

# 病历深度理解技术推动医疗人工智能应用

医渡云资深副总裁 胡健

# 国内现状: 医疗资源、医疗质量不足, 国家推动基于人工智能的医学应用



### 民生需求

- 医生资源缺口严重, 2.2/千人四
- 社区医生本科以下约占56%[2]

# ATA

### 国家政策

· 国家推动基于人工智能在医学 研发与应用

研发与应用

# --临床需求

- 减少误诊
  - 目前 误 诊 率 27.8%
- 提高医生效率
  - 缩短确诊时间
  - 提升输入效率
- 减少医疗差错
  - 医疗差错占医疗事故的20%[3]

### 经济需求

• 控制医疗费用: 美国糖尿病CDSS 应用后每年节省343亿美元

多样化的

刚性需求

• 节约医疗资源[4]

- 2019中国人工智能医疗白皮书
- · 2016年卫生部统计年第HIMA2021会议资料
- · 新华社《广东推出"基表医疗积外风疗毒伤的转去数
- · OMAHA CDSS白皮书www. chima. org. cn

# 国内现状: 亟需利用信息化手段来提高治疗规范性



- 全国人大代表, 北京大学首钢医院院长 顾晋表示:
  - 1/3的临床医生不了解肿瘤的指南与规范;
  - 在大规模的调研数据中看到,临床医生对指南的答题正确率为67%,四分之一的医生正确率不足50%。
  - 不了解临床指南就很难保证疾病管理与治疗的规范性。
- 利用信息化手段来辅助诊疗,目的是提升疾病管理与治疗规范性的比例。
  - 提供指南要求的规范诊疗行为与标准;
  - 评价医疗行为是否符合要求。

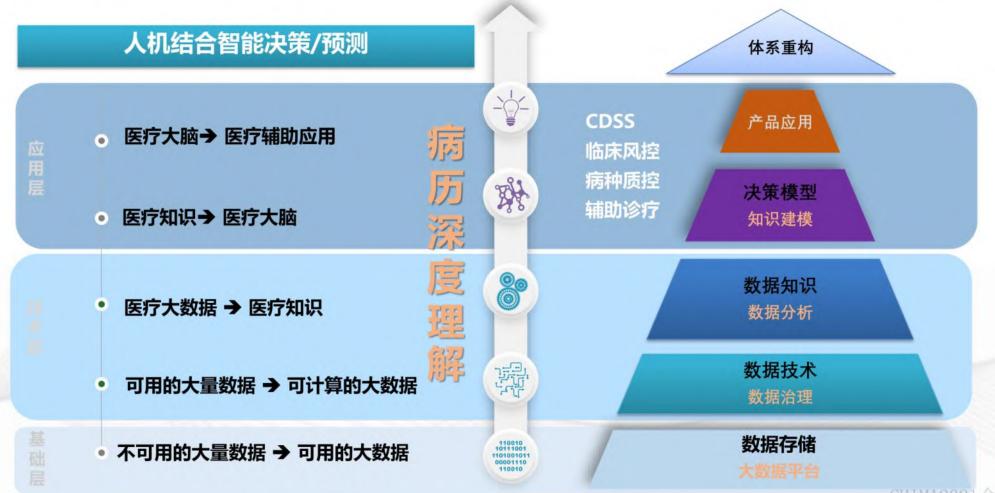


恶性肿瘤治疗的四大基石

CHIMA2021会议资料 未经授权请勿转载

# 医疗人工智能发展路线





# 患者诊疗模型



### 全面覆盖医院各系统各场景的可扩展的患者诊疗模型



#### 模型重构

HIS	EMR	手麻	重症	放疗	输血	护理	病案
LIS	RIS	超声	内镜	病理	心电	核医学	基因

# 疾病模型



	恶性肿瘤			
相	基本信息			
同	一诉	五史		
信	体格	检查		
息	专科	检查		
	诊断			
	肿瘤标志物	分子病理		
	CT检查	MRI检查		
	PET-CT	ECT		
疾	内镜检查	病理检查		
病	手术治疗			
差异	化学治疗	放射治疗		
信	内分泌治疗	靶向治疗		
息	介入治疗	免疫治疗		
	随访			
	其他			

冠心病					
基本	信息				
一诉	五史				
体格	检查				
专科	检查				
诊	断				
心肌酶谱	心肌酶谱 肌钙蛋白				
BNP	电解质				
冠脉造影    冠脉CT					
超声心动图 心电图					
心肌核	素显像				
手术	治疗				
介入治疗	溶栓治疗				
抗凝治疗	治疗 抗血小板				
扩张血管治疗  营养心肌治疗					
随访					
其他					

糖尿病				
一诉	五史			
体格	检查			
<b>专科</b>				
诊	断			
糖化血红蛋白	血糖			
胰岛素释放试验	C肽释放试验			
肾功能    血气分析				
胰岛细胞抗体	胰岛素抗体			
糖耐量	量试验			
眼底检查				
口服降糖药治疗				
胰岛素治疗				
调脂治疗				
随访				
其他				

已有50类病种 100多典型疾病 迅速扩展中

# 归一规范,真正形成"标准可用数据集"



数据种类	原始词数量(个)	参考标准 (部分)	标准词数量级(个)	数据标签
		北京临床版ICD-10-V6.01		编码 – 临床诊断标准词
诊断词典规范	14,587,447	ICD-10-O-3	3,340,034 4 降77%	形态学编码 – 肿瘤病理诊断
		MESH (医学主题词表)		编码 – 主题词
手术及操作词典规范	853,801	北京临床版ICD-9-V6.01	66,532   降92%	编码 – 手术及操作标准词
影像学检查及其他治 疗性操作词典规范	545,248	北京临床版ICD-9-V6.01	7,733 👢 降99%	编码 – 诊断性和治疗性操作标准 词
		LOINC		成分
	131,623			属性类型
<b>松砂丛木油曲</b> 切世			2 522 ■ 12001/	时间特征
检验检查词典规范			2,532 <b>降</b> 98%	样本类型
				标尺类型
				方法类型
	177,701	CFDA		商品名
<b>#</b> 12 13 14 11 14 14 14 14 14 14 14 14 14 14 14			12,924   降93%	通用名
药品词典规范			12,924   降93%	INN名
		ATC分类		ATC编码
科室词典规范	9,790	国家卫计委 – 医疗机构诊疗科目名录	95 👢 降99%	编码 – 诊疗科目

# AI赋能,真正形成"标准可用数据集"



以肺癌为例: L2/L3

#### 现病史节选

患者于2011年3月13日因"反复干咳、右胸闷、气促、盗汗1月余"入住我院胸外科,入院后行VATS下右上肺楔形切除术。术中无不适,术后病理:肺中分化腺癌胸膜转移。 医GFR基因突变检测阴性。ALK无融合。诊断:右上肺中分化腺癌伴胸膜转移T2N2M1alV期。于2011-6-15、2011-7-6、2011-7-26、2011-8-15分别于我院行培美曲塞650mgd1+顺铂46mg d1-3化疗4个疗程,过程顺利,过程无恶心呕吐。复查CT未见异常。现因进一步诊治收入我科。

手术路径	电视辅助胸腔 镜手术
手术名称	右上肺楔形切 除术
病理组织学分 型	腺癌
组织学分级	中分化
基因突变	EGFR阴性
基因突变	ALK阴性
T分期	T2
N分期	N2
M分期	1a
临床分期	IV期
是否化疗	是
化疗方案	培美曲塞+顺铂
化疗疗程数	4

糖尿病,(Ⅱ型) 糖尿病,(2型) 糖尿病、(II型) 糖尿病॥型 糖尿病II型 2型糖尿病 ||型糖尿病 糖尿病二型 2-型糖尿病 二型糖尿病 2型糖尿病(可能性大) 糖尿病(2型) 2型糖尿病可能性大 2-糖尿病 || 型糖尿病 糖尿病2型 2型糖尿病(非肥胖) (川型)糖尿病 (II型)糖尿病 乙型糖尿病

多份资料聚合

(2型)糖尿病

ICD10编码: E11.901

归一算法

ICD10诊断名称: 2型糖尿病

多家医院或分支机构录入 干差万别的同义关键词

# 病历认知能力: 医学自然语言处理平台





# MKG+预训练+DNN混合病历结构化模型

基于医学文档预训练模型



## 预训练/深度学习



基于医学逻辑的知识图谱

基于Intel OpenVINO BERT 推理优化性能提升50%+

intel



### 医疗文书结构化建模能力



### 以结直肠癌为例

### 体格检查

直肠指诊: 膝胸位, 肛门括约肌稍紧张, 肛门周围粘膜光滑, 未见明显红肿、破溃及外痔皮赘等。入指约6cm, 于11点至6点处可明显触及环形肿物, 占据肠腔2/3周, 质地较硬, 不能推动, 基底广。直肠壶腹部空虚, 直肠壁无明显触痛。出指指套少量染血。肛门镜检查: 膝胸位11点至6点处, 距齿状线约6cm处可见肿物隆起, 表面少量渗血, 观察不清。



#### 病理检查

直肠癌切除标本: 直肠溃疡型中分化腺癌, 肿瘤大小3.5x3x1cm; 癌浸润至直肠周围组织; 可见脉管癌栓及血管壁、神经侵犯; 肠周5个淋巴结可见癌转移 (5/12); 切缘 (临床送检远、近切缘及环周切缘)未见癌; 肿瘤病理分期pT3N2a。免疫组化结果: CDX-2 (+), Ki-67 (70%+), MSH2 (+), MSH6 (+), MLH1 (+), PMS2 (+), β-catenin (膜+)。

### 结构化处理

-1121010		
直肠指检体位		膝胸位
直肠肿瘤位置		11点至6 点
直肠肿瘤占肠腔周径		2/3
直肠肿块活动度	11	不能推动
直肠壁压痛部位		无
肿块下缘距肛距离		6CM
指套有无血染		是

#### 结构化处理

结构化字段能力

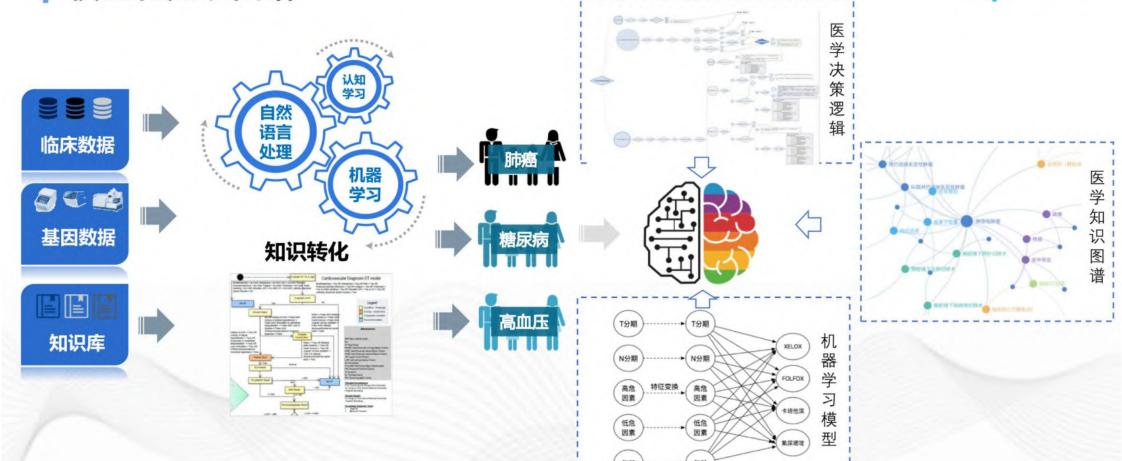
结构化字段能

CDX-2		阳性	
Ki-67	11	阳性	
MSH2		阳性	
MLH1	-	阳性	
PMS2		阳性	
神经侵犯		是	
病理类型-分级		中分化	
送检淋巴结数目		12	
阳性淋巴结数目		5	
组织学分型		腺癌	
脉管癌栓	CHEMA	婦014	
TNM病理分期-N分期	土奴士	N2a	
	/ ST	XIXIII	

www.chima.org.cr

# 模型化知识图谱

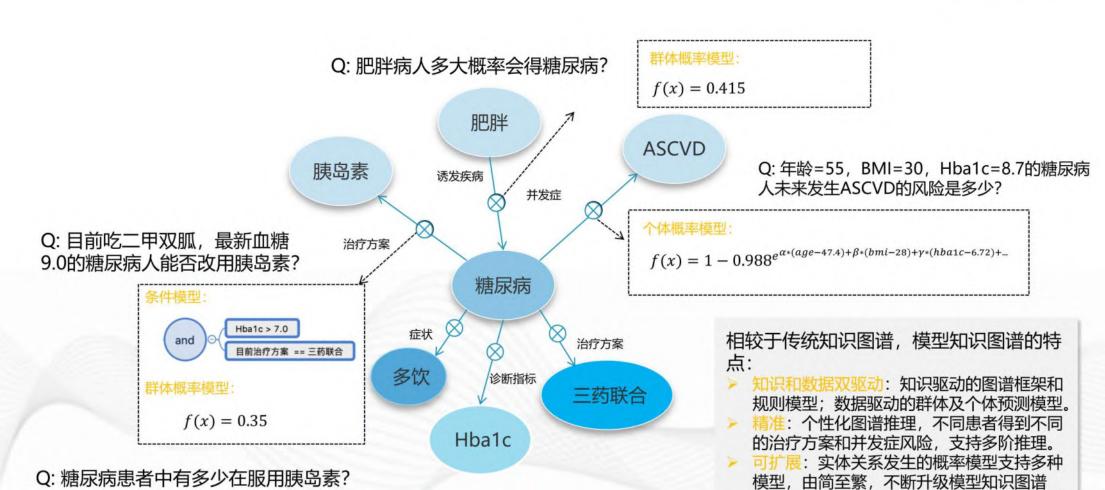




知识引擎

# 示例: 糖尿病模型化知识图谱





未经授权请勿转载

www.chima.org.cr

# 糖尿病模型化知识图谱系统应用



#### 模型化知识图谱一站式解决各类决策支持:

- ▶ 治疗方案推荐: 规则关系 (来自指南)
- > ASCVD风险评估:模型关系 (预测模型)
- 用药知识库:一阶关系(传统知识图谱)



# 病案智能质控





引入大数据技术及智能化工具,辅助临床医生正确选择诊断,将病历质控从终末前置到运行环节,从 源头提高首页质量

#### 医学同义词检索

基于NLP以及结构化归一。 引导医生快速定位诊断

#### 结构化辅助编码

根据搜索要求按照ICD编码轴 心辅助医生高效编码

#### 智能化自动编码

以编码规则库、疾病知识库等为基础、 采用AI技术构建深度学习算法模型、 通读病历完成诊断及诊断编码的推荐

#### 诊断推荐

构建院内诊断知识图谱,根据互联网推荐算法推荐诊断, 提升查询效率

面向 病案

引入智能化质控工具及专病知识库,结合病历对首页内容进行质控,提高工作效率,保证工作质量。

#### 多维度全面深度质控

提供基于临床知识及编码知识库的全面覆盖首页所有信息的完整性、一致性、有效性、规范性数据质量质控

#### 诊断依据快速定位

利用医学同义词及规则库, 自动标示诊断锚点,帮助 编码员快速找到诊断描述 位置

#### 诊断遗漏

根据算法模型推荐出的诊断 及诊断编码质控病案首页中 给出的诊断及诊断编码,解 决漏诊的问题

#### DRGs合理编码推荐

通过相似病例检索,结合 DRGs偏差分析,辅助编码 人员合理编码

#### 简洁方便的统计分析

从科室、编码员、问题等多 个维度为管理人员提供统计 分析

#### 大数据技术赋能

- -自然语言处理(NLP)
- -结构化、医学归-
- -知识图谱(KG)

### 深度临床知识支撑

-专病知识图谱和神经网络动态规则 自生成技术

### 多源异构数据治理

-电子病历数据,并提供经过医学术语 归一的标完整展现准化诊断与手术术 语检索。

www.chima.org.cn

版权© 2021 医渡云保留所有权利

# 多维病案智能质控



### 首页评分质控

- 遵循国家及地方病案首页 质控规范
- 审查首页完整性、规范性 并给出质量评分

### 编码规则质控

- 基于权威ICD编码指南的编码 规则,提示主要诊断和次要诊 断的编码合理性问题,
- 基于DIP付费下的主要诊断和 主要操作的质控校验



• 运用机器模型发现病案首页诊断 遗漏项,解决漏诊质控的问题

### 医学逻辑质控

• 将首页内容与病历内容结合进行 合理性检查, 发现数据矛盾

### 诊断寻证质控

基于机器学习得到的具有临床思维 的诊断寻证模型, 病历精准定位诊 断依据

# 智能化精准定位助力病案质控

### 医渡云 YIDUCLOUD

### 智能缺陷提示

- 着色显示缺陷字段,直观可见
- 浮动提示具体原因,简单明了

### 精准依据定位

- 提供诊断依据位置,一触即达
- 诊断依据标红提示,一目了然



按照病案室阅读病历顺序罗列归档 病历文书

按国家、地方及医院标准自动化对 首页数据填写质量进行评分

直观显示DRGs分组情况,并参考地方基准数据进行偏差分析与编码推荐,即避免"高码低编"也遏制"低码高编"

直接显示诊断依据病历文书锚点, 一触即达诊断依据语句位置,并着 色高亮显示。

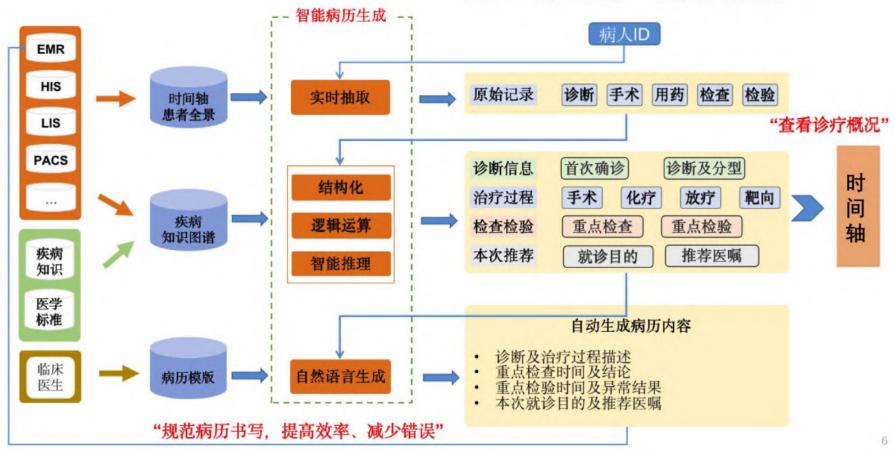
列举填写缺陷、逻辑冲突、编码缺陷等问题列表,点击直达对应首页检查项位置,并高亮着色显示。

未经授权请勿转载

# 智能生成病历

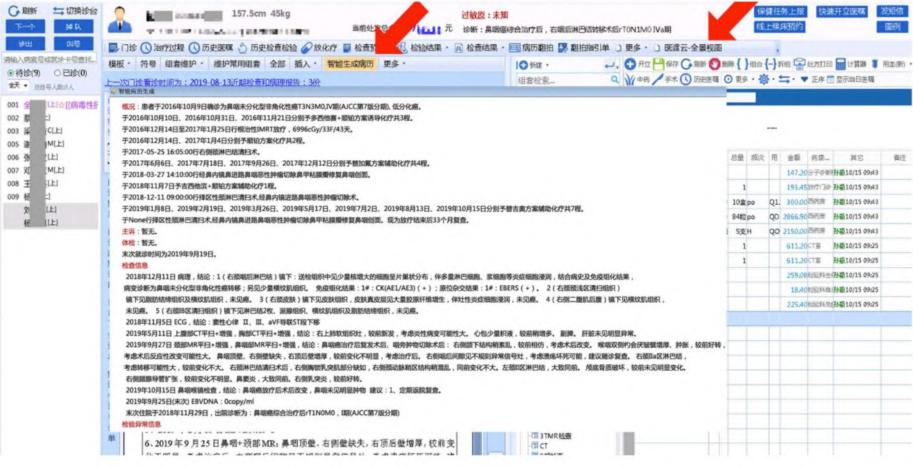


#### "记录 -> 事件 -> 病历, 一键生成, 提高效率"



# 智能生成病历





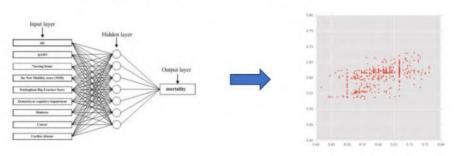
# 单病种CDSS: 基于真实世界研究的机器学习模型

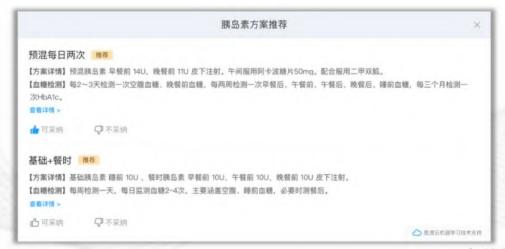




达成项目目标:基于真实世界研究和机器学习方法,学习 专家胰岛素方案及用量,做到患者的<mark>个性化精准治疗</mark>。

### 糖尿病胰岛素用量模型





## 提供面对风险因素/诊断/治疗/预后预测不同场景的预测模型



### 应用场景

#### 风险评估

- 风险因素识别和干预
- 综合风险评估模型构建

#### 早期诊断

- 新型诊断标志物探索
- 现有诊断标志物优化

#### 治疗指导

- · 分子分型与治疗决策
- · 治疗方案规范与优化

#### 预后评价

- 预后标志物与风险因素
- 不同疗法的预后比较



# 预测模型落地案例: 鼻咽癌预后生存期预测



### 鼻咽癌生存期预测应用

版权© 2021 医渡云保留所有权利

- 基于鼻咽癌专病库获取的10127位包含随访信息的患者数据,涵盖64个与鼻咽癌治疗和预后高相关指标。
- 为预测模型生成网页端及移动端应用,预测模型应用 网页随文章一同发表,推广科研成果,同时为医疗工 作者提供一个专业的学习分析平台。
- 研究发现,性别年龄,烟酒史,家族肿瘤史,放疗前肿瘤TN分期,放疗后血红蛋白和白蛋白是影响鼻咽癌预后的重要因素。



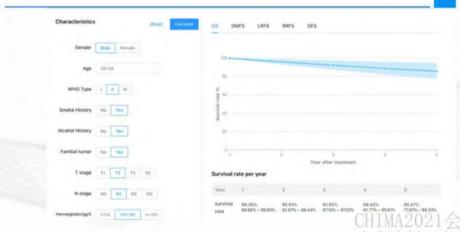
SSAGE

<u>Ther Adv Med Oncol</u>. 2019; 11: 1758835919833863. Published online 2019 Mar 21. doi: 10.1177/1758835919833863 PMCID: PMC6431774 PMID: 30923578

Do all patients with advanced N-stage nasopharyngeal carcinoma benefit from the addition of induction chemotherapy to concurrent chemoradiotherapy?

Ji-Jin Yao, Ya-Nan Jin, Zhi-Gang Liu, Qiao-Dan Liu, Xiao-Feng Pei, Huai-Li Zhou, Wang-Jian Zhang, Fan Zhang, Li Lin, Wayne R. Lawrence, Si-Yang Wang, Jun Ma, Guan-Qun Zhou, and Ying Sun

Author information - Article notes - Copyright and License information <u>Disclaimer</u>



# 预测模型落地案例: 住院患者静脉血栓 (VTE) 评估



### ▶ 目标:

针对院内患者进行自动VTE评估并且提供建议措施;同时,提出针对肿瘤病人的特定评估模型。

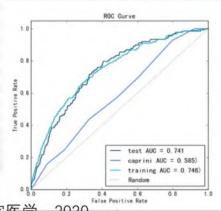
### ▶ 特点:

- ✔ 自动抽取特征进行评估,风险提醒。
- ✓ 标准Caprini评估模型
- ✓ 针对肿瘤病人优化模型

### > 效果

- ▶ 22个特征全自动抽取
- ▶ 耗时2分钟->3~5秒
- ▶ 评估率40%~60%->100%
- ▶ 医师确认率 60%->96%







#### 已于今年9月初实现全院外科科室、内科 科室的VTE风险自动评估功能



《大数据技术下静脉血栓风险智能预警系统设计与建设》,中国数字医学,202

未经授权请勿转载

# 预测模型落地案例:新冠死亡预测模型



### ▶ 目标:

针对患新冠肺炎的患者在住院期间进行风 险评估并且,从而针对性的对高风险患者 给予关注和治疗。

### ▶ 特点:

- ✓ 自动抽取指标并进行评估
- ✓ 评估准确度较高
- ✓ 选取常见指标, 方便易用

Table 2 Top ten features with highest predictive ability

Feature no.	Feature added	P value of coef	AUC on train	AUC on test
1.0	LDH	< 0.001	0.840	0.876
2.0	BUN	< 0.001	0.882	0.877
3.0	Lymphocyte (%)	< 0.001	0.895	0.903
4.0	Age	< 0.001	0.903	0.911
5.0	SPO <sub>2</sub>	< 0.001	0.915	0.917
6.0	Platelets	< 0.001	0.923	0.925
7.0	CRP	< 0.001	0.930	0.921
8.0	IL-10	0.001	0.932	0.930
9.0	HDL-C	0.005	0.934	0.932
10.0	SaO <sub>2</sub>	0.005	0.935	0.931

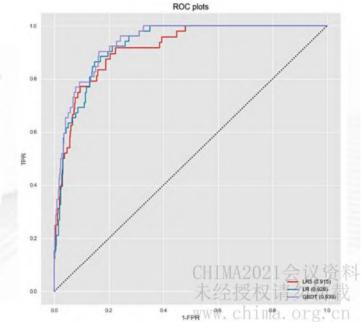
LDH, lactic dehydrogenase; BUN, blood urea nitrogen; CRP, C-reactive protein; HDL-C = high-density lipoprotein cholesterol; AUC, area under curve

#### **Al Algorithm Based Prediction**

 Risk Assessment for Severe Case
It is not necessarily to predict severity since the patient had already been severe or critical. Risk Accessment for Mortality



 Risk Factors Include LYM% abnormally low Sp02 abnormally low LDH abnormally high



Development and external evaluation of predictions models for mortality of COVID-19 patients using machine learning method, 《Neural Computing and Applications》, 2021.1



# 我们的发展历程

脱敏数据 可计算的大数据 可计算的大数据 医疗知识

医疗知识 医疗大脑

医疗大脑 应用落地

健康医疗 生态平台

2014

2015

2016

2017

2018

2019

2021

医渡云成立 自主研发医疗 人工智能平台 开启AI时代的 临床科研新形态 与国内多家知名 医疗机构战略合作

与地方政府合作, 构建区域医疗平台 多地部署医疗 人工智能研究中心 以科技之力赋能 大健康产业升级 致力于构建智慧 医疗全生态

50+类

疾病数据库

















М











10个

疾病标准数据集出版



近100篇

高质量论文

450 +

累计论文影响因子

THE LANCET Oncology

multicentre, cohort study

Development and validation of a gene

metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective,

expression-based signature to predict distant

35.386

medicine 30.641

The practical implementation of artificial intelligence technologies in medicine

u®, Sally L. Barter, Je-to, Jerong to bengton Zhou & Kang Shang ®



守护人民健康的云

本资料为保密信息,仅供CHIMA交流。未经医渡云书面同意,不得向任何第三方披露或提供本资料或其中任何信息。任何人士或机构不得使用或依赖本资料所包含的信息做出任何商业上的决定或作为证据。