



# 目录 Content

访问学者.....	1
研究院新成员.....	3
访问学生.....	6
实验室介绍.....	13



# 访问学者

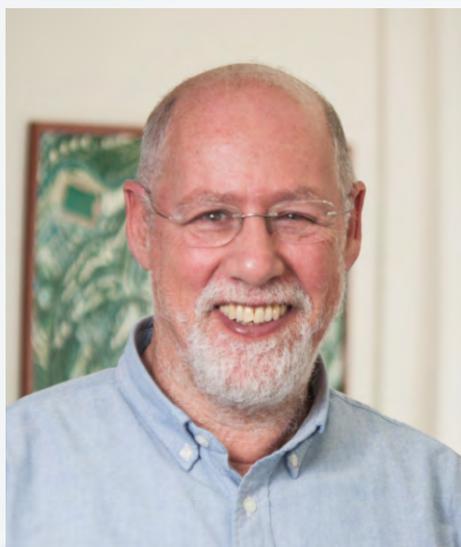
## Paul D. Feigin

Paul D. Feigin 于1972年在墨尔本大学取得统计学学士学位，并于1975年在澳大利亚国立大学取得统计学博士学位。

Feigin教授从1976年至今就职于以色列理工学院，于1999-2002年间担任以色列理工学院威廉·戴维森机械工程与管理学院院长。2007-2013年Feigin教授担任以色列理工学院高级执行副校长，在他的努力下，以色列理工学院通过积极建立与康奈尔等国际一流大学的联系提高了国际声誉。Feigin教授还参与了广东以色列理工学院的建设和，学校于2016年正式批准设立，并于2017年首度招生。



Feigin教授在统计设计和经验分析（如基因组和临床研究）领域享有声誉。他有三个主要的研究方向：随机过程推断（如时间序列或离散事件序列）；现代预测及数据挖掘方法，在该领域他设计的模型已经用于电力需求预测；行业测试和临床测试的设计与分析。他最近的研究项目是利用统计学分析研究遗传相关性和在呼叫中心的病人耐心程度。



## Avishai Mandelbaum

Avishai Mandelbaum是以色列理工学院内机械工业与管理学院的荣耀教授。他在以色列特拉维夫大学获得了数学与计算机学士学位和统计学硕士学位，并于康奈尔大学获得运筹学博士学位。博士毕业后，Mandelbaum教授先于1983年就职于斯坦福商学院的研究生院，随后于1991年回到以色列，并于以色列理工学院就职。Mandelbaum教授于2015-2018担任机械工业与管理学院院长，推动了整个以色列理工学院的“数据化”，特别是他参与创立了数据

科学及工程学士项目并获得广泛的认可。

Mandelbaum教授是运筹学和管理学研究协会（INFORMS）和生产与服务执行管理协会（MSOM）的

成员。他参与了很多业内期刊的编辑（如《数学与操作研究》等）。他的研究和教学水平得到广泛认可，他因此获得了许多奖项，如首届以色列理工学院学术成就Yanai奖。

他的研究领域包括随机模型（分析、渐进性、控制）和统计学，包括排队论、服务系统（如远程服务、医院等）的应用和数据科学。



# 宋彦

宋彦博士于2013年在香港城市大学获得计算语言学博士学位，现为华盛顿大学客座助理教授，场人工智能工程院的院长。2013年，宋彦博士前往华盛顿并在微软公司AI & Research担任研究员；2017-2019年，他在腾讯AI Lab担任专家（首席）研究员。

宋彦博士主攻自然语言理解及人机交互，研究方向涉及自然语言处理与机器学习，智能检索，语言资源及计算等。发表论文50余篇，并多次担任人工智能及计算语言学顶级会议（AAAI, ACL, NAACL, EMNLP, EACL, COLING等）程序委员会委员及高级程序委员会委员。

在微软期间作为创始团队成员启动了微软小冰项目，并深度参与了微软多项人工智能项目的研发及技术转化，涉及产品包括Cortana, Windows 10等。

宋彦博士的主要研究方向是基于自然语言处理的机器学习、自然语言理解、中文处理、表示法学习、社交网络及社交媒体。

# 研究院新成员



**王思逸**

电子科技大学  
专业计算机与信息工程  
合作博弈



**赵伟兵**

北京师范大学  
计算机科学与技术专业  
深度学习、计算机视觉



**李成刚**

哈尔滨建筑大学  
太学大数据平台研发



**吴毓双**

南京邮电大学  
CIE  
计算机视觉



**李兆群**

北京航空航天大学  
系统工程专业  
生物识别，计算机视觉



**刘 召**

香港中文大学（深圳）  
计算机与信息工程  
运营管理



**张策尧**

香港中文大学 (深圳)  
计算机与信息工程  
元学习和强化学习



**王帅**

西南大学  
计算机科学与技术  
大规模优化算法



**宁述亮**

香港中文大学 (深圳)  
计算机信息工程  
视频生成和预测



**葛颂阳**

香港中文大学 (深圳)  
计算机科学  
社会行为大数据应用



**李磊**

北京理工大学  
CIE  
无线通信中的优化与算法设计



**郑玉健**

哈尔滨工业大学(威海)  
计算机信息工程  
Geometric Learning



**王烟濛**

重庆邮电大学  
信息与通信工程  
无线通信



**闫子正**  
香港中文大学（深圳）  
计算机与信息工程  
深度学习



**王 钦**  
香港中文大学（深圳）  
计算机信息工程  
医学图像



**余湘港**  
香港中文大学（深圳）  
计算机与信息工程  
计算机视觉



**颜 旭**  
香港中文大学（深圳）  
计算机与信息工程  
3D点云分析



**朱明鹤**  
香港中文大学（深圳）  
专业：计算机与信息工程  
研究方向：机器学习、无线通信

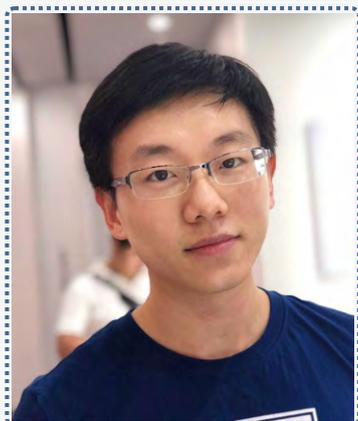


**肖福龙**  
悉尼新南威尔士大学  
专业：金融分析  
公司金融、公司治理、  
国际资本市场



**张南南**  
香港中文大学（深圳）  
计算机与信息工程  
模式识别、医学生物特征识别、  
机器学习

# 访问学生



**叶坤林**  
暨南大学  
影像医学与核医学  
影像诊断



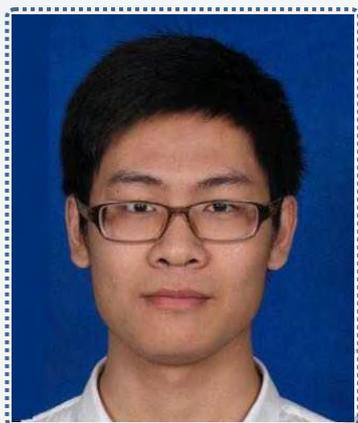
**阮博文**  
武汉大学  
遥感科学与技术类  
深度学习



**王心童**  
哈尔滨工业大学  
软件工程  
机器学习，联邦学习和优化



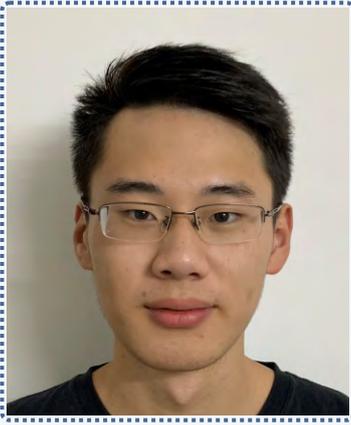
**黄鸿基**  
麻省理工学院  
EECS  
无线网络和统计机器学习



**高建焘**  
上海大学  
机械电子工程  
深度学习、计算机视觉



**肖玉婷**  
香港中文大学（深圳）  
数据科学专业  
大数据挖掘



**袁之浩**  
中山大学  
软件工程  
计算机视觉



**黄冕**  
爱丁堡大学 (英国)  
Data Science  
Bayesian Inference,  
Domain Adaptation/Meta -  
learning,  
Graph Neural Network



**林逸群**  
香港中文大学 (深圳)  
CIE  
3D scene understanding,  
point cloud analysis



**章海超**  
浙江大学  
空天信息技术  
(一级学科: 计算机科学与技术)  
计算机视觉



**曹冉**  
香港中文大学 (深圳)  
翻译专业



**陈肯**  
香港中文大学 (深圳)  
金融数学  
量化金融



**石 峰**

西南交通大学  
管理科学与工程  
流动性研究



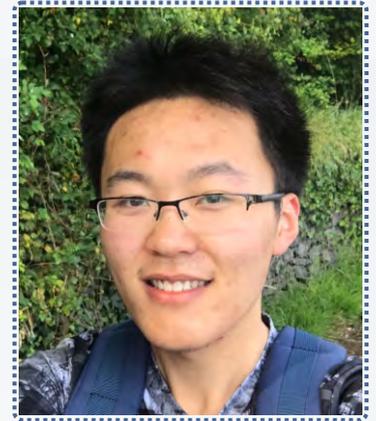
**徐志成**

香港中文大学（深圳）  
数据科学硕士  
机器学习



**万 里**

天津大学  
计算机技术专业  
多模态语音信号处理



**李竞雄**

Queen Mary University of London  
Artificial Intelligence  
人呼出气体的智能病理分析



**刘浩霖**

香港中文大学（深圳）  
Computer and information  
engineering  
Text-Shape Retrieval/  
Embodied Object Searching  
in Virtual Environment.

# 新闻动态

## 研究院与市卫健委签订战略合作框架协议

2019年11月6日，研究院与深圳市卫生健康委员会（以下简称“市卫健委”）正式签订战略合作框架协议。本次签署会议在市卫健委会议室进行，与会人员有市卫健委吴兵副主任及相关处室领导，研究院罗智泉院长、崔曙光常务副院长及相关人员。

市卫健委将向研究院提供深圳市医疗卫生相关数据，用于开展医疗数据结构化处理、高危人群智能筛查和早期风险预测预警以及人工智能医学影像辅助诊断合作。双方本着开放共赢、优势互补，共同发展、的原则，将共同推动以大数

据、人工智能为代表的新一代信息技术与卫生健康行业的融合创新，推进新一代医学人工智能的发展，提升新一代人工智能在深圳医疗卫生领域的应用水平。



双方本次合作，将有利于贯彻落实党中央、国务院关于发展健康医疗大数据以及新一代人工智能的决策部署，抢占中国特色社会主义先行示范区和粤港澳大湾区的战略发展机遇，更好地推动深圳市新一代医学人工智能创新服务发展。

## 抚州市市长来访深圳市大数据研究院

2019年11月12日，江西省抚州市张鸿星市长对研究院进行了参观交流。同行人员有商务局胡会华局长、教育体育局吴建发局局长、大数据发展管理局伍旭鸿局长及相关人员。研究院罗智泉院长主持了本次会议，参与人员为院务张会元经理。

中国已步入数字经济时代，数字经济既可以改变生活方式，又对行业带来了新的挑战，要缩小地区之间、城乡之间的“数字鸿沟”，就必须推行“十三五规划”中有关数字经济发展战略，从核心技术的科研开发应用来缩小差距，调整产业结构，推



进数字经济发展。

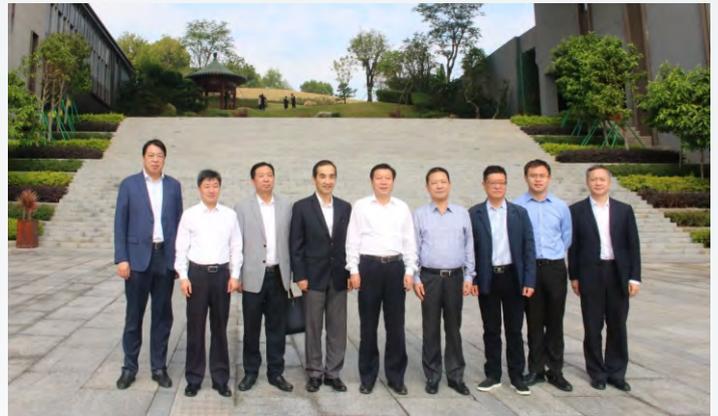
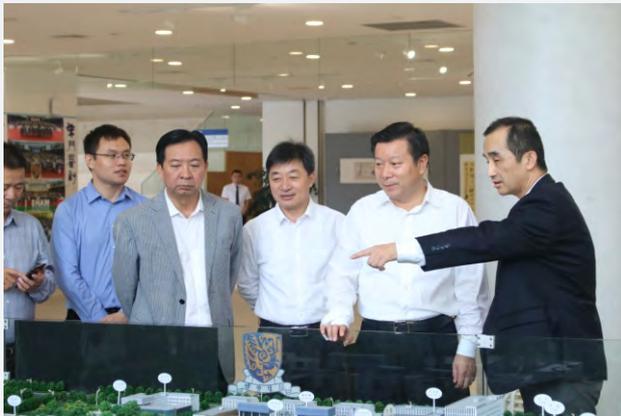


会议上，张鸿星市长介绍了抚州市现有的科技环境和未来长期的发展。抚州卓朗大数据中心项目是华中地区最大的数据中心项目，支撑我国互联网、云计算、大数据及智慧城市产业整体发展；50万台服务器的创世纪超算产业中心落户抚州高新区；抚州市积极建设了多个“数字化车间”及“智能工厂”。



罗智泉院长介绍了香港中文大学（深圳）和研究院的建设情况，并对研究院科研项目进展及阶段性的成果进行详细的介绍。双方就“大数据”方面的内容进行了讨论，寻找合作点，共同推进城市及地区的数字经济的发展。

会议结束后，罗智泉教授带领大家参观了校园景色，包括：行政楼、图书馆、体育馆等，并合影留念。



## 深圳市大数据研究院联合主办2019年数据科学国际研讨会

12月18日至20日，2019年数据科学国际研讨会在香港中文大学（深圳）完美落幕。本次研讨会由深圳市大数据研究院、香港中文大学（深圳）、西电综合业务网理论及关键技术国家重点实验室、深圳大学联合主办，腾讯和华为作为会议的赞助商。开幕式上，香港中文大学（深圳）副校长、深圳市大数据研究院院长罗智泉教授首先致欢迎辞，罗教授简要介绍了MIIS的发展史以及研究院与香港中文大学（深圳）的概况。随后，西安电子科技大学通信工程学院院长沈八中教授，深圳大学信息工程学院黄磊教授分别致辞。



深圳市大数据研究院院长  
罗智泉教授



西安电子科技大学通信工程学院院长  
沈八中教授



深圳大学信息工程学院  
黄磊教授致辞

本次研讨会聚集了行业知名学者，在三天半的议程中针对时下热门的数据科学研究领域的前沿问题进行报告和讨论。具体报告的学者和主题如下：

Alfred Hero 教授



Learning Sparse Spatio-temporal Data

Tie-Yan Liu 教授



Understanding and Improving the  
Transformer Architecture

Zhi Ding 教授



End-to-End Cloud-Based Machine Learning  
over Bandwidth-Constrained Networks

Stark Draper 教授



Mitigating Stragglers in Distributed  
Computation and Optimization

Hongyu Zhao 教授



Statistical Methods for Genetic Risk  
Prediction

Hua Xu 教授



Natural Language Processing for Clinical  
and Translational Research

R. Srikant 教授



The Loss Landscape of Neural Networks

Zhengyou Zhang 教授



Deeply-Supervised Nets (DSN)

Meisam Razaviyayn 教授



End-to-End Cloud-Based Machine Learning over Bandwidth-Constrained Networks

Ami Wiesel 教授



Multitask Regression and Flood Forecasting

Yang Wang 教授



Scattering Transform and Stylometry Analysis in Arts

Haizhou Li 教授



Speech Processing at Cocktail Party

Dong Zhang 教授



Optimization in Today's Telecom Technology Service- Applications and Challenges

Bo Chen 教授



Deep Probabilistic Autoencoder and Its Applications

Jiangzhou Wang 教授



Fog-networks and its Technologies

Mérouane Debbah 教授



Machine Learning for Wireless Communications

Pheng Ann Heng 教授



Medical Image Analysis and Surgical Simulation – AI and VR Applications for Medicine

Hongyuan Zha 教授



Machine Learning for Wireless Communications

# 实验室介绍

为促进科研绩效的提高，推动研究院科研工作的可持续发展，深圳市大数据研究院建立了3个研究所，研究所下又包含不同方向领域的实验室，组织科研人员进行具体的科研活动。

其中，研究部门3个研究所分别为：大数据基础理论研究所、大数据核心算法与技术研究所和大数据驱动的智能应用技术研究所。

大数据核心算法与技术研究所包含大数据安全计算实验室。大数据驱动的智能应用技术研究所包含了6个不同方向领域的实验室，分别为：医疗大数据实验室、信息系统大数据实验室、金融/企业大数据实验室、社会行为大数据实验室、政务大数据实验室以及智慧城市、交通、物流大数据实验室，由研究所和实验室负责人具体负责组织团队成员开展日常的科研活动。从本期开始，Newsletter将详细介绍这7个实验室的基本情况。

## 医疗实验室 Medical Big Data Lab

大数据在医疗健康的应用是研究院发展的重点方向之一。医疗健康大数据规模极大，超出传统数据库软件工具获取、存储、管理及分析能力范围。医疗健康大数据不仅数据体量巨大、数据种类多、



2019年医学人工智能研讨会参会人员

样，而且数据获取渠道多、数据变化快，并且具有极高的应用价值。医疗健康大数据若能及时获取妥善存储并有效分析，对于改善生活方式、提升诊断治疗效率、提高医疗服务质量、加速药品研发以及优化医疗体制等均有着十分重要的意义。

为了推进研究院在医疗健康大数据方面的应用研究工作，研究院于2018年9月专门成立了医疗大数据实验室并且提供了大量的人力物力资源，实验室目标是借助人工智能和机器学习技术，整合电子医疗病历，医学图像、生物医学及生物信息数据，开发新颖的智能模型来辅助复杂疾病的诊断和治疗。

目前实验室人员包括研究员、博士生、研究生及工程师总计42人。



医疗实验室的老师們

主要项目负责人员如下：

## 实验室主任



### 万翔 博士

本科毕业于中国人民大学信息学院，研究生和博士都毕业于加拿大艾伯塔大学计算机专业，博士毕业后分别在香港科技大学担任博士后和在香浸会大学担任研究助理教授。

目前在深圳市大数据研究院任职研究科学家及医疗大数据实验室主任，主要从事生物医学大数据的模式识别和分析工作，在运用人工智能和机器学习方法分析生物大数据方面有丰富的研究经验并且取得一定的研究成果。目前已在国际知名刊物上发表了将近50多篇期刊文章，包括国际顶尖的自然基因学期刊（影响因子：27.15）、美国人类基因学期刊（影响因子：12.03），生物信息学期刊和统计医学期刊。近期在循证医学方面的元分析工作，备受医学界的关注。

## 研究科学家



### 张纵辉 博士

台湾清华大学通讯工程博士。IEEE信号处理学会2018年度最佳论文奖，2015年IEEE通信学会亚太区杰出年轻学者奖。研究方向信号处理和优化理论在数据通信和医疗健康数据分析中的应用。



### 韩晓光 博士

2009年本科于南京航空航天大学毕业（系第一名），2011年获得浙江大学应用数学系硕士学位，2011年至2013年于香港城市大学创意媒体学院任研究助理，之后于2017年获得香港大学计算机专业博士学位。其研究方究方向包括计算机视觉、计算机图形学、虚拟

现实和医疗图像处理等，在该方向著名国际期刊和会议发表论文19篇，包括顶级会议和期刊SIGGRAPH, CVPR, ICCV, AAAI, ACM TOG, IEEE TVCG等。



### 李 镇 博士

2011年和2014年于中山大学获得学士和硕士学位，在2018年于香港大学获得博士学位。李镇博士同时在2016年和2018年于芝加哥大学，丰田芝加哥研究院进行访问学者研究工作。

李镇博士于2018年九月加入香港中文大学深圳和深圳大数据研究院。研究方向主要是利用数据挖掘和深度学习算法进行蛋白质结构预测，从序列层面到折叠层面。他是蛋白质结构预测竞赛CASP12冠军的主要成员，并获得PLOS CB 2018最新突破和创新奖项。他同时从事机器学习算法和三维计算机视觉问题的研究，例如RGB-D语义分割，形状补全等。

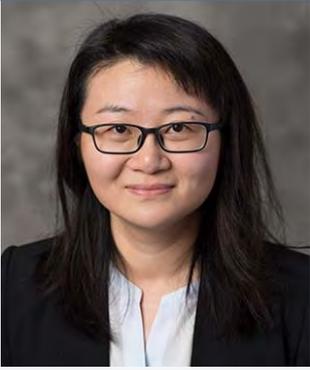


### 刘 李 博士

2018年9月获法国格勒诺布尔-阿尔卑斯大学电子工程博士学位。2018年9月-2019年9月在加拿大瑞尔森大学，电气与计算机工程系从事博士后研究。她目前的研究兴趣包括自动视听语音识别，模式融

合，Cued Speech的发明与研究，嘴唇和手势自动检测，医疗健康大数据整合分析。





**石朋忆 博士**

2014年佐治亚理工获得博士学位。主要研究方医院过程控制和优化。同中国、美国及新加坡多家主要医院及医疗机构有合作研究项目。

## 博士后研究员



**阎文文 博士**

2017年毕业于武汉大学计算机学院，获得工学博士学位。阎博士在机器学习、数据挖掘与生物信息学交叉领域，发表十几篇学术论文，其中包括Bioinformatics、IEEE Transactions on Knowledge and Data Engineering、IEEE/ACM Transactions on Computational Biology and Bioinformatics等顶级期刊。目前阎博士主持国家青年基金一项，省青年基金一项。



**王昌淼 博士**

博士毕业于中国科学院大学，研究方向：基于弱监督的智能医疗影像分析。参与过肺结节检测与分类，胸部骨抑制，肺裂分割，眼底等图像等相关项目研究，此外，还参与过多项国家和省级自然科学基金项目。致力于融合计算机视觉技术在医疗影像领域的应用，协助医生更高效的提高疾病筛查精度。



**林恩轩 博士**

本科毕业于大连理工大学应用数学专业，研究生毕业于美国南佛罗里达大学金融专

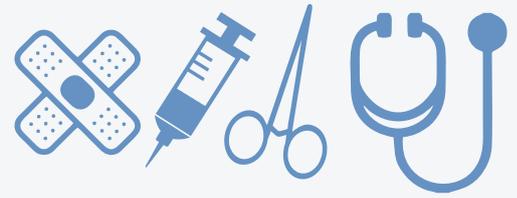


**胡湘红 博士**

本科华南农业大学统计系，研究生中国人民大学统计系，博士为香港浸会大学数学系，

业，博士毕业于香港浸会大学数学与统计学专业。曾参与2018年北京师范大学和香港浸会大学组织的万人计划。林恩萱博士的研究方向和兴趣主要包括循证医学的meta分析，网络药理学研究及高维数据分析。

博士研究方向为基于孟德尔随机化的复杂形状间的因果推断方法。



## 项目介绍

### 电子病历的精准分词级结构化

负责人：万翔，张纵辉

项目计划在目前的电子病历系统中引入中文分词技术，可以解决电子病历数据的结构化难题，实现电子病历用户在实际应用中通过自然语言进行自由文本方式的输入的同时，能够通过计算机的辅助进行病历内容的结构化，为日后的查询、统计、数据交换提供基础。通过将非结构化的自由文本病历转换为结构化电子病历文本应用外，还可以实现电子病历查询的快速索引、辅助结构化输入等实用功能。

### 电子病历的实体识别

负责人：万翔，张纵辉

项目主要解决如何从电子病历中的自由文本部分识别关键信息。目前，在医疗文本标注领域尚无成熟的系统，我们需要结合中文语言的特点，对现有的中文医疗文本标注方法做出改进，建立一套完整可行的标注准则并依据这个准则建立相应的词库，同时改进和提出新的机器学习算法开发一套中文电子病历自由文本自动标注系统，对医疗文本中的各类实体及断言(包括否定)进行自动标注。

### 电子病历的关系抽取及知识图谱构建

负责人：张纵辉，万翔

现有知识图谱构建技术在医学领域中普遍存在效率低，限制多，拓展性差等问题。本项目针对医疗数据跨语种，专业性强，结构复杂等特点，对构建医学知识图谱的关键技术进行了自底向上的全面解析。现阶段，国内医疗领域中基于知识图谱的疾病预测研究工

### 影像报告的自动生成

负责人：万翔，张纵辉

放射学图像和病理学图像等医疗图像在医院与诊所已有普遍的应用，但阅读与理解通常是由专业的医疗从业者完成，编写医疗图像报告艰难、乏味且耗时。该项目旨在从医疗影像中自动生成医疗诊断报告；研究涉及医疗影像的分析以及诊断文本的生成，通过分析特定影像（例如胸片）的关键区域并利用自然语言处理技术

作才刚起步，所以构建基于知识图谱的医疗知识系统对于智慧医疗的发展具有一定的辅助意义。

实现诊断报告的自动生成，以辅助医生进行高效准确的诊断；研究同样与医疗的实际项目相结合，在提升报告质量和减轻医生的工作负担方面都具有重要意义。

## 肾结石成分智能分析

负责人：万翔，张纵辉

肾和输尿管结石产生的梗阻可导致肾绞痛，这是一种重度、急性肋腹痛，可放射至同侧腹股沟，并常伴有恶心和呕吐。临床上，在取出结石后，对结石进行物理分析，得到结石成分，通过结石成分分析患者病因形成及制定相应预后方案。如果可在治疗之前获取结石的成分信息，为制定治疗方案提供相应的参考信息。这个项目将对通过肾结石患者CT影像、检查报告及历史病历预测结石种类及成分，有望替代传统的结石成分物理分析方法。

## 胸部骨结构抑制分析

负责人：王昌淼，韩晓光

该项目旨在从X光胸片中抑制骨结构（肋骨和锁骨），增强X光胸片中疾病的对比度；研究也涉及胸部肺结节疾病的检测与分类；骨抑制后的胸片更有利于肺结节的检测，辅助医生通过X光胸片早期筛查肺癌。

## 乳腺钼靶微钙化检测

负责人：王昌淼，韩晓光

该项目旨在从乳腺钼靶影像中自动检测微钙化区域，研究涉及小目标检测和异常检测方法；钙化点的检出是乳腺癌早期诊断中最重要的部分，为诸如肿块、钙化簇检测等问题提供了基础，这有利于早期筛查乳腺癌。

## 基于眼底图像的眼科疾病智能分类诊断

负责人：韩晓光

该项目关注眼底健康，基于深度学习实现眼底图像中青光眼，白内障，糖尿病视网膜病变等多种疾病的自动识别与分类工作，让常规眼底疾病筛查更加高效、准确、快速。

## 医学影像的多器官分割及肿瘤检测

负责人：李镇

该项目基于deep learning来实现CT，MRI格式

## 基于医学影像的子宫检测及阴道闭锁异常检测

负责人：李镇，刘李

该项目主要是医疗图像的多器官或病灶分割，旨在从CT或核磁共振等3D医疗图像中自动分割出关心的器官或病变区域；研究也涉及医疗诊断报告指引的器官分割，通过visual grounding实现从诊断报告定位病变区域；研究同样与医疗的实际项目相结

的医学图像的语意分割任务。包括多器官分割，肿瘤损伤分割等。只存在部分标柱样本的半监督和只有image-level标注的弱监督医学图像分割，及小样本的one-shot和few-shot的分割任务。

合，探索MRI图像中罕见女性生殖疾病的诊断，辅助医生阅读MRI图像。

## 用数据驱动的决定工具提高医院流程管理和病人体验

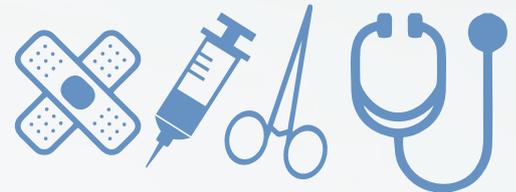
负责人：石朋忆

该项目旨在把这些方法和经验带到中国的医院，推动中国医院的数据驱动管理，提高医院的管理效率。基于这个目的，我们和深圳罗湖医院集团开展了深度合作。我们利用罗湖医院的数据进行验证和测试，并将在真实的医疗环境中实施（如妇产科门诊）。我们的最终目标是提供数据驱动的决定工具，实时地给出运营的最优决策。具体而言，我们从罗湖医院门诊部入手，利用大数据来改善该部门的运作流程并提高其效率。其中主要的两个项目包括优化患者流程、提高患者就诊体验及提高病人需求管理用以改善病人资源的利用。

## 融合多种癌症多组学数据提取癌症靶点

负责人：闵文文，万翔

随着生物医疗技术的快速发展，大量复杂的多癌症多尺度组学数据以前所未有的速度在被收集。从这些数据中，挖掘癌症异常生物分子表达模块，对理解癌症机制，发现癌症靶标具有重要意义。然而分析这些癌症组学数据并从中提取有用信息是一个大的挑战。在统计学上，这些复杂的高维数据可表示成高维张量。本项目拟开展寻求整合分析同一种癌症的多种组学数据预测临床风险和发现生物分子标志物的张量模型与算法。



## 基于基因组的多种疾病症状的因果关系构建

负责人：胡湘红，林恩轩，万翔

孟德尔随机化法（Mendelian Randomization, MR）是一种以基因变异（如APOE）作为工具变量来研究暴露因素（如血浆胆固醇水平）和结局变量（如癌症）之间的因果关系的方法，被广泛应用于复杂疾病或性状的因果研究中。该方法突破了传统的观察性流行病学在研究因果关系的多种限制如反向因果、混杂因素等，大量关于基因型和性状的GWAS概括型数据可公开获取。然而人类基因组的复杂性（多基因效应，基因多效型等问题）会造成模型的假设受到冲击。本课题将提出更加新颖及可靠的分析模型及算法。

