

自述文件

医学数据 / 自述文件



Andy-jqa更新README.md ...

历史

23位 贡献者



+11

生的

怪

复制文件

42.2 KB



652行 (397 sloc) | 42.2 KB

# 机器学习医学数据

这是用于机器学习的医疗数据的精选列表。  
提供此列表仅供参考，请确保您遵守此处列出的任何数据的所有使用限制。

## 1.医学影像数据

### EchoNet-Dynamic

来自斯坦福大学的用于医学机器学习的大型新型心脏运动视频数据资源。概述：

<https://echonet.github.io/dynamic/index.html> 访问：

<https://echonet.github.io/dynamic/index.html#access>

### 国家医学图书馆展示了MedPix®

数据库，其中包含来自13,000位患者的53,000张医学图像，并带有注释。需要注册。

信息：<https://medpix.nlm.nih.gov/home>

**注意：自闭症脑成像数据交换：对自闭症内在大脑结构的大规模评估。**

539例ASD患者和573例典型对照患者的功能MRI图像。这1112个数据集由结构和静态功能MRI数据以及大量的表型信息组成。**需要注册。**

论文：[http](http://www.ncbi.nlm.nih.gov/pubmed/23774715)

: [://www.ncbi.nlm.nih.gov/pubmed/23774715](http://www.ncbi.nlm.nih.gov/pubmed/23774715) 信息：[http](http://fcon_1000.projects.nitrc.org/indi/abide/) :

[//fcon\\_1000.projects.nitrc.org/indi/abide/](http://fcon_1000.projects.nitrc.org/indi/abide/)

预处理版本：[http](http://preprocessed-connectomes-project.org/) : [// preprocessed-connectomes-project](http://preprocessed-connectomes-project.org/)。组织/遵守/

---

**关于阿尔茨海默氏病患者和健康对照者的阿尔茨海默氏病神经影像学倡议（ADNI） MRI数据库。** 还具有临床，基因组和生物制造商数据。**需要注册。**

论文：[http](http://www.neurology.org/content/74/3/201.short) : [://www.neurology.org/content/74/3/201.short](http://www.neurology.org/content/74/3/201.short)

访问：[http](http://adni.loni.usc.edu/data-samples/access-data/) : [//adni.loni.usc.edu/data-samples/access-data/](http://adni.loni.usc.edu/data-samples/access-data/)

---

**CT结肠癌结肠造影术（癌症影像存档）** CT扫描用于诊断结肠癌。包括没有息肉，6-9mm息肉和大于10mm息肉的患者的数据。访问：

[https](https://wiki.cancerimagingarchive.net/display/Public/CT+COLONOGRAPHY#dc149b9170f54aa29e88f1119e25ba3e) : [//wiki.cancerimagingarchive.net/display/Public/CT+COLONOGRAPHY#dc149b9170f54aa29e88f1119e25ba3e](https://wiki.cancerimagingarchive.net/display/Public/CT+COLONOGRAPHY#dc149b9170f54aa29e88f1119e25ba3e)

---

**用于血管提取的数字视网膜图像（DRIVE）**

DRIVE数据库用于比较研究视网膜图像中血管的分割。它由40张照片组成，其中7张显示出轻度早期糖尿病性视网膜病变的迹象。

论文：[https](https://ieeexplore.ieee.org/document/1282003) : [://ieeexplore.ieee.org/document/1282003](https://ieeexplore.ieee.org/document/1282003)

访问：[http](http://www.isi.uu.nl/Research/Databases/DRIVE/download.php) : [://www.isi.uu.nl/Research/Databases/DRIVE/download.php](http://www.isi.uu.nl/Research/Databases/DRIVE/download.php)

---

**AMRG心脏图谱** AMRG心脏MRI图谱是使用奥克兰MRI研究小组的Siemens Avanto扫描仪采集的正常患者心脏的完整带标签MRI图像集。该数据集旨在为大学生和中学生，MR技术人员，临床医生...

**先天性心脏病（CHD）图集** 先天性心脏病（CHD）图集代表成人和患有各种先天性心脏病的儿童的MRI数据集，生理临床数据和计算机模型。数据已从包括Rady ...在内的多个临床中心获得。

**确定** 通过磁共振成像评估降低风险的除颤器，是一项针对冠心病和轻度至中度左心功能不全患者的前瞻性，多中心，随机临床试验。主要目标...

**MESA** 动脉粥样硬化多民族研究是一项在美国六个中心进行的大规模心血管疾病研究（参加人数超过6,500人）。目的是调查临床前亚临床到心血管疾病的表现。

---

**OASIS** 影像研究开放获取系列（OASIS）是一个旨在向科学界免费提供大脑的MRI数据集的项目。有两个数据集：横截面和纵向集。

- 年轻，中年，非痴呆和痴呆老年人的横断面MRI数据：该集合包括416名18至96岁的受试者的横断面集合。对于每位受试者，单次获得3或4次单独的T1加权MRI扫描包括扫描会话。受试者都是右撇子，包括男性和女性。在60岁以上的受试者中，有100名已被临床诊断为非常轻至中度的阿尔茨海默氏病（AD）。此外，还包括一个可靠性数据集，其中包含20位未痴呆的受试者，这些受试者在其初次就诊后90天内的后续访问中成像。
- 非痴呆和痴呆老年人的纵向MRI数据：该组纵向收集了150名60至96岁的受试者。两次或两次以上就诊时对每个受试者进行了扫描，相隔至少一年，共进行373次成像。对于每个受试者，包括在单次扫描会话中获得的3或4个单独的T1加权MRI扫描。受试者都是右撇子，包括男性和女性。在整个研究过程中，有72位受试者表现为非痴呆。纳入的受试者中有64位在初次就诊时就被诊断为痴呆，并在随后的扫描中一直如此，包括51位轻度至中度阿尔茨海默氏病患者。另有14名受试者在初次就诊时被定为非痴呆症，随后在以后的访视中被定为非痴呆症。

访问：<http://www.oasis-brains.org/>

---

**Isic档案库-黑色素瘤** 该档案库包含分类皮肤病变的23k图像。它包含了恶性和良性的例子。

每个示例均包含病变的图像，有关病变的元数据（包括闭锁和分割）以及有关患者的元数据。

可以在以下链接中查看数据：<https://www.isic-archive.com>（在画廊部分）

可以通过网站或使用此存储库下载：<https://github.com/GalAvineri/ISIC-存档下载器>

---

**SCMR共识数据** SCMR共识数据集是从不同的MR机器（4个GE，5个西门子，6个飞利浦）中获得的，涉及混合病理（5例健康，6例心肌梗塞，2例心力衰竭和2例肥大）的15例心脏MRI研究。）。主要目标...

**Sunnybrook心脏数据** Sunnybrook心脏数据（SCD），也称为2009心脏MR左心室分割挑战数据，由45种cine-MRI图像组成，这些图像来自以下患者和以下疾病：健康，肥大，有心梗的心衰和心脏。..

访问：<http://www.cardiacatlas.org/studies/>

---

## 肺图像数据库联盟（LIDC）

初步临床研究表明，肺部螺旋CT扫描可以改善高危人群的肺癌早期检测。图像处理算法有潜力协助螺旋CT研究中的病变检测，并评估系列CT研究中病变大小的稳定性或变化。这种计算机辅助算法的使用可以显着提高螺旋CT肺部筛查的敏感性和特异性，并通过减少医生解释所需的时间来降低成本。

肺影像数据库协会（LIDC）计划的目的是支持一个机构协会，为螺旋CT肺图像资源制定共识指南，并建立螺旋CT肺图像数据库。在此计划的资助下，研究人员为数据库使用以及将数据库开发为测试平台和这些方法的展示创建了一套准则和指标。该数据库可通过Internet供研究人员和用户使用，并广泛用作研究，教学和培训资源。

LIDC计划的具体目标是提供：

- 用于图像处理或CAD算法的相对评估的参考数据库，以及
- 灵活的查询系统，将为研究人员提供机会评估该数据库中可能对研究应用很重要的各种技术参数和未识别的临床信息。

该资源将刺激进一步的数据库开发，以用于图像处理和CAD评估，以用于包括癌症筛查，诊断以及图像指导的干预和治疗在内的应用。因此，NCI鼓励研究人员发起的在其研究中利用数据库的拨款申请。NCI还鼓励研究人员启动赠款申请，这些申请提供的工具或方法可以改善或补充LIDC的任务。

访问：<https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI#>

---

## TCIA收藏

跨各种癌症类型（例如癌，肺癌，骨髓瘤）和各种成像方式的癌症影像数据集。癌症影像档案库（TCIA）中的图像数据被组织成专门目的的受试者集合。受试者通常具有共同的癌症类型和/或解剖部位（肺，脑等）。下表中的每个链接均包含有关集合的科学价值的信息，有关如何获取任何可用的支持非图像数据的信息，以及用于查看或下载成像数据的链接。为了支持科学研究的可重复性，TCIA支持数字对象标识符（DOI），这些标识符使用户可以共享研究手稿中引用的TCIA数据的子集。

访问：<http://www.cancerimagingarchive.net/>

---

## 白俄罗斯结核病门户

结核病是白俄罗斯公共卫生的主要问题。最近，由于需要长期治疗的耐多药/广泛耐药结核和艾滋病毒/结核的出现和发展使情况变得复杂。通常，许多最严重的病例会在全国范围内传播给不同的结核病诊所。通过使用包含患者的放射影像，实验室工作和临床数据的通用数据库，将大大提高领先的白俄罗斯结核病专家追踪此类患者的能力。这还将显著提高对治疗方案的依从性，并导致更好的治疗结果记录。在门户网站的数据库中纳入临床病例的标准-肺部和肺结核RSPC的MDR-TB部门收治的被诊断或怀疑为MDR-TB的患者，

访问：<http://tuberculosis.by/>

---

## DDSM：乳腺钼靶筛查数字数据库

乳腺钼靶筛查数字数据库（DDSM）是由乳腺钼靶图像分析研究社区使用的资源。此项目的主要支持是美国陆军医学研究和Material司令部的乳腺癌研究计划提供的资金。DDSM项目是在马萨诸塞州综合医院（D. Kopans, R. Moore），南佛罗里达大学（K. Bowyer）和桑迪亚国家实验室（P. Kegelmeyer）的合作研究中共同努力的成果。华盛顿大学医学院的其他病例由放射学和内科医学助理教授Peter E. Shile博士提供。其他合作机构包括维克森林大学医学院（医学工程与放射学系），圣心医院和ISMD, Incorporated。该数据库的主要目的是促进计算机算法开发中的合理研究，以帮助进行筛选。数据库的次要目的可以包括开发有助于诊断的算法以及开发教学或培训教具。该数据库包含大约2500个研究。每个研究包括每个乳房的两张图像，以及一些相关的患者信息（研究时的年龄，ACR乳房密度等级，异常的细微等级，ACR异常的关键词描述）和图像信息（扫描仪，空间分辨率，...）。包含可疑区域的图像具有与可疑区域的位置和类型有关的像素级“地面真相”信息。

访问：<http://marathon.csee.usf.edu/Mammography/Database.html>

---

### **INbreast：数字化乳腺摄影数据库**

INbreast数据库是一个乳腺摄影数据库，其中的图像是从位于大学医院（葡萄牙波尔图的乳房中心的圣若昂医院）的乳房中心获取的。INbreast共有115例（410幅图像），其中90例来自双乳女性（每例4幅图像），而25例来自乳房切除术患者（每例2幅图像）。包括几种类型的病变（肿块，钙化，不对称和变形）。专家还以XML格式提供了精确的轮廓。

访问：[http://medicalresearch.inescporto.pt/breastresearch/index.php/Get\\_INbreast\\_Database](http://medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database)

---

### **mini-MIAS：MIAS最小化乳腺摄影数据库**

乳房X线照片图像分析协会（MIAS）是英国研究小组的组织，对理解乳房X线照片感兴趣，并且已经建立了数字乳房X线照片数据库。来自英国国家乳房筛查计划的胶片已通过Joyce-Loebl扫描微光密度计数字化至50微米像素边缘，该设备在光密度范围0-3.2内呈线性，并用8位字表示每个像素。该数据库包含322个数字化胶片，可在2.3GB 8mm（ExaByte）磁带上使用。它还在可能出现的任何异常位置上包括放射科医生的“真相”标记。该数据库的边缘已缩小到200微米，并已填充/剪切为所有图像均为1024x1024。

访问：<http://peipa.essex.ac.uk/info/mias.html>

---

### **前列腺**

前列腺癌（CaP）在世界范围内已被报告为第二大最常被诊断为男性的癌症，占13.6%（Ferlay et al.（2010））。据统计，2008年，新诊断病例数为899,000，死亡人数不少于258,100（Ferlay et al.（2010））。

磁共振成像（MRI）提供了可以诊断和定位CaP的成像技术。I2CVB提供了一个多参数MRI数据集，以帮助开发计算机辅助检测和诊断（CAD）系统。访问：<http://i2cvb.github.io/>

---

访问: <http://www.ehealthlab.cs.ucy.ac.cy/index.php/facilities/32-software/218-datasets>

- 多发性硬化症数据库中的MRI病变分割
- 紧急远程骨科X射线数字图书馆
- IMT细分
- 针EMG MUAP时域功能

---

**DICOM图像样本集** 这些数据集专用于研究和教学。您无权再分发或出售它们, 或将其用于商业目的。

所有这些DICOM文件都以JPEG2000传输语法压缩。

访问: <http://www.osirix-viewer.com/resources/dicom-image-library/>

---

### SCR数据库: 胸部X光片中的分割

胸部X光片中解剖结构的自动分割对于这些图像中的计算机辅助诊断非常重要。建立了SCR数据库, 以促进在标准前后胸片中对肺野, 心脏和锁骨的分割进行比较研究。

本着合作科学进步的精神, 我们自由共享SCR数据库, 并致力于维护有关这些分割任务的各种算法的结果的公共存储库。在这些页面上, 可以找到有关下载数据库和上载结果的说明, 并且可以检查各种方法的基准结果。

访问: <http://www.isi.uu.nl/Research/Databases/SCR/>

---

### 医学图像数据库和图书馆

访问: [http://www.omnimedicalsearch.com/image\\_databases.html](http://www.omnimedicalsearch.com/image_databases.html)

#### 一般类别

- e-Anatomy.org-交互式解剖图集-e-anatomy是一个解剖学在线学习网站。从正常的CT和MR检查中选择了1500多个切片, 以覆盖人体的整个解剖结构。图像使用解剖学标记。用户友好的界面允许通过结合交互式文本信息, 3D模型和解剖图的多层图像系列进行电影放映。
- 医学图片和定义-欢迎访问Internet上最大的医学图片和定义数据库。提供医疗信息的站点很多, 但是提供医疗图片的站点很少。据我们所知, 我们是唯一提供有关医学影像数据库的基本信息的人。编者注: 不错的网站, 可免费访问且没有讨厌的注册, 可定义1200多种与健康 and 医疗相关的图像。
- 核医学艺术-医学插图, 医学艺术。包括3D动画。“Nucleus Medical Art, Inc.是医学插图, 医学动画和交互式多媒体的领先创作者和发行商, 其作品在美国和国外的出版, 法律, 医疗保健, 娱乐, 制药, 医疗设备, 学术界和其他市场中均占有。注意: 很棒的网站。



- 互联网上的医学图像数据库（UTHSCSA库）-带有主题特定医学图像的网站链接的目录。
- 手术视频-国家医学图书馆MedlinePlus收集了100到100多种不同手术程序的链接。您必须在计算机上安装RealPlayer媒体播放器才能观看这些免费的视频。
- 带有插图的ADAM医学百科全书。《ADAM医学百科全书》也许是当今互联网上插图最好的医学作品之一，其中包括关于疾病，测试，症状，伤害和手术的4,000多篇文章。它还包含一个广泛的医学照片和插图资料库，以备份这4,000篇文章。这些插图和文章对公众免费。
- 哈丁医学博士-医学和疾病图片是爱荷华大学提供的免费且已建立的资源，已有相当长的一段时间了。主页采用目录样式，用户必须向下钻取以查找所需的图像，其中许多图像不在现场。尽管如此，Hardin MD还是通往数千张详细医学照片和插图的绝佳门户。
- 健康教育资产图书馆（HEAL）-网络健康基金会（HON）是总部位于瑞士的网络基金会健康画廊，它是一个国际组织，旨在鼓励以道德的方式提供在线健康信息。“HONmedia（图像库）是一个唯一的存储库，包含6800多种医学图像和视频，涉及1,700个主题和主题。这个无与伦比的数据库是HON手动创建的，全球范围内不断添加新的图像链接Web。HON鼓励用户通过“提交图像”链接使自己的图像链接可用。”图书馆包括解剖图像，疾病的视觉影响以及状况和程序。
- 公共卫生图像库（PHIL）由疾病控制与预防中心（CDC）的一个工作组创建，PHIL为CDC的图像提供了一个组织化的通用电子门户。我们欢迎公共卫生专业人员，媒体，实验室科学家，教育工作者，学生以及全世界的公众使用此材料作为参考，教学，演示和公共卫生信息。内容按人员，地点和科学的层次结构分类，并以单个图像，图像集和多媒体文件的形式呈现。
- 来自医学史的图像-通过该系统，可以访问美国国家医学图书馆（NLM）的医学史科（HMD）的照片和照片集中的近60,000张图像。该馆藏包括肖像，机构照片，漫画，流派场景以及各种媒体上的图形艺术，这些插图说明了医学的社会和历史方面。
- Pozemedicale.org-西班牙语，意大利语，葡萄牙语和意大利语的医学图像集合。
- 旧的医学图片：十九世纪末至20世纪初的数百幅有趣而有趣的古老但高质量的照片和图像。

## 学科专业影像图书馆和馆藏

- 亨利·格雷（Henry Gray）撰写的《人体解剖学》-格雷的《人体解剖学》Bartleby.com版载有1,247幅充满活力的版画，其中许多色彩来自于1918年的经典出版物。
- Crookston馆藏-John H. Crookston博士拍摄的医学幻灯片的馆藏已被数字化，可供公众和医生使用。
- DAVE Project-可搜索的胃肠道内窥镜视频素材库，涵盖广谱内镜成像。
- Dermnet-可浏览的8,000多种高质量皮肤病图像集。
- 交互式皮肤病学地图集-常见和罕见皮肤问题的图像参考来源。
- 多维人类胚胎是由美国国家儿童健康与人类发展研究所（NICHD）资助的一项合作项目，旨在基于磁共振成像技术制作并通过互联网提供人类胚胎的三维图像参考。

- GastroLab内窥镜档案馆始建于1996年，其目标是维护所有感兴趣的医疗保健人员免费使用的内窥镜影像馆。
- MedPix是放射和医学图片数据库资源工具。主页界面令人困惑，并且整个网站设计都不友好，并且具有1990年代中期的感觉。但是，如果您有时间（耐心），对于某些人来说，它可能是重要的资源。
- OBGYN.net图像库-该站点完全致力于提供对妇女健康感兴趣的图像的访问。除了为您提供对OBGYN.net图像的访问权限之外，我们还指向Internet上其他与女性健康相关的图像。由于材料的图形性质，某些人可能不希望查看这些图像，仅出于教育目的提供它们。

---

## 威盛集团公共数据库

记录的图像数据库对于开发定量图像分析工具至关重要，特别是对于计算机辅助诊断（CAD）任务而言。通过与I-ELCAP小组的合作，我们建立了两个公共图像数据库，其中包含DICOM格式的肺部CT图像以及放射科医生的异常记录。请访问以下链接以获取更多详细信息：

访问：<http://www.via.cornell.edu/databases/>

---

CVonline：图片数据库 访问：<http://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm>

---

**USC-SIPI图像数据库** USC-SIPI图像数据库是数字化图像的集合。维护它主要是为了支持图像处理，图像分析和机器视觉方面的研究。USC-SIPI图像数据库的第一版于1977年发行，此后又添加了许多新图像。

数据库根据图片的基本特征分为多个卷。每个体积中的图像大小各异，例如256x256像素，512x512像素或1024x1024像素。对于黑白图像，所有图像均为8位/像素，对于彩色图像，所有图像均为24位/像素。当前有以下卷：

Textures	Brodatz textures, texture mosaics, etc.
Aerials	High altitude aerial images
Miscellaneous	Lena, the mandrill, and other favorites
Sequences	Moving head, fly-overs, moving vehicles

访问：<http://sipi.usc.edu/database/>

---

## 组织学数据集：不同染色切片的图像配准

数据集由2D组织学显微镜检查组织切片组成，并用不同的染色剂染色，并在每个切片中标出代表关键点的地标。任务是图像配准-将特定图像集（连续的污点切割）中的所有切片对齐在一起，例如对齐到初始图像平面。这些图像的主要挑战如下：超大图像尺寸，外观差异以及缺乏独特的外观对象。数据集包含108个图像对和手动放置的地标，用于注册质量评估。



访问: <http://cmp.felk.cvut.cz/~borovji3/?page=dataset>

---

## 2.挑战/竞赛数据

---

**放射学中的视觉概念提取挑战** 手动注释来自几种不同成像方式（例如CT和MR）的几个解剖结构（例如肾脏，肺，膀胱等）的放射学数据。它们还提供了一个云计算实例，任何人都可以用来根据基准开发和评估模型。

访问: <http://www.visceral.eu/>

---

### 生物医学图像分析的巨大挑战

通过标准化评估标准，*促进生物医学成像挑战的收集，以促进新解决方案与现有解决方案之间的更好比较*。您也可以创建自己的挑战。撰写本文时，有92个挑战提供了可下载的数据集。

访问: <http://www.grand-challenge.org/>

---

### 梦想挑战

梦想挑战提出了有关系统生物学和转化医学的基本问题。由各种组织的研究人员社区设计和运营，我们的挑战邀请参与者提出解决方案-在此过程中促进协作并建立社区。Sage Bionetworks提供专业知识和机构支持，以及通过Synapse平台应对挑战的基础设施。我们共同拥有一个愿景，使个人和团体可以公开协作，以使“人群的智慧”对科学和人类健康产生最大的影响。

- 数字乳腺X射线摄影梦想挑战赛。
- ICGC-TCGA DREAM体细胞突变呼吁RNA挑战（SMC-RNA）
- 梦想创意挑战
- 这些是添加时的积极挑战，现在还有更多过去的挑战和即将到来的挑战！

访问: <http://dreamchallenges.org/>

---

### Kaggle糖尿病性视网膜病变

由临床医师在0-4严重性等级上标注的高分辨率视网膜图像，用于检测糖尿病性视网膜病变。该数据集是已完成的Kaggle竞赛的一部分，该竞赛通常是可公开获得的数据集的重要来源。

访问: <https://www.kaggle.com/c/diabetic-retinopathy-detection>

---

### 宫颈癌筛查

在这场kaggle竞赛中，您将开发算法来根据宫颈图像正确分类子宫颈类型。我们数据集中的这些不同类型的子宫颈都被认为是正常的（不是癌变的），但是由于转化区并不总是可见的，因此有些患者需要进一步检查，而有些则不需要。

访问: <https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/data>

---

## 多发性硬化病变分割

挑战2008。大脑MRI扫描的集合，以检测MS病变。

访问: <http://www.ia.unc.edu/MSseg/>

---

## 多峰脑肿瘤分割挑战

脑肿瘤磁共振扫描的大数据集。自2012年以来，他们每年都在扩展此数据集并提出挑战。

访问: <http://braintumorsegmentation.org/>

---

## 编码4癌症

美国国立卫生研究院和鼠尾草生物网络基金会提出了一项新倡议，以应对改善癌症筛查的一系列挑战。首先是用于数字化乳腺X线摄影的读数。第二个是用于肺癌检测。挑战尚未开始。

访问: <http://coding4cancer.org/>

---

## Kaggle上的EEG挑战数据集

- 墨尔本大学AES / MathWorks / NIH癫痫发作预测-预测人类长期颅内脑电图记录中的癫痫发作

访问: <https://www.kaggle.com/c/melbourne-university-seizure-prediction>

- 美国癫痫协会癫痫发作预测挑战-预测颅内EEG记录中的癫痫发作

访问: <https://www.kaggle.com/c/seizure-prediction>

- UPenn和Mayo诊所的癫痫发作检测挑战-检测颅内EEG记录中的癫痫发作

访问: <https://www.kaggle.com/c/seizure-detection>

- 抓握并抬起脑电图-从脑电图记录中识别手部动作

访问: <https://www.kaggle.com/c/grasp-and-lift-eeg-detection>

---

## MICCAI会议的挑战轨迹

医学图像计算和计算机辅助干预。大多数挑战本应由大型挑战之类的网站涵盖。您仍可以在会议站点的“卫星事件”选项卡下看到所有这些挑战。

- 2019年-<https://www.miccai2019.org/programme/workshops-challenges-tutorials/#tablepress-10>
- 2018年-<https://www.miccai2018.org/en/WORKSHOP---CHALLENGE---TUTORIAL.html>
- 2017年-<http://www.miccai2017.org/satellite-events>
- 2016年-<http://www.miccai2016.org/en/SATELLITE-EVENTS.html>
- 2015- [https://www.miccai2015.org/frontend/index.php?page\\_id=589](https://www.miccai2015.org/frontend/index.php?page_id=589)

访问: <http://www.miccai.org/ConferenceHistory>

---

## 国际生物医学影像研讨会 (ISBI)

IEEE国际生物医学影像研讨会 (ISBI) 是一个科学会议, 致力于涵盖所有观察规模的生物医学影像的数学, 算法和计算方面。这些挑战中的大多数将列在大挑战中。您仍然可以通过访问每年网站上“程序”下的“挑战”标签来访问它。

- 2019- <https://biomedicalimaging.org/2019/challenges/>
- 2018年-<https://biomedicalimaging.org/2018/challenges/>
- 2017年-<http://biomedicalimaging.org/2017/challenges/>
- 2016年-[http://biomedicalimaging.org/2016/?page\\_id=416](http://biomedicalimaging.org/2016/?page_id=416)

访问: <http://biomedicalimaging.org>

---

## 连续注册挑战赛 (CRC)

连续配准挑战 (CRC) 是受现代软件开发实践启发而进行的肺和脑图像配准的挑战。参与者使用开源SuperElastix C++ API实现其算法。挑战集中于肺和脑的成对注册, 这是临床环境中经常遇到的两个问题。他们收集了七个开放访问数据集和一个私有数据集 (3 + 1个肺部数据集, 4个脑部数据集)。挑战结果将在即将举行的生物医学图像注册研讨会 (WBIR 2018) 上进行介绍和讨论。

访问: <https://continuousregistration.grand-challenge.org/home/>

---

## 自动非刚性组织学图像配准 (ANHIR)

这项ANHIR挑战旨在针对用不同染料染色的组织病理学组织样本的2D全切片成像 (WSI) 显微镜图像进行自动非线性图像配准。由于影响组织样本的非线性变形, 每种污渍的不同外观, 重复的纹理以及整个幻灯片图像的大尺寸, 因此该任务很困难。

- 挑战: <https://anhir.grand-challenge.org/>
- 基准: <http://borda.github.io/BIRL>
- 参考: BIRL: 具有地标验证功能的图像配准方法的基准

## 使用MURA的骨骼X射线深度学习竞赛

MURA（肌肉骨骼X光片）是骨骼X射线的大型数据集。斯坦福大学ML组和AIMI中心将举办一项竞赛，其中的任务是确定X线检查是否正常。在207个肌肉骨骼研究的测试集上对算法进行了评估，其中每项研究均由6名获得董事会认证的放射科医生分别回顾性地标记为正常或异常。这些放射线医师中的三位被用来制定黄金标准，定义为放射线医师标签的多数票，另外三位放射线学家用于获得最佳放射线医师的表现，定义为三位放射线医师在黄金标准下的最高分作为地面真理。挑战排行榜是公开托管的，每两周更新一次。

访问：<https://stanfordmlgroup.github.io/competitions/mura/>

---

## 2019肾脏和肾脏肿瘤分割挑战赛（KiTS19）

KiTS19面临的挑战是在增强CT扫描中对肾脏和肾脏肿瘤的语义分割。该数据集包含300名术前动脉期腹部CT的患者，并由专家注释。其中210个（70%）作为训练集发布，其余90个（30%）作为测试集保留。该挑战与MICCAI 2019一起举行。

访问：<https://github.com/neheller/kits19/>

---

## 3.数据来自电子健康记录（EHR）

从数百万个临床叙事中构建医学图，

从1400万临床笔记和260,000患者中提取的医学术语共现统计。

论文：<http://www.nature.com/articles/sdata201432>

数据：<http://datadryad.org/resource/doi:10.5061/dryad.jp917>

---

### 学习医学概念的

低维表示使用索赔数据构建的医学概念的低维嵌入。请注意，本文利用了数以百万计的临床故事中的医学图构建数据。

论文：[http://cs.nyu.edu/~dsontag/papers/ChoiChiuSontag\\_AMIA\\_CRI16.pdf](http://cs.nyu.edu/~dsontag/papers/ChoiChiuSontag_AMIA_CRI16.pdf)

数据：<https://github.com/clinicalml/embeddings>

---

### MIMIC-III，可免费访问的重症监护数据库

匿名重症监护EHR数据库，包含38,597名患者和53,423次ICU入院。需要注册。

论文：<http://www.nature.com/articles/sdata201635>

数据：<http://physionet.org/physiobank/database/mimic3cdb/>

---

## 临床概念曲面嵌入从医学数据的海量渠道获悉

曲面嵌入了60万名患者，170万篇期刊论文，以及20万名患者的临床笔记了解到108477个医学概念纸业：<https://arxiv.org/abs/1804.01486>

曲面嵌入：<HTTPS://figshare.com/s/00d69861786cd0156d81>

交互式工具：<http://cui2vec.dbmi.hms.harvard.edu>

---

## 在癌症中心对患者的实验室测试代码的嵌入进行评估

在霍普市国家医疗中心（加利福尼亚州洛杉矶），从8,280,238份实验室订单中为79,081名患者训练了1098个实验室测试代码（LOINC）的200维Word2Vec嵌入。

论文：<https://arxiv.org/abs/1907.09600>

嵌入和代码：[https://github.com/elleros/DSHealth2019\\_loinc\\_embeddings](https://github.com/elleros/DSHealth2019_loinc_embeddings)

---

# 4.国家医疗保健数据

---

## 疾病预防控制中心（CDC）

来自CDC的许多方面的数据，包括：

- 生物监测
- 儿童疫苗接种
- 流感疫苗
- 健康统计
- 伤害与暴力
- 轻水堆
- 摩托车
- 美国国家健康科学中心
- 神经网络
- 怀孕和疫苗接种
- 性病
- 吸烟与抽烟
- 青少年疫苗接种
- 创伤性脑损伤
- 预防接种
- 网络指标

着陆页：<https://data.cdc.gov>

数据目录：<https://data.cdc.gov/browse>

---



## Medicare数据

来自Medicare和Medicaid Services中心（CMS）的有关医院，疗养院，医生，家庭保健，透析和设备提供商的数据。

登陆页面：<https://data.medicare.gov>

浏览器：<https://data.medicare.gov/data>

---

**得克萨斯州公共住院病人数据文件**该数据包含2006年至2009年间得克萨斯州的1100万人次就诊，诊断，程序代码和结果。

链接：<https://www.dshs.texas.gov/thcic/hospitals/Inpatientpdf.shtm>

---

Propublica**医生的美元**调查制药公司支付给医生的钱。

信息：<https://www.propublica.org/series/dollars-for-docs>

搜索工具：<https://projects.propublica.org/docdollars/>

数据请求：<https://projects.propublica.org/data-store/sets/health-d4d-national-2>

---

通过信息自由行为请求获得的**DocGraph** Physician交互网络。涵盖近一百万个实体。

主页：<http://www.docgraph.com>

信息：<http://thehealthcareblog.com/blog/2012/11/05/tracking-the-social-doctor-opening-up-physician-referral-data-and-更多/>

数据：<http://linea.docgraph.org>

---

## 5. UCI数据集

---

345例有和没有肝病的**肝病数据集**数据。特征是5种被认为与肝病有关的血液生物标志物。

数据：<https://archive.ics.uci.edu/ml/datasets/Liver+Disorders>

### 甲状腺疾病数据集

数据：<https://archive.ics.uci.edu/ml/datasets/甲状腺+疾病>

### 乳腺癌数据集

数据：<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>

### 心脏病数据集

数据：<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>

### 淋巴造影数据集

数据：<https://archive.ics.uci.edu/ml/datasets/Lymphography>

### 帕金森数据集

数据: <https://archive.ics.uci.edu/ml/datasets/parkinsons>

### 帕金森远程监控数据集

数据: <https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring>

### 具有多种录音类型的帕金森语音数据集数据集

数据:

<https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings>

### 帕金森氏病分类数据集

数据: <https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification>

原发性肿瘤数据集 数据: <https://archive.ics.uci.edu/ml/datasets/primary+tumor>

## 6. 生物医学文献

---

### PMC Open Access Subset

收集PubMed Central中所有全文本，开放获取文章的集合。

信息: <http://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/>

存档文件: [http://www.ncbi.nlm.nih.gov/pmc/tools/ftp/#Data\\_Mining](http://www.ncbi.nlm.nih.gov/pmc/tools/ftp/#Data_Mining)

### PubMed 200k RCT

来自随机对照试验（RCT）的公开摘要的集合。摘要中每个句子的注释均可用。

论文: <https://arxiv.org/abs/1710.06071>

数据: <https://github.com/Franck-Dernoncourt/pubmed-rct>

### PubMed文章的Web API

NLM还提供了Web API，用于访问PubMed中的生物医学文献。

获取PubMed文章的说明: <https://www.ncbi.nlm.nih.gov/research/bionlp/APIs/BioC-PubMed/>  
(不是全文，只是标题，摘要等)

对于PubMed Central中的文章，获取全文的说明: <https://www.ncbi.nlm.nih.gov/research/bionlp/APIs/BioC-PMC/>

### 循证医学

来自随机对照试验（RCT）的公开摘要的集合。提供了人口，干预和结果的注释（PICO元素）。

论文: <https://arxiv.org/abs/1806.04185>

数据: <https://ebm-nlp.herokuapp.com/annotations>

网站: <https://ebm-nlp.herokuapp.com/index>

## 证据推论

推断随机对照试验 (RCT) 结果的数据集。来自开放访问子集的已发布RCT的集合。提供 (干预, 比较干预, 结果, 显着性发现, 证据跨度) 的注释。

论文: <https://arxiv.org/abs/1904.01606>

数据: <https://github.com/jayded/evidence-inference/tree/master/annotations>

网站: <http://evidence-inference.ebm-nlp.com/>

## PubMedQA

生物医学研究问答的数据集。任务是使用“是/否/也许”来回答PubMed标题中自然出现的问题。

论文: <https://arxiv.org/abs/1909.06146>

数据: <https://github.com/pubmedqa/pubmedqa>

网站: <https://pubmedqa.github.io/>

## 6. TREC精准医学/临床决策支持轨道

---

文本检索会议 (TREC) 从2014年开始在精密医学/临床决策支持方面开展工作。

### 2014年临床决策支持追踪

重点: 检索与回答有关病历的一般临床问题有关的生物医学文章。

信息和数据: <http://www.trec-cds.org/2014.html>

### 2015年临床决策支持追踪

重点: 检索与回答有关病历的一般临床问题有关的生物医学文章。

信息和数据: <http://www.trec-cds.org/2015.html>

### 2016年临床决策支持追踪

重点: 检索与回答有关医疗记录的一般性临床问题有关的生物医学文章。使用实际的电子健康记录 (EHR) 患者记录代替综合病例。

信息和数据: <http://www.trec-cds.org/2016.html>

### 2017年临床决策支持追踪

重点: 向治疗癌症患者的临床医生检索有用的精密医学相关信息。

信息和数据: <http://www.trec-cds.org/2017.html>

## 7. 医学语音数据

---

## **TORGO数据库：构音障碍说话者的声音和发音语音表达构音障碍**

的TORGO数据库包含对齐的声音和来自脑瘫（CP）或肌萎缩性侧索硬化症（ALS）的说话者的测量3D发音特征，这是最严重的两种言语残疾的普遍原因（肯特和罗森，2004年）以及相匹配的对照。这个名为TORGO的数据库是多伦多大学计算机科学和言语语言病理学部门与多伦多Holland-Bloorview Kids Rehab医院合作的结果。

信息和数据：<http://www.cs.toronto.edu/~complingweb/data/TORGO/torgo.html>

纸：[链接](#)

---

## **NKI-CCRT语料库：同期放化疗治疗晚期头颈癌前后的语音清晰度。**

NKI-CCRT语料库将根据个人听众的判断，对55位接受过头颈癌治疗的演讲者的录音的清晰度进行评估，以供有限的科学用途。语料库包含三个评估时刻（治疗前和治疗后（10周和12个月））的语音清晰度的录音和知觉评估。通过放化疗（CCRT）进行治疗。

论文：[http://rec.elra.info/proceedings/lrec2012/pdf/230\\_Paper.pdf](http://rec.elra.info/proceedings/lrec2012/pdf/230_Paper.pdf)

访问：联系作者。

---

## **非典型情感话语子挑战**

BjörnSchuller, Simone Hantke及其同事正在提供EMOTASS语料库。这个独特的语料库是第一个可以访问残障人士的情感性言语记录的人，这些残障人士涉及各种精神，神经和身体上的残疾。它包含15位残疾人的记录（年龄从19至58岁，平均年龄31.6岁）。任务是从面对非典型展示的讲话中对五种情绪进行分类。在他们的日常工作环境中进行录音。总体而言，包括大约11k语音和大约9个小时的讲话时间。

论文：[http://emotion-research.net/sigs/speech-sig/is2018\\_compare.pdf](http://emotion-research.net/sigs/speech-sig/is2018_compare.pdf)

链接：[http://emotion-research.net/sigs/speech-sig/is18\\_compare](http://emotion-research.net/sigs/speech-sig/is18_compare)。

---

## **自闭症子挑战**

自闭症子挑战基于“儿童病理语言数据库”（CPSD）。它提供了在法国巴黎的两个儿童和青少年精神病学大学部门（皮埃尔·玛丽·居里大学/皮蒂·萨皮蒂埃医院和里内·笛卡尔大学/内克尔大学）录制的演讲。子挑战中使用的数据集包含来自99位6至18岁儿童的2.5 k个语音记录实例

论文：[http://emotion-research.net/sigs/speech-sig/is2013\\_compare.pdf](http://emotion-research.net/sigs/speech-sig/is2013_compare.pdf)

链接：[http://emotion-research.net/sigs/speech-sig/is13\\_compare](http://emotion-research.net/sigs/speech-sig/is13_compare)。