



(12)发明专利申请



(10)申请公布号 CN 110299205 A

(43)申请公布日 2019.10.01

(21)申请号 201910666786.3

(22)申请日 2019.07.23

(71)申请人 上海图灵医疗科技有限公司

地址 200241 上海市闵行区东川路555号丙
楼1129室

(72)发明人 徐赤坤 李伟

(74)专利代理机构 大连智高专利事务所(特殊
普通合伙) 21235

代理人 刘斌

(51)Int.Cl.

G16H 50/30(2018.01)

G16H 50/70(2018.01)

G06K 9/62(2006.01)

A61B 5/00(2006.01)

A61B 5/04(2006.01)

权利要求书4页 说明书16页

(54)发明名称

基于人工智能的生物医学信号特征处理及
评测方法、装置及应用

(57)摘要

基于人工智能的生物医学信号特征处理及
评测方法、装置及应用,属于诊断目的测量领
域,为了解决能对反映心脏疾病的数据挖掘,快
速反映采集数据与相关心脏疾病的关联性的问
题,采集人体健康状况分类数据;得到人体健康
状况分类数据的数据特征;计算人体健康状况分
类数据的数据特征的评估值,效果是实现了反
映心脏疾病的数据挖掘、匹配。

1. 一种基于人工智能的生物医学信号特征处理方法,其特征在于:包括采集人体健康状况分类数据;计算人体健康状况分类数据的数据特征的评估值。

2. 根据权利要求1所述的基于人工智能的生物医学信号特征处理方法,其特征在于,所述的人体健康状况分类数据的数据特征的评估值的计算方式如下:

$$Q = x \times \text{数据变量}$$

数据变量反映人体健康状况分类数据的数据特征,其是心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值中的一种;

其中: x 为权值,取值是在-100至100之间的实数,且 x 不为0。

3. 根据权利要求1所述的基于人工智能的生物医学信号特征处理方法,其特征在于,人体健康状况分类数据包括:心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值中的一种以上;

所述的人体健康状况分类数据的数据特征的评估值的计算方式如下: $Q = a \times \text{心电数据的异质度特征的量化数值} - b \times \text{心电向量数据的异质度特征的量化数值} + c \times \text{心电动力学数据的异质度特征的量化数值} - d \times \text{生物化学数据值} + e \times \text{人体生理信息数据值} - f \times \text{临床信息数据值}$;

其中: a 、 b 、 c 、 d 、 e 、 f 为权值,取值是在-100至100之间的实数。

4. 根据权利要求1所述的基于人工智能的生物医学信号特征处理方法,其特征在于,所述的人体健康状况分类数据的数据特征的评估值 Q 为: $Q = \frac{\sum_{i=1}^m \exp(x_i)}{m}$, $i = 1, 2, 3 \dots k$, k 的取值为不小于0的正整数, $\exp()$ 表示以自然基数 e 为底的指数函数, x_i 取自心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值、生物化学数据值、人体生理信息数据值和临床信息数据值中的一种, m 的值取自16~17之间的数值。

5. 一种人工智能的生物医学信号特征评测方法,其特征在于,包括以下步骤:

步骤一、构建人体健康状况分类数据集:对已知身体健康状况的人群,采集各个体的人体健康状况分类数据,并对采集数据依据健康或反应某种疾病给予相应标签,计算得到各个体的人体健康状况分类数据对应的人体健康状况的特征评估值;

步骤二、采集待测人员的人体健康状况分类数据,并计算待测人员的人体健康状况的特征评估值;

步骤三、将待测人员的人体健康状况的特征评估值,于人体健康状况分类数据集中进行特征评估值比对,找到与待测人员的特征评估值最相近或较相近的已知身体健康状况的个体样本,并以其身体健康状况作为待测人员健康状况参考;

其中:

所述的人体健康状况分类数据的数据特征的评估值的计算方式如下:

$$Q = x \times \text{数据变量}$$

数据变量反映人体健康状况分类数据的数据特征,其是心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值

中的一种;其中: x 为权值,取值是在-100至100之间的实数,且 x 不为0。

或者

人体健康状况分类数据包括:心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值、生物化学数据值、人体生理信息数据值、临床信息数据值中的一种以上;

所述的人体健康状况分类数据的数据特征的评估值的计算方式如下: $Q=a \times$ 心电数据的异质度特征的量化数值 $-b \times$ 心电向量数据的异质度特征的量化数值 $+c \times$ 心电动力学数据的异质度特征的量化数值 $-d \times$ 生物化学数据值 $+e \times$ 人体生理信息数据值 $-f \times$ 临床信息数据值;

其中: a 、 b 、 c 、 d 、 e 、 f 为权值,取值是在-100至100之间的实数。

6.如权利要求5所述的人工智能的生物医学信号特征评测方法,其特征在于,还包括:

步骤四、筛选与待测人员健康状况的特征评估值较相似的前 N 个样本数据集,相似度从高到低依次为: S_1 、 S_2 、 $S_3 \cdots S_N$,并得到 N 个样本数据的特征评估值: Q_1 、 Q_2 、 $Q_3 \cdots Q_N$,根据预设的回归模型计算待测人员的人体健康评估值 V_2 。

7.如权利要求6所述的人工智能的生物医学信号特征评测方法,其特征在于,所述的预设的回归模型包括线性回归模型、非线性回归模型或逻辑回归模型。

8.如权利要求7所述的人工智能的生物医学信号特征评测方法,其特征在于,所述的预设的逻辑回归模型,表述为: $\beta_0 + \beta_1 \times S_1 + \beta_2 \times S_2 + \beta_3 \times S_3 \cdots + \beta_N \times S_N$,其中, β_0 、 β_1 、 β_2 、 $\beta_3 \cdots \beta_N$ 的数值是在-1至1之间的实数,找到与待测人员的特征评估值一致的已知身体健康状况的个体样本,并以其身体健康状况作为待测人员健康状况参考;或者

所述的预设的逻辑回归模型,表述为: $\beta_0 + \beta_1 \times Q_1 + \beta_2 \times Q_2 + \beta_3 \times Q_3 \cdots + \beta_N \times Q_N$,其中, β_0 、 β_1 、 β_2 、 $\beta_3 \cdots \beta_N$ 的数值是在-1至1之间的实数,找到与待测人员的特征评估值一致的已知身体健康状况的个体样本,并以其身体健康状况作为待测人员健康状况参考。

9.如权利要求5至8任一权利要求所述的人工智能的生物医学信号特征评测方法,其特征在于,通过机器学习算法进行所述特征评估值比对。

10.如权利要求9所述的人工智能的生物医学信号特征评测方法,其特征在于,所述的机器学习算法是采用下述四种算法中的一种:

算法一:

用户的待测数据和人体健康状况分类数据集进行机器学习算法比对方法的操作步骤为:采集待测人员的健康状况数据($x_1(t)$, $x_2(t)$, $x_3(t)$, $\cdots$, $x_n(t)$)和健康状况数据集中的($y_1(t)$, $y_2(t)$, $y_3(t)$, $\cdots$, $y_n(t)$);所采集的心电数据都有 F 个周期;

$$S = \left(\sum_{k=1}^n |x_k(t) - y_k(t)|^p \right)^{\frac{1}{p}};$$

其中, $x_k(t)$ 为待测人员心电数据第 k 个向量的分量, $y_k(t)$ 为健康状况数据集中目标心电数据的第 k 个向量的分量; p 为非负实数, n 为健康状况数据的维度;

数据比对结果 S 的判别标准:0表示它们的指向是完全相同的,数值越大表示它们之间越独立;

算法二:

1)建立人体健康状况分类数据库,包括健康人员的健康状况数据库,和非健康人员的

健康状况数据库;

2) 将人体健康状况分类数据库的每个数据进行归一化处理,将处理完的数据张量输入到深度卷积神经网络加全连接层中提取特征;

3) 通过球面损失函数,利用梯度下降法寻找全连接层中的权重值和深度卷积神经网络中各个滤波器的权重值,在训练过程中,同时用健康人员的健康状况数据和非健康人员的健康状况数据验证数据库的统计结果;

4) 训练完成后,根据训练中所统计的数据,选择比对精度最高的深度卷积神经网络模型作为第一阶段模型;

5) 通过迁移学习复用第一阶段深度卷积神经网络模型以及第一阶段训练所得的所有权重值,通过二范数归一化球面损失函数,利用梯度下降法寻找全连接层中的权重值和深度卷积神经网络中各个滤波器的权重值,在训练过程中,同时全程记录健康人员的健康状况数据和非健康人员的健康状况数据验证数据库的统计结果;

6) 训练完成后,根据训练中所统计的数据,选择比对精度最高的深度卷积神经网络模型作为最终模型;

7) 将最终模型导入到将部署的目标嵌入式平台的测试程序中,将待测人员的人体健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 输入程序,通过模型计算特征向量,将特征向量与数据库中的每一人体健康状况分类数据 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$ 的特征向量进行余弦距离的计算,当余弦值小于阈值判断为同一种疾病;

算法三:

采集待测人员的健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 和人体健康状况分类数据集中的 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$, 然后进行比对,比对方法采用比对取值的方式进行:比对取值为将 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 变为 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$ 的最小替换次数,两数组比对的判别标准:0表示它们的指向是完全相同的,数值越大表示它们之间越独立;

算法四:

相似性比对方法A的操作步骤为:采集待测人员的心电数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 和健康状况数据集中的 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$;所采集的心电数据都有F个周期;

$$S_f = \frac{|n \sum_{k=1}^n x_k(t) \times y_k(t) - \sum_{k=1}^n x_k(t) \times \sum_{k=1}^n y_k(t)|}{\sqrt{n \sum_{k=1}^n x_k^2(t) - (\sum_{k=1}^n x_k(t))^2} \times \sqrt{n \sum_{k=1}^n y_k^2(t) - (\sum_{k=1}^n y_k(t))^2}};$$

$$S_s = \frac{1}{n} \sum_{k=1}^n (x_k(t) - y_k(t))^2;$$

$$S = a \times S_f + b \times S_s + T;$$

其中, $x_k(t)$ 为待测人员心电数据第k个向量的分量, $y_k(t)$ 为健康状况数据集中目标心电数据的第k个向量的分量, a和b取值为-1~1之间适应不同种心脏疾病的选择值, T的取值包括实数, S_f 为相对相似度、 S_s 为绝对相似度、a和b为适应不同种心脏疾病的权值系数, T为适应不同种心脏疾病的数据增强系数;

相似度S的判别标准:相似性范围从-1到1, -1意味着两个向量指向的方向正好截然相反, 1表示它们的指向是完全相同的, 0通常表示它们之间是独立的, 而在这之间的值则表

示中间的相似性。

11. 一种人工智能的生物医学信号特征评测方法,其特征在于:包括以下步骤:

步骤一、构建人体健康状况单一特征评估数据与人体健康状况记录数据之间相互对应的人体健康状况数据库,所述的人体健康状况单一特征评估数据取自心电向量数据的异质度特征的量化数据和心电动力学数据的异质度特征的量化数据中的一种,所述的人体健康状况记录数据包括反映人体健康状况的数据或反映人体疾病严重程度的数据;

步骤二、输入待测用户的单一特征评估数据到所述的人体健康状况数据库,使用机器学习算法得到与待测用户的单一特征评估数据相似度至少为S%的前M个数据库数据,各数据库中为疾病名称下对应的健康人的相应数据的范围以及患该疾病的病人的相应数据的范围,所述的S不小于10,所述的M为不小于 $\text{INT}(k+q \times m)$ 的数值,所述的INT函数表示取整,所述的 $k \geq 1$,所述的q值为进行过归一化处理的待测用户的单一特征评估数据的量化数值,m的数值是在-10至10之间的实数;

步骤三、将得到的M个数据库数据的信息输入到预设的人体健康评估模型,得到待测人员的M个特征评估值,与M个特征评估值一致的已知身体健康状况的个体样本的身体健康状况作为待测人员健康状况参考;

所述的人体健康评估模型:

$Q = a \times \text{心电向量数据的异质度特征的量化数值} + c \times \text{心电动力学数据的异质度特征的量化数值}$;其中:a、c为权值,取值是在-100至100之间的实数,a、c一个不为0时,另一个必为0。

12. 一种人工智能的生物医学信号特征评测方法,其特征在于,包括以下步骤:

步骤一、构建人体健康状况数据集,所述的人体健康状况数据集选自异质度特征的量化数据、心电向量数据和心电动力学数据中的一种;

步骤二、采集待测人员健康状况数据,输入到所述的人体健康状况数据集中,通过机器学习算法进行比对,获得与待测人员健康状况数据值相似度至少为S%的前k个数据,所述的待测人员健康状况数据选自异质度特征的量化数据、心电向量数据和心电动力学数据中的一种,所述的S不小于10,所述的 $k \geq 1$;

步骤三、依据所述的前k个数据的信息,输出待测人员健康状况的健康评估值,与待测人员健康状况数据值相似的前k个数据对应的已知身体健康状况的个体样本的身体健康状况作为待测人员健康状况参考。

13. 一种人体健康检测设备,其特征在于,存储有多条指令,所述指令适于处理器加载并执行:权利要求1至12任一权项所述的方法。

基于人工智能的生物医学信号特征处理及评测方法、装置及应用

技术领域

[0001] 本发明属于诊断目的测量领域,特别是涉及一种基于人工智能的生物医学信号特征处理方法、装置及应用。

背景技术

[0002] 随着人体健康智能终端技术的发展,各种智能终端技术开发产品推向了国内外人体健康状态监测设备市场,丰富了健康智能终端设备的健康监测性能,然而当前的健康智能终端设备采集的数据依然是单一或复合的病理数据进行人体健康状况的评测,无法根据用户不同时间段的人体重要的生理特征参数提供高效的病情早期预测和疾病早期筛查及评估,导致后期用户多易罹患如心肌梗死或脑梗等多种紧急突发疾病。因此,开发一种基于人工智能的生物医学信号特征处理方法,充分评估人体疾病患者的治疗前、治疗后及预后等各个阶段人体健康功状态显得十分重要。

发明内容

[0003] 为了解决能对反应心脏疾病的数据挖掘,快速反应采集数据与相关心脏疾病的关联性的问题,本发明提出如下技术方案:一种基于人工智能的生物医学信号特征处理方法,包括

[0004] 采集人体健康状况分类数据;

[0005] 计算人体健康状况分类数据的数据特征的评估值。

[0006] 进一步的,所述的人体健康状况分类数据的数据特征的评估值的计算方式如下:

[0007] $Q = x \times \text{数据变量}$

[0008] 数据变量反映人体健康状况分类数据的数据特征,其是心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值中的一种;

[0009] 其中: x 为权值,取值是在-100至100之间的实数,且 x 不为0。

[0010] 该方案的效果是:对于数据变量或该同类数据下的小类数据进行放大,统一量级,从而在后续数据分析中,能够将差异对比更为明显的表现,使得后续数据处理准确性更高,且处理速度更快,而选择的数据中,心电向量数据能对心电数据的信息总结,心电动力学数据能对心电数据的信息放大,能够对于心脏疾病具有信息表达,且较为准确。

[0011] 进一步的,人体健康状况分类数据包括:心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值中的一种以上;

[0012] 所述的人体健康状况分类数据的数据特征的评估值的计算方式如下: $Q = a \times \text{心电数据的异质度特征的量化数值} - b \times \text{心电向量数据的异质度特征的量化数值} + c \times \text{心电动力学数据的异质度特征的量化数值} - d \times \text{生物化学数据值} + e \times \text{人体生理信息数据值} - f \times \text{临}$

床信息数据值；

[0013] 其中： a 、 b 、 c 、 d 、 e 、 f 为权值，取值是在-100至100之间的实数。

[0014] 该方案的效果是：上述数据变量的选择和组合，是为了能够更为准确和全面的反应心脏疾病信息，得到准确的对比结果，特别是在应用机器学习算法中，对于不同的心脏疾病，能够反映其信息的主要数据变量具有差异，通过权值的调整，能够对于不同心脏疾病的选择数据占比进行调整，从而在机器学习算法中，对于多种疾病在作出参考数据的选择时能即时调整，准确性得以提高，并且，对于数据变量或该同类数据下的小类数据进行放大，统一量级，从而在后续数据分析中，能够将差异对比更为明显的表现，使得后续数据处理准确性更高，且处理速度更快，而选择的数据中，心电向量数据能对心电数据的信息总结，心电动力学数据能对心电数据的信息放大，能够对于心脏疾病具有信息表达，且较为准确。

[0015] 进一步的，所述的人体健康状况分类数据的数据特征的评估值 Q 为： $Q = \frac{\sum_{i=1}^m \exp(x_i)}{m}$ ， $i = 1, 2, 3 \cdots k$ ， k 的取值为不小于0的正整数， $\exp()$ 表示以自然基数 e 为底的指数函数， x_i 取自心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值、生物化学数据值、人体生理信息数据值和临床信息数据值中的一种， m 的值取自16~17之间的数值。该方案的效果同上。

[0016] 本发明还涉及一种人工智能的生物医学信号特征评测方法，包括以下步骤：

[0017] 步骤一、构建人体健康状况分类数据集：对已知身体健康状况的人群，采集各个体的人体健康状况分类数据，并对采集数据依据健康或反应某种疾病给予相应标签，计算得到各个体的人体健康状况分类数据对应的人体健康状况的特征评估值，进行数据存储；

[0018] 步骤二、采集待测人员的人体健康状况分类数据，并计算待测人员的人体健康状况的特征评估值；

[0019] 步骤三、将待测人员的人体健康状况的特征评估值，于人体健康状况分类数据集中进行特征评估值比对，找到与待测人员的特征评估值最相近或较相近的已知身体健康状况的个体样本，并以其身体健康状况作为待测人员健康状况参考；

[0020] 其中：

[0021] 所述的人体健康状况分类数据的数据特征的评估值的计算方式如下：

[0022] $Q = x \times \text{数据变量}$

[0023] 数据变量反映人体健康状况分类数据的数据特征，其是心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值中的一种；

[0024] 其中： x 为权值，取值是在-100至100之间的实数，且 x 不为0。

[0025] 或者

[0026] 人体健康状况分类数据包括：心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值、生物化学数据值、人体生理信息数据值、临床信息数据值中的一种以上；

[0027] 所述的人体健康状况分类数据的数据特征的评估值的计算方式如下： $Q = a \times \text{心电数据的异质度特征的量化数值} - b \times \text{心电向量数据的异质度特征的量化数值} + c \times \text{心电动力学数据的异质度特征的量化数值} - d \times \text{生物化学数据值} + e \times \text{人体生理信息数据值} - f \times \text{临床}$

床信息数据值；

[0028] 其中： a 、 b 、 c 、 d 、 e 、 f 为权值，取值是在-100至100之间的实数。

[0029] 该方法的效果是，能够准确的以数据对比方式，得到待测人员的的心脏信息。

[0030] 所述的人工智能的生物医学信号特征评测方法，还包括：

[0031] 步骤四、筛选与待测人员健康状况的特征评估值较相似的前 N 个样本数据集，相似度从高到低依次为： S_1 、 S_2 、 S_3 ... S_N ，并得到 N 个样本数据的特征评估值： Q_1 、 Q_2 、 Q_3 ... Q_N ，根据预设的回归模型计算待测人员的人体健康评估值 V_2 。

[0032] 该方法的效果是：选择一组均为相似的数据，再以回归模型计算，并不直接对应一个人的心脏信息，而是通过回归模型计算，将一组相似人员的信息进行综合考虑，得到其更为适合的心脏信息。

[0033] 进一步的，所述的预设的回归模型包括线性回归模型、非线性回归模型或逻辑回归模型。

[0034] 进一步的，所述的预设的逻辑回归模型，表述为： $\beta_0 + \beta_1 \times S_1 + \beta_2 \times S_2 + \beta_3 \times S_3 \cdots + \beta_N \times S_N$ ，其中， β_0 、 β_1 、 β_2 、 β_3 ... β_N 的数值是在-1至1之间的实数，找到与待测人员的特征评估值一致的已知身体健康状况的个体样本，并以其身体健康状况作为待测人员健康状况参考；或者

[0035] 所述的预设的逻辑回归模型，表述为： $\beta_0 + \beta_1 \times Q_1 + \beta_2 \times Q_2 + \beta_3 \times Q_3 \cdots + \beta_N \times Q_N$ ，其中， β_0 、 β_1 、 β_2 、 β_3 ... β_N 的数值是在-1至1之间的实数，找到与待测人员的特征评估值一致的已知身体健康状况的个体样本，并以其身体健康状况作为待测人员健康状况参考。

[0036] 进一步的，其特征在于，通过机器学习算法进行所述特征评估值比对。

[0037] 进一步的，所述的机器学习算法是采用下述四种算法中的一种：

[0038] 算法一：

[0039] 用户的待测数据和人体健康状况分类数据集进行机器学习算法比对方法的操作步骤为：采集待测人员的健康状况数据($x_1(t)$ ， $x_2(t)$ ， $x_3(t)$ ， $\cdots$ ， $x_n(t)$)和健康状况数据集的($y_1(t)$ ， $y_2(t)$ ， $y_3(t)$ ， $\cdots$ ， $y_n(t)$)；所采集的心电数据都有 F 个周期；

$$[0040] \quad S = \left(\sum_{k=1}^n |x_k(t) - y_k(t)|^p \right)^{\frac{1}{p}};$$

[0041] 其中， $x_k(t)$ 为待测人员心电数据第 k 个向量的分量， $y_k(t)$ 为健康状况数据集中目标心电数据的第 k 个向量的分量； p 为非负实数， n 为健康状况数据的维度；

[0042] 数据比对结果 S 的判别标准： 0 表示它们的指向是完全相同的，数值越大表示它们之间越独立；

[0043] 算法二：

[0044] 1) 建立人体健康状况分类数据库，包括健康人员的健康状况数据库，和非健康人员的健康状况数据库；

[0045] 2) 将人体健康状况分类数据库的每个数据进行归一化处理，将处理完的数据张量输入到深度卷积神经网络加全连接层中提取特征；

[0046] 3) 通过球面损失函数，利用梯度下降法寻找全连接层中的权重值和深度卷积神经网络中各个滤波器的权重值，在训练过程中，同时用健康人员的健康状况数据和非健康人员的健康状况数据验证数据库的统计结果；

[0047] 4) 训练完成后，根据训练中所统计的数据，选择比对精度最高的深度卷积神经网络

络模型作为第一阶段模型；

[0048] 5) 通过迁移学习复用第一阶段深度卷积神经网络模型以及第一阶段训练所得的所有权重值,通过二范数归一化球面损失函数,利用梯度下降法寻找全连接层中的权重值和深度卷积神经网络中各个滤波器的权重值,在训练过程中,同时全程记录健康人员的健康状况数据和非健康人员的健康状况数据验证数据库的统计结果;

[0049] 6) 训练完成后,根据训练中所统计的数据,选择比对精度最高的深度卷积神经网络模型作为最终模型;

[0050] 7) 将最终模型导入到将部署的目标嵌入式平台的测试程序中,将待测人员的人体健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 输入程序,通过模型计算特征向量,将特征向量与数据库中的每一人体健康状况分类数据 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$ 的特征向量进行余弦距离的计算,当余弦值小于阈值判断为同一种疾病;

[0051] 算法三:

[0052] 采集待测人员的健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 和人体健康状况分类数据集中的 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$, 然后进行比对,比对方法采用比对取值的方式进行:比对取值为将 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 变为 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$ 的最小替换次数,两数组比对的判别标准:0表示它们的指向是完全相同的,数值越大表示它们之间越独立;

[0053] 算法四:

[0054] 相似性比对方法A的操作步骤为:采集待测人员的心电数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 和健康状况数据集中的 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$;所采集的心电数据都有F个周期;

$$[0055] \quad S_f = \frac{|n \sum_{k=1}^n x_k(t) \times y_k(t) - \sum_{k=1}^n x_k(t) \times \sum_{k=1}^n y_k(t)|}{\sqrt{n \sum_{k=1}^n x_k^2(t) - (\sum_{k=1}^n x_k(t))^2} \times \sqrt{n \sum_{k=1}^n y_k^2(t) - (\sum_{k=1}^n y_k(t))^2}};$$

$$[0056] \quad S_s = \frac{1}{n} \sum_{k=1}^n (x_k(t) - y_k(t))^2;$$

$$[0057] \quad S = a \times S_f + b \times S_s + T;$$

[0058] 其中, $x_k(t)$ 为待测人员心电数据第k个向量的分量, $y_k(t)$ 为健康状况数据集中目标心电数据的第k个向量的分量, a和b取值为-1~1之间适应不同种心脏疾病的选择值, T的取值包括实数, S_f 为相对相似度、 S_s 为绝对相似度、a和b为适应不同种心脏疾病的权值系数, T为适应不同种心脏疾病的数据增强系数;

[0059] 相似度S的判别标准:相似性范围从-1到1, -1意味着两个向量指向的方向正好截然相反, 1表示它们的指向是完全相同的, 0通常表示它们之间是独立的, 而在这之间的值则表示中间的相似性。

[0060] 本发明还涉及一种人工智能的生物医学信号特征评测方法, 包括以下步骤:

[0061] 步骤一、构建人体健康状况单一特征评估数据与人体健康状况记录数据之间相对应的人体健康状况数据库, 所述的人体健康状况单一特征评估数据取自心电向量数据的异质度特征的量化数据和心电动力学数据的异质度特征的量化数据中的一种, 所述的人体健康状况记录数据包括反映人体健康状况的数据或反映人体疾病严重程度的数据;

[0062] 步骤二、输入待测用户的单一特征评估数据到所述的人体健康状况数据库, 使用

机器学习算法得到与待测用户的单一特征评估数据相似度至少为S%的前M个数据库数据,各数据库中为疾病名称下对应的健康人的相应数据的范围以及患该疾病的病人的相应数据的范围,所述的S不小于10,所述的M为不小于 $\text{INT}(k+q \times m)$ 的数值,所述的INT函数表示取整,所述的 $k \geq 1$,所述的q值为进行过归一化处理的待测用户的单一特征评估数据的量化数值, m的数值是在-10至10之间的实数;

[0063] 步骤三、将得到的M个数据库数据的信息输入到预设的人体健康评估模型,得到待测人员的M个特征评估值,与M个特征评估值一致的已知身体健康状况的个体样本的身体健康状况作为待测人员健康状况参考;

[0064] 所述的人体健康评估模型:

[0065] $Q = a \times \text{心电向量数据的异质度特征的量化数值} + c \times \text{心电动力学数据的异质度特征的量化数值}$;

[0066] 其中:a、c为权值,取值是在-100至100之间的实数,a、c一个不为0时,另一个必为0。

[0067] 本发明还涉及一种人工智能的生物医学信号特征评测方法,包括以下步骤:

[0068] 步骤一、构建人体健康状况数据集,所述的人体健康状况数据集选自异质度特征的量化数据、心电向量数据和心电动力学数据中的一种;

[0069] 步骤二、采集待测人员健康状况数据,输入到所述的人体健康状况数据集中,通过机器学习算法进行比对,获得与待测人员健康状况数据值相似度至少为S%的前k个数据,所述的待测人员健康状况数据选自异质度特征的量化数据、心电向量数据和心电动力学数据中的一种,所述的S不小于10,所述的 $k \geq 1$;

[0070] 步骤三、依据所述的前k个数据的信息,输出待测人员健康状况的健康评估值,与待测人员健康状况数据值相似的前k个数据对应的已知身体健康状况的个体样本的身体健康状况作为待测人员健康状况参考。

[0071] 本发明还涉及一种人体健康检测设备,存储有多条指令,所述指令适于处理器加载并执行所述的方法。

具体实施方式

[0072] 下列实施例中未注明具体条件的实验方法,通常按照国家标准测定。若没有相应的国家标准,则按照通用的国际标准、常规条件、或按照制造厂商所建议的条件进行。

[0073] 在本发明中,如果没有特别的说明,本发明所提到的所有实施方式以及优选实施方式可以相互组合形成新的技术方案。

[0074] 在本发明中,如果没有特别的说明,本发明所提到的所有技术特征以及优选特征可以相互组合形成新的技术方案。

[0075] 在本发明中,如果没有特别的说明,本发明所提到的心电数据包括但不限于心电图数据、预处理的心电图数据、截取的心电图数据、截取的心电数据、单导联的心电数据或多导联的心电数据。所述的心电数据包括但不限于特定的波、段或间期的心电数据;例如:截取P波、和/或R波、和/或QRS波群、和/或S波、和/或T波、和/或U波、和/或PR段、和/或ST段、和/或ST-T段、和/或PR间期、和/或ST间期、和/或QT间期等数据后的心电数据。

[0076] 在本发明中,如果没有特别的说明,本发明所提到的心电向量数据包括但不限于

反映心电大小和方向信息的数据。

[0077] 在本发明中,如果没有特别的说明,所述的心电向量数据包括但不限于完整的心电向量数据、预处理的心电向量数据、截取的心电向量数据、VCG数据、预处理的VCG数据或截取的VCG数据,所述的截取的VCG数据包括但不限于:存在VCG数据上任意大小的一部分数据。

[0078] 在本发明中,如果没有特别的说明,本发明所提到的心电向量数据包括但不限于单维度数据或多维度数据;所述单维度的心电向量数据包括但不限于心电向量数据在一个维度空间上投影的数据信息;所述的多维度的心电向量数据包括但不限于心电向量数据在多个维度空间上投影的数据信息。

[0079] 在本发明中,如果没有特别的说明,本发明所提到的心电向量数据包括但不限于将心电数据进行数学运算后得到的向量数据,所述的数学运算包括但不限于三维转换,比如12导联心电数据(或15导联心电数据或18导联心电数据)进行三维转换后得到的向量数据;所述的数学运算包括但不限于:Kors J.A.等在1990年发表在European Heart Journal 杂志的 11(12):1083的论文所述的计算方法。

[0080] 在本发明中,如果没有特别的说明,本发明所提到的心电向量数据是指将心电数据经过转换后获得的单维度数据或多维度数据,所述的转换包括但不限于将12导联(或多导联)心电数据转换为单维度数据或多维度数据。

[0081] 在本发明中,如果没有特别的说明,本发明所提到的心电向量数据包括但不限于:截取 P波、和/或R波、和/或QRS波群、和/或S波、和/或T波、和/或U波、和/或PR段、和/或ST段、和/或ST-T段、和/或PR间期、和/或ST间期、和/或QT间期等数据后的心电向量数据。

[0082] 在本发明中,如果没有特别的说明,本发明所提到的心电动力学数据是指心电数据或心电向量数据通过数学运算后获取的数据。所述的数学运算包括但不限于建模的方法,建模的方法包括但不限于自适应系统系统辨识方法,自适应系统系统辨识方法包括但不限于神经网络方法、调节函数法、最小二乘法、极大似然法、回归方程、或预报误差法等。

[0083] 在本发明中,如果没有特别的说明,本发明所提到的心电动力学数据包括但不限于CDG 数据。

[0084] 在本发明中,如果没有特别的说明,本发明所提到的心电动力学数据包括但不限于完整的心电动力学数据、预处理的心电动力学数据、截取的心电动力学数据、心电非线性系统动态图、预处理的心电非线性系统动态图、截取的心电非线性系统动态图、心电动力学图、预处理的心电动力学图或截取的心电动力学图;预处理的心电动力学图包括但不限于压缩的心电动力学图、改变对比度的心电动力学图、或放大的心电动力学图等;截取的心电动力学图包括但不限于存在于心电动力学图上任意大小和任意形状的一部分。截取的心电动力学数据,包括但不限于:存在于心电动力学数据上任意大小的一部分的数据。

[0085] 在本发明中,如果没有特别的说明,本发明所提到的心电动力学数据包括但不限于:完整的心电动力学数据、截取的心电动力学数据(包括但不限于P波、和/或R波、和/或QRS波群、和/或S波、和/或T波、和/或U波、和/或PR段、和/或ST段、和/或PR间期、和/或ST间期、和/或QT间期等数据信息)、预处理的心电动力学数据(包括但不限于P波、和/或R波、和/或QRS波群、和/或S波、和/或T波、和/或U波、和/或PR段、和/或ST段、和/或PR间期、和/或ST间期、和/或QT间期等。心电动力学数据是指将心电向量数据经过波、段或间期的截取后,然

后通过自适应系统系统辨识方法而获取的单维度数据或多维度数据。截取的心电动力学数据也包括但不限于：存在于心电动力学数据上任意大小的一部分的数据。

[0086] 在本发明中,如果没有特别的说明,本发明所提到的人体健康状况分类数据集(或待测人员的健康状况数据)包括但不限于异质度特征的量化数据、生物化学数据、人体生理信息数据和临床信息数据中的一种或几种。异质度特征的量化数据包括但不限于心电动力学数据的异质度特征的量化数据、心电向量数据的异质度特征的量化数据、心电数据的异质度特征的量化数据、肌电信号数据特征的量化数据、或脑电信号数据特征的量化数据。生物化学数据包括但不限于高敏C反应蛋白、心肌酶、天门冬氨酸氨基转移酶、乳酸脱氢酶、肌酸激酶及同工酶、 α -羟丁酸脱氢酶、肌红蛋白、肌酸激酶同工酶、C反应蛋白、甘三脂、高密度脂蛋白、低密度脂蛋白和乳酸脱氢酶等中的一种或几种的信息。人体生理信息数据包括但不限于性别、年龄、体重、吸烟史、饮酒史、运动锻炼状况、民族和/或地域等;人体生理信息数据也包括但不限于与心脏活动状态有关的病理数据:动态心电检测持续时间、最快(慢)心率、平均心率、总心搏数、室早总数、P波离散度、P波(Q波)时限、Q波深度、QRS 时限、QTc间期、QT/RR斜率、QT间期变异性、T波峰末时间、J波高度、J波离散度、房室传导阻滞情况、心脏病病史、冠心病史、心肌梗死病史、猝死家族史、心功能NYHA分级、高血压病史、瓣膜病病史、先心病病史、糖尿病病史、脑血管病病史、心血管病家族史、遗传病病史、收缩压、舒张压、体重指数、是否安装起搏器、冠脉搭桥状况、冠脉支架状况、ICD治疗状况、导消融状况、B受体阻滞剂使用状况、钙通道拮抗剂使用状况、ACEI/ARB 使用状况、利尿剂使用状况、抗心律失常药物使用史、洋地黄类药物使用史、降脂药使用史、非持续性室速、Lowns分级、、SDNN、SDANN、ASDNN、rMSSD、pNN50、pNN50a、TD、心率减速力、fQRS、RMS40、LAS、微伏级T波电交替、同型半胱氨酸、LDL、HDL、脑型利钠肽、NT-proBNP、肌酐、左房内径、左室舒张末内径、左室后壁厚度、左室间隔厚度、左室射血分数、FS、二尖瓣反流、三尖瓣反流、主动脉反流、阶段性室壁运动异常、LM、LAD、LCX、RCA、TIMI分级和内中膜厚度等;对应心脏疾病的判定标准参考医院诊断记录和公开的医学诊断规则。所述的临床信息数据包括但不限于心动过缓、血压降低、心悸、咽喉痛及烧灼感、咽喉紧缩感、牙痛、倦怠乏力、眩晕、胸部闷痛和/或气短等患者的状况信息,以及患者的病历或EMR(电子病历)、检测检验、影像、诊断、处方、治疗、评估表等数据。

[0087] 在本发明中,如果没有特别的说明,本发明所提到的异质度特征是指能表现出人体健康状况的非线性系统动态数据在空间分布和时间推演过程中的不均匀性和(或)复杂性的指标。

[0088] 在本发明中,如果没有特别的说明,本发明所提到的异质度特征是通过异质度分析的方法从目标数据中提取得到的;所述的目标数据包括但不限于人体生物电信号、心电信号、心电向量数据、心电动力学信号、心电动力学数据、肌电信号数据和脑电信号数据中的一种或几种;所述的异质度分析的方法包括但不限于几何特征分析的方法、模型的方法、非线性动力学分析法、时域分析法、频域分析法、时频分析法、高阶谱分析法或人工神经网络分析法。其中,所述的频域分析法包括但不限于快速傅里叶变换和离散傅里叶变换中的至少一种。所述的非线性动力学方法包括但不限于混沌综合、混沌分析、分形分析或孤子分析等;具体地,所述的非线性动力学方法包括但不限于相平面图、功率谱、庞加莱截面、散点图、复杂度、熵、李雅普诺夫指数谱、最大李雅普诺夫指数谱、分维数、相空间重构、非稳定周期

轨道、符号动力学、自相关函数、混沌学特征、孤立子特征、Hurst指数、单重分形和多重分形中的一种或几种。所述的时域分析方法但不限于均值、标准差、中值、最小值、最大值、最大最小值之差、最小值比率、最大值比率、差值均方的平方根、几何图形分析指标和直方图中的至少一种。所述的几何特征方法但不限于趋势、斜率、方向、形状、圆形度、球状度、均匀率、离心率、变异率和角度中的至少一种。所述的时频分析方法但不限于短时傅里叶变换、连续小波变换、希尔伯特-黄变换、Wigner-Ville分布、分数阶傅里叶变换、S变换和广义S变换中的一种或几种。所述的模型方法但不限于AR模型和TVAR模型中的一种或几种。

[0089] 在本发明中,如果没有特别的说明,本发明所提到的异质度特征包括但不限于人体生物电信号数据特征、心电动力学数据特征、心电向量数据特征、心电信号特征、肌电信号数据特征和脑电信号数据特征中的一种或几种。

[0090] 在本发明中,如果没有特别的说明,本发明所提到的心电信号特征(或肌电信号特征、或脑电信号特征)包括但不限于从心电数据(或肌电信号数据、或脑电信号数据)中提取的几何特征、非线性动力学特征、模型特征、时域特征、频域特征和时频特征等中的一种或几种。

[0091] 在本发明中,如果没有特别的说明,本发明所提到的心电向量数据特征包括但不限于从心电向量数据中提取的几何特征、非线性动力学特征、模型特征、时域特征、频域特征和时频特征等中的一种或几种。所述的几何特征包括但不限于趋势、斜率、方向、形状、圆形度、球状度、均匀率、离心率、变异率和角度中的一种或几种;所述的非线性动力学特征包括但不限于熵、复杂度、关联维、李雅普诺夫指数谱和最大李雅普诺夫指数谱中的一种或几种;所述的熵优选为信息熵、小波熵和近似熵中的一种或几种;所述的复杂度优选为 C_0 复杂度、Kolmogorov复杂度和LZ复杂度中的一种或几种;所述的模型特征包括但不限于AR模型系数和TVAR模型特征中的一种或几种;所述的时域特征包括但不限于均值和直方图等中的一种或几种;所述的时频特征包括但不限于短时傅里叶变换特征、小波变换特征、或二者的组合;所述的频域特征包括但不限于快速傅里叶变换特征。

[0092] 在本发明中,如果没有特别的说明,本发明所提到的心电动力学数据特征包括但不限于从心电动力学数据中提取的几何特征、非线性动力学特征、模型特征、时域特征、频域特征和时频特征等中的一种或几种。所述的几何特征包括但不限于趋势、斜率、方向、形状、圆形度、球状度、均匀率、离心率、变异率和角度中的一种或几种;所述的非线性动力学特征包括但不限于熵、复杂度、关联维、李雅普诺夫指数谱和最大李雅普诺夫指数谱中的一种或几种;所述的熵优选为信息熵、小波熵和近似熵中的一种或几种;所述的复杂度优选为 C_0 复杂度、Kolmogorov复杂度和LZ复杂度中的一种或几种;所述的模型特征包括但不限于AR模型系数和TVAR模型特征中的一种或几种;所述的时域特征包括但不限于均值和直方图等中的一种或几种;所述的时频特征包括但不限于短时傅里叶变换特征、小波变换特征、或二者的组合;所述的频域特征包括但不限于快速傅里叶变换特征。

[0093] 在本发明中,如果没有特别的说明,本发明所提到的心电动力学数据特征包括但不限于心电动力学数据内在的心脏静态病理特征、或心电动力学数据内在的心脏动态病理特征。

[0094] 在本发明中,如果没有特别的说明,本发明所提到的待测人员包括但不限于医院待测人员、体检人员和人体疾病患者等。

[0095] 在本发明中,如果没有特别的说明,本发明所提到的人体疾病包括糖尿病、心肌缺血、心肌梗塞、心源性猝死、心肌炎、心律失常、心肌梗死、心绞痛、冠心病、急性冠脉综合征、高血压、高胆固醇、睡眠呼吸暂停、冠状动脉疾病或心力衰竭。

[0096] 在本发明中,如果没有特别的说明,本发明所提到的相似性比对是指评定两个事物之间相近程度的一种度量。

[0097] 在本发明中,如果没有特别的说明,本发明所提到的相似性比对算法包括但不限于明氏距离、马氏距离、曼哈顿距离、切比雪夫距离、相关系数、欧氏距离和余弦相似度等方法中的一种或几种。

[0098] 在本发明中,如果没有特别的说明,本发明所提到的心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值、生物化学数据值、人体生理信息数据值或临床信息数据值在进行人体健康状况的特征评估值计算前,都需进行标准化处理,标准化处理方法包括归一化处理的方法。

[0099] 以下结合具体实施例,对本发明作进一步说明。应理解,以下优选的实施例仅用于说明本发明而非用于限定本发明的范围。

[0100] 实施例1. 人体健康状况分类数据集的构建

[0101] 本实施例主要解决对人体健康状况分类数据集的构建的问题,所采用的人体健康状况分类数据集的构建方法为:纳入N个临床已知身体健康的个体($N > 5000$)以及M个临床已知某种疾病的个体($M > 10000$)作为测试样本人群,所述的人体健康状况分类数据集包括采集的健康人员的心电数据、心电向量数据、心电动力学数据、心电动力学数据特征的量化数据、肌电信号数据特征的量化数据、或脑电信号数据特征的量化数据,并同时采集非健康人员的心电动力学数据特征的量化数据、肌电信号数据特征的量化数据、或脑电信号数据特征的量化数据;在此,以 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$ 表示人体健康状况分类数据集。

[0102] 心电动力学数据的获取:采集带标签的心脏电信号数据 $e(t)$, $t = 1, 2, \dots, T$;将采集的心脏电信号数据进行滤波和基线漂移等预处理,然后将心脏电信号数据或心电向量数据采用数学运算方法获得心脏电信号数据或心电向量数据的神经网络模型,进行计算得到心电动力学数据,所述心电动力学数据可表述为 $[\bar{W}_1^T S(X(t)), \bar{W}_2^T S(Y(t)), \bar{W}_3^T S(Z(t))]$;其中, $S(X(t))$ 、 $S(Y(t))$ 和 $S(Z(t))$ 都是高斯径向基函数; $\bar{W}_1$ 、 $\bar{W}_2$ 和 $\bar{W}_3$ 是常值神经网络权值向量。在本实施例中,所述数学运算方法包括但不限于自适应系统辨识方法;所述自适应系统辨识方法采用包括但不限于建模方法,所述建模方法包括但不限于自适应系统辨识方法建模,所述的自适应系统辨识方法包括但不限于神经网络方法、调节函数法、最小二乘法、极大似然法、回归方程、或预报误差法等。

[0103] 在本实施例中,所述的人体健康状况分类数据集的标签的设定为:采用人体疾病的金标准指标的指标数据以及专家共识作为人体健康状况分类数据集的标签。其中标签是机器学习算法中对数据进行标记的内容,在数据采集过程中就会对数据进行注释,表明数据的含义。

[0104] 在本实施例中,所述的心电动力学数据特征包括但不限于通过异质度分析的方法从心电动力学数据中提取的几何特征、和/或非线性动力学特征、和/或模型特征、和/或时域特征、和/或频域特征、和/或时频特征。

[0105] 在本实施例中,人体生理信息数据根据无、轻度、重度三级量化,分别赋值0、1、2。临床信息数据包括但不限于心动过缓、血压降低、心悸、咽喉痛及烧灼感、咽喉紧缩感、牙痛、倦怠乏力、眩晕、胸部闷痛和/或气短等患者的状况信息,以及患者的病历或EMR(电子病历)、检测检验、影像、诊断、处方、治疗、评估表等数据。在本实施例中,对于临床信息数据,将按照无、轻度、中度、重度4级量化,分别赋值0、1、2、3;嗜睡和口黏等也可进行二类判别,根据有无分别赋值1和0。

[0106] 在一种实施例中,本公开给出了一种基于人工智能的生物医学信号特征处理方法,包括:

[0107] 采集人体健康状况分类数据;提取人体健康状况分类数据的数据特征;计算人体健康状况分类数据的数据特征的评估值。

[0108] 所述的人体健康状况分类数据的数据特征的评估值的计算方式如下:

[0109] $Q = x \times \text{数据变量}$

[0110] 数据变量反映人体健康状况分类数据的数据特征,其是心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值中的一种;

[0111] 其中: x 为权值,取值是在-100至100之间的实数,且 x 不为0;更为优选的, x 取值是-100或100,增加权值目的是将数据差异进行放大,而以10的整数倍放大能够方便后续计算,对于当前数据的量级,当数据放大至100倍时,能够满足放大需求,绝大部分数据差异能够较好体现,且不会过大增加数据处理负担。

[0112] 选择上述心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值用于计算评估值,一方面可以更为全面地挖掘出心电数据内部蕴含的动态数据特征,另一方面,可以筛选携带不同人体疾病患者的微弱心电信号变化所产生的动态病理特征上变化和差异。

[0113] 在该实施例中可以知晓,心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值中的一种可以作为单独变量以计算人体健康状况分类数据的数据特征的评估值。

[0114] 在另一种实施例中:人体健康状况分类数据包括:心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值中的一种以上;

[0115] 所述的人体健康状况分类数据的数据特征的评估值的计算方式如下: $Q = a \times \text{心电数据的异质度特征的量化数值} - b \times \text{心电向量数据的异质度特征的量化数值} + c \times \text{心电动力学数据的异质度特征的量化数值} - d \times \text{生物化学数据值} + e \times \text{人体生理信息数据值} - f \times \text{临床信息数据值}$;

[0116] 其中: a 、 b 、 c 、 d 、 e 、 f 为权值,取值是在-100至100之间的实数。通过调整 a 、 b 、 c 、 d 、 e 、 f 的数值,可对疾病诊断产生协同作用。

[0117] 由该实施例可知, a 、 b 、 c 、 d 、 e 、 f 权值可以取0,基于上述方案的理解, a 、 b 、 c 、 d 、 e 、 f 中可以有一个以上为0,但是 a 、 b 、 c 、 d 、 e 、 f 权值不同时为0,从而形成评估值计算的不同人体健康状况分类数据的组合。如各个数据的单独使用计算评估值,或者两个、三个、四个、五个不同数据的组合计算评估值,或者使用全部数据计算评估值,形成心电数据的异质度特征

的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值的这些数据的不同使用组合,从而以不同数据组合方式计算评估值。

[0118] 在一种实施例中,所述的人体健康状况分类数据的数据特征的评估值Q为:

$$Q = \frac{\sum_{i=1}^m \exp(x_i)}{m}, \quad i=1, 2, 3 \cdots k, k \text{ 的取值为不小于0的正整数, } \exp() \text{ 表示以自然基数e为底的指数函数, } x_i \text{ 取自心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值、生物化学数据值、人体生理信息数据值和临床信息数据值中的一种, } m \text{ 的值取自16~17之间的数值。进行计算后得到的数值范围,让不同的数据具有统一的量化范围,上述方法能将健康人和人体疾病患者的数据特征的差异进行更为明显的对比。}$$

[0119] 实施例2.应用机器学习算法对用户罹患人体疾病的风险进行预测

[0120] 本实施例在实施例1构建的人体健康状况分类数据集基础上,阐述了如何对用户的待测数据和人体健康状况分类数据集进行机器学习算法比对。具体的机器学习算法包括但不限于:支持向量机、贝叶斯分类器、K近邻算法、K均值算法、线性回归、逻辑回归、多元非线性回归拟合方法、Adaboost算法、隐马尔可夫模型、极限学习机、随机森林算法、决策树算法、聚类算法卷积神经网络、循环神经网络、生成式对抗网络、堆叠自动编码器、全连接网络、无监督预训练网络、深度信念网络、深度玻尔兹曼机、神经张量网络、明氏距离(Minkowski Distance)、马氏距离(Mahalanobis Distance)、曼哈顿距离(Manhattan Distance)、切比雪夫距离(Chebyshev Distance)、欧氏距离(Euclidean Distance)、汉明距离(Hamming distance)、皮尔逊相关系数(Pearson correlation coefficient)、杰卡德相似系数(Jaccard similarity coefficient)、谷本系数(Tanimoto Coefficient)、信息熵(Information Entropy)和余弦相似度(Cosine similarity)中的一种或几种。

[0121] 机器学习算法一:

[0122] 用户的待测数据和人体健康状况分类数据集进行机器学习算法比对方法的操作步骤为:采集待测人员的健康状况数据($x_1(t), x_2(t), x_3(t), \cdots, x_n(t)$)和健康状况数据集中的($y_1(t), y_2(t), y_3(t), \cdots, y_n(t)$);所采集的心电数据都有F个周期。

[0123]
$$S = \left(\sum_{k=1}^n |x_k(t) - y_k(t)|^p \right)^{\frac{1}{p}};$$

[0124] 其中, $x_k(t)$ 为待测人员心电数据第k个向量的分量, $y_k(t)$ 为健康状况数据集中目标心电数据的第k个向量的分量;p为非负实数。

[0125] 数据比对结果S的判别标准:0表示它们的指向是完全相同的,数值越大表示它们之间越独立。

[0126] 机器学习算法二:

[0127] 1) 建立人体健康状况分类数据库,包括健康人员的健康状况数据库,和非健康人员的健康状况数据库;

[0128] 2) 将人体健康状况分类数据库的每个数据进行归一化处理,将处理完的数据张量输入到深度卷积神经网络加全连接层中提取特征;

[0129] 3) 通过球面损失函数,利用梯度下降法寻找全连接层中的权重值和深度卷积神经网络中各个滤波器的权重值,在训练过程中,同时用健康人员的健康状况数据和非健康人员的健康状况数据验证数据库的统计结果;

[0130] 4) 训练完成后,根据训练中所统计的数据,选择比对精度最高的深度卷积神经网络模型作为第一阶段模型;

[0131] 5) 通过迁移学习复用第一阶段深度卷积神经网络模型以及第一阶段训练所得的所有权重值,通过二范数归一化球面损失函数,利用梯度下降法寻找全连接层中的权重值和深度卷积神经网络中各个滤波器的权重值,在训练过程中,同时全程记录健康人员的健康状况数据和非健康人员的健康状况数据验证数据库的统计结果;

[0132] 6) 训练完成后,根据训练中所统计的数据,选择比对精度最高的深度卷积神经网络模型作为最终模型;

[0133] 7) 将最终模型导入到将部署的目标嵌入式平台的测试程序中,将待测人员的人体健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 输入程序,通过模型计算特征向量,将特征向量与数据库中的每一人体健康状况分类数据 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$ 的特征向量进行余弦距离的计算,当余弦值小于阈值判断为同一种疾病。

[0134] 机器学习算法三:

[0135] 采集待测人员的健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 和人体健康状况分类数据集中的 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$, 然后进行比对,比对方法采用比对取值的方式进行:比对取值为将 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 变为 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$ 的最小替换次数。

[0136] 两数组比对的判别标准:0表示它们的指向是完全相同的,数值越大表示它们之间越独立。

[0137] 机器学习算法四:

[0138] 相似性比对方法A的操作步骤为:采集待测人员的心电数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 和健康状况数据集中的 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$;所采集的心电数据都有F个周期;

$$[0139] \quad S_f = \frac{|n \sum_{k=1}^n x_k(t) \times y_k(t) - \sum_{k=1}^n x_k(t) \times \sum_{k=1}^n y_k(t)|}{\sqrt{n \sum_{k=1}^n x_k^2(t) - (\sum_{k=1}^n x_k(t))^2} \times \sqrt{n \sum_{k=1}^n y_k^2(t) - (\sum_{k=1}^n y_k(t))^2}};$$

$$[0140] \quad S_s = \frac{1}{n} \sum_{k=1}^n (x_k(t) - y_k(t))^2;$$

$$[0141] \quad S = a \times S_f + b \times S_s + T;$$

[0142] 其中, $x_k(t)$ 为待测人员心电数据第k个向量的分量, $y_k(t)$ 为健康状况数据集中目标心电数据的第k个向量的分量, a和b取值为-1~1之间适应不同种心脏疾病的选择值, T的取值包括实数;

[0143] 相似度S的判别标准:相似性范围从-1到1, -1意味着两个向量指向的方向正好截然相反, 1表示它们的指向是完全相同的, 0通常表示它们之间是独立的, 而在这之间的值则表示中间的相似性。

[0144] 在本实施例中,所述的待测人员的健康状况数据包括但不限于:心电数据、心电向量数据、心电动力学数据、心电动力学数据特征的量化数据、肌电信号数据特征的量化数据、脑电信号数据特征的量化数据、生物化学数据、人体生理信息数据或临床信息数据。

[0145] 实施例3.一种基于异质度特征的人体疾病风险预测方法

[0146] 本实施例主要解决如何对人体疾病风险进行判定识别的问题。在本实施例中,所

述的人体疾病包括但不限于糖尿病、心肌缺血、心肌梗塞、心源性猝死、心肌炎、心律失常、心肌梗死、心绞痛、冠心病、急性冠脉综合征、高血压、高胆固醇、睡眠呼吸暂停、冠状动脉疾病和心力衰竭。

[0147] 本实施例所述的基于异质度特征的人体疾病风险预测方法,包括以下步骤:

[0148] 步骤一、采集待测人员的健康状况数据:心电数据的异质度特征的量化数值、心电向量数据的异质度特征的量化数值、心电动力学数据的异质度特征的量化数值、生物化学数据值、人体生理信息数据值和临床信息数据值;

[0149] 步骤二、计算待测人员健康状况的特征评估值 $Q' = a' * \text{心电数据的异质度特征的量化数值} - b' * \text{心电向量数据的异质度特征的量化数值} + c' * \text{心电动力学数据的异质度特征的量化数值} - d' * \text{生物化学数据值} + e' * \text{人体生理信息数据值} - f' * \text{临床信息数据值}$; a' 、 b' 、 c' 、 d' 、 e' 或 f' 的数值在-100至100之间的实数。

[0150] 步骤三、将步骤二得到的待测人员健康状况的特征评估值 Q' 输入到实施例1所构建的人体健康状况分类数据集中进行机器学习算法比对,所述的机器学习算法采用实施例2所述的方法,得到与待测人员的最相似的个体数据;

[0151] 步骤四、根据步骤三得到的与待测人员的最相似的个体数据的人体健康指数分值 V ,输出待测人员的人体健康指数分值 V_1 ;

[0152] 步骤五、筛选与待测人员健康状况的特征评估值较相似的前 N 个样本数据集,相似度从高到低依次为: S_1 、 S_2 、 $S_3 \cdots S_N$,得到待测人员的人体健康指数分值 V_2 , $V_2 = \beta_0 + \beta_1 \times S_1 + \beta_2 \times S_2 + \beta_3 \times S_3 \cdots + \beta_N \times S_N$, β_0 、 β_1 、 β_2 、 $\beta_3 \cdots \beta_N$ 的数值在-1至1之间的实数。

[0153] 在一种实施例中,本公开记载一种人工智能的生物医学信号特征评测方法,包括以下步骤:

[0154] 步骤一、构建人体健康状况单一特征评估数据与人体健康状况记录数据之间相互对应的人体健康状况数据库,所述的人体健康状况单一特征评估数据取自心电向量数据的异质度特征的量化数据和心电动力学数据的异质度特征的量化数据中的一种,所述的人体健康状况记录数据包括反映人体健康状况的数据或反映人体疾病严重程度的数据;

[0155] 步骤二、输入待测用户的单一特征评估数据到所述的人体健康状况数据库,使用机器学习算法得到与待测用户的单一特征评估数据相似度至少为 $S\%$ 的前 M 个数据库数据,各数据库中为疾病名称下对应的健康人的相应数据的范围以及患该疾病的病人的相应数据的范围,所述的 S 不小于10,所述的 M 为不小于 $\text{INT}(k+q \times m)$ 的数值,所述的 INT 函数表示取整,所述的 $k \geq 1$,所述的 q 值为进行过归一化处理的待测用户的单一特征评估数据的量化数值, m 的数值是在-10至10之间的实数;经过实验发现, m 的取值在该区间,对最终检测效果的判定更为准确,该方法能让数据处于同一个量化区间,相互之间仅是程度大小的区别, M 的确定,提高检测的准确性,更全面的反应待测人员的健康状态。

[0156] 步骤三、将得到的 M 个数据库数据的信息输入到预设的人体健康评估模型,得到待测人员的 M 个特征评估值,与 M 个特征评估值一致的已知身体健康状况的个体样本的身体健康状况作为待测人员健康状况参考;

[0157] 所述的人体健康评估模型:

[0158] $Q = a \times \text{心电数据的异质度特征的量化数值} + c \times \text{心电动力学数据的异质度特征的量化数值}$;

[0159] 其中： a 、 c 为权值，取值是在-100至100之间的实数， a 、 c 一个不为0时，另一个必为0。将数据差异进行放大，调节 a 、 c 的值，使得检测结果更为准确。

[0160] 在一种实施例中，本公开记载一种人工智能的生物医学信号特征评测方法，包括以下步骤：

[0161] 步骤一、构建人体健康状况数据集，所述的人体健康状况数据集选自异质度特征的量化数据、心电向量数据和心电动力学数据中的一种；

[0162] 步骤二、采集待测人员健康状况数据，输入到所述的人体健康状况数据集中进行比对，获得与待测人员健康状况数据值相似度至少为 $S\%$ 的前 k 个数据，所述的比对采用的算法是机器学习算法，所述的待测人员健康状况数据选自异质度特征的量化数据、心电向量数据和心电动力学数据中的一种，所述的 S 不小于10，所述的 $k \geq 1$ ；

[0163] 步骤三、依据所述的前 k 个数据的信息，输出待测人员健康状况的健康评估值，与待测人员健康状况数据值相似的前 k 个数据对应的已知身体健康状况的个体样本的身体健康状况作为待测人员健康状况参考。可以去掉一些不必要的杂音数据。

[0164] 由上述各个方案，对于待测者，可以将其 N 种因不同疾病计算得到的 Q_1 、 Q_2 、 Q_3 、 $\dots$ 、 Q_N 赋以权值，并综合得到待测者的健康分数值。

[0165] 实施例4. 一种心脏疾病风险的判别方法

[0166] 本实施例主要解决如何对心脏疾病风险进行判定识别的问题。本实施例提供一种心脏疾病风险的判别方法，具体为：如表1所示，采集待测人员周某的健康状况数据：心电数据的 Hurst参数的量化数值为0.75、心电向量数据的复杂度的量化数值为0.0371、心电动力学数据的复杂度的量化数值为0.0368、高敏C反应蛋白为3.6mg/L、人体生理信息数据值为1和临床信息数据值为4，计算待测人员健康状况的特征评估值 $Q' = 942.39$ ， Q' 的计算公式为： $Q' = a' * \text{心电数据的异质度特征的量化数值} - b' * \text{心电向量数据的异质度特征的量化数值} + c' * \text{心电动力学数据的异质度特征的量化数值} - d' * \text{生物化学数据值} + e' * \text{人体生理信息数据值} - f' * \text{临床信息数据值}$ ； $a' = 100$ ， $b' = -100$ ， $c' = 100$ ， $d' = -100$ ， $e' = 100$ ， $f' = -100$ 。

[0167] 表1. 心脏异常患者的信息

[0168]	住院号	*****
	姓名	周**
	性别	女
	年龄	55
	主诉	反复发作性胸痛 2 年，加重 10 天
	医院诊断	心肌缺血引发自发型心绞痛，高脂血症，返流性食管炎。 造影：LAD 近段 90%狭窄，于 LAD 处植入支架一枚。

[0169] 将周某的特征评估值 Q' 输入到实施例1所构建的人体健康状况分类数据集中进行机器学习算法的比对，所述的机器学习算法采用实施例2所述的机器学习算法一， p 的取值为1，得到与待测人员的相似度为100%的数据库的个体数据No.1265，根据个体数据No.1265的人体健康指数分值42分，输出待测人员的人体健康指数分值42分。

[0170] 在一种实施例中，对于上述情形，筛选与待测人员健康状况的特征评估值较相似的前5 个样本数据集，相似度从高到低依次为：No.1265、No.1104、No.734、No.207、No.453

等,并得到待测人员的6个特征评估值,根据预设的回归模型计算人体健康评估值 $V_2=40$,得到的心脏信息是:心脏的心肌缺血风险较大、高脂血症、返流性食管炎,符合程度为高相关性。

[0171] 在一种实施例中,对于上述情形,使用机器学习算法得到与待测用户的单一特征评估数据相似度至少为S%的前M个数据库数据,M为10,INT(k+q×m)的数值分别是8,待测人员的人体健康指数分39.5分,得到的心脏信息是:心脏的心肌缺血风险较大、高脂血症、返流性食管炎,符合程度为高相关性。

[0172] 在一种实施例中, $Q = \frac{\sum_{i=1}^m \exp(x_i)}{m}$, $i=1,2,3\cdots k$,得到与待测人员的相似度为100%的数据库的个体数据是No.1265、No.1104,待测人员的人体健康指数分值37.5分,得到的心脏信息是:心脏的心肌缺血风险较大、高脂血症、返流性食管炎,符合程度为高相关性。

[0173] 在一种实施例中,仅对心电数据的Hurst参数的量化数值0.75进行评估、匹配,得到与待测人员的相似度为100%的数据库的个体数据是No.1265,待测人员的人体健康指数分值42分,得到的心脏信息是:心脏的心肌缺血风险大,符合程度为中度相关性。

[0174] 在一种实施例中,仅对心电向量数据的复杂度的量化数值0.0371进行评估、匹配,得到与待测人员的相似度为100%的数据库的个体数据是No.1265,待测人员的人体健康指数分43分,得到的心脏信息是:心脏的心肌缺血风险大,符合程度为中度相关性。

[0175] 在一种实施例中,仅对心电动力学数据的复杂度的量化数值0.0368进行评估、匹配,得到与待测人员的相似度为100%的数据库的个体数据是No.1265,待测人员的人体健康指数分42.5分,得到的心脏信息是:心脏的心肌缺血风险较大,符合程度为中度相关性。

[0176] 在一种实施例中,对心电数据的Hurst参数的量化数值为0.75、心电向量数据的复杂度的量化数值为0.0371、心电动力学数据的复杂度的量化数值为0.0368进行评估、匹配,得到与待测人员的相似度为100%的数据库的个体数据是No.1265,待测人员的人体健康指数分35分,得到的心脏信息是:心脏的心肌缺血风险较大、高脂血症、返流性食管炎,符合程度为高相关性。同理地,如表2所示,获取姬某的心电数据的样本熵的量化数值为1.71、心电向量数据的近似熵的量化数值为5.94、心电动力学数据的近似熵的量化数值为6.21、高敏C反应蛋白为2.9mg/L、人体生理信息数据值为0和临床信息数据值为0,计算姬某健康状况的特征评估值 $Q'=942.39$,特征评估值 $Q'=1676$,将姬某的特征评估值 Q' 输入到实施例1所构建的人体健康状况分类数据集中进行机器学习算法的比对,所述的机器学习算法采用实施例2所述的机器学习算法一,p的取值为1,得到与姬某的特征评估值相似度为100%的数据库的个体数据为0,筛选与姬某的特征评估值较相似的前50个样本数据集,相似度从高到低依次为:82%的12个、70%的26个、65%的11个、50%的1个,得到待测人员的人体健康指数分值 V_2 , $V_2=0.31+0.82\times 12+0.7\times 26+0.65\times 11+0.5\times 1=36$ 。

[0177] 表2.心脏异常患者的信息

[0178]	住院号	*****
	姓名	姬**
	性别	男
	年龄	52
	主诉	发作性胸痛
	医院诊断	冠状动脉粥样硬化性心脏病

[0179] 虽然在本公开中已提供数个实施例,但应理解所公开系统和方法可在不偏离本公开的精神或范围的情况下以许多其它特定形式体现。这些实例将视为的示例性而非限制性的,并且并不旨在限于本发明中给出的细节。例如,各种元件或组件可组合或整合于另一系统中,或者某些特征可省略或不实施。

[0180] 在一种实施例中,上述各个方法,是基于软件实现的,则该实施例具体记载了一种人体健康检测设备,存储有多条指令,所述指令适于处理器加载并执行上述任意种方法。

[0181] 而且,在各种实施例中描述且示出为分立或单独的技术、系统、子系统和方法可在不偏离本公开的范围的情况下与其它系统、模块、技术或方法组合或整合。示出或讨论为彼此直接耦合或通信的其它项可通过某一接口、设备或中间组件以电力方式、以机械方式或以其它方式间接耦合或通信。改变、替换和变更的其它实例可由所属领域的技术人员确定并且可在不偏离本发明中所公开精神和范围的情况下作出。

专利名称(译)	基于人工智能的生物医学信号特征处理及评测方法、装置及应用		
公开(公告)号	CN110299205A	公开(公告)日	2019-10-01
申请号	CN201910666786.3	申请日	2019-07-23
[标]发明人	徐赤坤 李伟		
发明人	徐赤坤 李伟		
IPC分类号	G16H50/30 G16H50/70 G06K9/62 A61B5/00 A61B5/04		
CPC分类号	A61B5/04012 A61B5/7267 G06K9/6215 G16H50/30 G16H50/70		
代理人(译)	刘斌		
外部链接	Espacenet SIPO		

摘要(译)

基于人工智能的生物医学信号特征处理及评测方法、装置及应用，属于诊断目的的测量领域，为了解决能对反映心脏疾病的数据挖掘，快速反映采集数据与相关心脏疾病的关联性的问题，采集人体健康状况分类数据；得到人体健康状况分类数据的数据特征；计算人体健康状况分类数据的数据特征的评估值，效果是实现了反映心脏疾病的数据挖掘、匹配。

$$S = \left(\sum_{k=1}^n |x_k(t) - y_k(t)|^p \right)^{\frac{1}{p}} ;$$