



医数云析
MEDICAL DATA

医学大数据与AI应用产品介绍

汇报人：Julianne

目录

01 医疗数据系统间偏差分析系统

02 数据标准化预处理框架

03 科研文章自动化生成工具

04 数字模型期刊/文章转软著专利

05 医数智算一体机解决方案

06 本地专业知识库+AI智能体

07 云存储+网安

08 体检报告AI分析系统



医数云析

MEDICAL DATA

01

医疗数据系统间偏差分析系统
(室内质控室间化比对系统)

系统概述

数据偏差问题

在多中心研究中，不同医院的数据因检测系统、方法学等因素存在偏差，直接合并会导致数据分布不均，影响模型训练效果。

解决方案

提供两种偏差消除方法：一是使用统一质控品检测并采集数据，通过统计学质控分析法+PBRTQC技术计算调整偏差；二是利用体检大数据，采用专利的RUV技术消除系统间偏差。

效果验证

在天津市100+医院测试，效果显著。如两家医院的7个检测指标平行质控数据实验，偏差矫正后交互系数 β 从0.05提升至0.07，效果明显。



系统优势

通用性强

适用于医院临床科室产生的所有定量、定性数据类型，能广泛应用于各类医疗数据处理场景。



方法灵活

两种方法结合使用最佳，用户可根据自身情况选择，在有无单独预算的情况下都可开展。



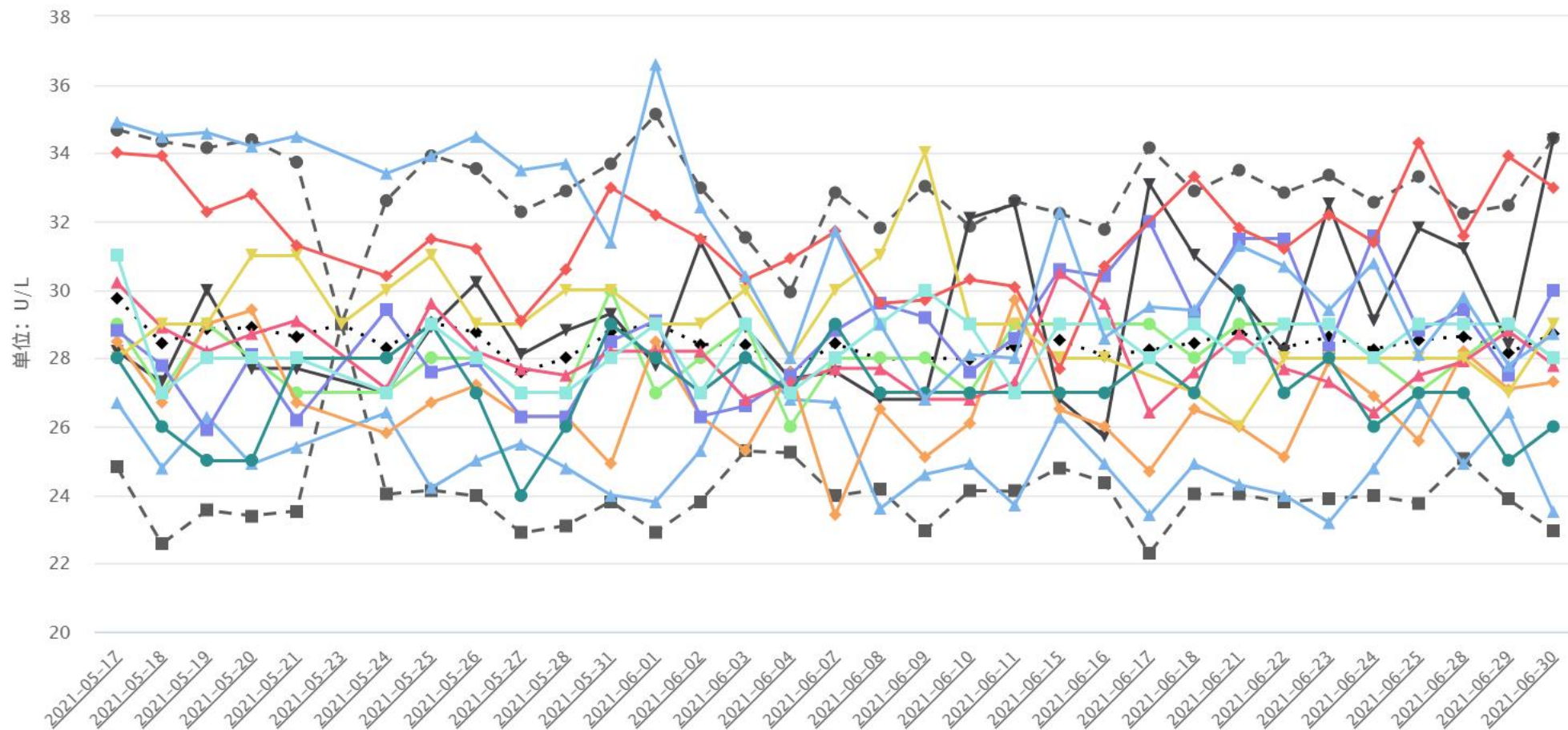
精确高效

基于统计学的数理模型计算效果精确，可解释性输出；结合PBRTQC方法双重检测更精准的识别差异点。



2020-06-28 to 2021-07-28ALT 项目室间比对

批号【jq_001】





医数云析
MEDICAL DATA

02

数据标准化预处理框架

解决的问题

- 1、数据从医院的数据库中导出的形式往往是不能直接用于分析使用的，需要治理、清洗、做标准化处理后才能成为可应用的格式。
- 2、成为应用格式的数据后，对特定的分析还需采用特定方法做进一步的矫正、纠偏、去中心化等一些列处理之后，用于模型训练统计分析才会得到理想的效果，其处理过程繁杂，能力要求高。

方法优势

- 1、兼容常见的数据库格式，如.sql、.dbf、.json、.xlsx及关联表形式等；以用户指定的格式导出。
- 2、多种标准化处理方法供选择，分析数据特征做出默认选择。
- 2、数据转换过程中，处理步骤和方式实时清晰呈现。
- 3、整个过程一键上传、一键导出，可按默认选项完全无干预完成。

效果验证

已做50+篇发表文章的公开数据集验证，通过我们标准化处理之后的结果比未应用处理的结果有明显提高的达到80%以上。



规则层

01

通用规则

通过外部通用规则库导入，把项目的方法学、单位、数据类型、精度、参考范围等所需的标准化项进行导入，以供处理时调用。

02

本地规则

针对原始数据的生产环境不同，如系统的校准模型、拟合方式、报告形式等，需将其调整为通用规则的要求形式，以便输出到下一层做预处理。

03

规则应用

规则层的应用对单家数据来说，是非必须的。但如研究者的文章或模型将来是想要多中心泛化的，则是力荐之推，

治理层

总体矫正

采用专利的“体检通用大模型量化确值法”，针对生产设备固有制约属性及不可控环境属性等问题进行矫正补偿，以修正一些本不该有的错误，如偏峰、多峰、数据断层等。

核心方法

多种方法选择，并针对不同数据类型提供默认组合：

方法名称	描述
auto	自动选择合适的标准化方法（基于数据分布特征）
log_transform	对数变换，适用于右偏分布且全正值的数据
min_max_scale	最小-最大缩放，将数据线性变换到 [0,1] 范围
z_score_standardize	Z 分数标准化，转换为均值为 0、标准差为 1 的分布
max_abs_scale	最大绝对值缩放，将数据除以该特征的最大绝对值
center_data	中心化处理，使数据均值为 0 但不缩放方差
scale_data	缩放处理，调整数据尺度但不改变中心位置
boxcox_transform	Box-Cox 变换，适用于正数数据的幂变换
yeojohnson_transform	Yeo-Johnson 变换，适用于含零/负值数据的广义幂变换

处理过程实时输出

✓ 处理完成

输入文件
data.xlsx

输出文件
1.xlsx

处理详情

数据质量评估与矫正：针对 453 个数据特征/列执行 /

[17:09:33] [INFO] 变量 'Alcohol'发现5个异常值（超出[-581.28, 730.23]范围）
[17:09:33] [INFO] 异常值处理：删除0个异常值（基于Alcohol的生理指标范围）
[17:09:33] [INFO] 自动选择：稳健缩放（基于中位数和IQR，抗异常值）| 原始中位数：4.20
选择依据：
- 样本量n=912，原始范围[0.42, 952.00]；偏度=2.58（右偏显著），峰度=7.73；异常值：120个（占比13.2%，显著）；0值占比0.0%（非稀疏）；正态性检验：Shapiro-Wilk检验（统计量=0.6206，p值=0.0000，未通过）；优先选择：显著异常值适合稳健缩放（抗极端值）
标准化后：范围[-0.0440, 11.0466]，均值≈0.8191，标准差≈1.5277
[17:09:33] [INFO] 4. 标准化处理：自动选择：稳健缩放（基于中位数和IQR，抗异常值）| 原始中位数：4.20
选择依据：
- 样本量n=912，原始范围[0.42, 952.00]；偏度=2.58（右偏显著），峰度=7.73；异常值：120个（占比13.2%，显著）；0值占比0.0%（非稀疏）；正态性检验：Shapiro-Wilk检验（统计量=0.6206，p值=0.0000，未通过）；优先选择：显著异常值适合稳健缩放（抗极端值）
标准化后：范围[-0.0440, 11.0466]，均值≈0.8191，标准差≈1.5277
[17:09:33] [INFO] 标准化后数据范围：[-0.0440, 11.0466]

清洗层

01

异常值处理

- 1、自动检测缺失值，进行中位均值填补，高于一定比例缺失则删除此维度。
- 2、异常值的修正：用本地模型训练出了对众多常见错误类型的修正，遇到不可修正的记录则删除，并将删除的记录合并呈现在结果界面，用户如想保留可自行修正。

02

非结构化数据处理

对非结构化数据，如文字描述段，需根据提示词和参考文调用通用大模型底座识别提取，一般常见的如部位、颜色、时间等有确切语义判断的，都能达到较理想效果。特殊意义的单独训练。

03

术语标准化

- 1、中文字段的列名转为英文（为减少很多软件和函数的不兼容性麻烦）
- 2、名称行业标准化：调用规则层中的标准名称命名，此为非必须，但推荐勾选。

数据预处理工具

上传您的数据文件，选择处理模型，然后进行分析。我们将帮助您完成数据预处理和术语标准化工作。

下载模板文件

点击下载数据模板，按照模板格式准备您的数据



拖放文件到此处上传

或者

选择文件

已上传: data.xlsx

上传成功

选择处理模型

数据预处理

☒ 包括变量分类、分类变量编码、高基数变量处理和连续变量标准化

术语标准化

☐ 将表头术语按照标准医学术语进行映射和标准化

自定义文件名: * 1

文件名会自动添加.xlsx扩展名，请勿包含特殊字符 (/:*? "<>|)

开始分析

处理完成

处理已完成

处理完成

输入文件

data.xlsx

输出文件

1.xlsx

处理详情

数据预处理 (注意: 255个0值被识别为大值)

[17:09:33] [INFO] 变量 'Alcohol' 发现5个异常值 (超出[-581.28, 730.23]范围)

[17:09:33] [INFO] 异常值处理: 删除0个异常值 (基于Alcohol的生理指标范围)

[17:09:33] [INFO] 自动选择: 稳健缩放 (基于中位数和IQR, 抗异常值) | 原始中位数: 4.20

选择依据:

- 样本量n=912, 原始范围[0.42, 952.00]; 偏度=2.58 (右偏显著), 峰度=7.73; 异常值: 120个 (占比13.2%, 显著); 0值占比0.0% (非稀疏); 正态性检验: Shapiro-Wilk检验 (统计量=0.6206, p值=0.0000, 未通过); 优先选择: 显著异常值适合稳健缩放 (抗极端值)

标准化后: 范围[-0.0440, 11.0466], 均值≈0.8191, 标准差≈1.5277

[17:09:33] [INFO] 4. 标准化处理: 自动选择: 稳健缩放 (基于中位数和IQR, 抗异常值) | 原始中位数: 4.20

选择依据:

- 样本量n=912, 原始范围[0.42, 952.00]; 偏度=2.58 (右偏显著), 峰度=7.73; 异常值: 120个 (占比13.2%, 显著); 0值占比0.0% (非稀疏); 正态性检验: Shapiro-Wilk检验 (统计量=0.6206, p值=0.0000, 未通过); 优先选择: 显著异常值适合稳健缩放 (抗极端值)

标准化后: 范围[-0.0440, 11.0466], 均值≈0.8191, 标准差≈1.5277

[17:09:33] [INFO] 标准化后数据范围: [-0.0440, 11.0466]

下载处理结果

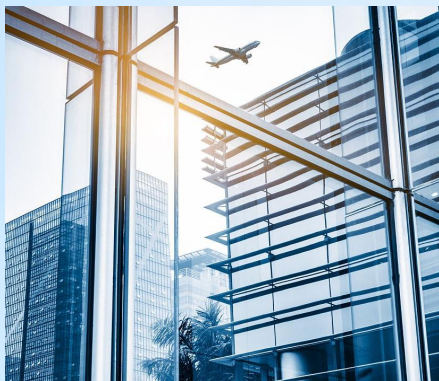


医数云析
MEDICAL DATA

03

科研文章自动化生成工具

数据预 处理模 块



兼容多种格式

兼容常见的数据库导出格式，如.sql、.dbf、.json、.xlsx及关联表形式等，方便用户导入数据，导入后自动完成表的关联，按照用户指定的格式导出。



标准化处理

将输入数据做分析标准化处理，输出为标准维度列形式，并进行清洗处理，为后续分析生成模块提供高质量数据。



数据特征分析

对数据特征、维度的意义，相关研究进展会做一个初步分析，提供给用户做研究思路参考。

AI分析生成模块

深度学习框架调用

web版通过API调用，单机版通过本地大模型底座调用深度学习框架，生成分析流的核心部分，包括文字、图表、计算结果等。

分析处理过程

分‘外脑’和‘内脑’，两部分AI来完成。外脑AI负责思路、流程和用户的交互选择。内脑AI负责操作本地数据库的调用处理。

流程模块化

工具给出些常见的分析模块，包括相关、回归、时间序列分析、机器学习模型等。用户只需按列名的选择来调用相应的模块完成计算即可。

版面导出模块

文章样式展示

以整篇完整的文章样式展示生成的科研文章，方便用户查看和修改。

01

可运行代码

提供分步可运行的python源代码部分，供中间修改再运行，用户可根据自身需求进行调整和优化。

02

文章导出

修改确认稿后可导出静态pdf文，方便用户保存和分享科研成果，满足不同场景下的使用需求。

03



医数云析
MEDICAL DATA

04

动态模型期刊/文章转软著专利

动态模型期刊

独家创念

是将真实的用户数据连接在前端文章页面调用中计算呈现，文章中的所有基于数据表的数字、图表都与后台数据库动态链接，之所谓‘动态模型期刊’。并可以让其他用户上传自己的同样维度的数据，验证是否能得到同样结果，能让众多用户为此文的模型做泛化验证。

社区互动

与天津医科大学vs天津医学检验学会联名推出，公开发行，接受业内评议。社区内按评论精华度、活跃度、贡献度给予积分，鼓励众多爱研究学习能力强的青年参与，用实力和技术比拼证明自己。

创新意义

动态模型期刊的推出，在AI时代颠覆传统静态paper式的文章，挑战真实、专业、技术，多重知识能力的运用，激发青年创新性。



动态模型期刊功能

模型生成

将文章写成软件模型形式，调用动态数据表生成文章里面的数据计算部分，实现数据的动态更新和展示。



结果复现

允许其他用户上传同样形式的自己的数据表来复现文章结果，验证模型的准确性和可靠性。



软著专利申报

动态模型期刊可作为软著专利申报的依据，为科研人员的知识产权保护提供支持，提升科研成果的价值。



血清γ-谷氨酰转肽酶与动脉粥样硬化关联研究论文生成工具

下载Excel数据模板

选择文件 未选择文件

分析数据

生成论文

下载为Word

Diastolic blood pressure (mm Hg)	0.48 (0.41 to 0.55)	<0.001	0.43 (0.33 to 0.53)	<0.001
Fasting plasma glucose (mg/dL)	0.24 (0.16 to 0.31)	<0.001	0.23 (0.13 to 0.34)	<0.001
Total cholesterol (mg/dL)	0.02 (−0.06 to 0.10)	0.64	0.28 (0.18 to 0.39)	<0.001
Triglycerides (mg/dL)	0.08 (0.00 to 0.16)	<0.05	0.22 (0.11 to 0.32)	<0.001
HDL-C (mg/dL)	−0.07 (−0.15 to 0.02)	0.11	−0.03 (−0.14 to 0.08)	0.59
LDL-C (mg/dL)	0.00 (−0.08 to 0.08)	0.96	0.26 (0.15 to 0.37)	<0.001
eGFR (mL/min/1.73 m ²)	−0.28 (−0.36 to −0.20)	<0.001	−0.35 (−0.45 to −0.24)	<0.001
Alcohol consumption (g/week)	0.02 (−0.06 to 0.10)	0.64	−0.05 (−0.16 to 0.06)	0.35
Current smoker	0.01 (−0.07 to 0.09)	0.77	−0.15 (−0.26 to −0.04)	<0.01
Regular exerciser	0.02 (−0.06 to 0.10)	0.67	0.05 (−0.06 to 0.16)	0.35
Fatty liver	0.07 (−0.02 to 0.15)	0.11	0.25 (0.14 to 0.35)	<0.001
Postmenopausal state	—	—	0.43 (0.33 to 0.52)	<0.001

baPWV, brachial-ankle pulse wave velocity; BMI, body mass index; eGFR, estimated glomerular filtration rate; GGT, γ -glutamyltranspeptidase; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

state did not greatly alter the association (model 3). Thus, $\log_2$ GGT was significantly associated with baPWV independently of covariates in women ($\beta=0.11$, 95% CI 0.02 to 0.19, $p<0.05$), but not in men ($\beta=-0.04$, 95% CI -0.10 to 0.03, $p=0.28$; model 3).

DISCUSSION

We examined the association between arterial stiffness measured by baPWV and GGT in the Japanese population. The present cross-sectional study clearly indicated that $\log_2$ GGT significantly correlated with the baPWV level independent of age, BMI, the volume of consumed alcohol and the presence of fatty liver in women, but not in men. Our data were in agreement with the results of a previous study¹⁴ showing that serum GGT was associated with baPWV only in women.

GGT is the enzyme responsible for the extracellular catabolism of glutathione.²⁷ GGT is found not only in the liver but also in other organ tissues including the kidney.

gene and certain mRNA subtypes are increased during oxidative stress in the inflammatory atheroma within the vascular endothelial wall,³² and produced GGT protein acts as strong reducing agent of the superoxide ion and hydrogen peroxide.¹ In fact, one isoform of the GGT protein detected in the supernatant of plaque homogenates was comparable with serum GGT and a significant correlation was observed between plaque GGT deposits and corresponding GGT levels.³¹ Those observations may indicate that elevated serum GGT includes GGT proteins produced in the plaque corresponding to oxidative stress. GGT proteins produced in the endothelium cell may have a significant effect on GGT levels, because the total weight of the endothelium cell of the whole body would be equal to that of the liver.³³

We considered that one possible mechanism of sexual discrepancy regarding GGT with baPWV was the difference in sexual hormones, especially oestradiol. In women, oestradiol produced by the granulosa cells of the ovaries is one of the steroid hormones and protects cells from the oxidative stress-induced cell ³⁴. Although



05

医数智算一体机解决方案

医数智算解决方案

算力主机 + 大模型底座 + 医疗数据标准化处理框架

产品概述

我们推出的"医院专用算力主机 + 大模型底座 + 医疗数据标准化处理框架"是一套完整的、专为医疗机构设计，旨在解决运行在AI医疗应用底层的算力、大脑、数据标准化的瓶颈和难题，帮助医院快速构建智能化应用场景。

医院专用算力主机

专为医疗场景优化的高性能计算设备，满足医疗AI模型训练和推理的算力需求。

- 医疗级安全与稳定性设计
- 支持多GPU并行计算
- 医疗数据本地化处理
- 符合医疗行业合规要求

医疗大模型底座

基于海量医疗数据训练的领域大模型，提供开箱即用的医疗AI能力。

- 医学知识图谱嵌入
- 多模态医疗数据处理
- 支持快速微调适配
- 持续更新迭代

数据标准化处理框架

经过验证的医疗数据质控体系，确保数据质量满足AI应用需求。

- 三层架构标准化处理
- 智能化数据清洗
- 专利纠偏算法
- 行业标准兼容

为什么选择我们的解决方案？

在AI医疗时代，数据质量直接决定智能系统的上限。我们的三位一体解决方案解决了医疗机构在AI落地过程中面临的三大核心问题：算力不足、模型缺乏和数据质量问题，为医院提供从基础设施到应用能力的完整支持。

核心优势

一站式解决方案

整合硬件、算法和数据处理的完整链条，避免医院在多个环节分别投入和集成的工作。

医疗场景优化

从硬件到软件都针对医疗场景特殊需求进行优化，确保符合医疗行业标准和规范。

快速部署应用

开箱即用的能力，帮助医院在短时间内构建AI应用场景，缩短价值实现周期。

持续进化能力

云端协同架构确保模型和数据标准持续更新，保持技术前沿性。

技术架构

医疗专用GPU服务器

容器化部署

医疗大模型底座

规则引擎

数据治理层

智能清洗层

知识图谱

联邦学习



医数云析
MEDICAL DATA

06

本地专业知识库训练、AI智能体定制

专业知识库训练

本地训练

在本地通过专有知识库训练AI智能体，利用医院所有的医生病例历史数据，使其具备专业的医疗知识和经验。



个性化定制

根据医院的实际情况和需求，定制个性化的AI智能体，满足不同科室、不同医生的使用习惯和需求。

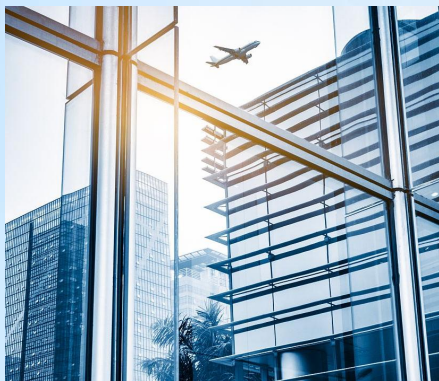


训练效果

训练后的AI智能体能够更好地理解和处理医疗数据，为医生提供更准确、更个性化的辅助诊断和治疗建议。



AI 智能 体应用



辅助诊断

AI智能体可辅助医生进行疾病诊断，提供诊断依据和建议，提高诊断的准确性和效率。



治疗方案推荐

根据患者的具体情况，推荐个性化的治疗方案，为医生的治疗决策提供参考，提升治疗效果。



知识更新

持续学习最新的医疗知识和研究成果，保持智能体的专业性和时效性，为医疗工作提供持续支持。

本地专业知识库训练与AI智能体定制优势

提升医疗服务质量

01

通过本地专业知识库训练和AI智能体定制，提高医疗诊断和治疗的准确性和效率，提升整体医疗服务质量。

满足个性化需求

02

个性化定制的AI智能体能够更好地适应不同医院、不同科室的需求，满足医疗工作的多样化需求。

推动医疗智能化发展

03

为医疗领域引入先进的AI技术，推动医疗智能化发展，提高医疗服务的可及性和便捷性。



医数云析
MEDICAL DATA

07

云存储服务

服务内容

医疗数据云存储

提供医疗数据云存储服务，满足医院对医疗数据存储的需求，确保数据的安全和完整。

网络安全套件

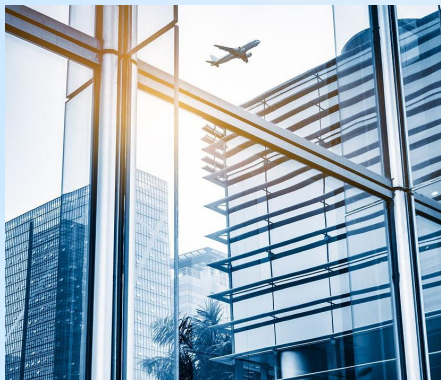
包含网络安全套件，保障数据传输和存储过程中的安全，防止数据泄露和被篡改。

数据标准化处理框架

提供数据质控标准化处理框架，对存储的数据进行质控和标准化处理，提高数据的质量和可用性。



服务特点



定制化交互调用页面

提供定制化的交互调用页面，方便医院根据自身需求进行数据查询和操作，提高工作效率。



VPN专线安全访问

通过VPN专线安全访问，确保数据访问的安全性和稳定性，保障医院的业务正常开展。



服务优势

云存储服务的提供，为医院解决了医疗数据存储和安全性问题，同时通过数据质控标准化处理框架，提高了数据的质量和可用性，为医疗数据的分析和应用提供了有力支持。



医数云析
MEDICAL DATA

08

体检报告AI分析系统

系统功能

传统体检报告升级

在传统的体检报告中加入AI分析模块，对体检结果进行解读和分析，为用户提供更全面、更深入的健康信息。

疾病风险预测

基于AI技术，对体检数据进行分析，预测潜在的疾病风险，提前发现健康问题，为用户提供预防和干预建议。

老年查体版特色

主推老年查体版，针对老年人的健康特点和需求，提供更精准的健康分析和疾病风险预测，为老年健康保驾护航。



体检报告分析系统

为您提供全面的健康数据分析与建议

编号: 15042712
最新报告年份: 2024

性别: 男; 年龄: 69; 身高: 174; 体重: 68; 体质指数: 22.46; 腰围: 74;

数据分析报告

各代谢指标检测结果

1 血液学检测项目

项目名称	结果	操作
白细胞数 (WBC)	5.53	详情
红细胞数 (RBC)		详情
血红蛋白 (HBG)	147	详情
淋巴细胞 (LYMPH)	29.90	详情
中间群细胞 (MID)	11.0	详情
血小板 (PLT)	212	详情
中性粒细胞		

AI智能分析

基于您的体检数据, AI将提供专业的健康分析和建议。

生成AI分析报告

AI分析结果

AI提示词为:

性别: 男; 年龄: 69; 身高: 174; 体重: 68; 体质指数: 22.46; 腰围: 74; HBG=147; PLT=212; WBC=5.53; NEUT=59.10; LYMPH=29.90; MID=11.0; uALB= (-) ; uGLU= (-) ; uKET= (-) ; BLD= (-) ; GLU=7.42; ALT=26.33; AST=26.94; TBIL=17.81; CREA=82.39; UREA=5.07; TG=1.58。以上是部分体检报告信息, 请给出总结分析、总体健康状态评估、健康指导, 并结合此年龄段人群的总体健康状态, 给出'优秀、良好、尚可、较差'的定级。

AI返回结果为:

好的, 我现在需要处理用户提供的体检报告数据, 并给出总结分析、健康评估和指导建议。首先, 用户是69岁的男性, 身高174cm, 体重68kg, BMI是22.46, 这个在正常范围内, 所以体重管理应该没问题。接下来查看各项指标。

系统应用

社区医院引流

01

为社区医院提供有力的引流项目，吸引更多老年人进行体检，提升社区医院的业务量和影响力。

老年产品植入

02

作为老年大赛道产品的植入项目，带动相关老年健康产品的推广和应用，促进老年健康产业发展。

系统价值

03

体检报告AI分析系统的应用，不仅提升了体检报告的价值和实用性，还为老年健康管理和疾病预防提供了有力支持，具有重要的社会和经济价值。

THANK YOU

感谢大家观看

汇报人：医数云析_Julianne