性化血糖管理及早期风险预警提供有效支持。

关键词: 1型糖尿病;多模态数据融合;交叉注意力机制;血糖预测;时序建模

中图分类号: R587.1 **文献标志码:** A

Blood glucose concentration prediction method for type 1 diabetes mellitus based on multi-modal cross-attention mechanism fusion

WANG Mengxing¹, XUE Fuzhong^{1,2,3}, YANG Fan^{1,2,3}

(1. Department of Medical Dataology, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250012, Shandong, China;

2. National Institute of Health and Medical Big Data, Jinan 250003, Shandong, China;

3. Qilu Hospital of Shandong University, Jinan 250012, Shandong, China)

Abstract: Objective To develop a blood glucose concentration prediction model for patients with type 1 diabetes mellitus (T1DM) by integrating multimodal information from flash glucose monitoring (FGM) data and structured electronic health records (EHR), so as to address the limitations of traditional unimodal models in capturing complex glucose fluctuation patterns and provide data support for personalized glycemic control strategies and early risk warning in clinical practice. **Methods** Based on the T1DiabetesGranada dataset, the study integrated multimodal features including FGM data, biochemical test indicators, demographic information, and diagnostic codes to construct a multimodal temporal prediction model, XCLA-Net. The model employed a one-dimensional convolutional neural network (1D-CNN) to extract short-term dynamic features of the glucose sequence, combined a long short-term memory (LSTM) network to capture long-term temporal dependencies, incorporated a cross-attention mechanism for multimodal

semantic alignment, and introduced a self-normalizing neural network (SNN) to enhance the stability of the fused representations. **Results** XCLA-Net significantly outperformed multiple baseline models in terms of error metrics such as root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The MAPE values for 1-hour and 3-hour prediction tasks were 19.64% and 37.81%, respectively, indicating strong predictive accuracy across different temporal scales. Clarke error grid analysis showed that the majority of prediction points fell within Zone A, reflecting good clinical consistency. Ablation experiments confirm the critical roles of the cross-attention mechanism, 1D-CNN, and LSTM in enhancing model performance. **Conclusion** The proposed XCLA-Net model effectively improves the accuracy and stability of blood glucose prediction through multimodal data fusion and temporal modeling. It demonstrates favorable clinical interpretability and practical value, providing reliable support for personalized glycemic management and early risk prediction in patients with T1DM.

Key words: Type 1 diabetes mellitus; Multimodal data fusion; Cross-attention mechanism; Blood glucose prediction; Time series modeling

1型糖尿病(type 1 diabetes mellitus, T1DM)是一种由自身免疫机制导致胰岛细胞破坏、胰岛素绝对缺乏的慢性代谢性疾病,需终生依赖外源性胰岛素治疗^[1]。根据国际糖尿病联盟(International Diabetes Federation, IDF)统计,截至2024年,全球T1DM患者总数已超过920万,预计2050年将升至1350万,增长约47%^[2]。

T1DM患者血糖管理的关键是将血糖水平长期维持在安全范围内,防止出现急性代谢紊乱事件,尤其是低血糖事件。作为T1DM患者最常见且最危险的急性代谢疾病之一,低血糖可在短时间内引起认知障碍、昏迷,严重时可导致猝死^[3]。但血糖波动的产生机制复杂,受治疗依从性、胰岛素吸收延迟、饮食和运动行为等多因素共同作用影响,呈现出显著非线性、时变性与个体异质性^[4]。因此,构建能够及时识别血糖异常趋势的个体化预测模型,实现对T1DM患者血糖异常事件的早期预警与精准干预,是提升其长期管理效率与生存质量的关键路径。

传统血糖预测方法多采用单一模态的时间序列建模技术,如自回归、支持向量回归和长短时记忆(long short-term memory, LSTM)网络等^[5]。但在真实T1DM个体中,血糖受多因素驱动,单一模态建模难以准确捕捉其波动规律^[6]。为提高模型对个体血糖调控机制的预测能力,近期研究开始尝试引入多模态输入。例如,PBI-CLA模型通过引入注意力机制对血糖与胰岛素序列进行权重建模,显著降低血糖浓度预测误差^[7];VMD-PSO-LSTM模型利用变分模态分解与粒子群优化算法提取多尺度特征以提升时序建模预测精度^[8];Wang等^[9]提出的双注意力耦合网络通过融合FGM时序特征与静态生理特征结合双层注意力机

制,实现多模态数据融合与糖尿病分类性能的显著提升。虽然现有研究在方法与性能上取得了一定进展,但现有血糖波动预测研究仍面临3个方面挑战:①多模态数据在结构、时间尺度和语义层次上存在显著异构性,缺乏统一对齐与协同机制,影响多模态数据融合效率与建模精度;②现有模型对短期剧烈波动与长期代谢风险因素缺乏统一建模路径,导致对低血糖等急性并发症的提前识别能力有限;③当前血糖预测模型在性能评估上多依赖数值型指标,临床可解释性有限,难以为医疗实践提供有效数据支撑与决策辅助。

针对上述问题,本文提出一种融合FGM与结构化EHR两类模态信息的血糖浓度预测模型—XCLA-Net。模型通过整合不同模态数据并联合时序特征进行建模,实现多模态语义对齐和融合,旨在提升T1DM患者未来血糖水平的预测精度,为临床精准控糖策略的制定及突发血糖异常的及时干预提供支持。

1 资料与方法

1.1 资料

1.1.1 数据来源

本研究所使用的数据来自公开获取的T1DiabetesGranada数据集。该数据集由西班牙格拉纳达大学信息与通信技术研究中心联合San Cecilio大学附属医院临床内分泌与营养科共同构建,并已通过当地伦理审查委员会审批(伦理编号:K134665CRL)。数据采集时间为2018年1月至2022年3月,共纳入736例确诊为T1DM的成年患者,累计观测时长达257780人日。该数据集是目前已公开的持续时间最长、样本量最大的T1DM纵向多模态数据集

之一^[10]。

T1DiabetesGranada数据集整合了患者的FGM数据、结构化EHR数据,具备高度时序化和个体化的特征信息。FGM数据提供分钟级的动态血糖浓度测量序列,总记录数超过2 260万条;结构化EHR数据包括静态的人口统计学变量及动态生化检测指标,同时包括每位患者在就诊期间获得的临床诊断信息ICD-9-CM。该数据集纳入患者信息见表1。

表1 T1DiabetesGranada纳入T1DM患者统计数据
Table 1 T1DiabetesGranada included statistical information of T1DM patients

数据模态	变量名称	变量类型
FGM	患者ID	String
	测量日期	Date
	测量时间	Time
	血糖浓度测量值	Integer
EHR	患者ID	String
	性别	String
	出生日期	Date
	生化检测日期	Date
	钾、氯、钠、高密度脂蛋白胆固醇、葡萄糖、甘油三酯、总胆固醇、肌酐(尿液)	String
	肌酐、尿酸、白蛋白	String
	γ -谷氨酰转氨酶、丙氨酸转氨酶	String
	促甲状腺激素、胰岛素、糖化血红蛋白	String
诊断编码	String	

1.1.2 数据预处理

本研究基于T1DiabetesGranada数据集,分别对FGM数据、生化检测指标、人口统计信息及诊断编码数据进行系统性清洗、时间对齐与特征构建,确保输入数据质量满足下游建模要求,实现对每个时间窗口样本未来1、2、3 h血糖浓度的连续预测建模。

1.1.2.1 样本筛选

为确保输入数据的观测密度与临床背景完整性,本研究设定如下入组条件:FGM有效观测天数 ≥ 14 d(即血糖记录数 $\geq 1\ 344$ 条),以确保其具备稳定的连续时序信息;ICD-9-CM诊断编码记录 ≥ 3 条,以保障模型输入的疾病背景信息完整性。根据上述入组条件,从原始736例患者中剔除不满足条件的226例患者,纳入510例患者参与模型训练样本构建。

1.1.2.2 FGM数据预处理

FGM数据来源于FreeStyle Libre设备,为每15 min采样一次的高频等间隔时间序列,该数据存在记录时间不均、观测点缺失的情况。本研究首先对血糖浓度值进行异常值检测,剔除血糖 < 40 mg/dL

或 > 500 mg/dL的极端测量值。以15 min为步长构建标准化的全局时间轴,并以此作为锚点进行时间对齐,完成不同模态数据的时序同步与缺失片段填补操作。本研究针对FGM数据的缺失情况设置预处理策略。对于缺失时间 ≤ 60 min的片段,采用线性插值^[11]进行填补,通过连接前后两个观测点构建线性变化趋势进行缺失值计算;若连续缺失时间 > 60 min,即超过4个连续时间点,认为是信息严重丢失,难以通过插值还原其真实波动趋势,故予以剔除。该策略主要基于以下三点考虑:①临床实践中,连续1 h以上无血糖观测常因设备脱落或患者脱离监测,具有非随机缺失特征;②线性插值在小时级缺失情况下已无法有效捕捉血糖的非线性剧烈波动,容易引入伪趋势信息;③模型训练更依赖连续、完整的输入序列以捕捉血糖波动的短期与中长期趋势,过多填补将显著增加模型偏差。因此,将缺失时间 > 60 min缺失段进行剔除。完成时序补全后,以6 h为窗口宽度,15 min为滑动步长构建样本滑动时间窗口;每个窗口包含一个完整的血糖历史序列,并以该窗口时间节点为基准,设置未来1、2、3 h位置处血糖浓度为预测标签,最终处理后的FGM数据包含43 230个样本窗口,每个样本包含24个血糖时间点输入、3个未来血糖浓度输出标签。

1.1.2.3 EHR数据预处理

在生化检测数据预处理中,各项生化检测指标在时间上为中低频分布,不具备固定采样频率。为实现EHR与FGM两种模态数据在时间维度上的对齐,以每个滑动窗口的中心时间点作为锚定参照,检索该时间点前最近一次生化检测记录,并将该次检测值作为当前窗口的生化特征输入;对于缺失指标,优先使用患者历史检测均值填补,若个体无检测历史,则以总体均值代替。所有生化变量在特征输入前统一采用Z-score标准化处理,以消除量纲影响,提升特征间的比较稳定性。最终处理后的生化检测数据包含16项生化特征的43 230个样本数据。

对于人口统计学信息,本研究提取患者性别与出生年份两项变量,其中性别变量为离散类别,通过独热编码^[12]转换为二元数值形式;年龄由窗口时间减去出生年份计算获得,随后使用最小-最大归一化^[13]。由于该类特征在时间上为静态变量,因此采用复制方式引入至所有时间窗口样本,确保静态特征在动态建模中的表达一致性。最终处理后的人口统计学信息为包含2项结构化变量的43 230个样

本数据。

对于临床诊断信息,依据 ICD-9-CM 标准对原始诊断记录进行编码,统计全体患者中出现频次 ≥ 10 的诊断项目,筛选保留 35 类高频诊断标签,构建 0/1 多标签表示结构。首先对多标签诊断向量构建嵌入层,将其映射为低维稠密语义空间,通过 MLP 对嵌入向量进行非线性特征转换。最终处理后的诊断编码数据包含 16 个诊断特征的 43 230 个样本数据。

所有模态数据统一按照 FGM 滑动窗口的时间节点进行对齐,确保每个样本在同一时间段内能获得对应血糖、生化检测指标、人口信息和诊断信息。最终共构建 43 230 条样本,每个样本包含 4 类输入特征:24 个连续血糖浓度值、16 项生化检测结果、性别与年龄两个变量以及 16 项诊断信息表示,共计 58 个特征值。对应预测标签为该窗口结束后 1、2 和 3 h 的血糖浓度值。每一类模态特征将分别输入独立特征提取子网络,保留不同模态数据原始结构特点,为多模态融合建模提供基础。

1.2 方法

模型架构见图 1。在特征提取部分,针对高频 FGM 血糖序列引入一维卷积神经网络(one-dimensional convolutional neural network, 1D-CNN)与 LSTM 联合结构,分别提取短期波动与长期依赖特征;对于结构化 EHR 数据,考虑其内部模态的异构性。对具有时序特征的生化检测数据,采用双层 1D-CNN 提取其局部动态变化与指标间依赖关系;对静态人口统计信息与多标签诊断编码数据,分别构建嵌入层^[14]与多层感知机(multi-layer perceptron, MLP)^[15]模块,提取其低维语义表示,实现不同类型 EHR 数据的特征提取。在模态融合部分,通过交叉注意力机制^[16]实现 FGM 主导的多模态语义对齐,引入自归一化神经网络(self-normalizing neural network, SNN)^[17]结构增强融合表示的稳定性。在完成多模态特征对齐与融合后,本研究进一步利用 1D-CNN 与 LSTM 的协同机制,对融合序列的时间动态变化趋势进行建模,以提升对未来血糖浓度的预测能力。

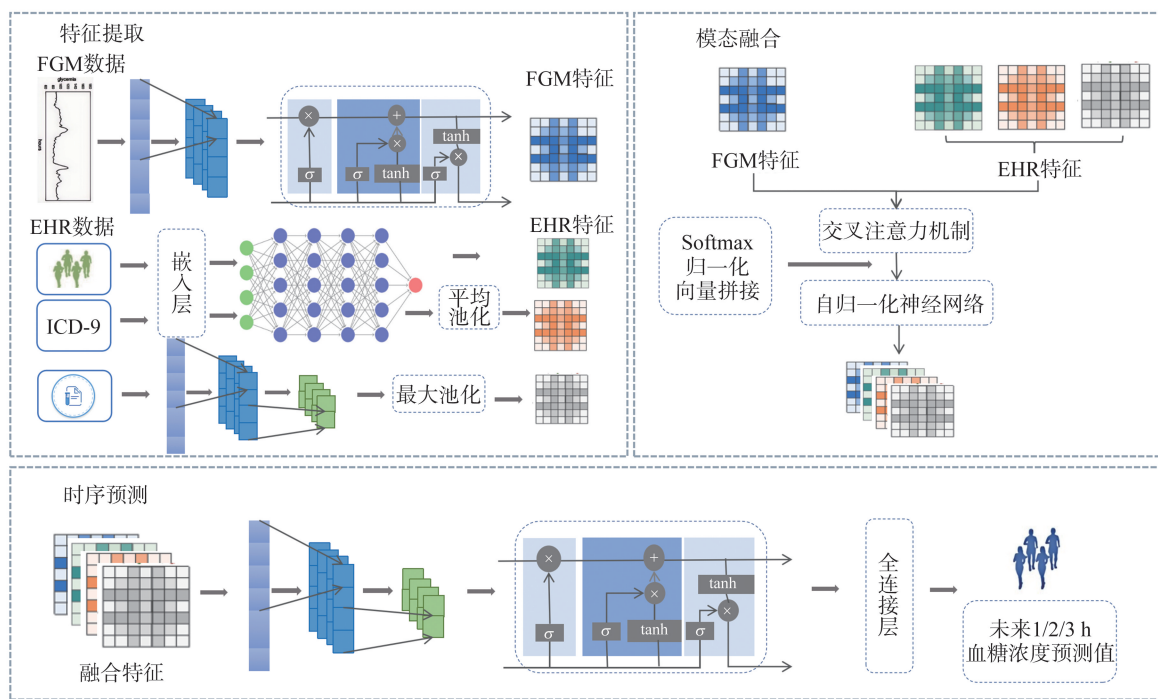


图1 XCLA-Net 模型框架

Figure 1 Framework of XCLA-Net model

1.2.1 特征提取

1.2.1.1 FGM 数据特征提取

FGM 数据为固定 15 min 采样的单变量时间序列。由于 FGM 具备高频时序结构,既包含短时突变,也包含跨时间窗口的长期代谢波动趋势,因此本研究采用 1D-CNN 与 LSTM 联合建模。其中,1D-CNN 是

专门用于处理一维序列数据的卷积神经网络结构,用于提取局部动态特征;LSTM 则用于整合前序时间片段的上下文依赖信息,形成连续时间上的隐状态表示,捕获长期时间依赖进行建模。最终选取时间窗口末时刻的 LSTM 层隐状态作为 FGM 模态序列编码表示;

$$\mathbf{H}_i^{\text{fgm}} = \text{ReLU}(\mathbf{W}_{\text{fgm}} * \mathbf{X}_i^{\text{fgm}} + \mathbf{b}_{\text{fgm}}) \in \mathbf{R}^{T \times d}, \quad (1)$$

$$\mathbf{h}_{i,t}^{\text{fgm}} = \text{LSTM}(\mathbf{H}_i^{\text{fgm}}), \quad t = 1, \dots, T, \quad (2)$$

$$\mathbf{z}_i^{\text{fgm}} = \mathbf{h}_{i,T}^{\text{fgm}} \in \mathbf{R}^{d_1}, \quad (3)$$

其中, $\mathbf{X}_i^{\text{fgm}}$ 表示第 i 个样本在 T 个时间步的原始 FGM 血糖序列输入; $\mathbf{W}_{\text{fgm}}, \mathbf{b}_{\text{fgm}}$ 表示一维卷积核权重与偏置项; 引入 $\text{ReLU}(\cdot)$ 作为激活函数增强模型表达能力; $\mathbf{H}_i^{\text{fgm}}$ 表示第 i 个样本在每个时间步下经过卷积变换后的特征表示; $\mathbf{h}_{i,t}^{\text{fgm}}$ 表示该样本在时间步 t 的隐藏状态, 包含当前时间点之前全部历史信息; $\mathbf{z}_i^{\text{fgm}}$ 表示该样本在当前窗口下整段 FGM 数据的特征表示。

1.2.1.2 EHR 数据特征提取

其中生化检测指标信息来自 EHR 数据中同步提取的 16 项生化检测指标, 其采样频率不固定, 每个窗口对齐取近 7 次记录作为时序输入。考虑到生化指标间存在明显的协同作用, 并具备长期变化趋势性, 因此设计双层 1D-CNN 网络建模捕获其指标间依赖及局部时序变化。提取过程如下:

$$\mathbf{H}_i^{\text{bio}} = \text{MaxPool} \{ \text{ReLU}[\mathbf{W}_2^{\text{bio}} \cdot \text{ReLU}(\mathbf{W}_1^{\text{bio}} \cdot \mathbf{X}_i^{\text{bio}} + \mathbf{b}_1) + \mathbf{b}_2] \}, \quad (4)$$

$$\mathbf{z}_i^{\text{bio}} = \mathbf{H}_i^{\text{bio}} \in \mathbf{R}^{d_2}, \quad (5)$$

其中, $\mathbf{X}_i^{\text{bio}} \in \mathbf{R}^{7 \times 16}$ 表示第 i 个样本在当前窗口内采集到生化检测指标的历史记录; $\text{MaxPool}(\cdot)$ 为最大池化操作, 用于压缩特征维度提升模型计算效率; $\mathbf{z}_i^{\text{bio}}$ 为第 i 个样本对应的生化检测数据最终嵌入表示。

人口统计信息数据包括性别与年龄两类数据, 性别采用 One-Hot 编码, 年龄经最小-最大归一化处理, 与性别编码拼接输入至两层 MLP 结构进行特征提取, 过程如下:

$$\mathbf{h}_i^{\text{demo}} = \text{ReLU}(\mathbf{W}_1^{\text{demo}} \cdot \mathbf{x}_i^{\text{demo}} + \mathbf{b}_1), \quad (6)$$

$$\mathbf{z}_i^{\text{demo}} = \text{ReLU}(\mathbf{W}_2^{\text{demo}} \cdot \mathbf{h}_i^{\text{demo}} + \mathbf{b}_2) \in \mathbf{R}^{d_3}, \quad (7)$$

其中, $\mathbf{x}_i^{\text{demo}} \in \mathbf{R}^{d_3}$ 表示第 i 个样本的人口统计学输入特征向量; $\mathbf{W}_1^{\text{demo}}, \mathbf{W}_2^{\text{demo}}$ 分别表示第 1 层与第 2 层 MLP 模块的权重矩阵, 对特征进行线性变换; $\mathbf{h}_i^{\text{demo}}$ 为第 1 层 MLP 模块输出的中间隐藏表示, 保留人口统计特征的低阶线性组合; $\mathbf{z}_i^{\text{demo}}$ 为最终人口统计学特征语义表示。

临床诊断编码数据先通过嵌入层将诊断标签转为稠密语义表示, 再用平均池化压缩维度, 输入 MLP 层形成向量化表达:

$$\mathbf{E}_i^{\text{diag}} = \text{Embedding}(\mathbf{x}_i^{\text{diag}}) \in \mathbf{R}^{35 \times d_e}, \quad (8)$$

$$\mathbf{h}_i^{\text{diag}} = \text{MeanPool}(\mathbf{E}_i^{\text{diag}}), \quad (9)$$

$$\mathbf{z}_i^{\text{diag}} = \text{ReLU}(\mathbf{W}^{\text{diag}} \cdot \mathbf{h}_i^{\text{diag}} + \mathbf{b}) \in \mathbf{R}^{d_4}, \quad (10)$$

其中, $\mathbf{x}_i^{\text{diag}} \in \{0, 1\}^{35}$ 表示第 i 个样本的诊断标签向量, 由 ICD-9 编码频次筛选得到的 35 个诊断标签构成; $\text{Embedding}(\cdot)$ 为词嵌入层, 将特征表示映射为连续的低维语义向量; $\mathbf{E}_i^{\text{diag}}$ 表示第 i 个样本所有诊断标签对应的嵌入向量集合; 引入 $\text{MeanPool}(\cdot)$ 平均池化层聚合所有嵌入向量的语义信息, 形成统一诊断特征表示; $\mathbf{W}^{\text{diag}}, \mathbf{b}$ 表示用于最终语义压缩的权重矩阵与偏置; $\mathbf{z}_i^{\text{diag}}$ 为诊断特征的最终语义表示向量。

1.2.2 模态融合

注意力机制通过对输入特征赋予动态权重, 使模型能够自动关注对当前任务最为重要的信息区域。标准注意力机制引入查询(Query)、键(Key)和值(Value)三组向量, 通过查询和键计算相关性得分, 经归一化后加权聚合值向量, 实现信息筛选与特征强化^[18]。基于标准注意力机制, 本研究引入交叉注意力机制, 实现不同来源模态间的动态语义对齐与深度融合。交叉注意力采用主模态特征作为 Query, 辅助模态特征作为 Key 和 Value, 学习二者间的语义关联, 从而在时间步上根据血糖动态状态灵活提取 EHR 辅助信息。

本研究以 FGM 数据为主模态, 引导 EHR 数据作为辅模态, 基于交叉注意力机制与 SNN 协同作用, 在统一时间尺度下实现跨模态特征的动态融合, 用于提升模态间语义一致性与模型鲁棒性。具体流程为, 对于第 i 个样本在时间步 t 时, 设主模态 FGM 的特征表示为 $\mathbf{z}_{i,t}^{\text{fgm}}$, 辅模态 EHR 的任一子模态特征为 $\mathbf{z}_{i,t}^{(m)}$, 其加权融合表示定义为:

$$\mathbf{a}_{i,t}^{(m)} = \alpha_{i,t}^{(m)} \cdot \mathbf{z}_{i,t}^{(m)} \in \mathbf{R}^{d_m}, \quad (11)$$

其中, 注意力权重 $\alpha_{i,t}^{(m)}$ 表示第 i 个样本在第 t 个时间步中, 主模态对该辅模态特征的关注强度, 计算方式如下:

$$\alpha_{i,t}^{(m)} = \text{Softmax}[\mathbf{w}_a^T \cdot \tanh(\mathbf{W}_q \mathbf{z}_{i,t}^{\text{fgm}} + \mathbf{W}_k \mathbf{z}_{i,t}^{(m)})], \quad (12)$$

其中, \mathbf{W}_q 与 \mathbf{W}_k 分别为主模态与辅模态特征的线性映射权重矩阵, \mathbf{w}_a 为注意力查询向量。该机制实现了以主模态为引导的动态注意力建模, 捕捉各模态之间的语义关联性。基于上述注意力加权特征, 将主模态与所有辅模态的对齐表示在特征维度上拼接, 得到当前时间步的融合特征向量:

$$\mathbf{z}_{i,t}^{\text{fused}} = \text{Concat}(\mathbf{z}_{i,t}^{\text{fgm}}, \mathbf{a}_{i,t}^{\text{bio}}, \mathbf{a}_{i,t}^{\text{demo}}, \mathbf{a}_{i,t}^{\text{diag}}) \in \mathbf{R}^{d_{\text{fusion}}}, \quad (13)$$

其中, $\text{Concat}(\cdot)$ 表示将不同模态数据在特征维度上进行拼接。将上述操作在所有时间步重复执行, 最终形成完整的融合特征序列表示为:

$$\mathbf{f}_i^{\text{fusion}} = [\mathbf{z}_{i,1}^{\text{fused}}, \mathbf{z}_{i,2}^{\text{fused}}, \dots, \mathbf{z}_{i,24}^{\text{fused}}] \in \mathbf{R}^{24 \times d_{\text{fusion}}}, \quad (14)$$

其中, f_i^{fusion} 表示第 i 个样本的最终融合特征序列, 包含 24 个时间点, 用于后续时序建模输入。

考虑到多模态数据融合会引起分布偏移与噪声累积, 为增强融合特征表示的稳定性与训练收敛性, 本研究在交叉注意力机制之后引入 SNN 层, 通过 SeLU 激活函数维持特征分布的零均值与单位方差, 从而提升模型在面对小样本波动与模态间噪声扰动下的鲁棒性。最终得到的特征融合表示 f_i^{fusion} 作为后续时序预测的输入, 实现对未来 1、2 和 3 h 血糖浓度值的连续预测建模。

1.2.3 时序预测

将第 i 个样本的模态融合表示为 f_i^{fusion} , 作为输入, 并加入 BatchNorm^[19] 与 Dropout^[20] 策略用于正则化, 缓解过拟合风险。将 1D-CNN 编码后的时间序列输入 LSTM, 以捕捉长期依赖关系并生成各时刻的隐状态; 随后通过全连接层映射, 输出未来特定时间窗口内的连续血糖浓度预测值。具体实现过程如下:

$$H_i = \text{ReLU}(W_{\text{cnn}} \cdot z_i^{\text{fusion}} + b_{\text{cnn}}) \in \mathbf{R}^{T \times d_h}, \quad (15)$$

$$h_{i,t} = \text{LSTM}(H_i), \quad t = 1, \dots, T, \quad (16)$$

$$z_i^{\text{final}} = h_{i,T} \in \mathbf{R}^{d_{\text{final}}}, \quad (17)$$

$$\hat{y}_i = W_p \cdot z_i^{\text{final}} + b_p, \quad (18)$$

其中, H_i 表示 1D-CNN 层的输出结果, 作为后续 LSTM 输入的嵌套时间特征表示; $h_{i,t}$ 表示第 i 个样本在时间步 t 上的 LSTM 隐状态, 编码了历史时间依赖与跨步趋势; $h_{i,T}$ 为 LSTM 最后一个时间窗口的隐状态, 表示整条序列在全局时间的编码向量; z_i^{final} 用于预测任务的最终时间依赖表示; \hat{y}_i 为样本 i 在未来 1、2、3 h 的血糖预测值。

模型训练过程中, 采用最小化均方误差作为损失函数, 以优化预测值与真实血糖浓度之间的数值偏差。优化器选用 Adam 优化器^[21]。初始学习率为 0.001, 批大小设置为 64, 训练轮次为 100。实验中采用 5 折交叉验证方式评估模型性能, 确保同一患者数据不会同时出现在训练与验证/测试集中, 从而避免信息泄露。具体划分方式为: 首先按患者层面随机划分为 5 个子集, 在每一折中选择其中 1 折作为测试集, 其余 4 折中划分 20% 作为验证集, 其他用于训练。

1.2.4 评估指标

本研究以 T1DM 患者血糖浓度预测作为研究目标, 其本质是一项多变量回归任务。由于血糖数据的动态性、非线性特征显著, 且实际观测中易受设备噪声、个体代谢异质性等因素干扰, 导致数据样本存在大量异常值。为确保模型在复杂场景下的预测

精度与临床适用性, 本研究选择均方根误差 (root mean square error, RMSE)、平均绝对误差 (mean absolute error, MAE) 以及平均绝对百分比误差 (mean absolute percentage error, MAPE) 作为评估指标^[22]。同时结合克拉克误差网格分析法 (error grid analysis, EGA)^[23], 其从临床可接受性角度划分预测结果的误差区域, 专门用于评估模型对血糖浓度预测的准确性, 因此加入 EGA 分析结果用于验证模型预测结果的临床一致性。

RMSE 通过平方运算放大预测误差的权重, 有效反映模型对异常样本的敏感性, 计算公式如下:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}. \quad (19)$$

MAE 通过计算预测值与真实值的绝对偏差均值, 对异常值具有较强鲁棒性, 计算公式如下:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|. \quad (20)$$

MAPE 通过相对误差百分比衡量预测偏差, 适用于不同量纲数据的横向对比, 计算公式如下:

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right|, \quad (21)$$

其中, \hat{y} 为模型预测血糖浓度值, y_i 为真实血糖浓度值, N 表示比较次数。上述指标值越小, 表明模型预测精度越高, 实际误差越小。

如图 2 所示, EGA 通过将预测血糖浓度值与真实值分别映射至纵轴与横轴, 并据此划分为五个区域 (A~E)。其中, 区域 A (偏差 $\leq 20\%$) 为临床精准预测, 区域 B (偏差 $> 20\%$ 但未导致误判) 仍属可接受范围, 区域 C 可能引发治疗风险, 区域 D/E 表示因严重偏差失去临床意义。

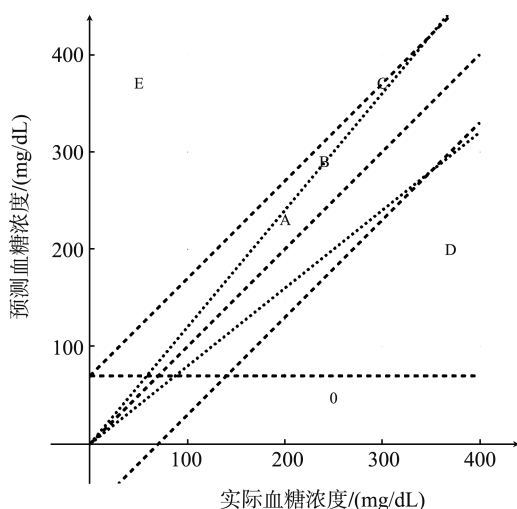


图 2 克拉克误差网格分析图
Figure 2 Clarke EGA

1.2.5 对比实验

为验证本研究提出的多模态融合模型 XCLA-Net 在 T1DM 患者血糖浓度预测任务中的性能表现,本研究设计并进行对比实验。对比模型包括 GRU^[24]、MLP 和 Autoencoder^[25] 三种常用预测模型,统一在 T1DiabetesGranada 数据集上进行建模训练与测试,并评估其在不同预测时间点(1/2/3 h)下的预测精度与误差表现。评价指标选用 RMSE、MAE 与 MAPE,以综合衡量模型在数值准确性与相对偏差控制方面的表现。

2 结果

2.1 对比实验

对比实验结果显示,XCLA-Net 在全部预测时间点上均取得了最优性能表现。在 1 h 预测中,XCLA-Net 的 RMSE、MAE 和 MAPE 分别为 37.21、21.37 和 19.64%,远低于其他对比模型,表示模型预测精度和相对误差控制能力显著提升。随着预测时间延长至 2、3 h,XCLA-Net 仍保持稳定优势,MAPE 分别为 25.93%与 37.81%,优于其他对比模型,表明该模型具备良好的趋势捕捉能力与长时序预测能力。此外,传统模型的误差随预测时间增长而迅速扩大,呈现出明显时间衰减特征,XCLA-Net 通过融合多模态数据特征信息,模型性能在长期预测场景下表现更好,能够有效挖掘模态间的跨模态关联,构建对血糖动态趋势具有高度敏感性的预测模型。见表 2。

表 2 对比实验结果
Table 2 Results of comparison experiments

模型	评价指标	预测时间/h		
		1	2	3
GRU	RMSE	47.68	52.72	65.93
	MAE	38.25	47.82	55.07
	MAPE/%	48.35	31.09	54.31
MLP	RMSE	49.68	54.91	68.71
	MAE	35.21	44.11	52.51
	MAPE/%	34.83	38.97	46.58
Autoencoder	RMSE	53.61	60.91	73.24
	MAE	43.21	58.21	61.74
	MAPE/%	53.82	41.03	57.74
XCLA-Net	RMSE	37.21	45.31	51.38
	MAE	21.37	26.84	46.73
	MAPE/%	19.64	25.93	37.81

2.2 消融实验

消融实验结果表明,逐步去除模型中的关键模块,模型在各项指标上均出现不同程度的性能下降,

说明三个模块均对血糖浓度的预测任务发挥了关键作用。其中,去除交叉注意力机制模块对性能影响最为显著:在 1 h 预测下, RMSE 从 38.79 提升至 54.86, MAPE 从 25.37% 上升至 35.68%; 在 3 h 预测下, RMSE 由 49.76 升高至 69.72, MAPE 从 38.09% 增至 52.24%, 表明缺失交叉注意力机制后模型对多模态语义的对齐能力明显下降, 难以有效挖掘模态间的协同信息。在去除 LSTM 模块后, 模型在各时间点的 RMSE、MAE 和 MAPE 均高于完整模型, 在 3 h 预测任务中, RMSE 升至 74.66, MAPE 达到 45.31%, 说明该模块在建模血糖变化的长期依赖性方面具有重要作用, 可有效捕捉血糖动态的趋势性特征。去除 1D-CNN 模块后, 模型同样出现性能下降, 尽管幅度略小于前两者, 但其短期动态建模能力减弱, 导致模型对突发血糖波动响应不及时, 表现为在 1 h 预测中 MAPE 上升至 31.25%, 在 3 h 预测中 MAE 上升至 55.21。见表 3。上述结果进一步证明了 1D-CNN 模块在提取时间局部特征方面的重要性。

表 3 消融实验结果
Table 3 Ablation experiment results

模型	评价指标	预测时间/h		
		1	2	3
XCLA-Net w/o 1D-CNN	RMSE	43.86	52.33	61.76
	MAE	32.81	42.76	55.21
	MAPE/%	31.25	36.56	48.27
XCLA-Net w/o cross-attention	RMSE	54.86	67.33	69.72
	MAE	39.45	45.76	63.29
	MAPE/%	35.68	43.05	52.24
XCLA-Net w/o LSTM	RMSE	45.74	54.81	74.66
	MAE	29.71	39.03	49.66
	MAPE/%	29.80	33.56	45.31
XCLA-Net	RMSE	38.79	45.51	49.76
	MAE	22.59	34.61	41.19
	MAPE/%	25.37	31.31	38.09

2.3 克拉克 EGA

见图 3。XCLA-Net 模型相较所有对比模型表现最为优异, 该模型的预测点高度集中于 A 区, 说明其预测值与真实血糖浓度偏差最小, 具备高度临床一致性与稳定性; 在 1 h 预测任务中, 有 85.6% 的预测点落在 A 区, 12.7% 落于 B 区, C~E 区合计占比 1.7%。在低血糖区间与高血糖区间预测中, 绝大多数预测结果位于 A 区, 提示模型具备较强的极端血糖值预测能力及临床可用性。相比之下, 虽然 MLP 与 GRU 模型在 A 区分布较为密集, 但预测点在 B 区的扩散更为明显, 个别点偏离对角线较远, 提示模型在特定血糖区间, 尤其是边缘状态下的预

测稳定性不足。Autoencoder 模型表现相对较差,预测点分布更为离散,A区覆盖度显著低于其他三种

模型,C/D区占比较高,说明其在多数样本上存在较大的预测偏差,临床可接受性明显不足。

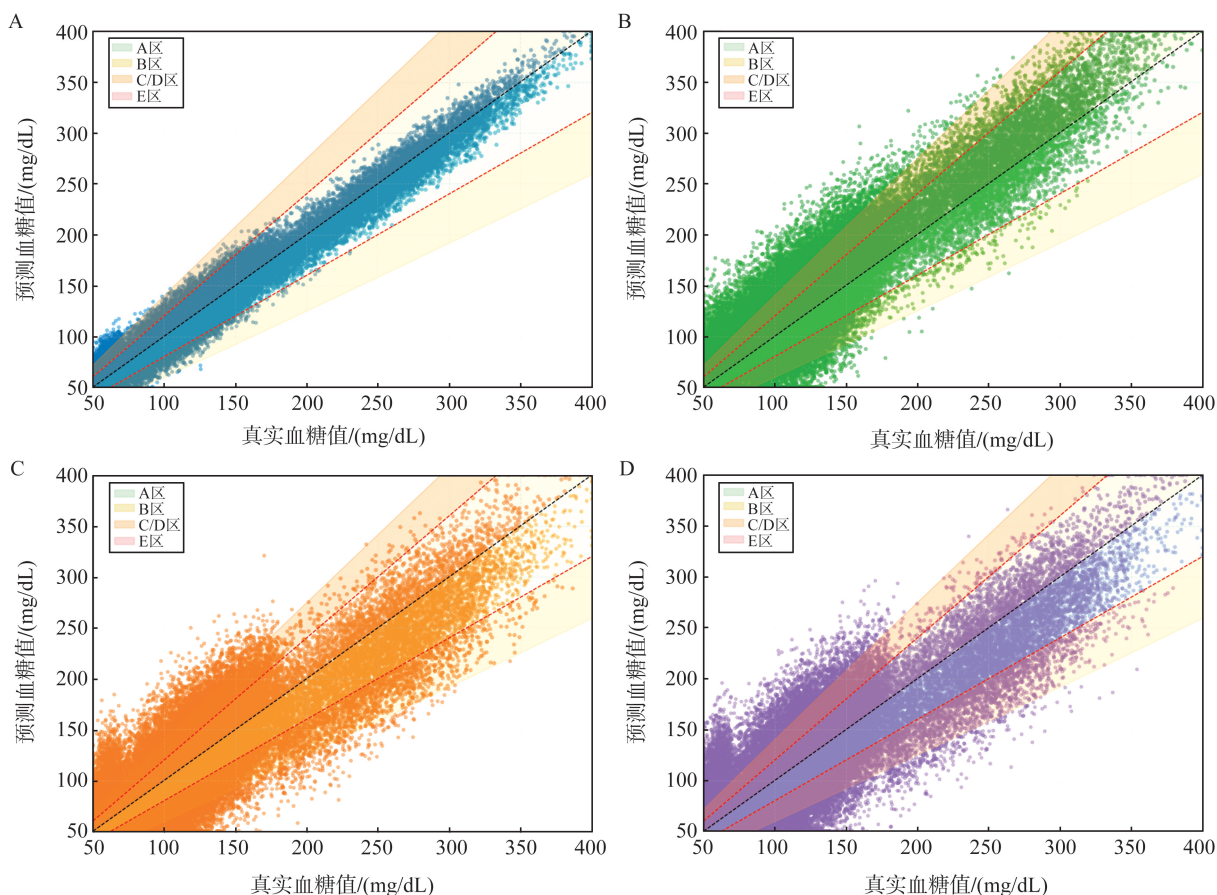


图3 不同模型1 h血糖浓度预测的克拉克误差网格分析

A: XCLA-Net 模型; B: GRU 模型; C: MLP 模型; D: Autoencoder 模型。

Figure 3 Clarke EGA of 1-hour blood glucose concentration prediction in different models

A: XCLA-Net model; B: GRU model; C: MLP model; D: Autoencoder model.

3 讨论

本研究提出了一种融合 FGM 与结构化 EHR 多模态数据的血糖浓度预测模型 XCLA-Net,实现了对 T1DM 患者血糖浓度的高精度预测。模型针对不同模态数据在采样频率、特征稠密性与语义结构等方面的显著差异,设计了由特征提取、模态融合、时序建模构成的结构化模型框架。利用交叉注意力机制实现主模态引导下的多模态语义对齐,在此基础上,构建以融合序列为输入的时序预测模块,联合 1D-CNN 与 LSTM 对血糖序列的短期与长期变化趋势进行建模,实现对未来连续血糖浓度的精准预测。实验结果表明,XCLA-Net 在 T1DM 患者血糖预测任务中,在准确性、稳定性及临床可接受性等多个维度上均优于对比模型。

对比实验结果显示,XCLA-Net 在 RMSE、MAE

和 MAPE 等数值误差指标上均取得最优表现,且在预测时间由 1 h 延长至 3 h 的过程中,误差增长幅度显著低于传统时序预测模型,体现了其优越的趋势捕捉能力与长时序建模能力。EGA 结果表明,该模型预测值高度集中于临床可接受的安全区域,表现出良好的临床一致性与应用安全性。综合上述结果,XCLA-Net 在准确性与稳定性方面均优于对比模型,验证了其在多模态特征融合机制支持下的预测有效性,并显示其在 T1DM 患者血糖管理的真实临床场景中具有可行的部署与应用潜力。

消融实验分析揭示 XCLA-Net 各子模块的重要性。交叉注意力模块通过引入模态间动态注意力机制,有效实现多模态语义对齐,是提升融合表示有效性的关键环节;LSTM 与 1D-CNN 模块分别侧重于长期趋势与短期波动的时序特征建模,使模型能够充分捕捉不同时间尺度的血糖动态信息。实验进一步揭示,移除交叉注意力机制将导致整体性能显著

下降,反映其在多模态交互中的核心地位;去除 LSTM 或 1D-CNN 模块则分别削弱了模型对长期趋势与短期波动的捕捉能力,说明二者在时序特征提取中的互补性对模型性能提升至关重要。

尽管 XCLA-Net 在血糖浓度预测任务中表现出显著优势,但仍存在一定局限性。① 本研究采用的 T1DiabetesGranada 数据集来自西班牙单中心队列,样本在种族分布、年龄结构及地域医疗实践中存在显著偏倚,该数据结构可能限制模型在不同地区和人群中的推广应用。因此在数据预处理过程中,本研究剔除了观测天数严重不足、关键变量严重缺失的个体。同时在模型训练过程中采用患者级别的随机分层方式,确保各子集中年龄、性别与血糖水平的分布大致一致,以减缓因样本结构偏倚带来的模型训练偏向问题。② 研究未对样本进行严格意义上的分层抽样,主要原因在于原始数据为回顾性收集,研究设计阶段不可控。未来可通过多中心协作,纳入跨地域、跨种族的多样化数据,并结合迁移学习^[26]和领域自适应技术^[27],缓解单中心数据带来的局限性。③ 本研究主要采用传统回归误差指标评估预测性能,虽能衡量总体数值偏差,但对于低血糖、高血糖等临床关键事件的预测能力评估尚不完善,未来研究可引入血糖风险区间指标,结合 Parkes 误差网格分析^[28]等方法,从临床安全性和决策辅助能力多维度综合评价模型性能。同时模型解释性方面仍有提升空间,考虑引入可视化注意力权重分析^[29]、SHAP^[30]等可解释性方法,增强模型预测结果在临床场景中的可理解性与信任度。

综上所述,XCLA-Net 通过多模态数据融合与时序建模,有效提升了 T1DM 患者血糖浓度预测的精度与稳定性,在处理结构复杂、动态变化明显且信息来源多样的临床数据场景中依然保持稳定的预测性能,体现出较强的泛化能力与适应性。未来研究可进一步拓展模型在不同人群与适应症中的适用性,并推动其在智能化糖尿病管理中的临床转化与实际应用。

参考文献:

- [1] Ilonen J, Lempainen J, Veijola R. The heterogeneous pathogenesis of type 1 diabetes mellitus [J]. *Nat Rev Endocrinol*, 2019, 15(11): 635-650.
- [2] International Diabetes Federation. IDF diabetes atlas, 11th edn [R]. Brussels, Belgium: International Diabetes Federation, 2025.
- [3] Martín-Timón I, Del Cañizo-Gómez FJ. Mechanisms of hypoglycemia unawareness and implications in diabetic patients [J]. *World J Diabetes*, 2015, 6(7): 912-926.
- [4] 师瑞, 冯磊, 唐灵通, 等. 糖尿病患者血糖波动评价指标研究进展 [J]. *中华全科医学*, 2022, 20(12): 2105-2109.
SHI Rui, FENG Lei, TANG Lingtong, et al. Research progress on evaluation indicators of blood glucose fluctuation in patients with diabetes [J]. *Chinese Journal of General Practice*, 2022, 20(12): 2105-2109.
- [5] Nemat H, Khadem H, Elliott J, et al. Data-driven blood glucose level prediction in type 1 diabetes: a comprehensive comparative analysis [J]. *Sci Rep*, 2024, 14: 21863. doi:10.1038/s41598-024-70277-x
- [6] Wang Q, Molenaar P, Harsh S, et al. Personalized state-space modeling of glucose dynamics for type 1 diabetes using continuously monitored glucose, insulin dose, and meal intake: an extended Kalman filter approach [J]. *J Diabetes Sci Technol*, 2014, 8(2): 331-345.
- [7] 肖泽秋, 李勇, 王霞. 基于 PBI-CLA 模型的糖尿病患者血糖浓度预测 [J]. *计算机工程*, 2025: 1-11. doi: 10.19678/j.issn.1000-3428.0070527
XIAO Zeqiu, LI Yong, WANG Xia. Prediction of blood glucose concentration in diabetic patients based on PBI-CLA model [J]. *China Industrial Economics*, 2025: 1-11. doi: 10.19678/j.issn.1000-3428.0070527
- [8] 童梦, 丁国荣, 余楠, 等. 一种基于 VMD-PSO-LSTM 的血糖预测方法 [J]. *计算机与数字工程*, 2023, 51(6): 1439-1443.
TONG Meng, DING Guorong, YU Nan, et al. Prediction method of blood glucose based on VMD-PSO-LSTM [J]. *Computer & Digital Engineering*, 2023, 51(6): 1439-1443.
- [9] Wang L, Pan ZL, Liu W, et al. A dual-attention based coupling network for diabetes classification with heterogeneous data [J]. *J Biomed Inform*, 2023, 139: 104300. doi:10.1016/j.jbi.2023.104300
- [10] Rodriguez-Leon C, Aviles-Perez MD, Banos O, et al. T1DiabetesGranada: a longitudinal multi-modal dataset of type 1 diabetes mellitus [J]. *Sci Data*, 2023, 10(1): 916. doi:10.1038/s41597-023-02737-4
- [11] Engels JM, Diehr P. Imputation of missing longitudinal data: a comparison of methods [J]. *J Clin Epidemiol*, 2003, 56(10): 968-976.
- [12] Dahouda MK, Joe I. A deep-learned embedding technique for categorical features encoding [J]. *IEEE Access*, 2021, 9: 114381-114391. doi:10.1109/ACCESS.2021.3104357

- [13] Jain YK, Bhandare SK. Min max normalization based data perturbation method for privacy protection[J]. *Int J Comput Commun Technol*, 2013; 233-238. doi: 10.47893/ijct.2013.1201
- [14] Palangi H, Deng L, Shen YL, et al. Deep sentence embedding using long short-term memory networks: analysis and application to information retrieval [J]. *IEEE/ACM Trans Audio Speech Lang Process*, 2016, 24(4): 694-707.
- [15] Shi J, Telesca D, Suchard MA. Statistical learning techniques applied to epidemiology: a simulated case-control comparison study with logistic regression [J]. *BMC Bioinformatics*, 2011, 12: 37. doi:10.1186/1471-2105-12-37
- [16] Zhang J, He XH, Liu Y, et al. Multi-modal cross-attention network for Alzheimer's disease diagnosis with multi-modality data[J]. *Comput Biol Med*, 2023, 162: 107050. doi:10.1016/j.combiomed.2023.107050
- [17] Klambauer G, Unterthiner T, Mayr A, et al. Self-normalizing neural networks [EB/OL]. (2017-09-07) [2025-04-26]. <https://doi.org/10.48550/arXiv.1706.02515>
- [18] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [EB/OL]. (2023-08-02) [2025-04-26]. <https://doi.org/10.48550/arXiv.1706.03762>
- [19] Santurkar S, Tsipras D, Ilyas A, et al. How does batch normalization help optimization? [EB/OL]. (2019-04-15) [2025-04-26]. <https://doi.org/10.48550/arXiv.1805.11604>
- [20] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. *J Mach Learn Res*, 2014, 15: 1929-1958. doi: 10.5555/2627435.2670313
- [21] 陶蔚, 陇盛, 刘鑫, 等. 深度学习步长自适应动量优化方法研究综述[J]. *小型微型计算机系统*, 2025, 46(2): 257-265.
TAO Wei, LONG Sheng, LIU Xin, et al. Review of adaptive stepsize momentum optimization methods in deep learning [J]. *Journal of Chinese Computer Systems*, 2025, 46(2): 257-265.
- [22] Felizardo V, Garcia NM, Pombo N, et al. Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction – a systematic literature review [J]. *Artif Intell Med*, 2021, 118: 102120. doi:10.1016/j.artmed.2021.102120
- [23] Clarke WL. The original Clarke error grid analysis (EGA) [J]. *Diabetes Technol Ther*, 2005, 7(5): 776-779.
- [24] Shiri FM, Perumal T, Mustapha N, et al. A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU [EB/OL]. (2025-03-17) [2025-04-26]. <https://doi.org/10.48550/arXiv.2305.17473>
- [25] Kamyshanska H, Memisevic R. The potential energy of an autoencoder [J]. *IEEE Trans Pattern Anal Mach Intell*, 2015, 37(6): 1261-1273.
- [26] Deng YX, Lu L, Aponte L, et al. Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients [J]. *NPJ Digit Med*, 2021, 4(1): 109. doi:10.1038/s41746-021-00480-x
- [27] Yang T, Yu X, Tao R, et al. Blood glucose prediction for type 2 diabetes using clustering-based domain adaptation [J]. *Biomed Signal Process Contr*, 2025, 105: 107629. doi:10.1016/j.bspc.2025.107629
- [28] Pfützner A, Klonoff DC, Pardo S, et al. Technical aspects of the Parkes error grid [J]. *J Diabetes Sci Technol*, 2013, 7(5): 1275-1281.
- [29] Yu X, Yang Z, Wang XZ, et al. A prior-knowledge-guided dynamic attention mechanism to predict nocturnal hypoglycemic events in type 1 diabetes [J]. *BMC Med Inform Decis Mak*, 2024, 24(1): 378. doi:10.1186/s12911-024-02761-3
- [30] Van den Broeck G, Lykov A, Schleich M, et al. On the tractability of SHAP explanations [J]. *Jair*, 2022, 74: 851-886. doi:10.1613/jair.1.13283

(编辑:相峰)