



2017中美智能医疗大数据峰会



王猛

睿至医疗大数据产品经理

拥有十余年医疗项目经验，在与客户的长期沟通中，积累了丰富的医疗行业实践经验，熟悉医疗行业信息化及数字化转型的真实需求和实际痛点，同时，更在近年来结合大数据行业技术与经验，形成了深厚的医疗行业大数据行业能力。



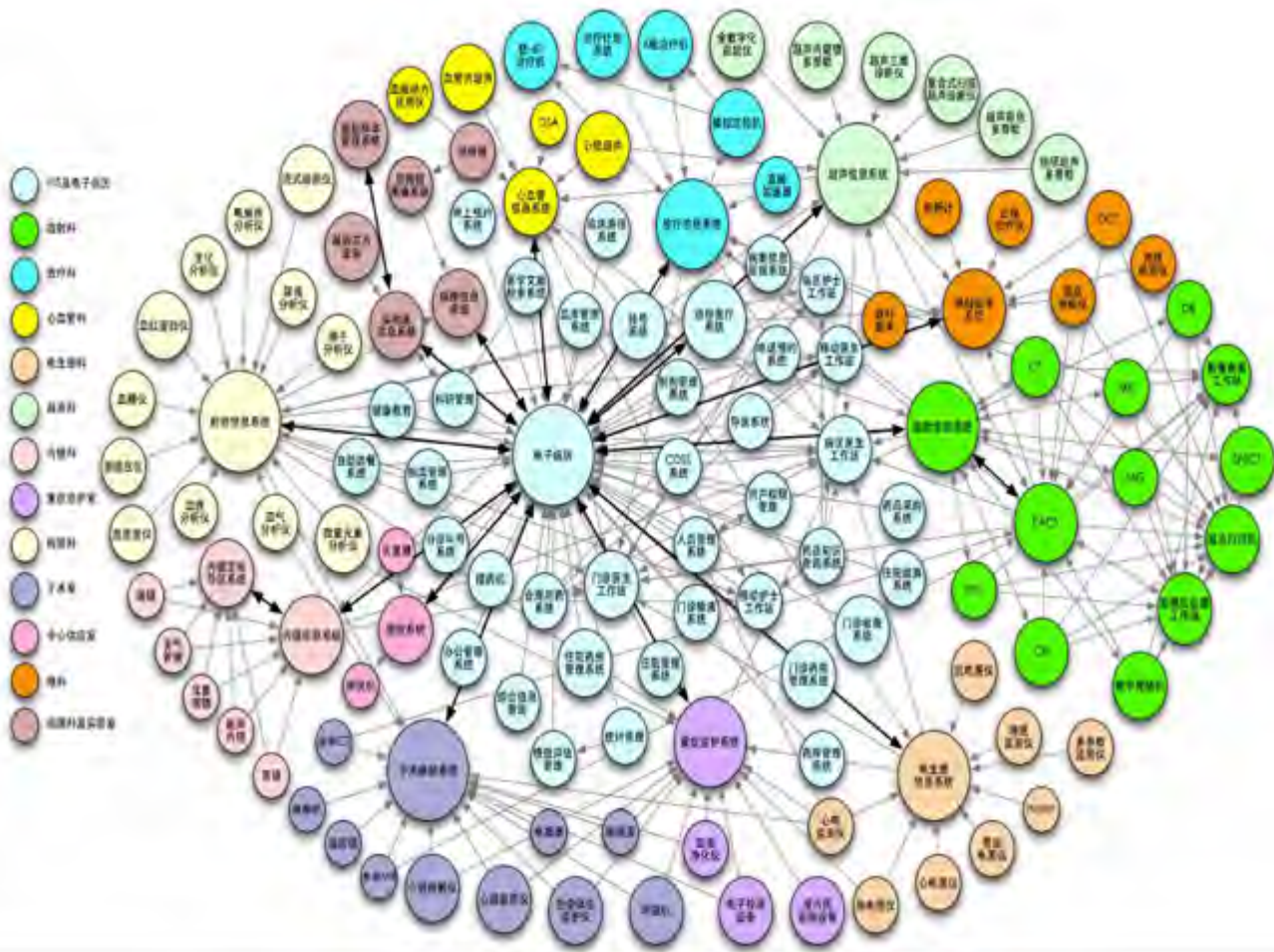


唤醒沉睡的 海量医疗数据

广义医疗大数据生态概览



医院临床大数据综述



结构化数据

患者建档信息、病历分类编目、用药记录、医疗费用记录、住出院记录...

半结构化数据

病历文书、检查报告、手麻记录、体征数据...

非结构化数据

医学影像学图像...

外部数据

移动医疗设备、可穿戴设备、外部知识库...

根本原因

数据非标准化

缺乏标准，各系统数据规范不一致，大量无效脏数据

病历非结构化

电子病历、诊断报告等非结构化数据，传统处理技术无法识别其含义

数据分散

院内业务系统由多家开发商提供，数据分散、共享性差

现状和挑战

病历搜索

缺乏基本的精准和高效的病历搜索技术和手段，医生无法充分利用病历数据为患者提供更精准更有效的治疗方案，甚至经常放弃对病历数据的使用，直接制约医生的工作质量和发展。



科研管理

临床科研数据的采集和分析仍以人工搜集、录入病历数据加简单统计软件的方式进行。耗时长、数据完整性和准确性差、分析结果不直观。科研历史数据和成果数据的沉淀积累和重复利用率低。



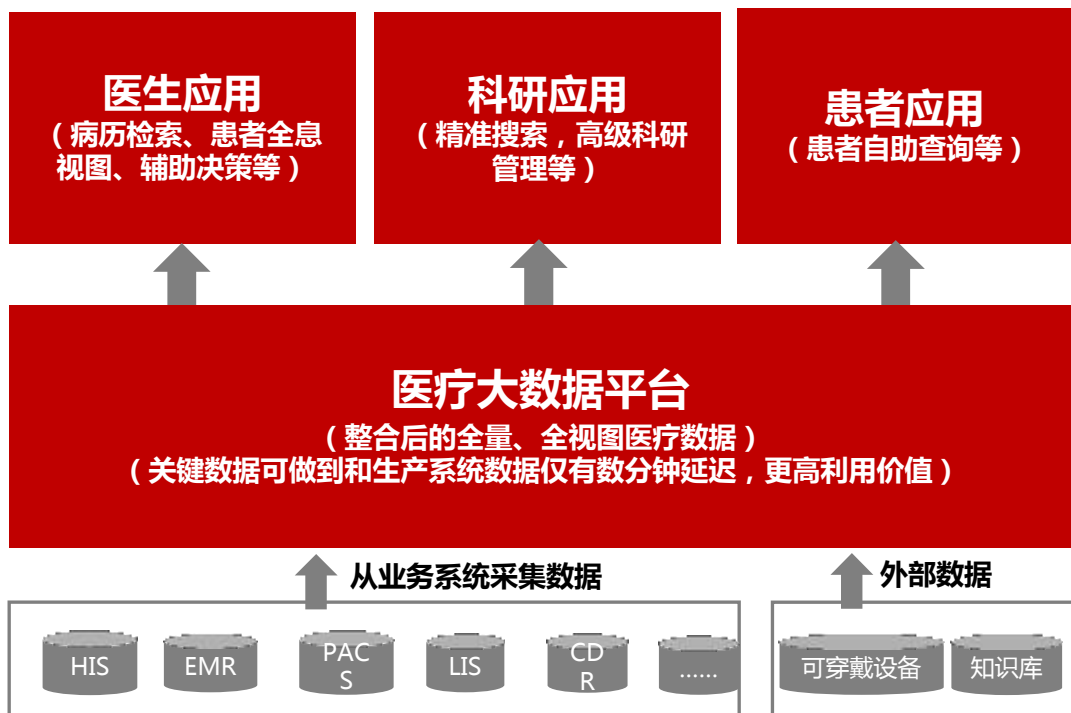
辅助决策

面对海量的、跨系统分散的、标准不一的数据，现在基于各个业务系统做的统计分析，分析结果的不准确、不全面，尚不能为医院管理层提供很好的辅助决策。



睿至医疗大数据平台产品定位

应用：基于医疗大数据平台，开发多个方向的应用系统，聚焦医生应用、科研应用、患者应用，更好的帮助医生、患者以及医院管理层进行医疗数据的深入分析、挖掘和利用。



平台特点：

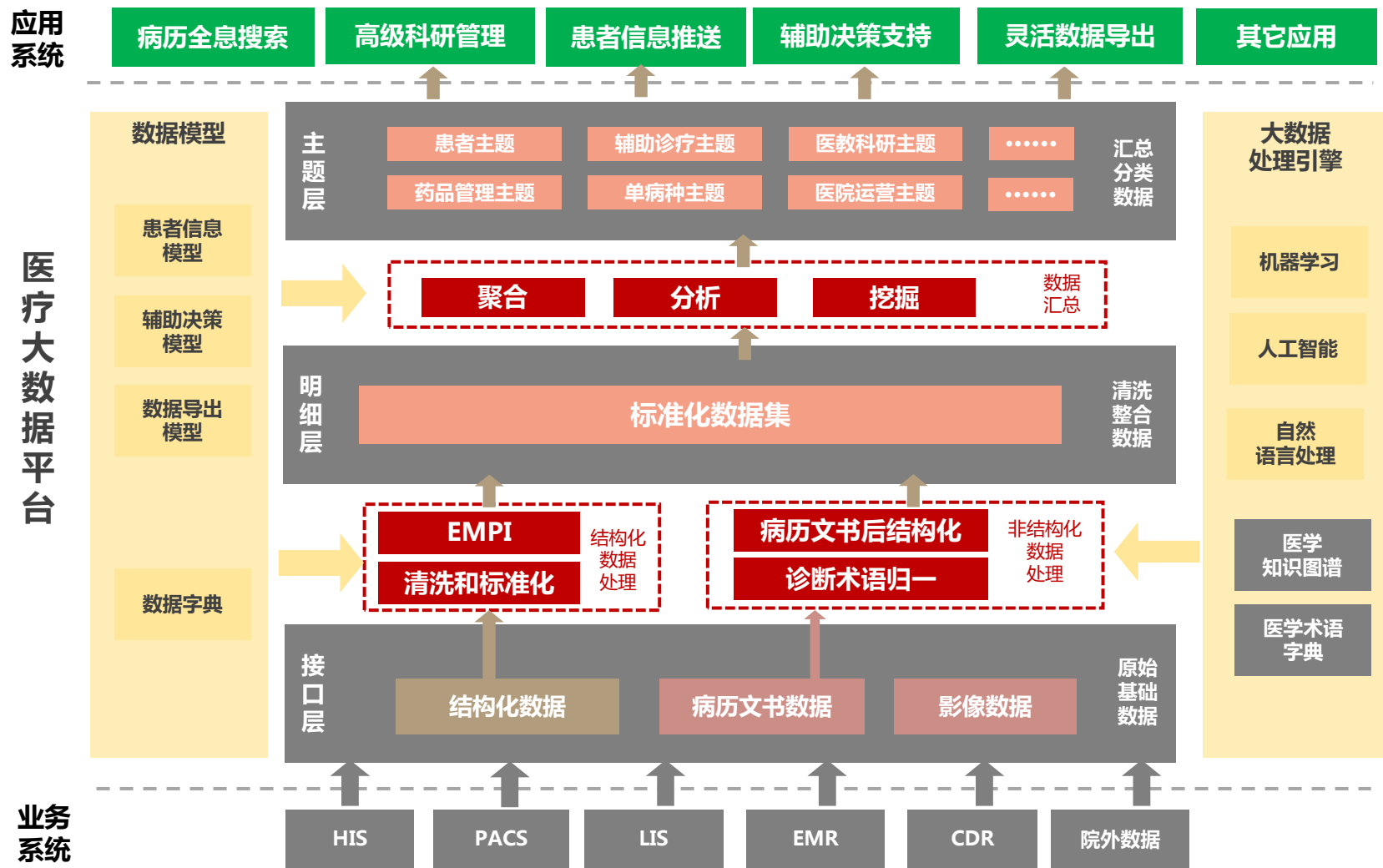
数据标准化：以患者为中心，按照标准数据格式整合医疗业务和医院管理的数据，建立全院级的信息资源中心，并在此基础上开放各种标准的数据访问服务。

高性价比影像存储：使用hadoop大数据存储技术，大量存储快速增长的影像数据，并能快速提取。

非结构化数据应用：通过机器学习、自然语言理解技术，对非结构化数据进行后结构处理，最终存储为结构化数据，并支持对它的搜索、利用。

采集：准实时抽取和采集医院内部各种业务系统数据，包括：HIS、PACS、LIS、CDR等院内院外系统数据源。

医疗大数据平台整体架构图



行业标准

健康档案相关卫生服务基本数据集标准（试行）

国际疾病分类ICD-10

手术与操作字典(ICD-9-CM-3)

医疗机构临床检验项目目录（2013年版）

HL7 Reference Information Model
Vers. 2.07

.....

汇聚医学知识
大量数据积累



参考国内外医疗数
据标准和顶层设计

医疗大数据平台标准

业务域
13

标准数据
字典表
68

数据字段
485

数据指标
6033

医疗数据模型构建完成后，可以形成统一的数据结构和数据内容标准。

- 数据结构的标准化可以更好的促进互联互通，更好的支持其他应用的扩展。
- 数据内容的标准化，能够保证基于一致规范的医疗数据进行更深层次挖掘、研究、分析。

在医疗数据结构模型中，针对临床数据，形成 13 个业务域分类，共485个数据字段，6000+个指标

业务域分类 (13个)	数据字段	指标	临床数据内容说明
患者	34	27	患者的姓名、性别、病人ID、证件、联系人、住址、医保、过敏及不良反应等
就诊记录	29	16	患者的历次就诊信息
症状	2	2	患者的症状相关信息
诊断	12	5	患者的诊断相关信息
用药	17	14	患者使用过的所有药物，包括门诊处方和住院药物医嘱等信息
手术	37	4	患者接受过的所有手术和麻醉信息
治疗类	55	4	患者接受过的所有治疗信息
检验	42	2405	患者接受过的所有检验及检验结果信息
检查	35	2607	患者接受过的所有检查及检查报告信息
分子病理	25	350	患者接受过的所有分子病理检查信息
病理	30	500	患者接受过的所有一般病理检查信息
生命体征	22	19	患者采集并记录的所有生命体征信息
医嘱	79	14	患者就诊过程中产生的所有门诊及住院医嘱信息
病历	66	66	医疗业务需要的病历文本结构化后信息
总数	485	6033	

在数据元值域和术语字典方面（共68张），3个国际标准、7个国内标准，形成58个院内标准。

国际标准字典表 (3)	国内标准字典表 (7)
国际疾病分类ICD-10扩展	性别字典
国际疾病分类ICD-10	国际字典
手术与操作字典(ICD-9-CM-3)	民族字典
	婚姻字典
	行政区字典
	职业分类字典
	教育程度字典
社会关系字典	工作人员分组情况
血型字典	手术等级字典
疾病字典	就诊类型字典
科室字典	手术顺利标志字典
诊断类别字典	号类字典
检查诊断字典	费别字典
临床诊疗项目字典	检查类别字典
诊疗项目分类字典	入院方式字典
标本类型字典	检查结果标识
检查结果状态字典	出院方式字典
检验项目与分类对照	住院目的字典
检验报告项目字典	身份字典
检验报告与申请项目对照	科室临床属性字典
检验结果标识字典	药物字典
检验联机仪器字典	检查分组字典
医保类别字典	长临标识字典
工作人员字典	药品处方属性字典
工作类别字典	药品名称字典
工作组字典	药品类别字典
	药品库存单位库房字典
	剂量单位字典
	药品供应商目录
	医嘱执行频率字典
	药品毒理分类字典
	给药途径字典
	医生职位字典
	时间单位字典
	医嘱类别字典
	医嘱状态字典
	会计科目字典
	手术操作字典
	切口愈合情况字典
	手术类型字典
	入院病情字典

院内字典表 (58)

病历后结构化

后结构化处理，可以将病历文书中的症状、体征等非结构化数据进行结构化存储，提取完成后通过明确的症状和体征表现，可以更有针对性的搜索病历，更好的帮助医生进行诊断决策。

自然语言文本

患者男，69岁，体温：37.0℃，脉搏：98次/分，呼吸：22次/分，血压：130/80 mmHg。患者4小时前无明显诱因出现上腹部持续性绞痛，无发热，无恶心、呕吐，无反酸烧心等不适。急查血常规示：WBC $11.59 \times 10^9/L$ ，NE%88.8%，HGB155g/L，PLT $198 \times 10^9/L$ 。立位腹平片示：消化道穿孔，膈下积气，积液可能。考虑病人病情危重，现为行急诊手术经急诊以“消化道穿孔”收入院。

■ 生理名词 ■ 数字词 ■ 判断词 ■ 方位词 ■ 解剖部位词
■ 症状词 ■ 检验词 ■ 检查词 ■ 诊断词 ■ 症状性质词

计算机不可识别、不可利用的半结构化医疗数据

以医疗信息学为基础

使用机器学习

使用知识图谱

结构化数据

患者症状

症状	逻辑	持续时间	部位	性质
绞痛	有	4小时	上腹部	持续性
发热	无	-	-	-
恶心	无	-	-	-
呕吐	无	-	-	-
反酸	无	-	-	-
烧心	无	-	-	-

患者生命体征

体温	37.0℃
脉搏	98次/分
呼吸	22次/分
血压	130/80 mmHg

物理检查

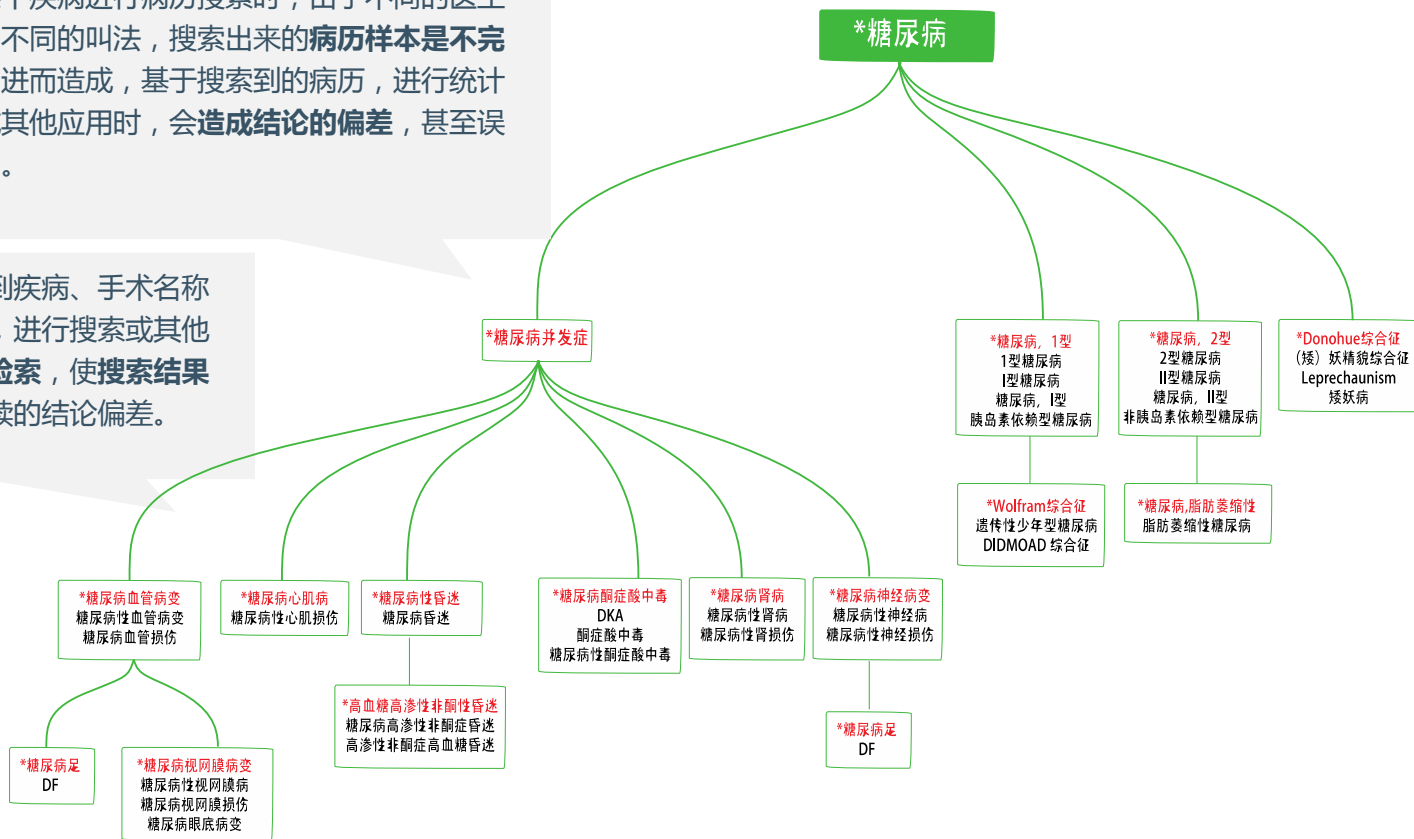
检查项目	解剖部位	异常所见
立位腹平片	消化道	穿孔
立位腹平片	腹腔	积气
立位腹平片	腹腔	积液

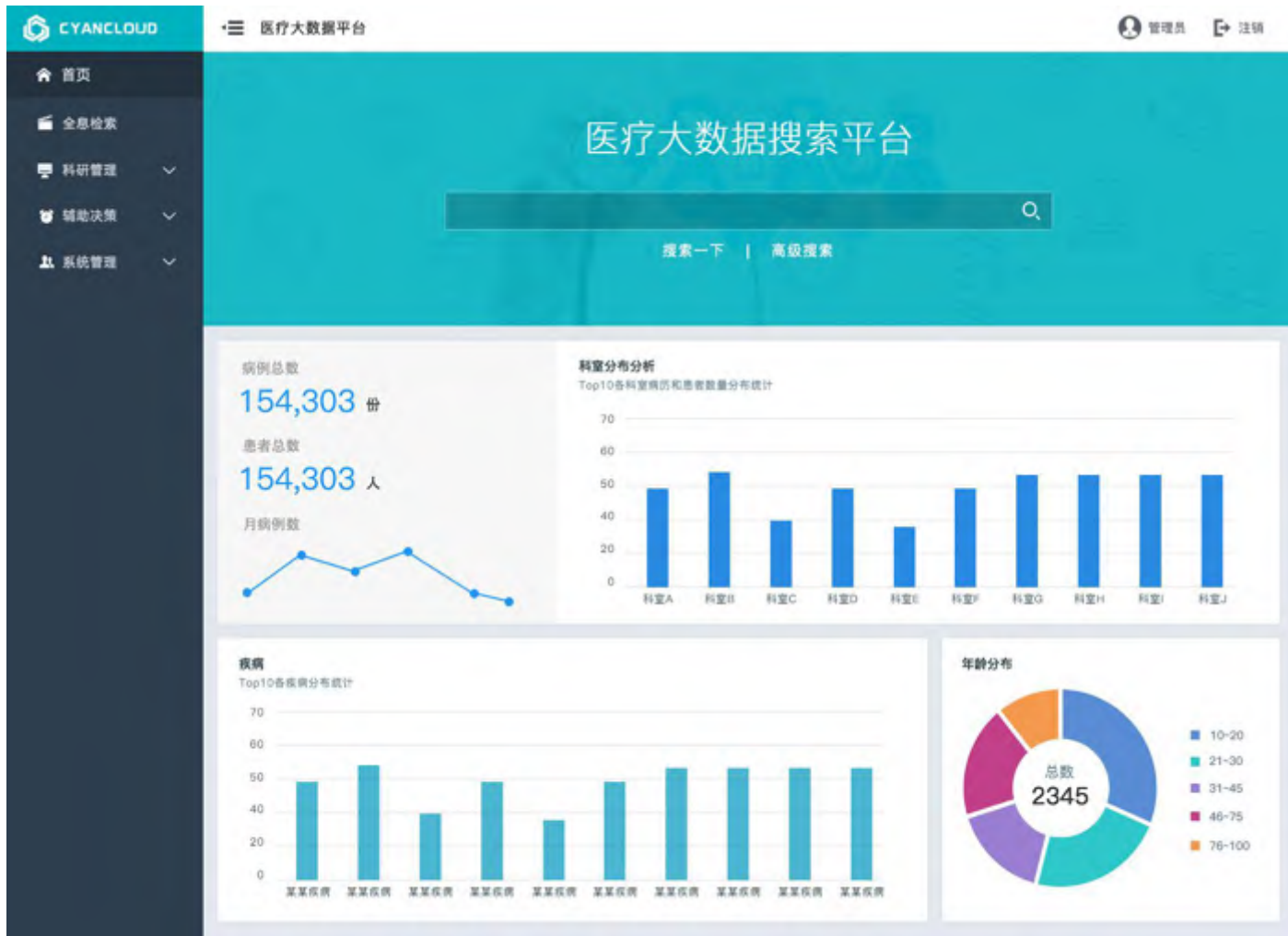
可存储、查询、统计、分析和挖掘的数据结构

通过人工智能加人工标注的方式，对存在大量同义词的诊断等专业术语，进行识别同义词并构建同义词库，从而避免按诊断搜索时数据不完整等问题。

当对某个疾病进行病历搜索时，由于不同的医生会使用不同的叫法，搜索出来的病历样本是不完整的。进而造成，基于搜索到的病历，进行统计分析或其他应用时，会造成结论的偏差，甚至误导用户。

完成同义词库构建后，对使用到疾病、手术名称等存在大量同义词的医疗数据，进行搜索或其他应用时，会对全部同义词进行检索，使搜索结果完整，避免遗漏，防止出现后续的结论偏差。



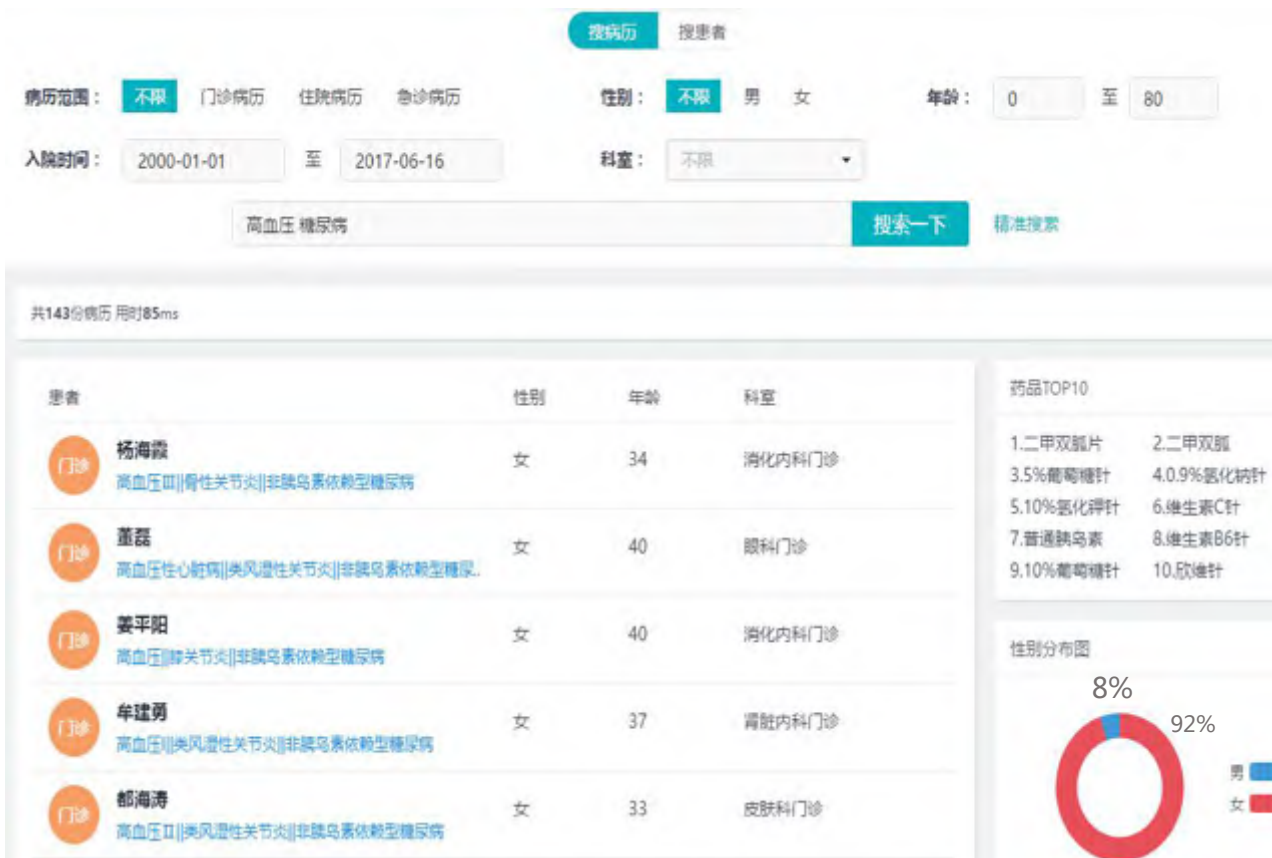


功能简介

- **搜索入口：**提供病历搜索和精准搜索入口
- **统计指标展示：**提供病历数量，每月就诊数量折线图，科室分布 TOP 10，疾病分布 TOP 10和年龄段分布图表。帮助用户直观快速了解病历概况。

病历搜索-关键词及条件组合搜索

示例说明：能搜索到既有高血压又有糖尿病等多种疾病的病历列表，并可增加年龄、性别、科室等条件缩小检索范围，从而最快速的找到想要的病历。



搜索病历 搜患者

病历范围：不限 门诊病历 住院病历 急诊病历

性别：不限 男 女

年龄：0 至 80

入院时间：2000-01-01 至 2017-06-16

科室：

高血压 糖尿病

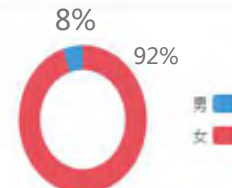
共143份病历 用时85ms

患者	性别	年龄	科室
<input checked="" type="radio"/> 门诊 杨海霞 高血压III 骨性关节炎 非胰岛素依赖型糖尿病	女	34	消化内科门诊
<input checked="" type="radio"/> 门诊 董磊 高血压性心脏病 类风湿性关节炎 非胰岛素依赖型糖尿病	女	40	眼科门诊
<input checked="" type="radio"/> 门诊 姜平阳 高血压 骨关节炎 非胰岛素依赖型糖尿病	女	40	消化内科门诊
<input checked="" type="radio"/> 门诊 牟建勇 高血压 类风湿性关节炎 非胰岛素依赖型糖尿病	女	37	肾脏内科门诊
<input checked="" type="radio"/> 门诊 郝海涛 高血压II 类风湿性关节炎 非胰岛素依赖型糖尿病	女	33	皮肤科门诊

药品TOP10

1.二甲双胍片	2.二甲双胍
3.5%葡萄糖针	4.0.9%氯化钠针
5.10%氯化钾针	6.维生素C针
7.普通胰岛素	8.维生素B6针
9.10%葡萄糖针	10.欣维针

性别分布图



8%	92%
<input type="checkbox"/> 男	<input checked="" type="checkbox"/> 女

支持医生进行病历分析及临床科研的病历搜索

功能简介

- **统一入口**：整合多个医疗业务系统数据，提供统一检索入口
- **数据完整**：集合院内结构化数据及非结构化数据
- **秒级检索**：采用大数据架构存储及提取技术，实现秒级检索。

精准搜索-结构化数据组合搜索

组合**多个搜索条件**，对病历进行**精准搜索**，查找**相似病历**



示例说明

通过左侧精准搜索条件设定，能够搜索到的患有糖尿病的女性，服用二甲双胍，血糖控制依然不理想的所有病历。然后医生可以查看某一个病历详细数据，分析这些病历的共性，血糖控制不理想的原因。

精准搜索-结构化及后结构化指标组合搜索

保存 删除 搜索历史 2017/6/20 9:19 清空历史 立即搜索

且 或

- 诊断名称 包含 肺炎
- 性别 = 女
- 转归情况 = 痊愈 **来源** 出院小结
- 20 ≤ 就诊发生时年龄 ≤ 35

且 或

- 主诉：咳嗽 **来源** 入院记录-现病史
- 体温 大于 37.3℃ **来源** 入院记录-体格检查

+
-
刷新
删除
收藏
?

示例说明

通过左侧精准搜索条件设定，可搜索到患过肺炎具典型肺炎特征并痊愈的中青年女性所有病历。医生可以分析每份病历的整体治疗方案，学习和优化原有治疗方法。

患者全息视图-诊疗信息查看



将不同业务系统中的数据进行**有机整合**，以**患者及患者单次就诊为维度**，进行集中展示。实现患者**最广（涵盖多系统数据）最深（历次就诊所有诊疗数据）**的诊疗信息呈现。



功能简介

- **全信息展示**：把当前患者历次诊疗记录整合展现
- **全维度展示**：每一次诊疗的检查、检验、医嘱、电子病历统一展示。免去跨多系统查询的不便

患者自助查询-多维度查询

患者**使用APP自主查看**在院内就诊记录、用药记录及所有诊疗单据结果。节省医院资源投入，提高就医效率，**提升患者与医院的粘性**及**患者就诊感受**。



首页

就诊记录

按检查日期浏览

按检查类别浏览

按检查部位浏览

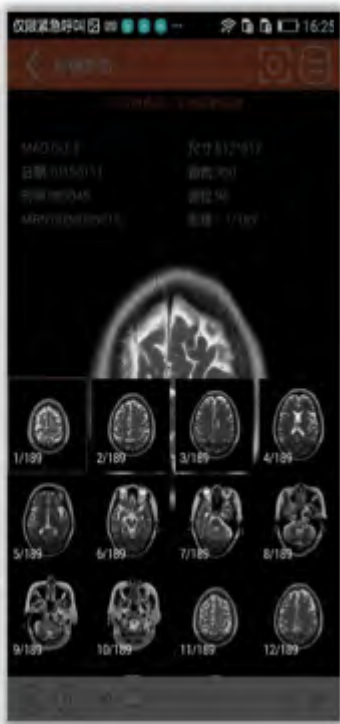
患者自助查询-诊疗结果展示



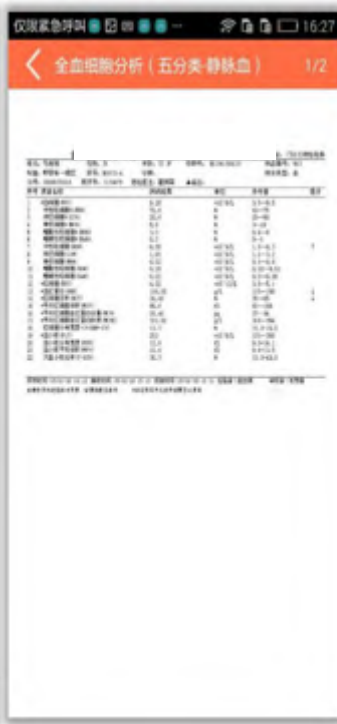
病理检查报告



影像报告浏览



影像缩略图浏览



检验报告浏览



门诊病历报告

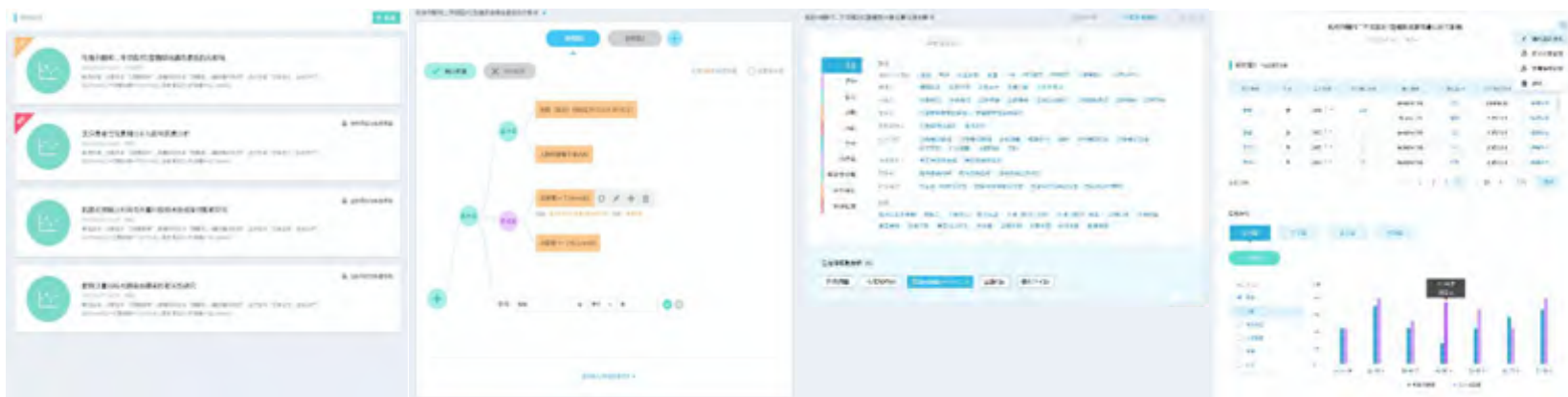
整合所有结构化和后结构化数据，**实时获取**准确病历数据，**直观实用**的统计分析，并以**图表形式呈现**。帮助医生**优化科研耗时**，极大**提升科研效率**。

1 秒级入组

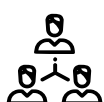
2 数据挖掘

3 指标观测

4 在线分析



队列设计



患者入组



观察指标设置



私有项目管理



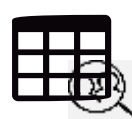
权限管理



智能CRF



随访管理



描述性分析表格
Tableone



可视化统计
分析结果



数据导出



生存周期

高级科研管理-病历快速入组

✓ 纳入标准 X 排除标准 共有349名研究对象 设置样本数

且

- 住院(就诊)时间在2010/3/1-2014/3/1
- 入院科室等于骨内科

或

- 血尿素 >= 7.0mmol/L
时间: 从 2015-01-01 到 2016-01-01 次数: 大于2次
- 血肌酐 >= 150.1μmol/L

新增: 性别 + 等于 男

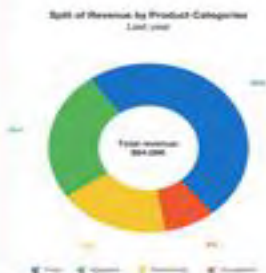
搜索观测指标

患者	患者
基础个人信息	性别 年龄 出生日期 体重 BMI ABO血型 RH血型 有无输血史 有无过敏史
婚育史	婚姻状态 是否怀孕 是否流产 生育个数 是否绝育 ✓
症状	月经史 初潮时间 月经周期 是否规律 是否痛经 是否有血凝块 是否有颜色异常 是否绝经 是否停经
诊断	家族史 有无家族聚集性疾病 家族聚集性疾病名称
用药	免疫接种史 有无接种过疫苗 疫苗名称
手术	生活习惯 过去是否吸烟 现在是否吸烟 日吸烟量 戒烟时间 烟龄 现在是否饮酒 过去是否饮酒 饮酒类型 日饮酒量 戒酒时间 酒龄
治疗类	传染病史 有无传染性疾病 传染性疾病名称
实验室检查	既往史 既往患病时间 既往疾病名称 既往疾病治疗效果
生命体征	就诊信息 历史总门诊就诊次数 历史总住院就诊次数 历史总急诊就诊次数 历史总诊疗费用
物理检查	病历 病历发生时年龄 病案号 入院方式 就诊类型 入院(就诊)时间 入院(就诊)科室 出院时间 出院科室 是否转科 住院天数 是否CCU/ICU 总金额 公费金额 自费金额 实付金额 医保类型

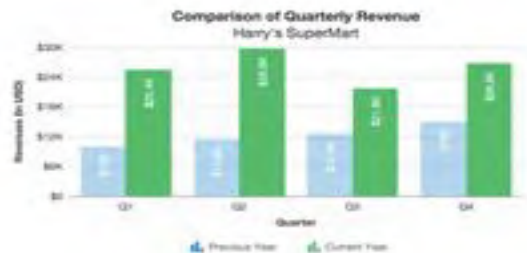
已选择观测指标 (6)

日吸烟量 有无输血史 空腹血糖值(mmol/L) X 空腹C肽 餐后2h C肽

性别



年龄



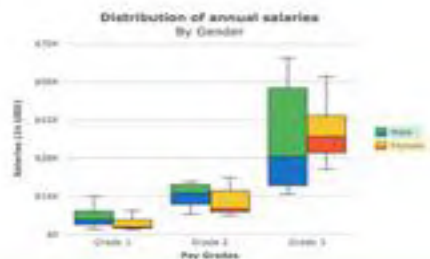
BMI



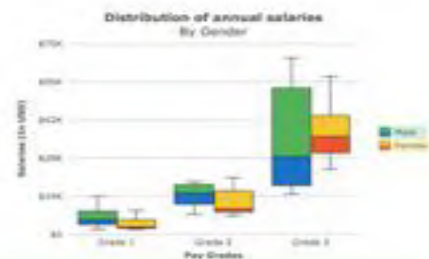
医嘱用药



TC (mmol/L)



TG (mmol/L)



通过灵活动态丰富的**指标数据展示**，帮助**医院决策层、管理层**更清晰的了解医院**当前运营现状、目标达成情况**。帮助医院更好的**识别风险**，实现**提早控制**和**辅助决策**。



功能简介

- **生动展现**：通过对某个时间段内医院数据的统计，分析，以仪表盘、饼图等形式动态展现
- **指标丰富**：可以查看挂号人次、在线人数、门诊收入、住院收入的详细信息。
- **决策支持**：院长驾驶舱数据可以帮助医院管理层及时掌握医院运营状况，为医院管理层提供数据决策支持

多维度可定制实时统计报表展示


医疗大数据平台

管理员
注销

统计报表

时间范围

至

确定

就诊人次总量

	就诊人次	占比率%
门诊	12535	84.6
急诊	1296	8.8
住院	578	6.6
总计	14809	100

再入院率分析

住院科室: 全部科室

	就诊人次	占比率%
7日再入院	19	6.6
30日再入院	31	10.7
90日再入院	31	10.7

门诊初、复诊统计

门诊科室: 全部科室

	就诊人次	占比率%
初诊	10018	71
复诊	4091	29
总计	14109	100

门诊药品收入

	收入金额	占比率(%)
西药	4,140,079	84.6
中成药	13,78,547	22
草药	768,984	12.2
总计	6,287,610	100

住院药品收入

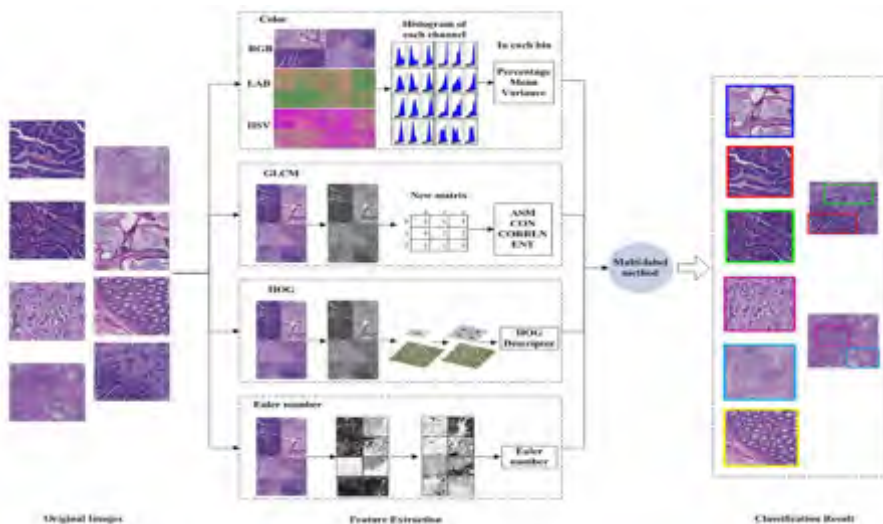
	收入金额	占比率(%)
西药	9,567,743	79.8
中成药	2,078,712	13.4
草药	344,478	2.8
总计	11,990,933	100

日常观测指标

	数值
门诊患者均次费用(元)	368.86
住院患者均次费用(元)	3756.15
平均住院日(天)	17
手术总量(例)	369

1、机器阅片：

医疗数据平台传输影像原图辅助诊疗系统，辅助诊疗系统结合历史影像人工智能和深度学习的技术，智能提取特征信息，识别病灶，推荐疑似诊断，并传输回医疗数据平台做为影像辅助诊断帮助医生进行诊断决策。



2、辅助诊疗：

医疗数据平台传输症状、检查、检验结果等诊疗数据给辅助诊疗系统，辅助诊疗系统给出下一步检查建议、处置建议、用药建议等疾病整体治疗方案，并回传医疗数据平台，作为医生决策的重要参考。

分期	0期	I期	II期	III期	IV期
治疗方案	手术治疗为主，可能需要化疗，酌情使用放疗和内分泌治疗或靶向治疗	手术治疗为主，可能需要化疗，根据具体情况使用放疗和内分泌治疗或靶向治疗	先化疗再手术，需要放疗或靶向治疗	先化疗再手术，需要放疗或靶向治疗	不能手术，依靠其它治疗方法延长生存
生存率	可治愈，长期存活	20年无瘤生存率68-80%	5年生存率约80%，10年生存率60%	5年生存率约50%，10年生存率20%	5年生存率15%

治疗方法	手术	放疗	化疗	内分泌治疗	靶向治疗
治疗时间	几小时	持续一个月，每周5天休息两天	共6-8周期	持续5-8年	持续1年，三个星期用一次，用十四次左右
费用	1万-1.5万	1万左右	几百元-2万元/周期	30元-2000元/月	约40万



北京睿至大数据有限公司（简称“睿至大数据”），是企业数字化转型战略级伙伴，致力打造中国卓越的云计算及大数据运营商。睿至大数据以“客户需求”为先，秉承“极致创新”理念，专注于云计算、大数据与智慧城市等领域的创新研究与行业实践。2014年成立以来，实现了跨越式发展：以产业前瞻研究和行业深度理解为双引擎，基于云计算、大数据等先进技术形成了**贯穿企业数字化转型生命周期的“全栈式”产品及解决方案**，包括IT战略咨询、规划、设计、实施、运营与商业模式创新等，并由来自IBM、Vmware、HDS、华为、神州数码等国内外一线企业的专家团队，通过遍布全球的服务网络，为用户输出同步国际的“云+数据”体系，实现IT云化和数据价值激活，提升整体核心竞争力。目前，政府、能源、金融、电信、医疗和教育等行业客户已与睿至大数据开展深入的合作。



2017
THANKS

感谢聆听！