



(12)发明专利申请

(10)申请公布号 CN 110881969 A

(43)申请公布日 2020.03.17

(21)申请号 201911184324.4

(22)申请日 2019.11.27

(71)申请人 太原理工大学

地址 030024 山西省太原市迎泽西大街79
号

(72)发明人 李灯熬 赵菊敏 付健

(74)专利代理机构 太原高欣科创专利代理事务
所(普通合伙) 14109

代理人 冷锦超 邓东东

(51)Int.Cl.

A61B 5/0402(2006.01)

A61B 5/00(2006.01)

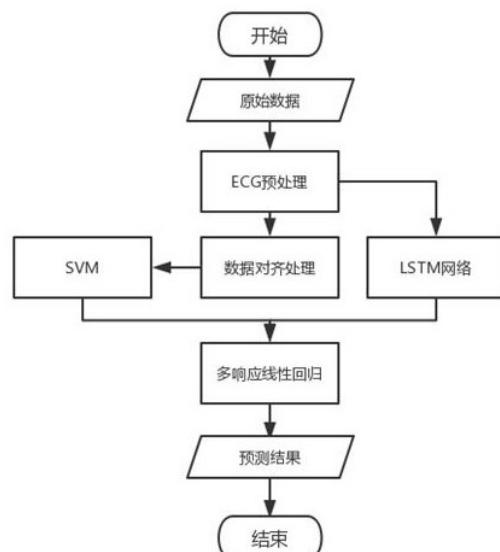
权利要求书2页 说明书5页 附图2页

(54)发明名称

一种基于Stacking集成学习的心力衰竭预
警方法

(57)摘要

本发明一种基于Stacking集成学习的心力衰竭预警方法,对患者是否患有心力衰竭疾病进行预警分析,从而辅助医生诊断患者,弥补单一模型带来的缺陷;具体包括步骤:将患者心电信号进行预处理放入LSTM中训练;对患者临床信息和人口统计学信息进行数据对齐放入SVM模型中训练;然后利用产生次级训练器的样本来训练多响应线性回归MLR模型,产生最终的预测结果;利用stacking集成学习算法综合考虑所有判别结果,对患者是否患有心力衰竭疾病进行预警分析,从而辅助医生诊断患者,以达到辅助医生诊断患者是否患有心衰的目标。



1. 一种基于Stacking集成学习的心力衰竭预警方法,其特征在于,具体包括以下步骤:

a) 心电信号预处理:对从患者得到的心电信号依次去除噪声、基线漂移,之后进行标准化处理,得到带有标记的心电信号数据;

b) 相关信息数据对齐:将该患者的临床信息和人口统计学信息进行数据对齐,做离散化处理,作为支持向量机模型的输入;所述数据对齐是将所述临床信息和人口统计学信息整合成一个m*n的样本数据,其中m为样本个数,n为特征维度;

c) 心力衰竭预警:将处理好的心电信号放入长短期记忆神经网络中训练,将数据对齐后的数据,放入SVM模型中训练,得到两个基分类器的分类结果,采用九折交叉验证的方式,用初级训练器未使用的样本来产生次级训练器的样本;然后利用产生次级训练器的样本来训练多响应线性回归模型,产生最终的预测结果。

2. 根据权利要求1所述的一种基于Stacking集成学习的心力衰竭预警方法,其特征在于,所述除噪声是使用IIR数字滤波器进行滤波处理,使用零相移滤波器去除基线漂移。

3. 根据权利要求1所述的一种基于Stacking集成学习的心力衰竭预警方法,其特征在于,对去除噪声的心电信号进行标准化处理:采用z-score标准化方法:

$$y_i = \frac{x_i - \bar{x}}{s}$$

其中 x_i 为第i个样本, \bar{x} 为样本均值, s 为样本标准差, \bar{x} 与 s 的计算公式如下:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

4. 根据权利要求1所述的一种基于Stacking集成学习的心力衰竭预警方法,其特征在于,一个LSTM神经单元的处理步骤如下:

遗忘门控制t-1时刻有多少信息被遗忘:

$$f_t = \sigma(W_f^1 h_{t-1} + W_f^2 x_t + b_f)$$

其中 h_{t-1} 表示上一时刻即t-1时刻的输出, x_t 为t时刻的输入, W_f^1, W_f^2 为遗忘门中 h_{t-1} 和 x_t 对应的更新权重系数, b_f 为遗忘门的偏置;

输入门为t-1时刻的输出与t时刻的输入:

$$i_t = \sigma(W_i^1 h_{t-1} + W_i^2 x_t + b_i)$$

其中 h_{t-1} 表示上一时刻即t-1时刻的输出, x_t 为t时刻的输入, W_i^1, W_i^2 为输入门中 h_{t-1} 和 x_t 对应的更新权重系数, b_i 为输入门的偏置;

c_t' 表示当前的输入单元状态:

$$c_t' = \sigma(W_c^1 h_{t-1} + W_c^2 x_t + b_c)$$

其中 h_{t-1} 表示上一时刻即t-1时刻的输出, x_t 为t时刻的输入, W_c^1, W_c^2 为当前输入的状态 h_{t-1} 和 x_t 对应的更新权重系数, b_c 为当前输入状态偏置;

c_t 表示当前单元状态,综合遗忘门和输入门更新t时刻状态:

$$c_t = c_{t-1} f_t + i_t \circ c_t'$$

其中 c_{t-1} 表示上一时刻t-1时刻的单元状态, c_t' 为t时刻的输入状态, f_t 为遗忘门的计算权重, i_t 为输入门的输入计算权重;

输出门状态由t-1时刻输出、t时刻输入与t时刻状态共同决定：

$$o_t = \sigma(W_o^1 h_{t-1} + W_o^2 x_t + b_o)$$

$$h_t = o_t \circ \tanh(c_t)$$

其中 W_o^1, W_o^2 为输出门对应于 h_{t-1} 和 x_t 的权重, b_o 为输出门的偏置, o_t 为当前时刻最终输出的状态, h_t 为当前时刻最终的输出;

其中 \circ 表示按位乘, $\sigma(\cdot)$ 为sigmoid函数, $\tanh(\cdot)$ 为双曲线函数。

一种基于Stacking集成学习的心力衰竭预警方法

技术领域

[0001] 本发明属于计算机医疗软件技术领域,具体涉及一种基于Stacking集成学习的心力衰竭预警方法。

背景技术

[0002] 心力衰竭(简称心衰,Heart Failure, HF)是由于任何原因的初始心肌损伤(心肌梗死、血液动力学负荷过重、炎症等),引起心肌结构和功能的变化,导致心室充盈和射血的障碍而引起的一组临床综合征。心力衰竭是全球慢性心血管疾病的重要组成部分,是各种心脏疾病发展的最终阶段,具有高患病率、高花费、预后差的特点,已成为全球范围内的重大的公共卫生问题。此外由于人们生活水平提高、生活习惯改变、人口老龄化等原因,心力衰竭的发病率、死亡率呈逐年上升趋势。早期发现和早期预测不良事件是获得良好治疗效果的基础。

[0003] Stacking算法就是一种利用结合策略去组合单一模型方法,由于Stacking算法,常用于异质集成,且可以构造多层个体分体器的集成,因此被广泛应用于各种领域。如用组合Stacking进行文本情感分类,基于Stacking利用ANN和自适应神经模糊推理系统(ANFIS)预测每年河流封冻日期,将Sentinel TOPS模式应用Stacking技术监测淮南矿区沉降等。

[0004] 利用stacking集成学习算法综合考虑所有判别结果,以达到辅助医生诊断患者是否患有心衰的目标,为临床的诊断提供了一种新思路。

发明内容

[0005] 本发明克服了现有技术的不足,提出一种基于Stacking集成学习的心力衰竭预警方法,使用stacking集成学习方法综合基分类器对患者是否患有心力衰竭疾病进行预警分析,从而辅助医生诊断患者,弥补单一模型带来的缺陷。

[0006] 为了达到上述目的,本发明是通过如下技术方案实现的。

[0007] 一种基于Stacking集成学习的心力衰竭预警方法,具体包括以下步骤:

[0008] a) 心电信号预处理:对从患者得到的心电信号依次去除噪声、基线漂移,之后进行标准化处理,得到带有标记的心电信号数据。

[0009] b) 相关信息数据对齐:将该患者的临床信息和人口统计学信息进行数据对齐,做离散化处理,作为支持向量机(Support Vector Machine, SVM)模型的输入;所述数据对齐是将所述临床信息和人口统计学信息整合成一个m*n的样本数据,其中m为样本个数,n为特征维度。

[0010] c) 心力衰竭预警:将处理好的心电信号放入长短期记忆神经网络(Long Short-Term Memory Neural Network, LSTM)中训练,将数据对齐后的数据,放入SVM模型中训练,得到两个基分类器的分类结果,采用九折交叉验证的方式,用初级训练器未使用的样本来产生次级训练器的样本;然后利用产生次级训练器的样本来训练多响应线性回归(Multi-response Linear Regression, MLR)模型,产生最终的预测结果。

[0011] 优选的,所述除噪声是使用IIR数字滤波器进行滤波处理,使用零相移滤波器去除基线漂移。

[0012] 优选的,对去除噪声的心电信号进行标准化处理:采用z-score标准化方法:

$$[0013] y_i = \frac{x_i - \bar{x}}{s}$$

[0014] 其中 x_i 为第i个样本, \bar{x} 为样本均值, s 为样本标准差, \bar{x} 与 s 的计算公式如下:

$$[0015] \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

[0016] 进一步的,一个LSTM神经单元的处理步骤如下:

[0017] 遗忘门控制t-1时刻有多少信息被遗忘:

$$[0018] f_t = \sigma(W_f^1 h_{t-1} + W_f^2 x_t + b_f)$$

[0019] 其中 h_{t-1} 表示上一时刻即t-1时刻的输出, x_t 为t时刻的输入, W_f^1, W_f^2 为遗忘门中 h_{t-1} 和 x_t 对应的更新权重系数, b_f 为遗忘门的偏置。

[0020] 输入门为t-1时刻的输出与t时刻的输入:

$$[0021] i_t = \sigma(W_i^1 h_{t-1} + W_i^2 x_t + b_i)$$

[0022] 其中 h_{t-1} 表示上一时刻即t-1时刻的输出, x_t 为t时刻的输入, W_i^1, W_i^2 为输入门中 h_{t-1} 和 x_t 对应的更新权重系数, b_i 为输入门的偏置。

[0023] c_t' 表示当前的输入单元状态:

$$[0024] c_t' = \sigma(W_c^1 h_{t-1} + W_c^2 x_t + b_c)$$

[0025] 其中 h_{t-1} 表示上一时刻即t-1时刻的输出, x_t 为t时刻的输入, W_c^1, W_c^2 为当前输入的状态 h_{t-1} 和 x_t 对应的更新权重系数, b_c 为当前输入状态偏置。

[0026] c_t 表示当前单元状态,综合遗忘门和输入门更新t时刻状态:

$$[0027] c_t = c_{t-1} f_t + i_t \circ c_t'$$

[0028] 其中 c_{t-1} 表示上一时刻t-1时刻的单元状态, c_t' 为t时刻的输入状态, f_t 为遗忘门的计算权重, i_t 为输入门的输入计算权重。

[0029] 输出门状态由t-1时刻输出、t时刻输入与t时刻状态共同决定:

$$[0030] o_t = \sigma(W_o^1 h_{t-1} + W_o^2 x_t + b_o)$$

$$[0031] h_t = o_t \circ \tanh(c_t)$$

[0032] 其中 W_o^1, W_o^2 为输出门对应于 h_{t-1} 和 x_t 的权重, b_o 为输出门的偏置, o_t 为当前时刻最终输出的状态, h_t 为当前时刻最终的输出。

[0033] 其中 \circ 表示按位乘, $\sigma(\cdot)$ 为sigmoid函数, $\tanh(\cdot)$ 为双曲线函数。

[0034] 本发明相对于现有技术所产生的有益效果为。

[0035] 本发明基于患者的心电信号和其带有的其他基本属性包括临床信息和人口统计学信息,首先对收集到的心电信号进行预处理,包括去噪、去基线漂移等,然后将心电信号数据对齐,针对不同数据格式的数据建立不同的学习模型,得出结论,再利用stacking集成学习算法综合考虑所有判别结果,对患者是否患有心力衰竭疾病进行预警分析,从而辅助医生诊断患者,以达到辅助医生诊断患者是否患有心衰的目标,弥补单一模型带来的缺陷。

附图说明

- [0036] 图1为本发明所零相移滤波器原理图。
- [0037] 图2为stacking集成策略图。
- [0038] 图3为LSTM单元结构图。
- [0039] 图4为本发明所述心力衰竭预警方法流程图。

具体实施方式

[0040] 为了使本发明所要解决的技术问题、技术方案及有益效果更加清楚明白,结合实施例和附图,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。下面结合实施例及附图详细说明本发明的技术方案,但保护范围不被此限制。

[0041] 一种基于Stacking集成学习的心力衰竭预警方法,具体为以下步骤:

[0042] 1、心电信号预处理

[0043] 由于心电信号(ECG)检测的实际环境情况并非理想状态,因此ECG信号中总会夹杂着各种随机噪声。这些噪声的存在使得心电信号模糊而且严重地影响了医疗诊断的可靠度,因此需要对心电信号进行预处理。

[0044] 预处理按照下面三个步骤分别进行:

[0045] 1.1 使用IIR数字滤波器进行滤波处理去除噪声:

[0046] IIR数组滤波器是一类递归型的线性时不变的因果系统,其差分方程可以写为:

$$[0047] y(n) = \sum_{i=0}^M a_i x(n-i) + \sum_{i=1}^N b_i y(n-i)$$

[0048] 其中M、N为滤波器的阶, a_i 、 b_i 为滤波器的系数, $x(n-i)$ 、 $y(n-i)$ 以n时刻为初始, 移动i位。

[0049] 进行z变换, 可得:

$$[0050] Y(z) = \sum_{i=0}^M a_i z^{-i} X(z) + \sum_{i=1}^N b_i z^{-i} Y(z)$$

[0051] 其中, M、N为滤波器的阶, a_i 、 b_i 为滤波器的系数, $X(z)$ 、 $Y(z)$ 为z变换。

[0052] 即 $X(z) = \sum_{m=-\infty}^{\infty} x(m) z^{-m}$, 其中 $z = e^{\sigma+j\omega} = e^{\sigma} (\cos \omega + j \sin \omega)$, σ 为实变数, ω 为实变

量, z 为幅值为 e^{σ} , 相位为 ω 的复变量。

[0053] 于是IIR数字滤波器的系统函数为:

$$[0054] H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^M a_i z^{-i}}{1 - \sum_{i=1}^N b_i z^{-i}}$$

[0055] 1.2 使用零相移滤波器去除基线漂移: 零相移滤波器原理如图1所示:

[0056] 最终此滤波器的输入和输出可以表示为:

$$[0057] Y(e^{j\omega}) = X(e^{j\omega}) |H(e^{j\omega})|^2$$

[0058] 其中, $H(e^{j\omega})$ 为频率响应函数, $X(e^{j\omega})$ 为离散时间傅里叶变换, $Y(e^{j\omega})$ 为最终滤波器的输出。

[0059] 1.3对去除噪声的心电信号进行标准化处理:

[0060] 此处采用z-score标准化方法:

$$[0061] y_i = \frac{x_i - \bar{x}}{s}$$

$$[0062] \text{其中 } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

[0063] 最终得到的数据是去除噪声,去除基线漂移,经过标准化处理,带有标记的心电信号。

[0064] 2、数据对齐

[0065] 除心电信号信息外,患者本身还有很多其他信息,主要是临床信息和人口统计学信息。临床信息主要包括患者的血液检查记录,如红细胞压积、白细胞数、总胆红素、丙氨酸氨基转移酶等十几种指标。人口统计学信息主要包括性别、年龄和BMI三项指标。无论是临床信息还是人口统计学信息这些指标值都是在一定的范围内的连续型数值信息,为了让这些指标作为SVM模型的输入,所以需要将数据对齐,做离散化处理。

[0066] 以年龄为例,可将年龄按照合适的区间(0-6岁为婴幼儿;7-12岁为少儿;13-17岁为青少年;18-65岁为青年;66-79岁为中年;80-99岁为老年)划分为0,1,2,...特征。利用上述阐述的方法,就可以将这些患者本身信息,整合成一个m*n的样本数据,其中m为样本个数,n为特征维度,即临床信息和人口学信息指标个数。

[0067] 3、心力衰竭预警

[0068] 心力衰竭预警模型建立与实施的步骤分为两步如下:

[0069] (1) 将处理好的心电信号放入LSTM中训练,将数据对齐后的数据,放入SVM模型中训练,得到两个基分类器的分类结果,本文采用九折交叉验证的方式,用初级训练器未使用的样本来产生次级训练器的样本。

[0070] (2) 然后利用产生次级训练器的样本来训练多响应线性回归(MLR)模型,产生最终的预测结果。

[0071] 如图2所示,stacking集成学习策略的处理步骤流程,其中基学习器(Base Learner)为LSTM和SVM分别处理心电信号和人口统计学资料;元学习器(Meta Learner)综合基学习器的训练结果和原始标签,利用MLR模型训练预测输出最终训练结果。

[0072] 在基学习器中,涉及到了一个用于处理心电信号的神经网络:长短期记忆神经网络,Long Short-Term Memory neural network简称LSTM,LSTM网络结构单元如图3所示。

[0073] 一个LSTM神经单元的处理步骤如下:

[0074] 遗忘门控制t-1时刻有多少信息被遗忘:

$$[0075] f_t = \sigma(W_f^1 h_{t-1} + W_f^2 x_t + b_f)$$

[0076] 输入门为t-1时刻的输出与t时刻的输入:

$$[0077] i_t = \sigma(W_i^1 h_{t-1} + W_i^2 x_t + b_i)$$

[0078] c_t' 表示当前输入状态:

$$[0079] c_t' = \sigma(W_c^1 h_{t-1} + W_c^2 x_t + b_c)$$

[0080] c_t 表示当前单元状态,综合遗忘门和输入门更新t时刻状态:

$$[0081] \quad c_t = c_{t-1} f_t + i_t \circ c_t'$$

[0082] 输出门状态由t-1时刻输出、t时刻输入与t时刻状态共同决定：

$$[0083] \quad o_t = \sigma (W_o^1 h_{t-1} + W_o^2 x_t + b_o)$$

$$[0084] \quad h_t = o_t \circ \tanh(c_t)$$

[0085] 其中 \circ 表示按位乘， $\sigma(\cdot)$ 为sigmoid函数， $\tanh(\cdot)$ 为双曲线函数。

[0086] 经过以上3个步骤就可以建立综合了心电信号和人口统计学信息等不同类别信息的心衰预警模型，图4为预警方法的流程示意图。

[0087] 本发明通过对患者的心电信号和其带有的其他属性，如心电信号本人的性别、年龄等属性，进行分析，弥补单一模型带来的缺陷，使用stacking集成学习方法综合基分类器对患者是否患有心力衰竭疾病进行预警分析，从而辅助医生诊断患者。

[0088] 以上内容是结合具体的优选实施方式对本发明所做的进一步详细说明，不能认定本发明的具体实施方式仅限于此，对于本发明所属技术领域的普通技术人员来说，在不脱离本发明的前提下，还可以做出若干简单的推演或替换，都应当视为属于本发明由所提交的权利要求书确定专利保护范围。

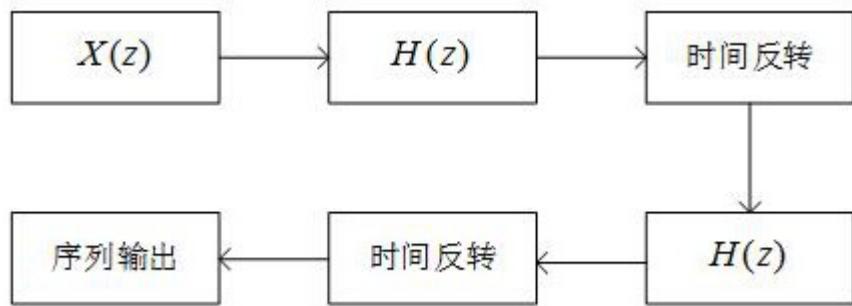


图1

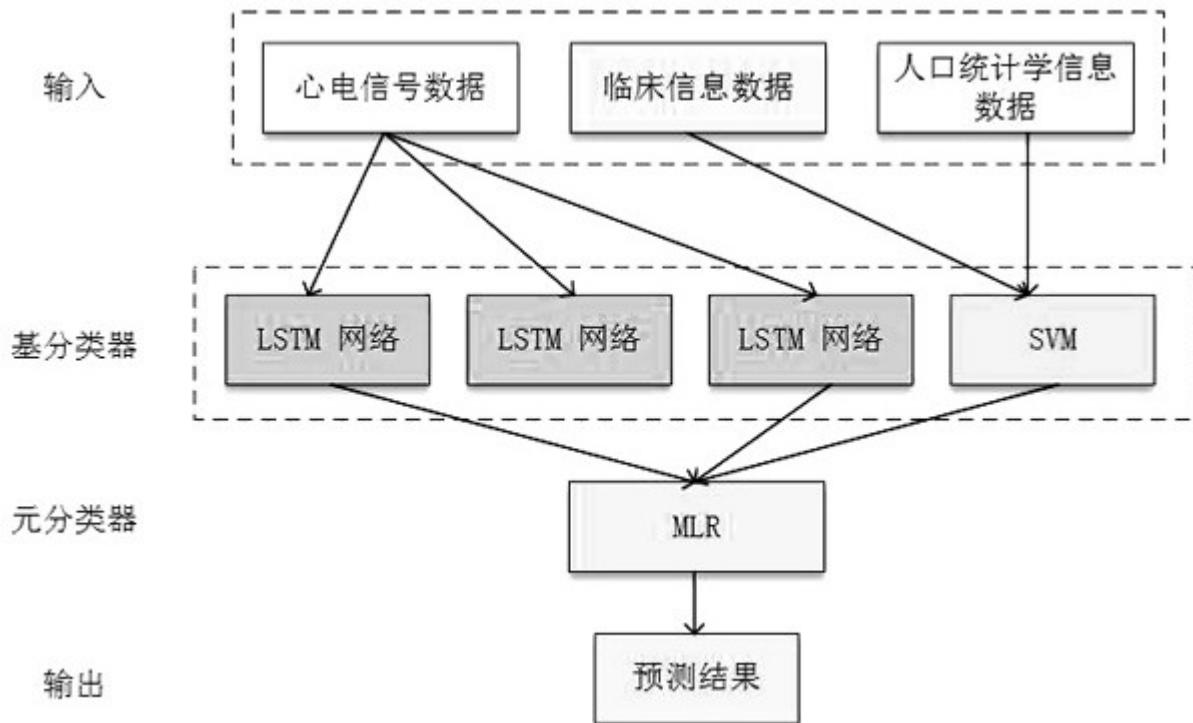


图2

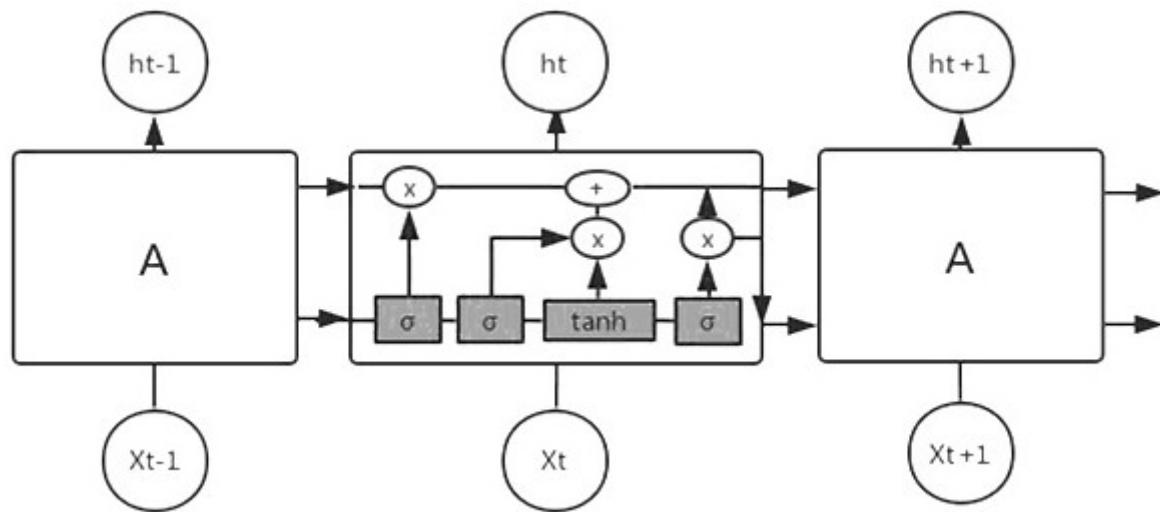


图3

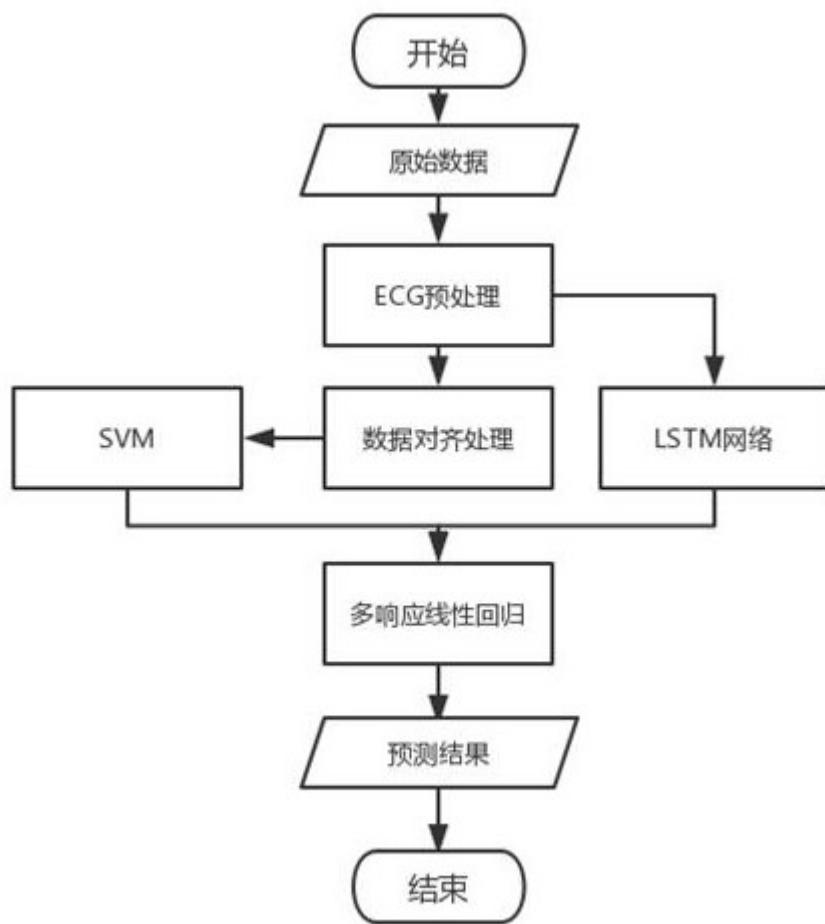


图4

| | | | |
|----------------|---|---------|------------|
| 专利名称(译) | 一种基于Stacking集成学习的心力衰竭预警方法 | | |
| 公开(公告)号 | CN110881969A | 公开(公告)日 | 2020-03-17 |
| 申请号 | CN201911184324.4 | 申请日 | 2019-11-27 |
| [标]申请(专利权)人(译) | 太原理工大学 | | |
| 申请(专利权)人(译) | 太原理工大学 | | |
| 当前申请(专利权)人(译) | 太原理工大学 | | |
| [标]发明人 | 李灯熬 赵菊敏 付健 | | |
| 发明人 | 李灯熬 赵菊敏 付健 | | |
| IPC分类号 | A61B5/0402 A61B5/00 | | |
| CPC分类号 | A61B5/04012 A61B5/0402 A61B5/7203 A61B5/7225 A61B5/725 A61B5/7267 | | |
| 代理人(译) | 冷锦超 邓东东 | | |
| 外部链接 | Espacenet Sipo | | |

摘要(译)

本发明一种基于Stacking集成学习的心力衰竭预警方法，对患者是否患有心力衰竭疾病进行预警分析，从而辅助医生诊断患者，弥补单一模型带来的缺陷；具体包括步骤：将患者心电信号进行预处理放入LSTM中训练；对患者临床信息和人口统计学信息进行数据对齐放入SVM模型中训练；然后利用产生次级训练器的样本来训练多响应线性回归MLR模型，产生最终的预测结果；利用stacking集成学习算法综合考虑所有判别结果，对患者是否患有心力衰竭疾病进行预警分析，从而辅助医生诊断患者，以达到辅助医生诊断患者是否患有心衰的目标。

