



(12)发明专利申请

(10)申请公布号 CN 109938695 A

(43)申请公布日 2019.06.28

(21)申请号 201910175676.7

(22)申请日 2019.03.08

(71)申请人 度特斯(大连)实业有限公司

地址 116011 辽宁省大连市长兴岛经济区
何屯村(原何屯小学)

(72)发明人 王云霞 何毅钊

(51)Int.Cl.

A61B 5/00(2006.01)

A61B 5/0402(2006.01)

权利要求书1页 说明书11页

(54)发明名称

一种基于异质度指标的人体疾病风险预测
方法及设备

(57)摘要

本发明公开了一种基于异质度指标的人体疾病风险预测方法及设备,具体流程为:采用异质度分析的方法分别提取健康人和特定人体疾病患者的心电非线性系统动态数据(或心电向量数据)中的异质度指标,获取异质度指标的量化数据,构建人体健康评估数据集;输入待测人员的异质度指标的量化数据到构建的人体健康评估数据集中进行相似性比对,获得待测人员的相似性比对结果,实现对待测人员的健康状况的预测或评估。

1. 一种基于异质度指标的人体疾病风险预测方法,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,所述的人体健康状况分类数据集包括异质度指标的量化数据;优选地,所述的异质度指标是通过异质度分析的方法提取得到;

步骤二、输入待测人员的健康状况数据到步骤一所述的人体健康状况分类数据集中进行相似性比对,所述的待测人员的健康状况数据包括待测人员的异质度指标的量化数据。

2. 如权利要求1所述的人体疾病风险预测方法,其特征在于,所述的步骤二还包括:设定特定人体疾病相似性比对的监测阈值 T ,并将待测人员的相似性量化指标 S_I 的量化值和监测阈值 T 进行对比,完成对待测人员疾病风险的评估;

其中,所述的 S_I 的量化值 = $a \times$ 明氏距离的量化值 $- b \times$ 马氏距离的量化指标 $+ c \times$ 曼哈顿距离的量化值 $- d \times$ 切比雪夫距离的量化值 $+ e \times$ 相关系数的量化值 $- f \times$ 欧氏距离的量化值 $+ g \times$ 余弦相似度的量化值 $- h \times$ 汉明距离的量化值 $+ i \times$ 信息熵的量化值 $- j$;其中, a 、 b 、 c 、 d 、 e 、 f 、 g 、 h 和 i 的取值为 $-1 \sim 1$ 之间适应不同种人体疾病的选择值, j 的取值包括实数。

3. 如权利要求1所述的人体疾病风险预测方法,其特征在于,步骤一所述的异质度指标的量化数据包括心电动力学信号特征的量化指标数据、心电信号特征的量化指标数据、心电图向量数据特征的量化指标数据、肌电信号数据特征的量化指标数据、或脑电信号数据特征的量化指标数据中的一种或几种;优选地,所述的心电动力学信号特征的量化指标数据包括心电非线性系统动态数据特征的量化指标数据。

4. 如权利要求1所述的人体疾病风险预测方法,其特征在于,所述相似性比对的方法包括明氏距离、马氏距离、曼哈顿距离、切比雪夫距离、欧氏距离、汉明距离、相关系数、信息熵和余弦相似度中的一种或几种。

5. 如权利要求1所述的人体疾病风险预测方法,其特征在于,所述的步骤二还包括:筛选与待测人员的健康状况数据的相似度 $\geq K\%$ 的前 N 个样本数据集进行人体疾病风险的判别;其中,所述的 K 为不小于20的实数,所述的 N 为不小于2的整数。

6. 如权利要求5所述的人体疾病风险预测方法,其特征在于,所述的判别方法还包括对不同的相似度比对结果赋以权值进行人体疾病风险的判别。

7. 一种人体疾病的检测产品,其特征在于,使用了权利要求1-6任一权项所述的人体疾病风险预测方法进行人体疾病的检测。

8. 一种人体疾病的检测产品,其特征在于,包括以下模块:

人体健康状况分类数据集模块,存储人体健康状况分类数据HHCV,所述的人体健康状况分类数据HHCV可表述为: $W \times$ 异质度指标的量化数据 $+ X \times$ 生物化学数据 $+ Y \times$ 人体生理信息数据 $+ Z \times$ 临床信息数据的叠加运算,所述的 W 、 X 、 Y 和 Z 取值为非负整数;

数据采集模块,用于采集待测人员的健康状况数据;以及

人体疾病检测模块:将数据采集模块得到的数据集输入到所述的人体健康状况分类数据集模块进行相似性比对。

9. 一种医学治疗效果的评估方法,其特征在于,所述的评估方法包括:使用权利要求1-6任一权项所述的人体疾病风险预测方法;所述的医学治疗包括手术治疗或药物治疗。

10. 一种如权利要求1-6任一权项所述的人体疾病风险预测方法或权利要求9所述的评估方法在疾病早期筛查、病人监护、手术治疗、药物治疗或健康管理上的应用。

一种基于异质度指标的人体疾病风险预测方法及设备

技术领域

[0001] 本发明涉及用于诊断目的的测量技术领域,特别是涉及一种基于异质度指标的人体疾病风险预测方法及设备。

背景技术

[0002] 在我国,多数人群对于自身的健康状况没有准确的了解,对于自身罹患疾病风险的生理数据的认知不足,导致很多患者错过了治疗最佳期;而医患之间因为医疗知识和文化水平等方面的差异造成的沟通障碍,也进一步地影响了医学治疗的效果;因此,每年因为人体重大疾病死亡的人数一直居高不下,一些人体微弱疾病症状的患者猝死的比率也没有显著下降,很多人体重大疾病患者都无明显的临床症状。因此,开发一种罹患人体疾病风险的判别方法和检测设备具有的一定的临床应用价值。譬如,心脏疾病治疗后的疾病风险的判别,心脏疾病患者经过手术和药物等治疗后,患者通常对治疗后的健康管理不能进行充分地实施,而且治疗后各种手术并发症依然会发生,支架内血栓形成和再狭窄也相应的形成;因此,充分评估疾病患者的治疗前后的风险及预后等各个阶段心脏功能的状况显得十分重要。

[0003] 尽管现有技术对患者静态病理特征的检测已取得一些进展;然而,到目前为止,现有技术对早期疾病患者的动态病理特征的检测依然存在一些难以避免的缺陷:在疾病患者治疗前后的健康评估过程中,主要依赖于医生的从业经验,诊断结果主观性强,诊断耗时比较长。

[0004] 因此,有必要提供改进的技术方案以克服现有技术中存在的技术问题。

发明内容

[0005] 为解决上述问题,本发明公开了一种基于异质度指标的人体疾病风险预测方法及设备,具体流程为:采用异质度分析的方法分别提取健康人和特定人体疾病患者的心电数据(或心电向量数据、或心电非线性系统动态数据)中的异质度指标,获取异质度指标的量化数据,接着构建健康评估数据集;然后输入待测人员的异质度指标的量化数据到构建的健康评估数据集中进行相似性比对,获得待测人员的相似性比对结果,实现对待测人员的健康状况的预测判定。

[0006] 本发明第一方面提供了一种基于异质度指标的人体疾病风险预测方法,包括以下步骤:

步骤一、构建人体健康状况分类数据集,所述的人体健康状况分类数据集包括异质度指标的量化数据;优选地,所述的异质度指标是通过异质度分析的方法提取得到;

步骤二、输入待测人员的健康状况数据到步骤一所述的人体健康状况分类数据集中进行相似性比对,所述的待测人员的健康状况数据包括待测人员的异质度指标的量化数据。

[0007] 在至少一个实施例中,在如前所述的的人体疾病风险预测方法中,步骤二还包括:设定特定人体疾病相似性比对的监测阈值 T ,并将待测人员的相似性量化指标 S_I 的量化值

和监测阈值T进行对比,完成对待测人员疾病风险的评估;其中,所述的 S_I 的量化值 = $a \times$ 明氏距离的量化值 $- b \times$ 马氏距离的量化指标 $+ c \times$ 曼哈顿距离的量化值 $- d \times$ 切比雪夫距离的量化值 $+ e \times$ 相关系数的量化值 $- f \times$ 欧氏距离的量化值 $+ g \times$ 余弦相似度的量化值 $- h \times$ 汉明距离的量化值 $+ i \times$ 信息熵的量化值 $- j$;其中, a 、 b 、 c 、 d 、 e 、 f 、 g 、 h 和 i 的取值为-1~1之间适应不同种人体疾病的选择值, j 的取值包括实数。

[0008] 在至少一个实施例中,在如前所述的的人体疾病风险预测方法中,步骤一所述的异质度指标的量化数据包括心电动力学信号特征的量化指标数据、心电信号特征的量化指标数据、心电向量数据特征的量化指标数据、肌电信号数据特征的量化指标数据、或脑电信号数据特征的量化指标数据中的一种或几种;优选地,所述的心电动力学信号特征的量化指标数据包括心电非线性系统动态数据特征的量化指标数据。

[0009] 在至少一个实施例中,在如前所述的的人体疾病风险预测方法中,所述相似性比对的方法包括明氏距离、马氏距离、曼哈顿距离、切比雪夫距离、欧氏距离、汉明距离、相关系数、信息熵和余弦相似度中的一种或几种。

[0010] 在至少一个实施例中,在如前所述的的人体疾病风险预测方法中,步骤二还包括:筛选与待测人员的健康状况数据的相似度 $\geq K\%$ 的前N个样本数据集进行人体疾病风险的判别;其中,所述的K为不小于20的实数,所述的N为不小于2的整数。

[0011] 在至少一个实施例中,在如前所述的的人体疾病风险预测方法中,所述的判别方法还包括对不同的相似度比对结果赋以权值进行人体疾病风险的判别。

[0012] 本发明第二方面提供了一种人体疾病的检测产品,使用了前述的人体疾病风险预测方法进行人体疾病的检测。

[0013] 本发明第三方面提供了一种人体疾病的检测产品,包括以下模块:

人体健康状况分类数据集模块,用于存储人体健康状况分类数据HHCV,所述的人体健康状况分类数据HHCV可表述为: $W \times$ 异质度指标的量化数据 $+ X \times$ 生物化学数据 $+ Y \times$ 人体生理信息数据 $+ Z \times$ 临床信息数据的叠加运算,所述的 W 、 X 、 Y 和 Z 取值为非负整数;

数据采集模块,用于采集待测人员的健康状况数据;以及

人体疾病检测模块:用于将数据采集模块得到的数据集输入到所述的人体健康状况分类数据集模块进行相似性比对。

[0014] 本发明第四方面提供了一种医学治疗效果的评估方法,所述的评估方法包括:使用前述的人体疾病风险预测方法;所述的医学治疗包括手术治疗或药物治疗。

[0015] 本发明第五方面提供了一种前述的人体疾病风险预测方法或评估方法在疾病早期筛查、病人监护、手术治疗、药物治疗或健康管理上的应用。

[0016] 有益效果:本发明提供一种基于异质度指标的人体疾病风险预测方法及设备,对待测人员罹患特定人体疾病风险的进行预测,或对待测人员接受医学治疗的效果进行评估判定,提供一种新的人体疾病风险预测方法;对人体损伤小、准确率高、检测效率高,可有效地为人体疾病的早期检测提供较为准确的参考内容,并为后期医生进行疾病的精准治疗提供便利。

具体实施方式

[0017] 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。

[0018] 在本发明中,如果没有特别的说明,本文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。

[0019] 在本发明中,如果没有特别的说明,本文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。

[0020] 在本发明中,如果没有特别的说明,本文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。

[0021] 在本发明中,如果没有特别的说明,本文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。

[0022] 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。

[0023] 在本发明中,如果没有特别的说明,本文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。

[0024] 在本发明中,如果没有特别的说明,本文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。

[0025] 在本发明中,如果没有特别的说明,本文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。

[0026] 在本发明中,如果没有特别的说明,本文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。

[0027] 在本发明中,如果没有特别的说明,本文所提到的心电非线性系统动态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非线性系统动态数据。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。

[0028] 在本发明中,如果没有特别的说明,本文所提到的心电非线性系统动态数据是指将心电向量数据经过(或不经过)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。

[0029] 在本发明中,如果没有特别的说明,本文所提到的心电非线性系统动态数据包括但不限于心电动力学数据或CDG数据。

[0030] 在本发明中,如果没有特别的说明,本文所提到的心电非线性系统动态数据包括但不限于完整的心电非线性系统动态数据、预处理的心电非线性系统动态数据、截取的心电非线性系统动态数据、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。截取的心电非线性系统动态数据,包括但不限于:存在于心电非线性系统动态数据上任意大小的一部分的数据。

[0031] 在本发明中,如果没有特别的说明,本文所提到的心电非线性系统动态数据包括但不限于:心电非线性系统动态数据(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)、预处理的心电非线性系统动态数据(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)、心电动力学数据、预处理的心电动力学数据或截取的心电动力学数据。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,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。

[0032] 在本发明中,如果没有特别的说明,本文所提到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。

[0033] 在本发明中,如果没有特别的说明,本文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指标的量化数据、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。异质度指标的量化数据包括但不限于心电动力学信号特征的量化指标数据、心电信号特征的量化指标数据、心电向量数据特征的量化指标数据、肌电信号数据特征的量化指标数据、或脑电信号数据特征的量化指标数据。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、心肌病病史、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、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、Lown's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分级和內中膜厚度等；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。

[0034] 在本发明中，如果没有特别的说明，本文所提到的异质度指标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。

[0035] 在本发明中，如果没有特别的说明，本文所提到的异质度指标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、心电非线性系统动态数据、肌电信号数据和脑电信号数据中的一种或几种。

[0036] 在本发明中，如果没有特别的说明，本文所提到的异质度指标包括但不限于人体生物电信号数据特征、心电非线性系统动态数据特征、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。

[0037] 在本发明中，如果没有特别的说明，本文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。

[0038] 在本发明中，如果没有特别的说明，本文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、圆弧度、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。

[0039] 在本发明中,如果没有特别的说明,本文所提到的心电非线性系统动态数据特征包括但不限于从心电非线性系统动态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。

[0040] 在本发明中,如果没有特别的说明,本文所提到的心电非线性系统动态数据特征包括但不限于心电非线性系统动态病理特征、心电非线性系统静态病理特征或心脏电活动非线性系统动态数据内在的心脏静态病理特征、或心脏电活动非线性系统动态数据内在的心脏动态病理特征。

[0041] 在本发明中,如果没有特别的说明,本文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。

[0042] 在本发明中,如果没有特别的说明,本文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。

[0043] 在本发明中,如果没有特别的说明,本文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。

[0044] 在本发明中,如果没有特别的说明,本文所提到的相似性比对方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。

[0045] 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。

[0046] 实施例1. 人体健康状况分类数据集的构建

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、心电非线性非线性系统动态数据、心电非线性系统动态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非线性系统动态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。

[0047] 心电非线性系统动态数据的获取: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,进行计算得到心电非线性系统动态数据,所述心电非线性系统动态数据可表述为 $[\bar{W}_1^T S(X(t)), \bar{W}_2^T S(Y(t)), \bar{W}_3^T S(Z(t))]$;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。

[0048] 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。

[0049] 在本实施例中,所述的心电非线性系统动态数据特征包括但不限于通过异质度分析的方法从心电非线性系统动态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。

[0050] 实施例2. 用于人体疾病风险判别的相似性比对方法

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,阐述了待测数据和人体健康状况分类数据集进行相似性比对的方法,具体的相似性比对方法如下:

本实施例采用的相似性比对方法包括但不限于: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。

[0051] 相似性比对方法A

相似性比对方法A的操作步骤为:采集待测人员的健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 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。

$$[0052] \quad S_f = \frac{|n \sum_{k=1}^n x_k(t) \times y_k(t) - \sum_{k=1}^n x_k(t) \times \sum_{k=1}^n y_k(t)|}{\sqrt{n \sum_{k=1}^n x_k^2(t) - (\sum_{k=1}^n x_k(t))^2} \times \sqrt{n \sum_{k=1}^n y_k^2(t) - (\sum_{k=1}^n y_k(t))^2}};$$

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

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

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之间适应不同种人体疾病的选择值,T的取值包括实数。

[0053] 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。

[0054] 相似性比对方法B

相似性比对方法B的操作步骤为:采集待测人员的健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 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。

$(t), \dots, x_n(t))$ 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。

$$[0055] \quad S = \sqrt{\sum_{k=1}^n \left((x_k(t) - y_k(t))^T V^{-1} (x_k(t) - y_k(t)) \right)};$$

其中, V^{-1} 为协方差矩阵, $x_k(t)$ 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。

[0056] 相似度 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。

[0057] 相似性比对方法 C

相似性比对方法 C 的操作步骤为: 采集待测人员的健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 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。

$$[0058] \quad S = (\sum_{k=1}^n |x_k(t) - y_k(t)|^p)^{\frac{1}{p}};$$

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。

[0059] 相似度 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。

[0060] 相似性比对方法 D

相似性比对方法 D 的操作步骤为: 采集待测人员的健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 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。

[0061] 相似性比对方法 D 的取值为将 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 变为 $(y_1(t), y_2(t), y_3(t), \dots, y_n(t))$ 的最小替换次数。

[0062] 相似度 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。

[0063] 相似性比对方法 E

相似性比对方法 E 的操作步骤为: 采集待测人员的健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 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。

$$[0064] \quad S = \frac{\sum_{k=1}^n x_k(t) \times y_k(t) - \frac{\sum_{k=1}^n x_k(t) \times \sum_{k=1}^n y_k(t)}{n}}{\sqrt{\left(\sum_{k=1}^n x_k^2(t) - \frac{(\sum_{k=1}^n x_k(t))^2}{n} \right) \times \left(\sum_{k=1}^n y_k^2(t) - \frac{(\sum_{k=1}^n y_k(t))^2}{n} \right)}};$$

相似度 S 的判别标准: 相似性范围从 -1 到 1, 接近 0 的变量被成为无相关性, 接近 1 或者 -1 被称为具有强相关性。

[0065] 相似性比对方法 F

相似性比对方法 F 的操作步骤为: 采集待测人员的健康状况数据 $(x_1(t), x_2(t), x_3(t), \dots, x_n(t))$ 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。

$(t), \dots, x_n(t))$ 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。

$$[0066] \quad S = \frac{x(t) \cap y(t)}{x(t) \cup y(t)};$$

相似度 S 的判别标准: 0 表示二者是完全独立的; 数值越大, 表示二者越具有相关性。

[0067] 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、心电非线性非线性系统动态数据、心电非线性系统动态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。

[0068] 实施例3. 一种基于异质度指标的人体疾病风险预测方法

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。

[0069] 待测人员治疗前的相似性比对

治疗前, 采集待测人员的健康状况数据, 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相似性比对, 所述的相似性比对方法采用实施例2所述的方法, 得到治疗前待测人员的相似性比对结果 $S_{\text{治疗前}}$ 。

[0070] 待测人员治疗后的相似性比对

治疗前, 采集待测人员的健康状况数据, 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相似性比对, 所述的相似性比对方法采用实施例2所述的方法, 得到治疗前待测人员的相似性比对结果 $S_{\text{治疗后}}$ 。

[0071] 待测人员的人体疾病风险的评估

评估步骤1: 如果 $S_{\text{治疗后}}$ 大于健康人群的相似度阈值 C , 则认为待测人员的治疗前后的健康度有明显提高; 如果 $S_{\text{治疗后}}$ 小于健康人群的相似度阈值 C , 则认为待测人员的治疗前后的健康度未有改善。如果 $S_{\text{治疗后}}$ 等于阈值 C , 则转到评估步骤2进行操作。

[0072] 评估步骤2: 评估结果 $M_i = S_{\text{治疗前}} / S_{\text{治疗后}}$; 治疗后, 如果待测人员的 M_i 大于1, 则认为待测人员的治疗前后的健康度有明显提高; 如果待测人员的 M_i 小于1, 则认为待测人员治疗后的健康度发生下降。如果待测人员的 M_i 等于1, 则转到评估步骤3进行操作。

[0073] 评估步骤3: 筛选与待测人员治疗前相关的前 N 个样本数据集, 得到治疗前待测人员的健康状况评定值 $S'_{\text{治疗前}} = (N - X_k) / N$, 同理地, 得到待测人员治疗后相似度 $S'_{\text{治疗后}} = (N - X'_k) / N$, 其中, X_k 和 X'_k 为健康人群的数量; 健康状况评估结果为 $M'_i = S'_{\text{治疗前}} / S'_{\text{治疗后}}$ 。

[0074] 健康人群的相似度阈值

纳入 W 个临床已知心脏健康个体 ($W > 500$), 与实施例1构建的健康状况数据集分别进行比较, 采集相似度 $\geq 60\%$ 为前 V 个样本数据进行 M_i 值的计算, 得到 $M_1, M_2, \dots, M_V$; 健康人群的相似度阈值 $C = \frac{1}{V} \sum_{i=1}^V M_i$ 。

[0075] 实施例4. 一种心脏疾病风险的判别方法

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,在待测人员进行经皮冠状动脉介入治疗(PCI治疗)前,将其心电非线性系统动态数据特征的量化数据输入到实施例1所述的人体健康状况分类数据集中进行相似性比对;相似性比对方法采用实施例2中的相似性比对方法,所述的心脏疾病风险的判别包括采用相似度量指标 S_I 进行评估,且设定相似度量指标的监测阈值 T ,将所述 S_I 的量化值和 T 对比进行心脏疾病风险的评估;其中,所述的 S_I 的量化值 = $a \times$ 明氏距离的量化值 $- b \times$ 马氏距离的量化指标 $+ c \times$ 曼哈顿距离的量化值 $- d \times$ 切比雪夫距离的量化值 $+ e \times$ 皮尔逊相关系数的量化值 $- f \times$ 欧氏距离的量化值 $+ g \times$ 余弦相似度的量化值 $- h \times$ 汉明距离的量化值 $+ i \times$ 信息熵的量化值 $- j$;其中, a 、 b 、 c 、 d 、 e 、 f 、 g 、 h 和 i 的取值为 $-1 \sim 1$ 之间适应不同种人体疾病的选择值, j 的取值包括实数。优选地, a 、 b 、 c 、 d 、 e 、 g 、 h 和 i 的取值为0, f 的取值为 -1 , p 的取值为2;得到治疗前健康指数值 $S_{I\text{治疗前}}$ 为32%。然后将待测人员进行PCI治疗,在待测人员进行PCI治疗,再次采集待测人员的心电非线性系统动态数据特征的量化数据,输入到实施例1所述的人体健康状况分类数据集中进行相似性比对;相似性比对方法采用实施例2中的相似性比对方法,所述的心脏疾病风险的判别包括采用相似度量指标 S_I 进行评估,且设定相似度量指标的监测阈值 T ,将所述 S_I 的量化值和 T 对比进行心脏疾病风险的评估;其中,所述的 S_I 的量化值 = $a \times$ 明氏距离的量化值 $- b \times$ 马氏距离的量化指标 $+ c \times$ 曼哈顿距离的量化值 $- d \times$ 切比雪夫距离的量化值 $+ e \times$ 皮尔逊相关系数的量化值 $- f \times$ 欧氏距离的量化值 $+ g \times$ 余弦相似度的量化值 $- h \times$ 汉明距离的量化值 $+ i \times$ 信息熵的量化值 $- j$;其中, a 、 b 、 c 、 d 、 e 、 f 、 g 、 h 和 i 的取值为 $-1 \sim 1$ 之间适应不同种人体疾病的选择值, j 的取值包括实数。优选地, a 、 b 、 c 、 d 、 e 、 g 、 h 、 i 和 j 的取值为0, f 的取值为 -1 , p 的取值为2;得到治疗后健康指数值 $S_{I\text{治疗后}}$ 为35%。因此,认为待测人员的治疗前后的健康度有了提高。

[0076] 在一些实施方式中,在接受治疗前后,可筛选与待测人员相似值 $>K\%$ 的 N 个个体,进行实施例3所述的人体疾病风险的判别,所述的 N 最小值为2;优选地,所述的 N 值越大,准确率越高。在接受治疗前后, K 取值为不小于20,待测人员治疗前的健康状况评估结果为0.07,待测人员治疗后的健康状况评估结果为0.4,因此,认为待测人员的治疗前后的健康度有明显提高。

[0077] 在一些实施方式中,所述的相似性比对方法采用实施例2中的相似性比对方法进行,所述的心脏疾病风险的判别包括采用相似度量指标 S_I 进行评估,且设定相似度量指标的监测阈值 T ,将所述 S_I 的量化值和 T 对比进行心脏疾病风险的评估;其中,所述的 S_I 的量化值 = $a \times$ 明氏距离的量化值 $- b \times$ 马氏距离的量化指标 $+ c \times$ 曼哈顿距离的量化值 $- d \times$ 切比雪夫距离的量化值 $+ e \times$ 皮尔逊相关系数的量化值 $- f \times$ 欧氏距离的量化值 $+ g \times$ 余弦相似度的量化值 $- h \times$ 汉明距离的量化值 $+ i \times$ 信息熵的量化值 $- j$;其中, a 、 b 、 c 、 d 、 e 、 f 、 g 、 h 和 i 的取值为 $-1 \sim 1$ 之间适应不同种人体疾病的选择值, j 的取值包括实数。优选地, b 、 e 、 f 、 g 、 h 、 i 和 j 的取值为0, a 或 c 的取值为 -1 或 1 , d 的取值为 -1 或 1 , p 的取值为2。

[0078] 实施例5.一种糖尿病风险的判别方法

本实施例主要解决如何对糖尿病风险进行判定识别的问题。本实施例提供一种糖尿病疾病风险的判别方法,具体为,将待测人员的心电非线性系统动态数据特征的量化数据输入到实施例1所述的人体健康状况分类数据集中进行相似性比对;相似性比对方法采用实施例2中的相似性比对方法,所述的糖尿病风险的判别包括采用相似度量指标 S_I 进行评估,且设定相似度量指标的监测阈值 T ,将所述 S_I 的量化值和 T 对比进行糖尿病风险的评

估;其中,所述的 S_I 的量化值 = $a \times$ 明氏距离的量化值 $-b \times$ 马氏距离的量化指标 $+c \times$ 曼哈顿距离的量化值 $-d \times$ 切比雪夫距离的量化值 $+e \times$ 皮尔逊相关系数的量化值 $-f \times$ 欧氏距离的量化值 $+g \times$ 余弦相似度的量化值 $-h \times$ 汉明距离的量化值 $+i \times$ 信息熵的量化值 $-j$;其中, $a、b、c、d、e、f、g、h$ 和 i 的取值为 $-1 \sim 1$ 之间适应不同种人体疾病的选择值, j 的取值包括实数。优选地, $a、c、d、e、f、h、i$ 和 j 的取值为 0 , b 的取值为 -1 或 1 , g 的取值为 1 ;得到相似性比对比对结果 A' 。然后,继续采用实施例3所述的疾病风险的判别方法继续进行待测人员罹患糖尿病风险的判别,具体为:筛选与待测人员治疗前相似度 $\geq K\%$ 的前 N 个样本数据集,待测人员相似度 A' 。如果 A' 大于或等于健康人群的相似度阈值 C ,则认为待测人员倾向于健康人;如果 A' 小于健康人群的相似度阈值 C ,则认为待测人员有罹患糖尿病的风险。

[0079] 在一些实施方式中,所述的心电非线性系统动态数据特征的量化数据还可为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。

[0080] 实施例6.一种高血压风险的判别方法

本实施例主要解决对高血压风险进行判定识别的问题。本实施例提供一种高血压疾病风险的判别方法,具体为,将待测人员的心电非线性系统动态数据特征的量化数据输入到实施例1所述的人体健康状况分类数据集中进行相似性比对;相似性比对方法采用实施例2中的相似性比对方法,所述的高血压风险的判别包括采用相似度量指标 S_I 进行评估,且设定相似度量指标的监测阈值 T ,将所述 S_I 的量化值和 T 对比进行待测人员罹患高血压风险的评估;其中,,所述的 S_I 的量化值 = $a \times$ 明氏距离的量化值 $-b \times$ 马氏距离的量化指标 $+c \times$ 曼哈顿距离的量化值 $-d \times$ 切比雪夫距离的量化值 $+e \times$ 相关系数的量化值 $-f \times$ 欧氏距离的量化值 $+g \times$ 余弦相似度的量化值 $-h \times$ 汉明距离的量化值 $+i \times$ 信息熵的量化值 $-j$;其中, $a、b、c、d、e、f、g、h$ 和 i 的取值为 $-1 \sim 1$ 之间适应不同种人体疾病的选择值, j 的取值包括实数。优选地, $a、b、d、c、f、g、h、i$ 和 j 的取值为 0 , $e、g、h$ 和 i 的取值为 -1 或 1 ,所述的相关系数包括但不限于皮尔逊相关系数、杰卡德相似系数或谷本系数;得到相似性比对比对结果 A' ;具体为:筛选与待测人员治疗前相似度 $\geq K\%$ 的前 N 个样本数据集,得到待测人员相似度 A' 。如果 A' 大于或等于健康人群的相似度阈值 T ,则认为待测人员倾向于健康人;如果 A' 小于健康人群的相似度阈值 T ,则认为待测人员有罹患高血压的风险。

[0081] 在一些实施方式中,所述的心电非线性系统动态数据特征的量化数据还可为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。

[0082] 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文中给出的细节。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。

[0083] 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文中所公开精神和范围的情况下作出。

专利名称(译)	一种基于异质度指标的人体疾病风险预测方法及设备		
公开(公告)号	CN109938695A	公开(公告)日	2019-06-28
申请号	CN201910175676.7	申请日	2019-03-08
[标]发明人	王云霞 何毅钊		
发明人	王云霞 何毅钊		
IPC分类号	A61B5/00 A61B5/0402		
外部链接	Espacenet SIPO		

摘要(译)

本发明公开了一种基于异质度指标的人体疾病风险预测方法及设备，具体流程为：采用异质度分析的方法分别提取健康人和特定人体疾病患者的心电非线性系统动态数据（或心电向量数据）中的异质度指标，获取异质度指标的量化数据，构建人体健康评估数据集；输入待测人员的异质度指标的量化数据到构建的人体健康评估数据集中进行相似性比对，获得待测人员的相似性比对结果，实现对待测人员的健康状况的预测或评估。