

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022 年 8 月 11 日 (11.08.2022)



(10) 国际公布号
WO 2022/166158 A1

(51) 国际专利分类号:
G16H 50/30 (2018.01) *G06N 3/04* (2006.01)
G16H 50/20 (2018.01) *G06N 3/08* (2006.01)

(21) 国际申请号: PCT/CN2021/113135

(22) 国际申请日: 2021 年 8 月 18 日 (18.08.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110179779.8 2021 年 2 月 8 日 (08.02.2021) CN

(71) 申请人: 之江实验室 (ZHEJIANG LAB) [CN/CN];
中国浙江省杭州市余杭区文一西路 1818 号, Zhejiang 310023 (CN)。

(72) 发明人: 李劲松 (LI, Jingsong); 中国浙江省杭州市余杭区文一西路 1818 号, Zhejiang 310023 (CN)。
王丰 (WANG, Feng); 中国浙江省杭州市余杭区文一西路 1818 号, Zhejiang 310023 (CN)。
朱世强 (ZHU, Shiqiang); 中国浙江省杭州市余杭区文一西路 1818 号, Zhejiang 310023 (CN)。

雨 (TIAN, Yu); 中国浙江省杭州市余杭区文一西路 1818 号, Zhejiang 310023 (CN)。
周天舒 (ZHOU, Tianshu); 中国浙江省杭州市余杭区文一西路 1818 号, Zhejiang 310023 (CN)。

(74) 代理人: 杭州求是专利事务所有限公司 (HANGZHOU QIUSHI PATENT OFFICE CO., LTD.); 中国浙江省杭州市西湖区古墩路 701 号绿城紫金广场 C 座 1506 室 / 刘静, Zhejiang 310030 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(54) Title: SYSTEM FOR PERFORMING LONG-TERM HAZARD PREDICTION ON HEMODIALYSIS COMPLICATIONS ON BASIS OF CONVOLUTIONAL SURVIVAL NETWORK

(54) 发明名称: 一种基于卷积生存网络的血透并发症长期风险预测系统

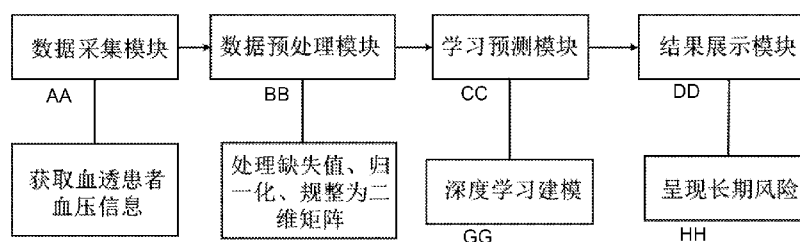


图 1

AA Data collection module

BB Data preprocessing module

CC Learning prediction module

DD Result presentation module

EE Acquire blood pressure information of hemodialysis patients

FF Process missing values, and normalize and arrange same into a two-dimensional matrix

GG Perform deep learning modeling

HH Present long-term hazards

(57) Abstract: Disclosed is a system for performing long-term hazard prediction on hemodialysis complications on the basis of a convolutional survival network. The system comprises a data collection module, a data preprocessing module, a learning prediction module and a result presentation module. In the present invention, a convolutional neural network is used to process multi-dimensional hemodialysis timing sequence features; a convolutional survival network is proposed by combining the convolutional neural network with a Cox proportional hazard assumption; and on the basis of using the convolutional survival network, a baseline hazard function is estimated by using Breslow, so as to calculate long-term hazard change situations for patients. By means of the present invention, common truncated data in medical research can be fully utilized; the main architecture of a convolutional neural network is used for facilitating visual analysis in order to produce interpretable and heuristic results; and long-term hazard change situations for patients can be predicted.

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 本发明公开了一种基于卷积生存网络的血透并发症长期风险预测系统, 该系统包括数据采集模块、数据预处理模块、学习预测模块和结果展示模块; 本发明利用卷积神经网络处理多维度的血透时序特征; 卷积神经网络结合Cox比例风险假设, 提出卷积生存网络; 在利用卷积生存网络的基础上, 采用Breslow估计基准风险函数, 计算患者的长期风险变化情况。本发明可以充分利用医学研究中常见的截尾数据; 应用卷积神经网络的主体架构, 便于可视化分析, 做出具有可解释性、具有启发性的结果; 可以预测患者长期的风险变化情况。

一种基于卷积生存网络的血透并发症长期风险预测系统

技术领域

本发明属于医疗及机器学习技术领域，具体地，涉及一种基于卷积生存网络的血透并发症长期风险预测系统。

背景技术

终末期肾病在世界范围内的发病呈上升趋势，造成巨大疾病负担。大部分患者需要依靠血液透析（血透）维持生命。长期血透过程中可能发生的血管通路感染、高血压、冠心病等并发症，严重影响患者生存状况，对患者、患者家庭以及社会都造成了巨大的影响和负担。因此，对血透并发症进行长期的风险预测以及早期的预防治疗，对于提高终末期肾病患者生存质量至关重要。长期血透过程中积累了大量时序数据，为相关研究带来了挑战的同时也带来了大量机会。近年来，随着信息技术的飞速发展，兴起了许多基于机器学习方法的时序分析方法。其中深度学习应用广泛且性能强大，包括循环神经网络以及卷积神经网络等。循环神经网络和卷积神经网络已经被广泛应用于医疗领域内的时序分析场景。

循环神经网络是一类以序列数据为输入，在序列的演进方向进行递归且所有节点按链式连接的递归神经网络。卷积神经网络是一类包含卷积计算且具有深度结构的前馈神经网络，具有表征学习能力，能够按其阶层结构对输入信息进行平移不变分类，因此也被称为“平移不变人工神经网络”。然而循环神经网络识别的特征复杂抽象，难以可视化解理解、提供启发性结果。卷积神经网络往往仅能预测截止单个时间点的风险值，无法对并发症发生风险进行长期、连续的风险预测，难以为临床医生提供准确有效的决策支持。除此之外，传统的循环神经网络以及卷积神经网络仅能分析简单的数据结构，而无法处理临床分析中常见的截尾数据。

发明内容

深度学习是人工智能领域的热门研究方向，在搜索、机器翻译、自然语言处理、多媒体学习、推荐和个性化技术，以及其他相关领域都取得了很多成果，但目前深度学习还很少应用到血透并发症预后分析领域。

本发明的目的在于针对传统深度学习方法的不足，提出一种基于卷积生存网络的血透并发症长期风险预测系统；本发明利用深度学习领域的卷积神经网络算法，基于其表征学习能

力以及卷积核的可视化方法，得到具有可解释性、具有启发性的结果；本发明改进卷积神经网络结构，使其可以充分利用截尾数据，预测患者的相对风险，进而进行长期、连续的风险预测。

本发明的目的是通过以下技术方案来实现的：一种基于卷积生存网络的血透并发症长期风险预测系统，该系统包括：用于采集血透患者血压信息的数据采集模块；用于对原始数据进行缺失值处理、归一化处理并规整化为二维矩阵的数据预处理模块；用于深度学习建模的学习预测模块；用于可视化输出、呈现长期风险变化情况的结果展示模块；

所述数据预处理模块的处理过程具体为：将连续血透的血压数据规整成二维矩阵，按血透治疗顺序排列，每行数据对应于一次血透治疗过程，使用 Min-Max 归一化所述二维矩阵以确保血压波形得以保留；

所述学习预测模块的处理过程包括两部分：

(1) 基于卷积生存网络的并发症相对风险预测：使用卷积神经网络架构并结合 Cox 比例风险损失函数训练卷积生存网络；所述卷积生存网络由多个卷积层和全连接层叠加而成，输入为数据预处理模块输出的二维矩阵，各卷积层将特征层层抽象，最终抽象为若干模式特征，各模式特征经全连接层输出为一个节点表示发生事件的相对风险；通过 Cox 比例风险损失函数优化网络参数；

风险函数 $h(t, x)$ 表示个体在给定时刻发生事件的概率，公式如下：

$$h(t, x) = h_0(t) \exp(\beta^T x) \quad (1)$$

其中 t 表示时间， $x = (x^1, \dots, x^p)^T$ 是 p 维协变量用于表征患者， p 为卷积生存网络最后一层全连接层的神经元个数， $\beta = (\beta^1, \dots, \beta^p)^T$ 是回归参数， $h_0(t)$ 是基准风险函数， $\beta^T x$ 是相对风险，即卷积生存网络的输出；

Cox 比例风险损失函数公式如下：

$$L(\beta) = -\sum_{i=1}^n \delta_i (\beta^T x_i - \log \sum_{j: T_j > T_i} e^{\beta^T x_j}) \quad (2)$$

其中 n 指患者个数； δ_i 是结局事件 E 的指示符，患者个体 x_i 发生事件则 $\delta_i = 1$ ，未发生事件则 $\delta_i = 0$ ； T_i 及 T_j 分别表示患者个体 x_i 和 x_j 的生存时间；

(2) 基于 Breslow 法结合相对风险，计算长期风险变化情况，具体地：基于 Breslow 法得到 t_i 时刻的基准累积风险函数 $H_0(t_i)$ ，根据 $H_0(t_i)$ 得到 $h_0(t)$ ， $H_0(t_i)$ 的估计表达式如下：

$$H_0(t_i) = \int_0^{t_i} h_0(t) dt = \sum_{t_j < t_i} \left[\frac{\delta_j}{\sum_{j \in R_i} \exp(\beta^T x_j)} \right] \quad (3)$$

其中 R_i 为 t_i 时刻处于风险中的样本集；

根据基准风险函数 $h_0(t)$ 结合卷积生存网络计算得到的相对风险 $\beta^T x$ ，根据式(1)计算描述

长期风险的风险函数 $h(t, x)$ 。

进一步地, 所述数据预处理模块首先筛除收缩压小于 60mmHg 或大于 250mmHg 的数据, 之后筛选连续 36 次血透的收缩压数据规整成二维矩阵, 按血透治疗顺序排列, 每行数据对应于一次时长 5 小时的血透治疗过程, 列与列之间间隔 10 分钟, 总计 36 行, 每行 30 个点。

进一步地, 所述学习预测模块中, 卷积生存网络接收数据预处理模块输出的二维矩阵, 首先识别一维血压模式, 然后将感知野补成方形, 将特征层层抽象后接全局均值池化层获得若干模式特征, 各模式特征经全连接层输出为一个节点表示发生事件的相对风险; 最后一层激活函数为 Linear, 其余层激活函数为 Relu。

进一步地, 所述学习预测模块中, 卷积生存网络的输入为数据预处理模块输出的 1 通道 36*30 大小的二维矩阵, 通过 16 个 1*5 的卷积识别一维血压模式, 然后通过 32 个纵向 5*1 的卷积操作将感知野补成方形, 之后接 2 层 16 个 5*5 的卷积核、1 层 9 个 3*3 的卷积核将特征层层抽象, 再之后接全局均值池化层获得 9 个模式特征值, 9 个模式特征值经全连接层输出为一个节点表示发生事件的相对风险。

进一步地, 所述学习预测模块中, Cox 比例风险损失函数的计算流程如下:

(a) 将患者的特征数据 X 和生存结局事件 E 统一按生存时间 T 降序排列, 形成矩阵 M ;
 (b) y_{i_pred} 为卷积生存网络输出, 对应于 $\beta^T x_i$; 由于 M 按 T 降序排列, $\sum_{j:T_j > T_i} e^{\beta^T x_j}$ 为矩阵 M 的前 i 行通过卷积生存网络后输出的指数累加和, 即前 i 行 $e^{y_{i_pred}}$ 的累加和, 令累计求和函数记为 cumsum;

(c) δ_i 即矩阵 M 第 i 行个体生存结局事件 E 的取值, 记作 E_i ;

(d) 令求和函数记为 sum, 由此损失函数表示为:

$$\text{loss} = -\text{sum}(E_i * (y_{i_pred} - \log(\text{cumsum}(y_{i_pred}))))。$$

进一步地, 所述学习预测模块中, 为了使训练集和测试集对应的损失函数取值为同一个量级, 将求和函数 sum 替换为求均值函数 mean, 损失函数表达为:

$$\text{loss} = -\text{mean}(E_i * (y_{i_pred} - \log(\text{cumsum}(y_{i_pred}))))。$$

进一步地, 所述学习预测模块中, 通过 Cox 比例风险损失函数优化网络参数时, 将训练集随机分层划分成 10 个 batch, 每个 batch 中生存结局事件 E 的比例相等, 并将每个 batch 的生存数据按生存时间 T 倒序排列用于计算损失函数, 由此遍历一遍数据集更新 10 次网络参数。

本发明的有益效果是: 本发明利用卷积神经网络处理多维度的血透时序特征; 卷积神经网络结合 Cox 比例风险假设, 提出卷积生存网络; 在利用卷积生存网络的基础上, 采用 Breslow 估计基准风险函数, 计算患者的长期风险变化情况。本发明可以充分利用医学研究中常见的

截尾数据；应用卷积神经网络的主体架构，便于可视化分析，做出具有可解释性、具有启发性的结果；可以预测患者长期的风险变化情况。

附图说明

图 1 为本发明基于卷积生存网络的血透并发症长期风险预测系统结构框图；

图 2 为基于卷积生存网络的血透并发症长期风险预测流程图；

图 3 为卷积生存网络模型示意图。

具体实施方式

为使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图对本发明的具体实施方式做详细的说明。

在下面的描述中阐述了很多具体细节以便于充分理解本发明，但是本发明还可以采用其他不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本发明内涵的情况下做类似推广，因此本发明不受下面公开的具体实施例的限制。

本发明中所述的卷积生存网络：应用于生存分析的卷积神经网络，可以处理时序、图像类的数据并进行生存分析与风险预测；长期风险预测：区别于截止一个时间点的风险预测，“长期风险预测”预测较长一段时间内连续的风险变化情况；截尾数据：如果在规定的结束时间没有出现结果事件的数据称之为截尾数据，从起点到截尾的时间称为截尾时间。

如图 1 所示，本发明提出的一种基于卷积生存网络的血透并发症长期风险预测系统，该系统包括：用于采集血透患者血压信息的数据采集模块；用于对原始数据进行缺失值处理、归一化处理并规整化为二维矩阵的数据预处理模块；用于深度学习建模的学习预测模块；用于可视化输出、呈现长期风险变化情况的结果展示模块。

数据预处理模块对血透血压数据进行预处理，具体如下：

血透过程中记录的血压数据与患者心脑血管并发症等不良临床事件具有相关性。本发明筛选收缩压小于 60mmHg 或大于 250mmHg 的数据，筛选连续 36 次（约 3 个月）血透的收缩压数据规整成二维矩阵的格式，按血透治疗顺序排列，每行数据对应于一次时长 5 小时的血液透析治疗过程（不足 5 小时以 0 填充），列与列之间间隔 10 分钟（线性插值），总计 36 行，每行 30 个点。使用 Min-Max 归一化该二维矩阵以确保血压波形得以保留。该二维矩阵作为卷积生存网络的输入，二维矩阵见图 3 中最左侧的输入矩阵。

学习预测模块的处理过程包括两部分：基于卷积生存网络的并发症相对风险预测；基于 Breslow 法结合相对风险，计算长期风险变化情况，如图 2 所示，具体步骤如下：

(1) 基于卷积生存网络的并发症相对风险预测

本发明使用卷积神经网络的架构并结合 Cox 比例风险损失函数训练卷积生存网络。卷积生存网络由多个卷积层和全连接层叠加而成，输入为上述 1 通道 36*30 大小的二维矩阵，通过 16 个 1*5 的卷积识别一维血压模式，然后通过 32 个纵向 5*1 的卷积操作将感知野补成方形，之后接 2 层 16 个 5*5 的卷积核、1 层 9 个 3*3 的卷积核将特征层层抽象，再之后接全局均值池化层获得 9 个模式特征值，9 个模式特征值全连接输出为一个节点表示发生事件（如心血管并发症发生）的相对风险，即下文 $\beta^T x$ ；最后一层激活函数为 Linear，其余层激活函数为 Relu。本发明通过 Cox 比例风险损失函数优化网络参数，网络输出为相对风险。卷积生存网络模型结构见图 3。

卷积生存网络的关键是它可以通过使用 Cox 比例风险损失函数来处理截尾数据并优化网络参数。风险函数 $h(t, x)$ 表示个体在给定时刻发生事件的概率，风险函数如式(1)所示，表示患者 x 在 t 时刻发生事件的概率：

$$h(t, x) = h_0(t) \exp(\beta^T x) \quad (1)$$

其中 t 表示时间， $x = (x^1, \dots, x^p)^T$ 是 p 维协变量， p 为卷积生存网络最后一层全连接层的神经元个数（本实施例中对应于全局均值池化后获得的 9 个模式特征值，即 p 取 9）用于表征患者 x ， $\beta = (\beta^1, \dots, \beta^p)^T$ 是回归参数， $h_0(t)$ 是基准风险函数， $\beta^T x$ 是相对风险（即卷积生存网络的输出）。

Cox 比例风险损失函数为：

$$L(\beta) = -\sum_{i=1}^n \delta_i (\beta^T x_i - \log \sum_{j: T_j > T_i} e^{\beta^T x_j}) \quad (2)$$

其中 n 指 ESRD 患者个数， δ_i 是结局事件 E 的指示符；患者个体 x_i 发生事件则 $\delta_i = 1$ ，未发生事件则 $\delta_i = 0$ ； T_i 及 T_j 分别表示患者个体 x_i 和 x_j 的生存时间。式(2)值越小表示参数 β （相当于最后一层全连接层的参数）越能拟合患者的相对风险。

通常损失函数由 y_{pred} 和 y_{true} 计算得到。 y_{pred} 为模型输出的预测值， y_{true} 则表示其真值。例如 y_{pred} 和 y_{true} 的差方和取均值即为均方误差损失函数。但是式(2)中仅 $\beta^T x_i$ 和 $\beta^T x_j$ 作为模型预测输出的相对风险能用 y_{pred} 表示，而患者真实并发症发生风险未知。为了计算式(2)的损失函数，本发明采用以下流程：

(1.1) 将患者的特征数据 X （表示所有患者的特征数据）和生存结局事件 E （表示所有患者的生存结局），统一按生存时间 T （表示所有患者的生存时间）降序排列，形成矩阵 M ；

(1.2) $y_{i\text{-pred}}$ 为卷积生存网络输出，表示相对风险，对应于 $\beta^T x_i$ 。由于 M 按 T 降序排列， $\sum_{j: T_j > T_i} e^{\beta^T x_j}$ 为矩阵 M 的前 i 行通过卷积生存网络后输出的指数累加和，即前 i 行 $e^{y_{i\text{-pred}}}$

的累加和, 令累计求和函数记为 cumsum ;

(1.3) δ_i 即矩阵 M 第 i 行个体生存结局事件 E 的取值, 又可记作 E_i ;

(1.4) 令求和函数记为 sum , 由此损失函数可表示为:

$$\text{loss} = -\text{sum}(E_i * (y_{i_pred} - \log(\text{cumsum}(y_{i_pred}))));$$

(1.5) 由于训练集和测试集样本量通常有差异, 为了使得二者损失函数取值为同一个量级, 本发明将步骤 (1.4) 的求和改为求均值; 令求均值函数记为 mean , 损失函数可表达为

$$\text{loss} = -\text{mean}(E_i * (y_{i_pred} - \log(\text{cumsum}(y_{i_pred}))))。$$

可以观察到, 与普通损失函数使用 y_pred 和 y_true 计算不同的是, Cox 比例风险损失函数在利用生存时间 T 倒序排列特征数据 X 和生存结局事件 E 后, 使用模型输出 y_pred 和生存结局事件 E 进行计算, 并且由于需要计算累加和函数 cumsum , 因此只能批量计算损失函数。

使用 Cox 比例风险损失函数优化网络参数时, 通常是使用整个数据集进行计算。但是如此 1 个 epoch (遍历一遍数据集) 只能更新一遍网络参数, 时间成本较高。本发明将训练集随机分层划分成 10 个 batch, 每个 batch 中生存结局事件 E 的比例相等, 并将每个 batch 的生存数据按生存时间 T 倒序排列用于计算损失函数。因此 1 个 epoch 有 10 个 batch 的数据用于更新网络参数, 遍历一遍数据集更新 10 次网络参数, 效率显著提升。

(2) 基于 Breslow 法结合相对风险, 计算长期风险变化情况

式(1)中等号左边 $h(t, x)$ 即描述长期风险的风险函数; 等号右边 $\beta^T x$ 即相对风险, 已由卷积生存网络计算得到; 因此, 仅需估算 $h_0(t)$ 即可结合 $\exp(\beta^T x)$ 计算描述长期风险的风险函数 $h(t, x)$ 。

在已知 β 的情况下估算基准风险函数 $h_0(t)$, 最常用的方法即 Breslow 法。Breslow 法提出 t_i 时刻的基准累积风险函数 $H_0(t_i)$ 的估计表达式:

$$H_0(t_i) = \int_0^{t_i} h_0(t) dt = \sum_{t_j < t_i} [\delta_i / \sum_{j \in R_i} \exp(\beta^T x_j)]$$

其中 R_i 为 t_i 时刻处于风险中的样本集。

本实施例使用某三甲医院终末期肾病患者连续 36 次血透过程中的收缩压记录训练模型, 使用分层的十折交叉验证评估模型准确度 (C-Index)。本实施例将终末期肾病患者 36 次血透血压记录求最大值、最小值、均值、血压变异性, 用于 DeepSurv 及传统 Cox 比例风险回归模型的输入, 并且进行十折交叉验证用于对比评估模型准确度。传统 Cox 比例风险回归模型 C-Index 为 0.646 ± 0.065 (0.95CI), DeepSurv 模型 C-Index 为 0.658 ± 0.038 (0.95CI), 本发明模型 C-Index 为 0.839 ± 0.039 (0.95CI)。可见 DeepSurv 模型较传统 Cox 比例风险回

归模型稳定且准确度 C-Index 相差无几，而本发明模型准确度 C-Index 明显优于前两者，能够实现血透并发症长期风险预测。

以上所述仅是本发明的优选实施方式，虽然本发明已以较佳实施例披露如上，然而并非用以限定本发明。任何熟悉本领域的技术人员，在不脱离本发明技术方案范围情况下，都可利用上述揭示的方法和技术内容对本发明技术方案做出许多可能的变动和修饰，或修改为等同变化的等效实施例。因此，凡是未脱离本发明技术方案的内容，依据本发明的技术实质对以上实施例所做的任何的简单修改、等同变化及修饰，均仍属于本发明技术方案保护的范围内。

权利要求书

1、一种基于卷积生存网络的血透并发症长期风险预测系统，其特征在于，该系统包括：用于采集血透患者血压信息的数据采集模块；用于对原始数据进行缺失值处理、归一化处理并规整化为二维矩阵的数据预处理模块；用于深度学习建模的学习预测模块；用于可视化输出、呈现长期风险变化情况的结果展示模块；

所述数据预处理模块的处理过程具体为：将连续血透的血压数据规整成二维矩阵，按血透治疗顺序排列，每行数据对应于一次血透治疗过程，使用 Min-Max 归一化所述二维矩阵以确保血压波形得以保留；

所述学习预测模块的处理过程包括两部分：

(1) 基于卷积生存网络的并发症相对风险预测：使用卷积神经网络架构并结合 Cox 比例风险损失函数训练卷积生存网络；所述卷积生存网络由多个卷积层和全连接层叠加而成，输入为数据预处理模块输出的二维矩阵，各卷积层将特征层层抽象，最终抽象为若干模式特征，各模式特征经全连接层输出为一个节点表示发生事件的相对风险；通过 Cox 比例风险损失函数优化网络参数；

风险函数 $h(t, x)$ 表示个体在给定时刻发生事件的概率，公式如下：

$$h(t, x) = h_0(t) \exp(\beta^T x) \quad (1)$$

其中 t 表示时间， $x = (x^1, \dots, x^p)^T$ 是 p 维协变量用于表征患者， p 为卷积生存网络最后一层全连接层的神经元个数， $\beta = (\beta^1, \dots, \beta^p)^T$ 是回归参数， $h_0(t)$ 是基准风险函数， $\beta^T x$ 是相对风险，即卷积生存网络的输出；

Cox 比例风险损失函数公式如下：

$$L(\beta) = - \sum_{i=1}^n \delta_i (\beta^T x_i - \log \sum_{j: T_j > T_i} e^{\beta^T x_j}) \quad (2)$$

其中 n 指患者个数； δ_i 是结局事件 E 的指示符，患者个体 x_i 发生事件则 $\delta_i = 1$ ，未发生事件则 $\delta_i = 0$ ； T_i 及 T_j 分别表示患者个体 x_i 和 x_j 的生存时间；

(2) 基于 Breslow 法结合相对风险，计算长期风险变化情况，具体地：基于 Breslow 法得到 t_i 时刻的基准累积风险函数 $H_0(t_i)$ ，根据 $H_0(t_i)$ 得到 $h_0(t)$ ， $H_0(t_i)$ 的估计表达式如下：

$$H_0(t_i) = \int_0^{t_i} h_0(t) dt = \sum_{t_j < t_i} \left[\frac{\delta_i}{\sum_{j \in R_i} \exp(\beta^T x_j)} \right] \quad (3)$$

其中 R_i 为 t_i 时刻处于风险中的样本集；

根据基准风险函数 $h_0(t)$ 结合卷积生存网络计算得到的相对风险 $\beta^T x$ ，根据式(1)计算描述

长期风险的风险函数 $h(t, x)$ 。

2、根据权利要求1所述的一种基于卷积生存网络的血透并发症长期风险预测系统，其特征在于，所述数据预处理模块首先筛除收缩压小于 60mmHg 或大于 250mmHg 的数据，之后筛选连续 36 次血透的收缩压数据规整成二维矩阵，按血透治疗顺序排列，每行数据对应于一次时长 5 小时的血透治疗过程，列与列之间间隔 10 分钟，总计 36 行，每行 30 个点。

3、根据权利要求1所述的一种基于卷积生存网络的血透并发症长期风险预测系统，其特征在于，所述学习预测模块中，卷积生存网络接收数据预处理模块输出的二维矩阵，首先识别一维血压模式，然后将感知野补成方形，将特征层层抽象后接全局均值池化层获得若干模式特征，各模式特征经全连接层输出为一个节点表示发生事件的相对风险；最后一层激活函数为 Linear，其余层激活函数为 Relu。

4、根据权利要求2所述的一种基于卷积生存网络的血透并发症长期风险预测系统，其特征在于，所述学习预测模块中，卷积生存网络的输入为数据预处理模块输出的 1 通道 36×30 大小的二维矩阵，通过 16 个 1×5 的卷积识别一维血压模式，然后通过 32 个纵向 5×1 的卷积操作将感知野补成方形，之后接 2 层 16 个 5×5 的卷积核、1 层 9 个 3×3 的卷积核将特征层层抽象，再之后接全局均值池化层获得 9 个模式特征值，9 个模式特征值经全连接层输出为一个节点表示发生事件的相对风险。

5、根据权利要求1所述的一种基于卷积生存网络的血透并发症长期风险预测系统，其特征在于，所述学习预测模块中，Cox 比例风险损失函数的计算流程如下：

(a) 将患者的特征数据 X 和生存结局事件 E 统一按生存时间 T 降序排列，形成矩阵 M ；

(b) y_{i_pred} 为卷积生存网络输出，对应于 $\beta^T x_i$ ；由于 M 按 T 降序排列， $\sum_{j: T_j > T_i} e^{\beta^T x_j}$ 为矩阵 M 的前 i 行通过卷积生存网络后输出的指数累加和，即前 i 行 $e^{y_{i_pred}}$ 的累加和，令累计求和函数记为 cumsum；

(c) δ_i 即矩阵 M 第 i 行个体生存结局事件 E 的取值，记作 E_i ；

(d) 令求和函数记为 sum，由此损失函数表示为：

$$\text{loss} = -\text{sum}(E_i * (y_{i_pred} - \log(\text{cumsum}(y_{i_pred}))))。$$

6、根据权利要求5所述的一种基于卷积生存网络的血透并发症长期风险预测系统，其特征在于，所述学习预测模块中，为了使训练集和测试集对应的损失函数取值为同一个量级，将求和函数 sum 替换为求均值函数 mean，损失函数表达为：

$$\text{loss} = -\text{mean}(E_i * (y_{i_pred} - \log(\text{cumsum}(y_{i_pred}))))。$$

7、根据权利要求1所述的一种基于卷积生存网络的血透并发症长期风险预测系统，其特

征在于，所述学习预测模块中，通过 Cox 比例风险损失函数优化网络参数时，将训练集随机分层划分成 10 个 batch，每个 batch 中生存结局事件 E 的比例相等，并将每个 batch 的生存数据按生存时间 T 倒序排列用于计算损失函数，由此遍历一遍数据集更新 10 次网络参数。

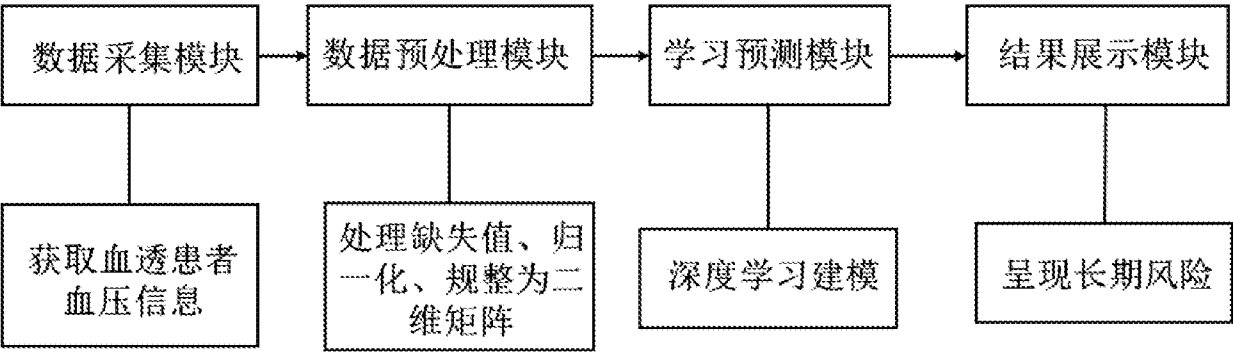


图 1

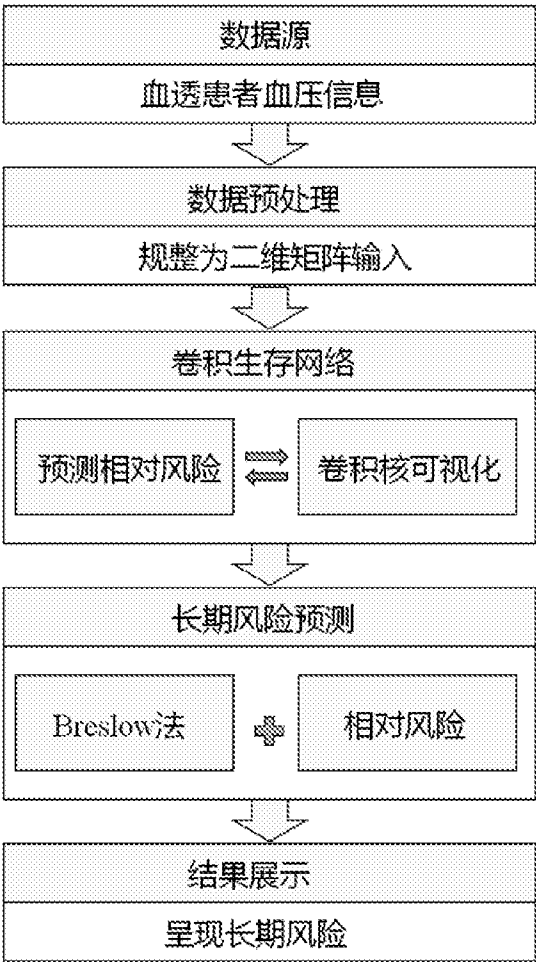


图 2

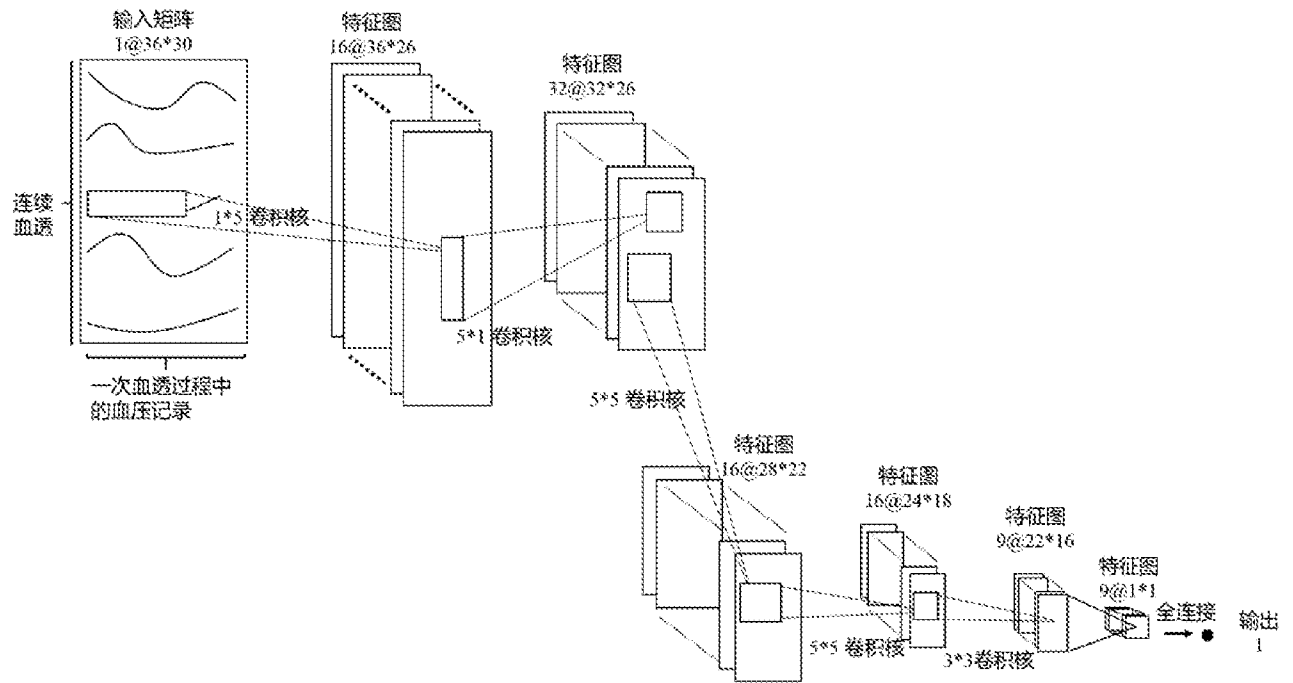


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/113135

A. CLASSIFICATION OF SUBJECT MATTER

G16H 50/30(2018.01)i; G16H 50/20(2018.01)i; G06N 3/04(2006.01)i; G06N 3/08(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G16H50/-;G06N3/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, CNKI: 卷积神经网络, cox, 比例风险, 回归, breslow, 血透, 血压; VEN, USTXT, IEEE: CNN, convolutional, proportional hazards, cox, risk, regression, breslow, hemodialysis, blood pressure

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 112530594 A (ZHEJIANG LAB) 19 March 2021 (2021-03-19) claims 1-7	1-7
Y	王丰 (WANG, Feng). "终末期肾病血透血压模式识别及预后分析方法研究 (Study on Hemodialysis Blood Pressure Patterns Recognition and Prognosis Analysis in Patients with End Stage Renal Disease)" 《中国优秀博士学位论文全文数据库(博士) 医药卫生科技辑》 (Chinese Doctoral Dissertations Full-text Database, Medicine & Public Health), No. 01, 15 January 2021 (2021-01-15), ISSN: 1674-022X, chapters 4.2.3, 5	1-7
Y	CN 111582577 A (SI-TECH INFORMATION TECHNOLOGY LTD.) 25 August 2020 (2020-08-25) description, paragraphs [0052]-[0055]	1-7
A	CN 112289455 A (WANG, Zhi) 29 January 2021 (2021-01-29) entire document	1-7
A	WO 2020077352 A1 (HUMAN LONGEVITY INC.) 16 April 2020 (2020-04-16) entire document	1-7



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

26 September 2021

Date of mailing of the international search report

26 October 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088, China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/113135

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112530594	A	19 March 2021	CN	112530594	B	11 May 2021
CN	111582577	A	25 August 2020	None			
CN	112289455	A	29 January 2021	None			
WO	2020077352	A1	16 April 2020	AU	2019356597	A1	20 May 2021
				CA	3115991	A1	16 April 2020

国际检索报告

国际申请号

PCT/CN2021/113135

A. 主题的分类

G16H 50/30(2018.01)i; G16H 50/20(2018.01)i; G06N 3/04(2006.01)i; G06N 3/08(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G16H50/-;G06N3/-

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS、CNTXT、CNKI: 卷积神经网络、cox、比例风险、回归、breslow、血透、血压; VEN、USTXT、IEEE: CNN、convolutional、proportional hazards、cox、risk、regression、breslow、hemodialysis、blood pressure

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 112530594 A (之江实验室) 2021年 3月 19日 (2021 - 03 - 19) 权利要求1-7	1-7
Y	王丰. “终末期肾病血透血压模式识别及预后分析方法研究” 《中国优秀博硕士学位论文全文数据库(博士) 医药卫生科技辑》, 第01期, 2021年 1月 15日 (2021 - 01 - 15), ISSN: 1674-022X, 第4.2.3、5章	1-7
Y	CN 111582577 A (北京思特奇信息技术股份有限公司) 2020年 8月 25日 (2020 - 08 - 25) 说明书第[0052]-[0055]段	1-7
A	CN 112289455 A (王智) 2021年 1月 29日 (2021 - 01 - 29) 全文	1-7
A	W0 2020077352 A1 (HUMAN LONGEVITY INC) 2020年 4月 16日 (2020 - 04 - 16) 全文	1-7

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2021年 9月 26日

国际检索报告邮寄日期

2021年 10月 26日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)
中国 北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

朱颖

电话号码 (86-28) 62967642

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/113135

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	112530594	A	2021年 3月 19日	CN 112530594 B	2021年 5月 11日
CN	111582577	A	2020年 8月 25日	无	
CN	112289455	A	2021年 1月 29日	无	
WO	2020077352	A1	2020年 4月 16日	AU 2019356597 A1	2021年 5月 20日
				CA 3115991 A1	2020年 4月 16日