

(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
2020 年 11 月 26 日 (26.11.2020)



(10) 国际公布号
WO 2020/233254 A1

- (51) 国际专利分类号:
G16H 50/70 (2018.01) **G16H 50/20** (2018.01)
- (21) 国际申请号: PCT/CN2020/083584
- (22) 国际申请日: 2020 年 4 月 7 日 (07.04.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910629798.9 2019年7月12日 (12.07.2019) 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)。
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)。 吴君雅(WU, Junya); 中国浙江省杭州市余杭区文一西路 1818 号, Zhejiang 310023 (CN)。 胡佩君(HU, Peijun); 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玉古路 147 号黄鸿年科技综合大楼 215 室/刘静, Zhejiang 310013 (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, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,

(54) Title: MEDICAL DATA ANALYSIS SYSTEM INTEGRATING STRUCTURED IMAGE DATA

(54) 发明名称: 一种融合结构化影像数据的医疗数据分析系统

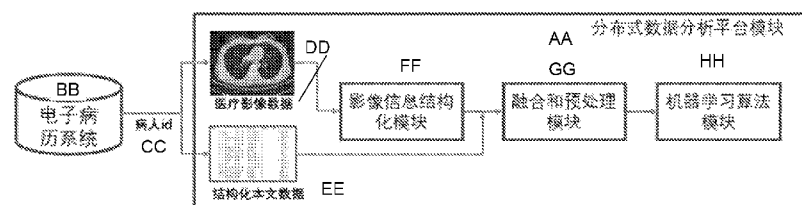


图 1

AA DISTRIBUTED-TYPE DATA ANALYSIS PLATFORM MODULE
BB ELECTRONIC MEDICAL RECORDS SYSTEM
CC PATIENT ID
DD MEDICAL IMAGE DATA
EE STRUCTURED TEXT DATA
FF IMAGE INFORMATION STRUCTURING MODULE
GG INTEGRATION AND PREPROCESSING MODULE
HH MACHINE LEARNING ALGORITHM MODULE

(57) Abstract: A medical data analysis system integrating structured image data. The system comprises an image information structuring module, an integration and preprocessing module, and a machine learning algorithm module. According to the invention, on the basis of a distributed data analysis platform using Spark and Hadoop, and to address the difficulty of both integrating medical image data with structured data residing in electronic medical records and analyzing same, an image information structuring module is used to analyze medical image data by means of computer vision technology and used to perform structuring conversion on key information. Said data is then integrated with structured data of the relevant patient, the structured data including relevant diagnostic information and demographic information from an electronic medical records system; missing value processing and classification data conversion is performed by means of a data preprocessing module; and data analysis and results visualization are carried out using a machine learning algorithm module. A highly efficient medical data analysis system is thus formed. The invention increases the utilization rate of various types of multi-dimensional medical data, and satisfies the various topical research requirements of researchers.

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。

(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))。
- 在修改权利要求的期限届满之前进行, 在收到该
修改后将重新公布(细则48.2(h))。
- 根据申请人的请求, 在条约第21条(2)(a)所规定的
期限届满之前进行。

(57) 摘要: 一种融合结构化影像数据的医疗数据分析系统, 该系统包括影像信息结构化模块、融合与预处理模块、机器学习算法模块, 其基于Spark和Hadoop实现的分布式数据分析平台, 针对医学影像数据难以和医疗电子病历中结构化数据融合分析的难点, 利用影像信息结构化模块通过计算机视觉技术对医学影像数据分析并进行关键信息的结构化转化, 融合电子病历系统中同一病人的其他诊断信息、人口统计学信息等结构化数据, 通过数据预处理模块进行缺失值处理和分类型数据转化, 结合机器学习算法模块进行数据分析和结果可视化, 形成高效率的医疗数据分析系统, 提高多种类多维度医疗数据的利用率, 可满足研究人员不同的课题研究需求。

一种融合结构化影像数据的医疗数据分析系统

技术领域

本发明属于医疗数据分析领域，尤其涉及一种融合结构化影像数据的医疗数据分析系统。

背景技术

生物医学数据广泛涉及人类健康相关的各个领域，随着新型生物/医学技术应用的普及，医疗电子病历数据的量增长迅速，呈现种类繁多、高维度多模式、内涵复杂丰富等特点，特别是生命科学组学数据、医疗影像数据等高维度非结构化医疗数据要求快速特异性分析处理能力。现有技术方案针对大数据平台实现基于 Hadoop 开源软件的分布式存储和并行汇总计算，提升存取能力，提高分析任务的有效调度和处理效率。

医疗电子病历数据包括人口统计学数据、临床诊断数据、检查数据、影像数据等，具有维度高、种类复杂多样、结构复杂等特点，故现有处理方法大多只是针对单一类型数据（如眼底视网膜检查数据、特定部位 CT 扫描数据等）进行针对性地疾病科学研究，可能存在数据量小、分析速度慢等缺点，且数据收集周期较长，很难满足准确快速分析的要求。

发明内容

本发明的目的在于针对现有生物医学数据分析模式的局限性，提出了一种融合结构化影像数据的医疗数据分析系统。

本发明是通过以下技术方案来实现的：一种融合结构化影像数据的医疗数据分析系统，该系统包括影像信息结构化模块、融合与预处理模块和机器学习算法模块：

影像信息结构化模块：该模块根据病人id从医院电子病历系统中筛选相对应的医疗影像数据，采用计算机视觉技术，利用U-Net卷积神经网络框架实现端到端的图像分割模型，以单张图像作为模型输入并得到与原图像大小相同的分割结果图，其中目标分割区域的像素点值为1，其余像素点值为0，以等距离划分原图像的长和宽，得到等面积的多个子区域，并对每个子区域赋予方位编号；根据分割结果图判断是否有目标区域，计算各个目标区域 R_i 的面积 S_i 、中心点坐标 (x_i, y_i) 、横向最长距离 W_i 和纵向最长距离 H_i ，并获取目标区域 R_i 的方位编号 P_i ，将 S_i 、 (x_i, y_i) 、 W_i 、 H_i 、 P_i 作为特征与病人id进行关联，形成影像信息结构化数据表。

融合与预处理模块：该模块根据影像信息结构化数据表中存在的病人id，从电子病历系统中筛选出这些病人id对应的结构化数据，以id对应的方式将结构化数据拼接到影像信息结构化数据表中，实现同一个病人的非结构化数据和结构化数据的融合；对于融合后的影像信息结构化数据表中的每个病人id，计算该病人id所有特征的缺失率，对缺失率超过阈值的病人id

及其对应数据进行剔除，对未超过阈值的缺失特征进行补全。

机器学习算法模块：该模块预设若干机器学习算法，将机器学习算法封装成函数形式，由用户自行选择机器学习算法并设定算法参数；该模块接收融合与预处理模块输出的数据表，将数据表转换为dataframe格式数据，和用户设定算法参数共同作为用户选择函数的输入，完成数据分析，并以图表形式对分析结果进行可视化展示。

进一步地，所述影像信息结构化模块，以原图像中心点为中心，将原图像等面积分割成四个子区域，并给定方位编号1、2、3、4，分割结果图映射到原图像上，按目标区域 R_i 的中心点坐标 (x_i, y_i) 所在子区域的方位编号作为该目标区域 R_i 的方位编号。

进一步地，所述影像信息结构化模块中，利用像素点坐标信息计算各个目标区域 R_i 的 S_i 、 (x_i, y_i) 、 W_i 、 H_i ，计算公式如下：

$$S_i = \frac{p^2 * N_i}{H * W}$$

$$x_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$$

$$y_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}$$

$$W_i = \frac{p * N_{wi}}{W}$$

$$H_i = \frac{p * N_{hi}}{H}$$

其中 i 为目标区域的编号 $i = 1, 2, 3, \dots$ ， N_i 为目标区域 R_i 的像素数， (x_{ij}, y_{ij}) 为 R_i 中每个像素点的坐标， j 为 R_i 中像素点编号， H ， W 分别为原图像的高和宽， N_{wi} 为目标区域 R_i 的横向最长像素数， N_{hi} 为目标区域 R_i 的纵向最长像素数， p 为图像中像素与真实长度比例尺。

进一步地，所述影像信息结构化模块中，加入图形分析单元（GPU）加速图像分析速度，并通过python高性能计算库Numba实现Spark集群和GPU的联合使用。

进一步地，所述融合与预处理模块中，对未超过阈值的缺失特征进行补全具体为：

对于数值型特征，生成随机数 x 实现数据插值补全， $x \in [x_{\min}, x_{\max}]$ ， $x_{\min}$ 、 $x_{\max}$ 分别为该特征在数据表中的最小值和最大值；

对于分类型特征，随机取该特征在数据表中出现的类别进行数据插值补全。

进一步地，所述融合与预处理模块中，对于分类型特征，在特征补全后，采用OneHotEncoder函数映射，将该特征的每个元素转化为一个可以用来计算的数值，即OneHot型数据，使之适用于通用机器学习算法。

进一步地，所述机器学习算法模块中的机器学习算法包括线性回归、逻辑回归、支持向量机、随机森林等。

进一步地，该系统还包括分布式数据分析平台：该平台由三台服务器构成，通过部署Hadoop和Spark集群实现一台主服务器（Master）和两台从服务器（Slave）间数据的分布式存取和任务调度及并行化计算。

本发明的有益效果是：本发明在实际应用过程中具有计算效率高、数据利用率大、操作简单且普遍适用于医疗数据分析课题的研究，可以达到实用的水平。具体如下：

1) 根据研究者对生物医学数据可能存在的数据要求和分析要求，本发明针对医院电子病历系统中的疾病检查诊断、人口统计学等结构化数据和医疗影像数据等非结构化数据综合，可提高生物医学数据的利用率。

2) 针对医学影像数据难以和医疗电子病历中结构化数据融合分析的难点，利用计算机视觉技术将医学图像关键信息进行结构化转化，并融合电子病历系统中相同id的其他数据，通过数据预处理模块，利用机器学习算法综合分析数据，形成高效率高利用率的医疗数据分析系统，可满足研究人员不同的课题研究需求。

3) 采用Spark和Hadoop开源软件实现数据分布式存取和并行汇总计算以及计算资源有效调度，在影像信息结构化模块加入图形处理单元（GPU）提高高维度高容量数据处理速度，综合提高分析处理速度，更好地实现数据处理的高效性。

附图说明

图 1 医疗数据分析系统流程图；

图 2 影像信息结构化模块示意图；

图 3 原图像方位编码示意图；

图 4 融合与预处理模块示意图；

图 5 机器学习算法模块示意图。

具体实施方式

下面结合附图和具体实施例对本发明作进一步详细说明。

如图 1 所示，本发明提供一种融合结构化影像数据的医疗数据分析系统，该系统包括：

一、影像信息结构化模块

如图2所示，该模块根据病人id从医院电子病历系统中筛选相对应的医疗影像数据（非结构化数据），采用计算机视觉技术，利用基于Olaf Ronneberger的U-Net卷积神经网络框架实现端到端的图像分割模型，以单张图像作为模型输入并得到与原图像大小相同的分割结果图，其中目标分割区域的像素点值为1，其余像素点值为0，以等距离划分原图像的长和宽，得到

等面积的多个子区域，并对每个子区域赋予方位编号；以下为一个方位编号的实例：

如图3所示，以原图像中心点为中心，将原图像等面积分割成四个子区域，并给定方位编号1、2、3、4，分割结果图映射到原图像上，按目标区域 R_i 的中心点坐标 (x_i, y_i) 所在子区域的方位编号作为该目标区域 R_i 的方位编号。

根据分割结果图判断是否有目标区域，计算各个目标区域 R_i 的面积 S_i 、中心点坐标 (x_i, y_i) 、横向最长距离 W_i 和纵向最长距离 H_i ，并获取目标区域 R_i 的方位编号 P_i ，将 S_i 、 (x_i, y_i) 、 W_i 、 H_i 、 P_i 作为特征与病人id进行关联，形成影像信息结构化数据表。

利用像素点坐标信息计算各个目标区域 R_i 的 S_i 、 (x_i, y_i) 、 W_i 、 H_i ，计算公式如下：

$$S_i = \frac{p^2 * N_i}{H * W}$$

$$x_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$$

$$y_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}$$

$$W_i = \frac{p * N_{wi}}{W}$$

$$H_i = \frac{p * N_{hi}}{H}$$

其中 i 为目标区域的编号 $i = 1, 2, 3, \dots$ ， N_i 为目标区域 R_i 的像素数， (x_{ij}, y_{ij}) 为 R_i 中每个像素点的坐标(j 为 R_i 中像素点编号)， H 、 W 分别为原图像的高和宽， N_{wi} 为目标区域 R_i 的横向最长像素数， N_{hi} 为目标区域 R_i 的纵向最长像素数， p 为图像中像素与真实长度比例尺。

为克服图像数据容量大、处理速度慢的缺点，并保证在研究人员的单次数据分析研究中图像分析速度与结构型数据分析速度差在可接受范围内，该模块加入图形分析单元（GPU）加速图像分析速度，并通过python高性能计算库Numba实现Spark集群和GPU的联合使用。

二、融合与预处理模块

如图4所示，该模块根据影像信息结构化数据表中存在的病人id，从电子病历系统中筛选出这些病人id对应的结构化数据（结构化数据可以包括多个特征，例如分类型特征、数值型特征），以id对应的方式将结构化数据拼接到影像信息结构化数据表中，实现同一个病人的非结构化数据和结构化数据的融合；对于融合后的影像信息结构化数据表中的每个病人id，计算该病人id所有特征的缺失率，对缺失率超过阈值（阈值可设置为60%）的病人id及其对应数据进行剔除，对未超过阈值的缺失特征进行补全。

对未超过阈值的缺失特征进行补全可以采用以下方式：

对于数值型特征，生成随机数 x 实现数据插值补全， $x \in [x_{\min}, x_{\max}]$ ， $x_{\min}, x_{\max}$ 分别为该特征在数据表中的最小值和最大值；

对于分类型特征，随机取该特征在数据表中出现的类别进行数据插值补全。

对于分类型特征，在特征补全后，可以采用OneHotEncoder函数映射，将该特征的每个元素（即每个病人id对应该特征的数据）转化为一个可以用来计算的数值，即OneHot型数据，使之适用于通用机器学习算法。

三、机器学习算法模块

如图5所示，该模块预设若干机器学习算法，机器学习算法包括线性回归、逻辑回归、支持向量机、随机森林等；将机器学习算法封装成函数形式，由用户自行选择机器学习算法并设定算法参数；该模块接收融合与预处理模块输出的数据表，将数据表转换为dataframe格式数据，和用户设定算法参数共同作为用户选择函数的输入，在分布式数据分析平台完成数据分析，并以图表形式对分析结果进行可视化展示。

四、分布式数据分析平台

该系统还可以包括分布式数据分析平台，该平台由三台服务器构成，通过部署Hadoop和Spark集群实现一台主服务器（Master）和两台从服务器（Slave）间数据的分布式存取和任务调度及并行化计算。

以下给出一个本发明医疗数据分析系统的具体使用场景：

当用户发起分析请求时如研究药物A对肺结核治疗的影响，即预测用药持续一年后肺结核是否痊愈，其过程如下：

（1）首先在医院电子病历系统及相关联数据库系统中筛选诊断为肺结核且使用了药物A和相同条件下未使用药物A的病人，根据病人id号整理成相对应的结构化文本数据和胸部CT影像数据，存储在分布式数据分析平台；

（2）影像信息结构化模块：该类型胸部CT图像大小为512*512，每幅图像经过已训练好的U-Net卷积神经网络模型，输出相应的分割结果图，根据像素点计数， S_1 像素个数 N_1 为1144， S_2 像素个数 N_2 为8876， p 为21.4mm计算可得：

$$S_1 = \frac{p^2 * N_1}{H * W} = \frac{458 * 1144}{512 * 512} = 2mm^2,$$

$$S_2 = \frac{p^2 * N_2}{H * W} = \frac{458 * 8876}{512 * 512} = 9mm^2$$

选择最大分割区域面积， $N_{w2}=72$ ， $N_{h2}=84$ ，继续计算：

$$W_2 = \frac{p * N_{w2}}{W} = \frac{21.4 * 72}{512} = 3mm$$

$$H_i = \frac{p * N_{hi}}{H} = \frac{21.4 * 84}{512} = 3.5mm$$

对于同一个病人 id 的不同 CT 图像分析所得结果，选择 S_i 最大的分析结果，根据设计好的特征名称，对所有病人 id 的图像数据分析完成后形成结构化信息数据。

(3) 融合与预处理模块：相同病人 id 对应的结构化文本数据以 dataframe 格式读取，数据内容包括 patientId (病人 id)、是否用药、性别、年龄、是否吸烟、是否喝酒等，根据病人 id 匹配融合影像结构化信息数据和结构化文本数据，计算各个病人 id 对应的数据缺失率，对于缺失率超过 60%的数据进行剔除，缺失率低于 60%的缺失数据进行中位数插值，对于如是否用药等字符型数据根据“是”——“[1,0]”、“否”——“[0,1]”规则进行 OneHotEncoder 函数映射，形成统一数值型的可计算数据。

(4) 机器学习算法模块：针对预测用药持续一年后肺结核痊愈的准确性问题，选择逻辑回归算法，设定最大迭代次数为1000等参数，将预处理后的数据以dataframe的格式读取，并通过接口调用的方式将数据、算法模型、算法参数等传入分析函数中，待计算完成后，得到分析结果，如性别统计饼图、年龄分布统计柱状图、预测准确率0.94，和可视化图像PR曲线、POC曲线，至此完成系统分析流程。

以上仅为本发明的实施实例，并非用于限定本发明的保护范围。凡在本发明的精神和原则之内，不经过创造性劳动所作的任何修改、等同替换、改进等，均包含在本发明的保护范围内。

权利要求书

1、一种融合结构化影像数据的医疗数据分析系统，其特征在于，该系统包括：

(1) 影像信息结构化模块：该模块根据病人id从医院电子病历系统中筛选相对应的医疗影像数据，采用计算机视觉技术，利用U-Net卷积神经网络框架实现端到端的图像分割模型，以单张图像作为模型输入并得到与原图像大小相同的分割结果图，其中目标分割区域的像素点值为1，其余像素点值为0，以等距离划分原图像的长和宽，得到等面积的多个子区域，并对每个子区域赋予方位编号；根据分割结果图判断是否有目标区域，计算各个目标区域 R_i 的面积 S_i 、中心点坐标 (x_i, y_i) 、横向最长距离 W_i 和纵向最长距离 H_i ，并获取目标区域 R_i 的方位编号 P_i ，将 S_i 、 (x_i, y_i) 、 W_i 、 H_i 、 P_i 作为特征与病人id进行关联，形成影像信息结构化数据表。

(2) 融合与预处理模块：该模块根据影像信息结构化数据表中存在的病人id，从电子病历系统中筛选出这些病人id对应的结构化数据，以id对应的方式将结构化数据拼接到影像信息结构化数据表中，实现同一个病人的非结构化数据和结构化数据的融合；对于融合后的影像信息结构化数据表中的每个病人id，计算该病人id所有特征的缺失率，对缺失率超过阈值的病人id及其对应数据进行剔除，对未超过阈值的缺失特征进行补全。

(3) 机器学习算法模块：该模块预设若干机器学习算法，将机器学习算法封装成函数形式，由用户自行选择机器学习算法并设定算法参数；该模块接收融合与预处理模块输出的数据表，将数据表转换为dataframe格式数据，和用户设定算法参数共同作为用户选择函数的输入，完成数据分析，并以图表形式对分析结果进行可视化展示。

2、根据权利要求1所述一种融合结构化影像数据的医疗数据分析系统，其特征在于：所述影像信息结构化模块，以原图像中心点为中心，将原图像等面积分割成四个子区域，并给定方位编号1、2、3、4，分割结果图映射到原图像上，按目标区域 R_i 的中心点坐标 (x_i, y_i) 所在子区域的方位编号作为该目标区域 R_i 的方位编号。

3、根据权利要求1所述一种融合结构化影像数据的医疗数据分析系统，其特征在于：所述影像信息结构化模块中，利用像素点坐标信息计算各个目标区域 R_i 的 S_i 、 (x_i, y_i) 、 W_i 、 H_i ，计算公式如下：

$$S_i = \frac{p^2 * N_i}{H * W}$$

$$x_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$$

$$y_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}$$

$$W_i = \frac{p * N_{wi}}{W}$$

$$H_i = \frac{p * N_{hi}}{H}$$

其中 i 为目标区域的编号 $i = 1, 2, 3, \dots$ ， N_i 为目标区域 R_i 的像素数， (x_{ij}, y_{ij}) 为 R_i 中每个像素点的坐标(j 为 R_i 中像素点编号)， H ， W 分别为原图像的高和宽， N_{wi} 为目标区域 R_i 的横向最长像素数， N_{hi} 为目标区域 R_i 的纵向最长像素数， p 为图像中像素与真实长度比例尺。

4、根据权利要求1所述一种融合结构化影像数据的医疗数据分析系统，其特征在于：所述影像信息结构化模块中，加入图形分析单元（GPU）加速图像分析速度，并通过python高性能计算库Numba实现Spark集群和GPU的联合使用。

5、根据权利要求1所述一种融合结构化影像数据的医疗数据分析系统，其特征在于：所述融合与预处理模块中，对未超过阈值的缺失特征进行补全具体为：

对于数值型特征，生成随机数 x 实现数据插值补全， $x \in [x_{\min}, x_{\max}]$ ， $x_{\min}, x_{\max}$ 分别为该特征在数据表中的最小值和最大值；

对于分类型特征，随机取该特征在数据表中出现的类别进行数据插值补全。

6、根据权利要求1所述一种融合结构化影像数据的医疗数据分析系统，其特征在于：所述融合与预处理模块中，对于分类型特征，在特征补全后，采用OneHotEncoder函数映射，将该特征的每个元素转化为一个可以用来计算的数值，即OneHot型数据，使之适用于通用机器学习算法。

7、根据权利要求1所述一种融合结构化影像数据的医疗数据分析系统，其特征在于：所述机器学习算法模块中的机器学习算法包括线性回归、逻辑回归、支持向量机、随机森林等。

8、根据权利要求1所述一种融合结构化影像数据的医疗数据分析系统，其特征在于：该系统还包括分布式数据分析平台：该平台由三台服务器构成，通过部署Hadoop和Spark集群实现一台主服务器（Master）和两台从服务器（Slave）间数据的分布式存取和任务调度及并行化计算。

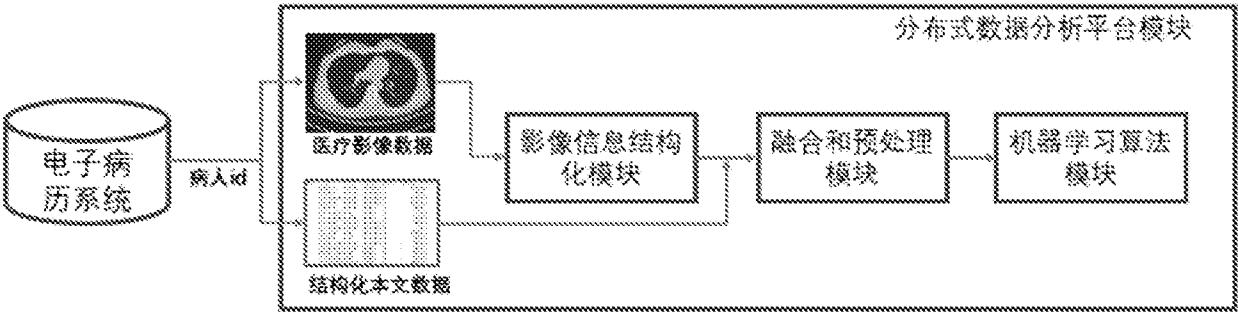
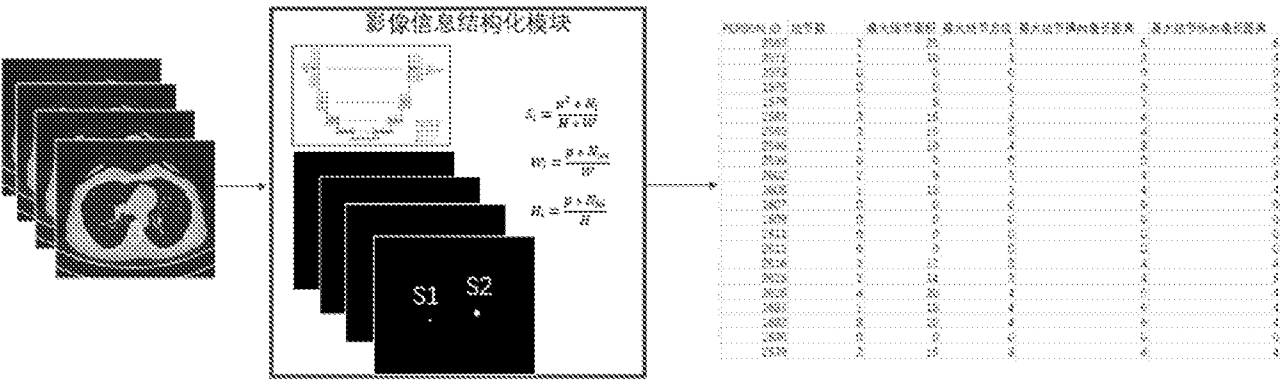


图 1



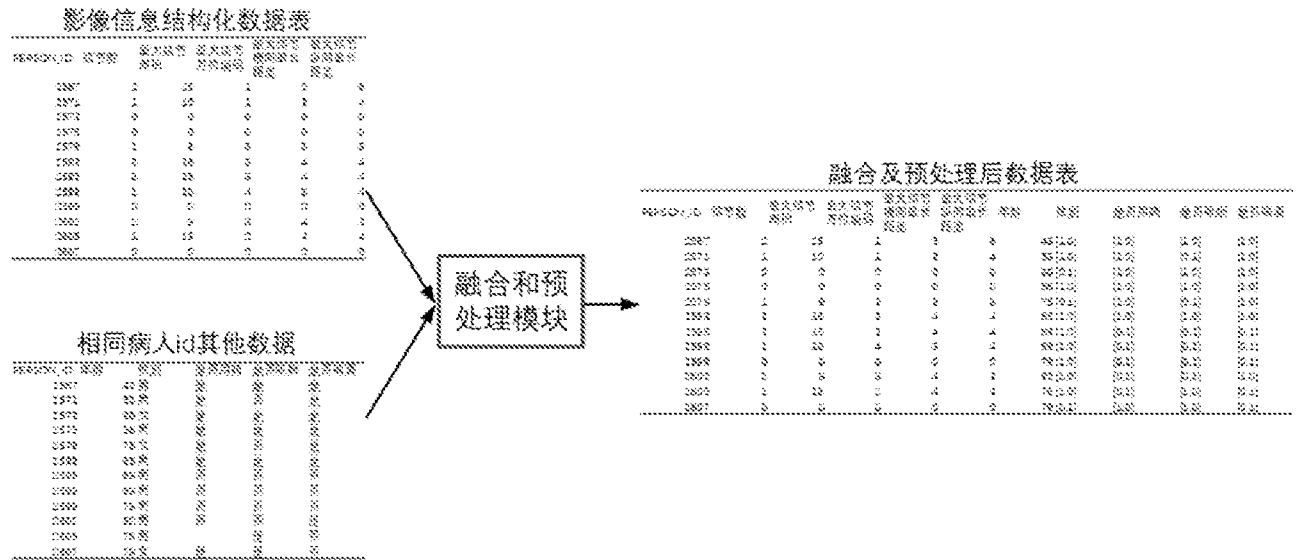


图 4

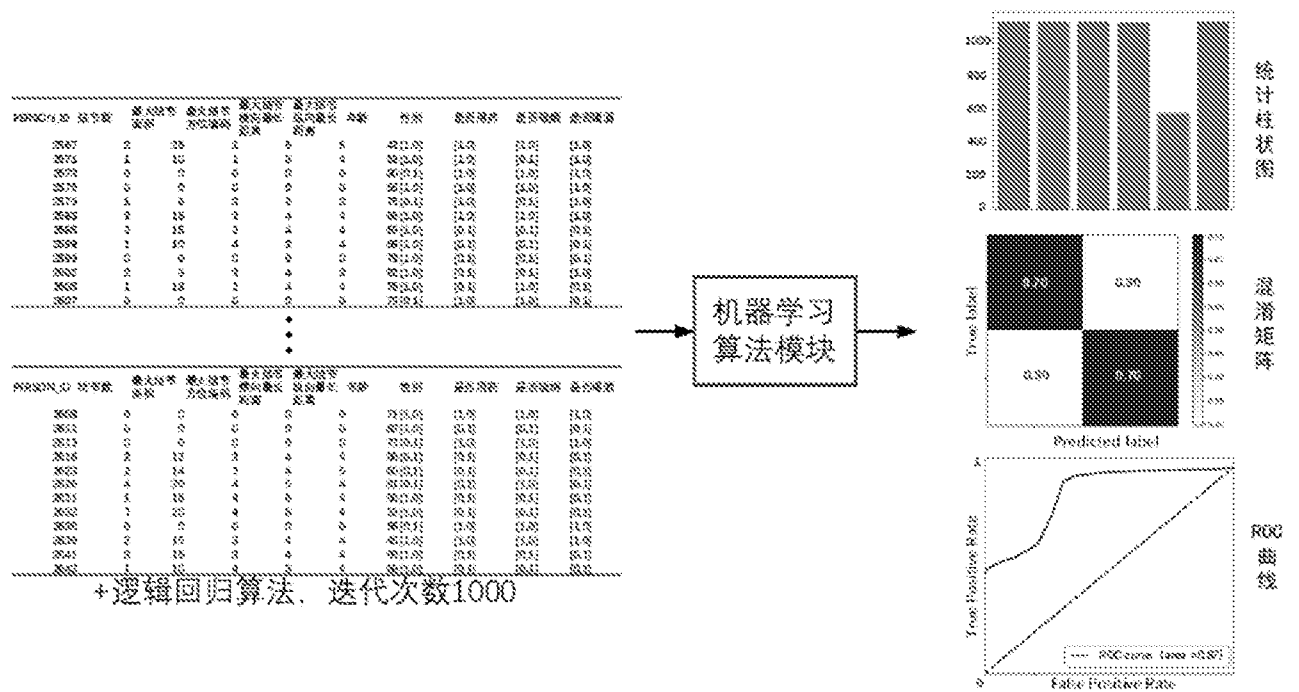


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/083584

A. CLASSIFICATION OF SUBJECT MATTER

G16H 50/70(2018.01)i; G16H 50/20(2018.01)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)

G06H; G06N; G06T

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)

EPODOC, WPI, CNPAT, IEEE, CNKI: 之江实验室, 融合, 合并, 医, 影像, 图像, 数据, 分析, 结构化, 格式化, 标准化, 卷积神经网络, 分割, 目标, 方位, 面积, 横向, 纵向, 最, 像素, 预处理, 算法, merging, fusion, medical, image, data, analysis, structured, RRN, divid+, target

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110349652 A (ZHEJIANG LAB) 18 October 2019 (2019-10-18) claims 1-8	1-8
Y	CN 109830303 A (SHANGHAI TRIMAN INFORMATION & TECHNOLOGY CO., LTD.) 31 May 2019 (2019-05-31) description, paragraphs [0007]-[0029], [0037]-[0103], [0123]-[0139], [0150]-[0163]	1-8
Y	CN 106846317 A (LINKINGMED CORP) 13 June 2017 (2017-06-13) description, paragraphs [0075]-[0127]	1-8
A	CN 1613070 A (SIEMENS MEDICAL SOLUTIONS USA, INC. et al.) 04 May 2005 (2005-05-04) entire document	1-8
A	CN 109544561 A (HANGZHOU DIYINGJIA TECH CO., LTD.) 29 March 2019 (2019-03-29) entire document	1-8



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

15 June 2020

Date of mailing of the international search report

23 June 2020

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/CN2020/083584

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110349652	A	18 October 2019	None			
CN	109830303	A	31 May 2019	None			
CN	106846317	A	13 June 2017	None			
CN	1613070	A	04 May 2005	EP	1440412	A2	28 July 2004
				US	2003120514	A1	26 June 2003
				EP	1440409	A2	28 July 2004
				JP	2005508544	A	31 March 2005
				WO	03040965	A2	15 May 2003
				WO	03040964	A3	22 January 2004
				CN	1636210	A	06 July 2005
				US	9165116	B2	20 October 2015
				WO	03040966	A2	15 May 2003
				CN	100449531	C	07 January 2009
				WO	03040965	A3	19 February 2004
				US	7744540	B2	29 June 2010
				JP	2005508557	A	31 March 2005
				US	8280750	B2	02 October 2012
				US	7711404	B2	04 May 2010
				US	2003120134	A1	26 June 2003
				CA	2465533	A1	15 May 2003
				JP	2005509217	A	07 April 2005
				EP	1440385	A2	28 July 2004
				WO	03040878	A8	01 April 2004
				CA	2465712	A1	15 May 2003
				US	2003120133	A1	26 June 2003
				US	2003125988	A1	03 July 2003
				CA	2465531	A1	15 May 2003
				CN	1613087	A	04 May 2005
				WO	03040987	A2	15 May 2003
				JP	2005534082	A	10 November 2005
				CN	1613069	A	04 May 2005
				WO	03040990	A2	15 May 2003
				US	2003130871	A1	10 July 2003
				US	2003120458	A1	26 June 2003
				EP	1442415	A2	04 August 2004
				WO	03040988	A2	15 May 2003
				WO	03040990	A3	08 April 2004
				CA	2465725	A1	15 May 2003
				EP	1440387	A2	28 July 2004
				EP	1440390	A2	28 July 2004
				WO	03040987	A3	06 May 2004
				WO	03040988	A3	12 February 2004
				CA	2465706	A1	15 May 2003
				CA	2465760	A1	15 May 2003
				WO	03040989	A2	15 May 2003
				US	8214225	B2	03 July 2012
				US	2009259487	A1	15 October 2009
				CA	2464374	A1	15 May 2003
				EP	1440388	A2	28 July 2004
				US	2003125985	A1	03 July 2003

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/083584

Patent document cited in search report		Publication date (day/month/year)		Patent family member(s)		Publication date (day/month/year)	
				US	2015100352	A1	09 April 2015
				US	8214224	B2	03 July 2012
				JP	2005523490	A	04 August 2005
				EP	1440389	A2	28 July 2004
CN	109544561	A	29 March 2019	None			

国际检索报告

国际申请号

PCT/CN2020/083584

A. 主题的分类

G16H 50/70(2018.01)i; G16H 50/20(2018.01)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)

G06H; G06N; G06T

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))

EP0DOC, WPI, CNPAT, IEEE, CNKI: 之江实验室, 融合, 合并, 医, 影像, 图像, 数据, 分析, 结构化, 格式化, 标准化, 卷积神经网络, 分割, 目标, 方位, 面积, 横向, 纵向, 最, 像素, 预处理, 算法, merging, fusion, medical, image, data, analysis, structured, RRN, divid+, target

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 110349652 A (之江实验室) 2019年 10月 18日 (2019 - 10 - 18) 权利要求1-8	1-8
Y	CN 109830303 A (上海众恒信息产业股份有限公司) 2019年 5月 31日 (2019 - 05 - 31) 说明书第[0007]-[0029]、[0037]-[0103]、[0123]-[0139]、[0150]-[0163]段	1-8
Y	CN 106846317 A (北京连心医疗科技有限公司) 2017年 6月 13日 (2017 - 06 - 13) 说明书第[0075]-[0127]段	1-8
A	CN 1613070 A (美国西门子医疗解决公司 等) 2005年 5月 4日 (2005 - 05 - 04) 全文	1-8
A	CN 109544561 A (杭州迪英加科技有限公司) 2019年 3月 29日 (2019 - 03 - 29) 全文	1-8

☐ 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

2020年 6月 15日

国际检索报告邮寄日期

2020年 6月 23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

李铎

电话号码 86-(10)-53961393

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/083584

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	110349652	A	2019年 10月 18日	无			
CN	109830303	A	2019年 5月 31日	无			
CN	106846317	A	2017年 6月 13日	无			
CN	1613070	A	2005年 5月 4日	EP	1440412	A2	2004年 7月 28日
				US	2003120514	A1	2003年 6月 26日
				EP	1440409	A2	2004年 7月 28日
				JP	2005508544	A	2005年 3月 31日
				WO	03040965	A2	2003年 5月 15日
				WO	03040964	A3	2004年 1月 22日
				CN	1636210	A	2005年 7月 6日
				US	9165116	B2	2015年 10月 20日
				WO	03040966	A2	2003年 5月 15日
				CN	100449531	C	2009年 1月 7日
				WO	03040965	A3	2004年 2月 19日
				US	7744540	B2	2010年 6月 29日
				JP	2005508557	A	2005年 3月 31日
				US	8280750	B2	2012年 10月 2日
				US	7711404	B2	2010年 5月 4日
				US	2003120134	A1	2003年 6月 26日
				CA	2465533	A1	2003年 5月 15日
				JP	2005509217	A	2005年 4月 7日
				EP	1440385	A2	2004年 7月 28日
				WO	03040878	A8	2004年 4月 1日
				CA	2465712	A1	2003年 5月 15日
				US	2003120133	A1	2003年 6月 26日
				US	2003125988	A1	2003年 7月 3日
				CA	2465531	A1	2003年 5月 15日
				CN	1613087	A	2005年 5月 4日
				WO	03040987	A2	2003年 5月 15日
				JP	2005534082	A	2005年 11月 10日
				CN	1613069	A	2005年 5月 4日
				WO	03040990	A2	2003年 5月 15日
				US	2003130871	A1	2003年 7月 10日
				US	2003120458	A1	2003年 6月 26日
				EP	1442415	A2	2004年 8月 4日
				WO	03040988	A2	2003年 5月 15日
				WO	03040990	A3	2004年 4月 8日
				CA	2465725	A1	2003年 5月 15日
				EP	1440387	A2	2004年 7月 28日
				EP	1440390	A2	2004年 7月 28日
				WO	03040987	A3	2004年 5月 6日
				WO	03040988	A3	2004年 2月 12日
				CA	2465706	A1	2003年 5月 15日
				CA	2465760	A1	2003年 5月 15日
				WO	03040989	A2	2003年 5月 15日
				US	8214225	B2	2012年 7月 3日
				US	2009259487	A1	2009年 10月 15日
				CA	2464374	A1	2003年 5月 15日
				EP	1440388	A2	2004年 7月 28日
				US	2003125985	A1	2003年 7月 3日

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/083584

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)
		US 2015100352 A1	2015年 4月 9日
		US 8214224 B2	2012年 7月 3日
		JP 2005523490 A	2005年 8月 4日
		EP 1440389 A2	2004年 7月 28日
CN 109544561 A	2019年 3月 29日	无	