**二型糖尿病(T2D)的预测分类模型**

## **一、背景与意义**

提高机器学习模型在个体层面准确预测二型糖尿病(T2D)风险的能力，为实施循证生活方式或药物干预提供了坚实的基础，从而延缓或预防T2D及其并发症的发展。目前，大规模生物医学数据库提供了丰富数据，包含来自参与者的遗传、生活方式和电子医疗记录等丰富信息。利用这些数据进行机器学习分析，可以帮助医生提前预测患病风险，更好地理解T2D的发展趋势，并为个性化的预防和治疗提供支持。因此，将机器学习方法应用于T2D预测任务上具有重要的实际意义，有望改善临床决策的准确性和效率。然而，这些研究存在一些局限性，包括模型可解释性、数据异质性和模型特异性。

**挑战一：**模型可解释性

在医疗领域，模型的可解释性是至关重要的，因为医生需要了解模型的决策依据。然而，当前的机器学习模型，特别是深度学习模型，通常被认为是黑盒子，难以解释其预测结果。因此，如何提高模型的可解释性是一个亟待解决的难题。

**挑战二：**数据异质性

由于不同医疗评估中心之间的测量标准不一致，以及患者在提供个人信息时的主观性和不确定性，导致数据存在噪声和不完整性。如何有效地处理这些异质性数据，提取有用的信息，是一个具有挑战性的任务。

**挑战三：**模型特异性

在医疗领域，模型的特异性这一挑战尤为突出。由于每位患者的生理特征、基因组变异和生活方式因素不同，以及医疗数据的分布可能因患者人群、地理位置和医疗实践的不同而异，导致了模型的泛化能力不足。因此，如何深入挖掘丰富的多源医疗数据，整合生物标志物、影像学数据、遗传信息和临床记录等多维度数据，以提高模型的预测能力和个性化治疗的精准性，是当前亟需解决的问题。

## **二、题目**

* **所属领域：**生物医药领域。
* **任务描述：**本次竞赛分为三个子任务。

**子任务1：单模态数据患者分类**

根据输入的个体单模态数据，包括基础信息、生活习惯判断个体是否患有T2D；

**子任务2：多模态数据患者分类**

根据输入的个体多模态数据，包括基础信息、生活习惯、心电数据判断个体是否患有T2D；

**子任务3：患者发病时间预测**

根据输入的个体多模态数据，预测患者发病时间。

赛题讲解：<http://www.aisccc.cn/course/course-details?id=117>

## **三、数据来源**

收集了多个来自不同医疗评估中心的个体样本数据集，该数据集包括数3000位健康样本及1000位二型糖尿病患者基础数据，心电数据及生活习惯等数据。按照8:2随机划分成训练集和测试集，训练集包括2400位健康样本及800位糖尿病患者的信息，测试集包括600位健康样本及200位糖尿病患者的信息。该数据集覆盖了不同年龄、性别的样本，提供了丰富的生物标本、心电组学数据、生活方式信息和疾病诊断记录等多种数据类型，有助于从多个角度研究 T2D 预测因素。主要难点在于：T2D 是一种遗传和环境因素相互作用导致的疾病，需要综合考虑临床特征，环境因素和生活方式等多方面因素，对疾病进行准确预测；且不同特征之间可能存在复杂的关联性，需要处理特征之间的共线性和交互作用，以避免过拟合和提高模型的解释性。

数据详情及文件下载链接: <http://www.aisccc.cn/database/data-details?id=122>

## **四、评价标准**

**1. 主观（50%）：**AUROC,AUPRC, F1 score, R2, PCC

**2. 客观（50%）：**数据处理、特征选择、运行时间、模型可解释性等

## **五、提交内容**

文章、方案设计思路、源代码

## **六、参考文献**

1. Deberneh, Henock M., and Intaek Kim. "Prediction of type 2 diabetes based on machine learning algorithm." International journal of environmental research and public health 18.6 (2021): 3317.
2. Kumar, Mukkesh, et al. "Machine learning–derived prenatal predictive risk model to guide intervention and prevent the progression of gestational diabetes mellitus to type 2 diabetes: Prediction model development study." JMIR diabetes 7.3 (2022): e32366.
3. Deng, Yixiang, et al. "Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients." NPJ Digital Medicine 4.1 (2021): 109.
4. Das, Bihter. "A deep learning model for identification of diabetes type 2 based on nucleotide signals." Neural Computing and Applications 34.15 (2022): 12587-12599.