

人工智能赋能医疗健康行业

中国信息院医疗健康大数据和网络研究中心
闵栋

厚德實學 興業致遠

2019年8月

中国信通院智慧医疗总体情况



工作基础：《新一代人工智能产业创新重点任务揭榜工作方案》

工业和信息化部办公厅关于印发《新一代人工智能产业创新重点任务揭榜工作方案》的通知

发布时间：2018-11-14

工信厅科〔2018〕80号

各省、自治区、直辖市及计划单列市、新疆生产建设兵团工业和信息化主管部门，有关单位：

为加快推动我国新一代人工智能产业创新发展，落实《促进新一代人工智能产业发展三年行动计划（2018-2020年）》，我部制定了《新一代人工智能产业创新重点任务揭榜工作方案》（以下简称《工作方案》），现印发你们。

请各省、自治区、直辖市及计划单列市、新疆生产建设兵团工业和信息化主管部门、中央企业集团、人工智能相关行业组织（以下统称推荐单位）积极组织企业、高校、科研院所等申报，并做好推荐工作。具体流程如下：

4.医疗影像辅助诊断系统

揭榜任务：包括医学影像数据采集标准化与规范化，脑、肺、眼、骨、心脑血管、乳腺等典型疾病领域的医学影像辅助诊断技术研发，医疗影像辅助诊断系统的产品化及临床辅助应用。

13.行业训练资源库

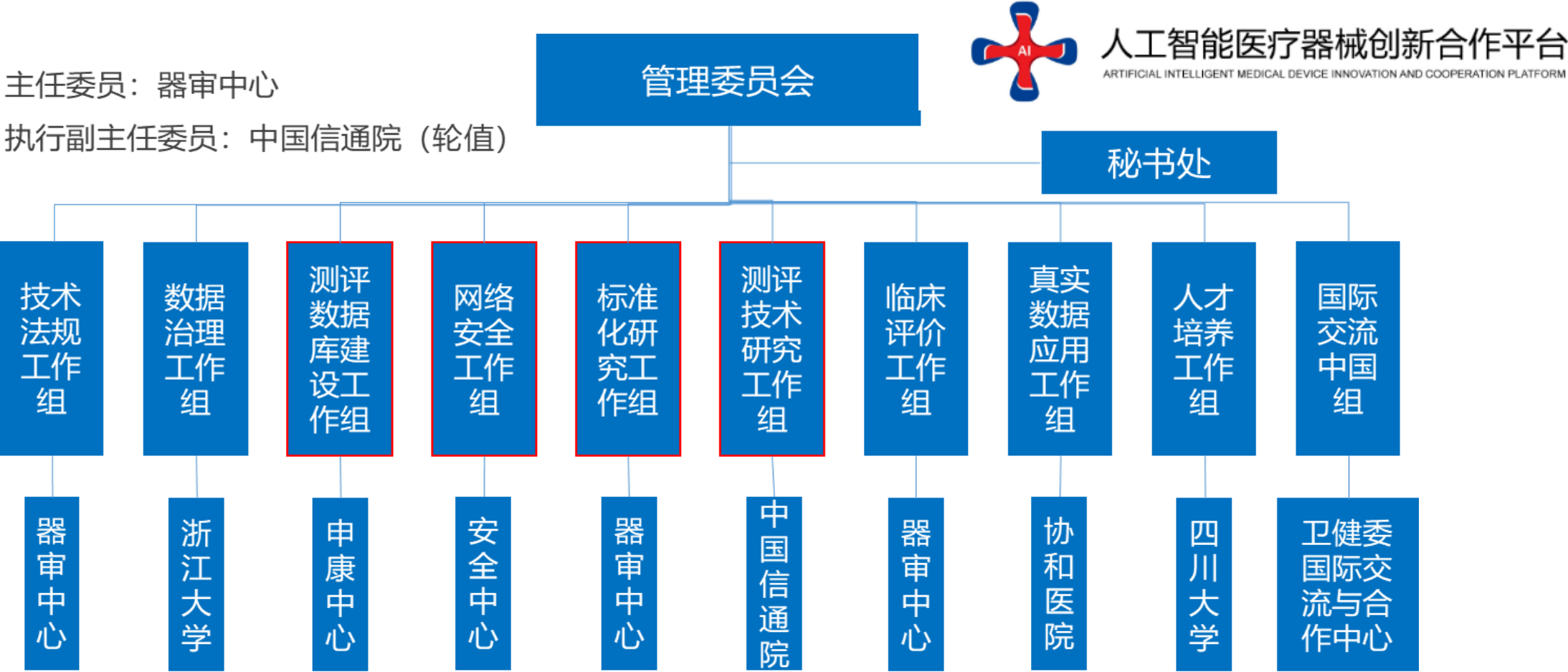
揭榜任务：包括面向语音识别、视觉识别、自然语言处理等重点产品研发，以及工业、医疗、金融、交通等行业应用的高质量人工智能训练资源库、标准测试数据集和云服务平台。

工作基础：发起成立人工智能医疗器械创新合作平台

- 2019年7月17日，国家药监局器审中心、中国信通院、国家卫健委国际交流与合作中心等14家单位联合发起的“人工智能医疗器械创新合作平台”在京正式成立。

主任委员：器审中心

执行副主任委员：中国信通院（轮值）



工作基础：联合医疗机构共建医疗AI临床评价联合实验室



2019年4月1日，国际医学人工智能论坛，发布智能化医疗器械技术评价联合实验室倡议书。

国家卫健委国际合作与交流中心、国家药监局器审中心、中国信通院联合医疗机构发起智能化医疗器械临床评价联合实验室倡议，推动医学人工智能、医疗健康可穿戴设备、5G、医用机器人等的临床评价研究，并探索医疗器械风险评估及临床试验保险机制。

人工智能辅助诊断产品临床评价：联合中华医学会四大主委（放射、超声、放疗、核医学），北京协和医院眼科、皮肤科、同仁医院眼科、301医院普外科（甲状腺）、上海申康医院发展中心等。

5G智慧医疗临床评价：联合中日医院、郑大一附院、北航等，开展5G环境下的远程手术机器人、远程影像、远程病理、远程心电、远程智能监护、VR远程急救等临床评价研究。

健康可穿戴设备临床评价：联合中南大学湘雅医院、郑大一附院等开展可穿戴设备临床评价研究。

工作基础：发起成立ITU与WHO健康医疗人工智能焦点组

支撑国际电信联盟（ITU）和世界卫生组织（WHO）正式成立**健康医疗人工智能焦点组（FG-AI4H）**，构建健康医疗人工智能的国际交流平台，共同推进健康医疗人工智能的国际标准、开源工具、模型测试、效果评估、应用示范、个人隐私保护等工作。**中国信通院为副主席单位。**

焦点组架构

Working Groups:

- Operations (Fraunhofer HHI)
- Health Requirements (WHO)
- **Administration**
- **3 Ad-Hoc Groups (AHGs)**
- Data handling & data acceptance
- Test data set assessment
- Thematic classification scheme

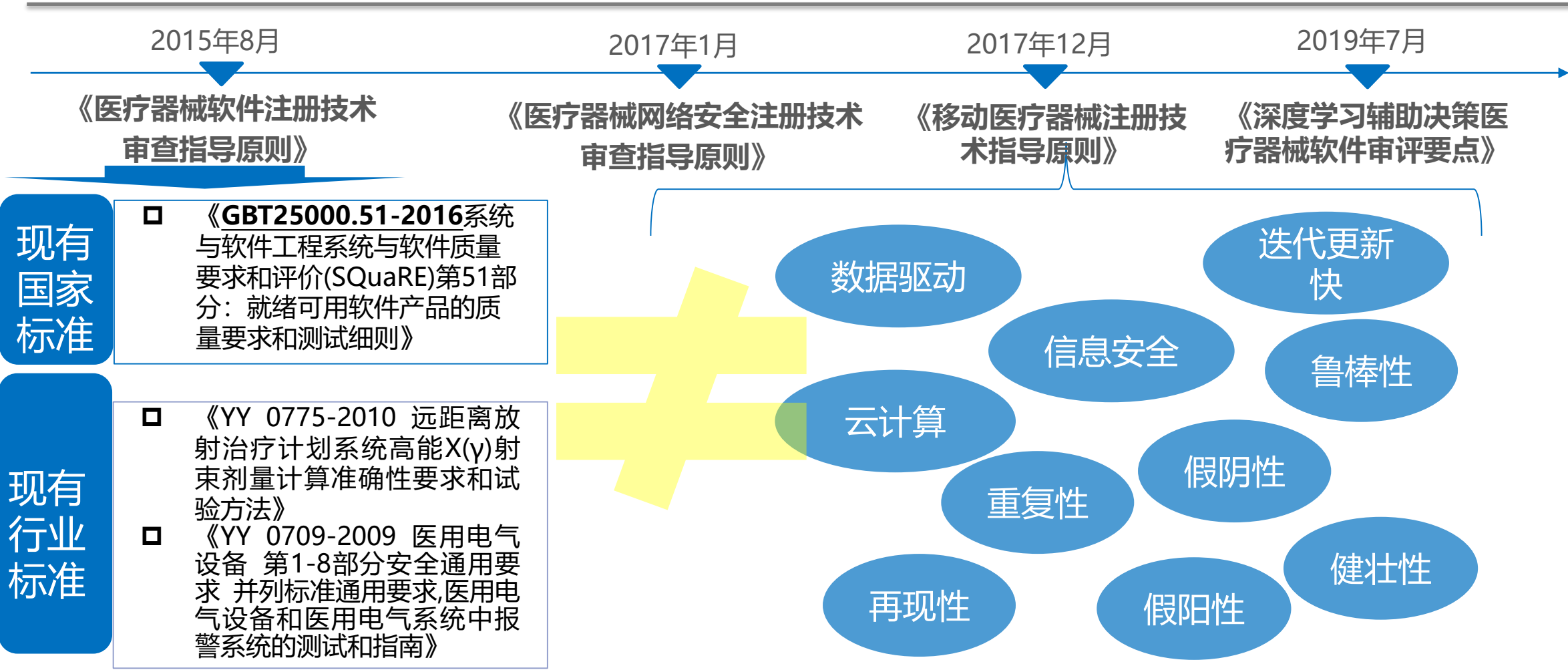
焦点组项目立项

- Breast cancer (乳腺癌)
- Alzheimer disease (阿尔茨海默病)
- Anti-venom selection with snake-identification (蛇毒)
- Autism through analysis of brain imagery (自闭症)
- Evaluating the accuracy of "symptom checker" (症状检查评估)
- Skin lesions (皮肤病变)
- Elderly' s risk of falling (老年人跌倒)
- Diabetic retinopathy (糖网病变)

AI4H监管工作组:

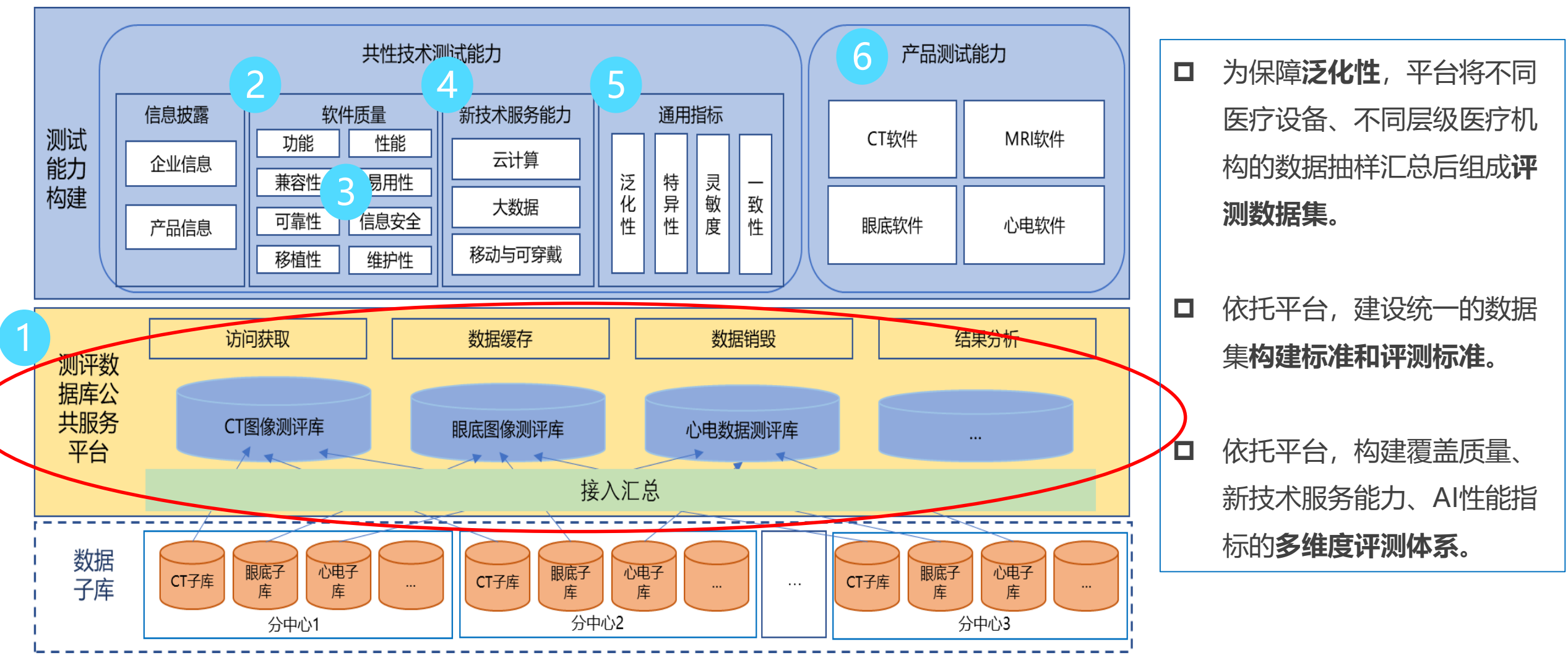
- 中国国家药品监督管理局 (CMDE)
- 美国食品药品监督管理局 (FDA)
- 欧洲其他国家医疗健康监管机构

存在的问题：医学人工智能产品亟需完善评测体系

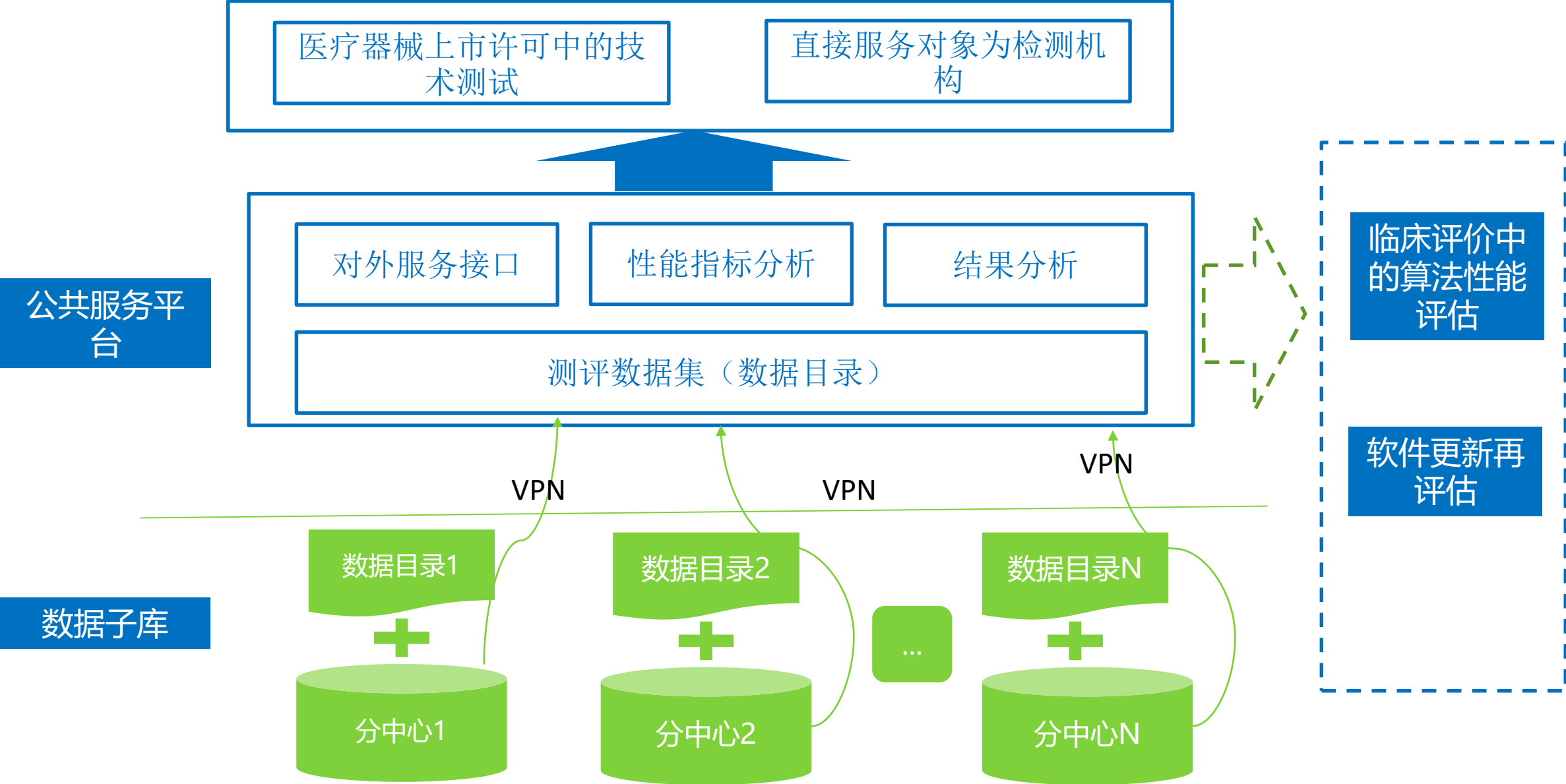


总结：现有的测试标准缺少云计算、大数据等新技术服务能力、AI性能指标等内容，同时信息安全25000中有提及，但没有具体的具体的技术要求和测试细则，现有测试体系远远不能满足新兴医疗器械软件的技术监管要求，亟需完善医疗器械软件评测体系。

工作计划：配合卫健委医管中心构建医学人工智能评测体系



构建测试能力1：配合医管中心建设数据库公共服务平台（1）



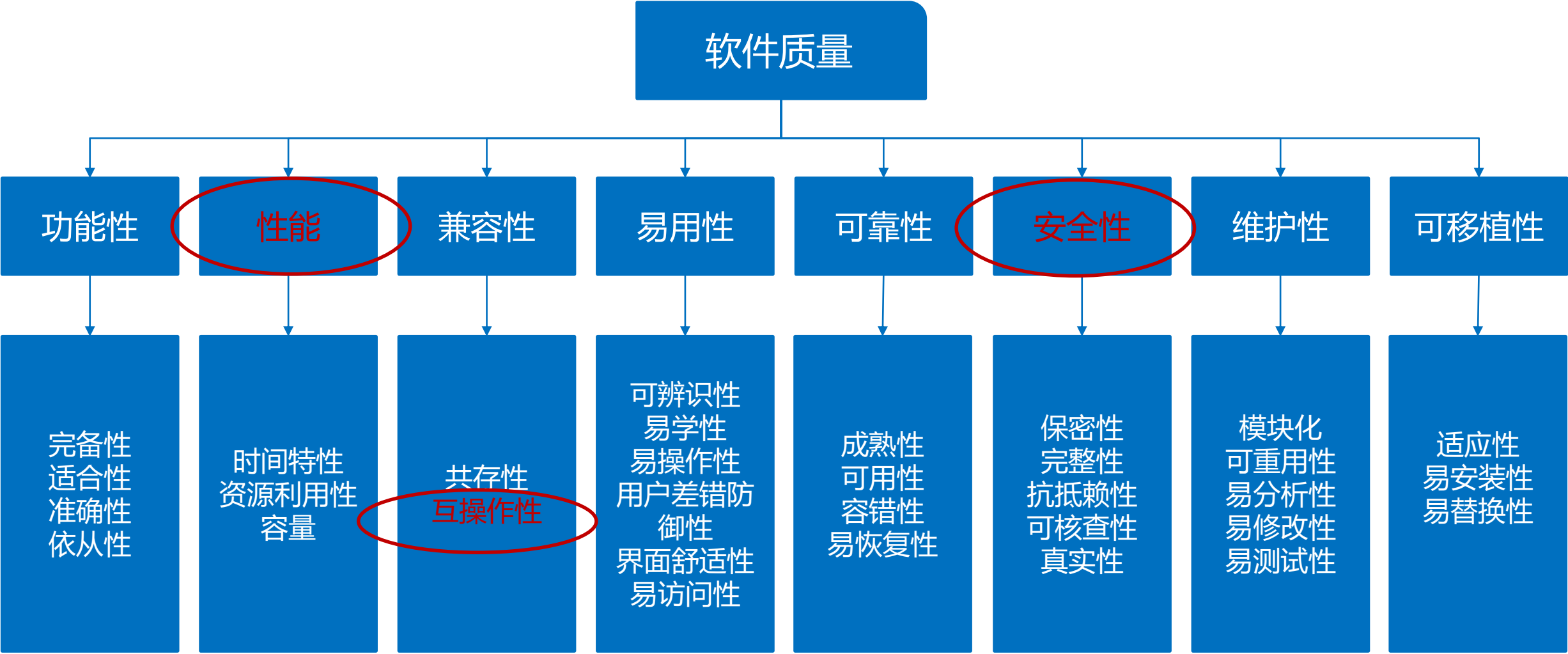
构建测试能力1：配合医管中心建设数据库公共服务平台（2）

配合医管中心推动医学人工智能数据库标准制定：

- 医学人工智能数据质量与标注通用规范
- 医学人工智能数据采集通用规范（对采集设备的性能要求和临床采集方法的要求）
- 医学人工智能数据库构架及产品测试接口通用规范（数据库构架，包括临床流行病学构架、数据库IT构架、产品测试时的接口规范）



构建测试能力2：夯实质量测试能力



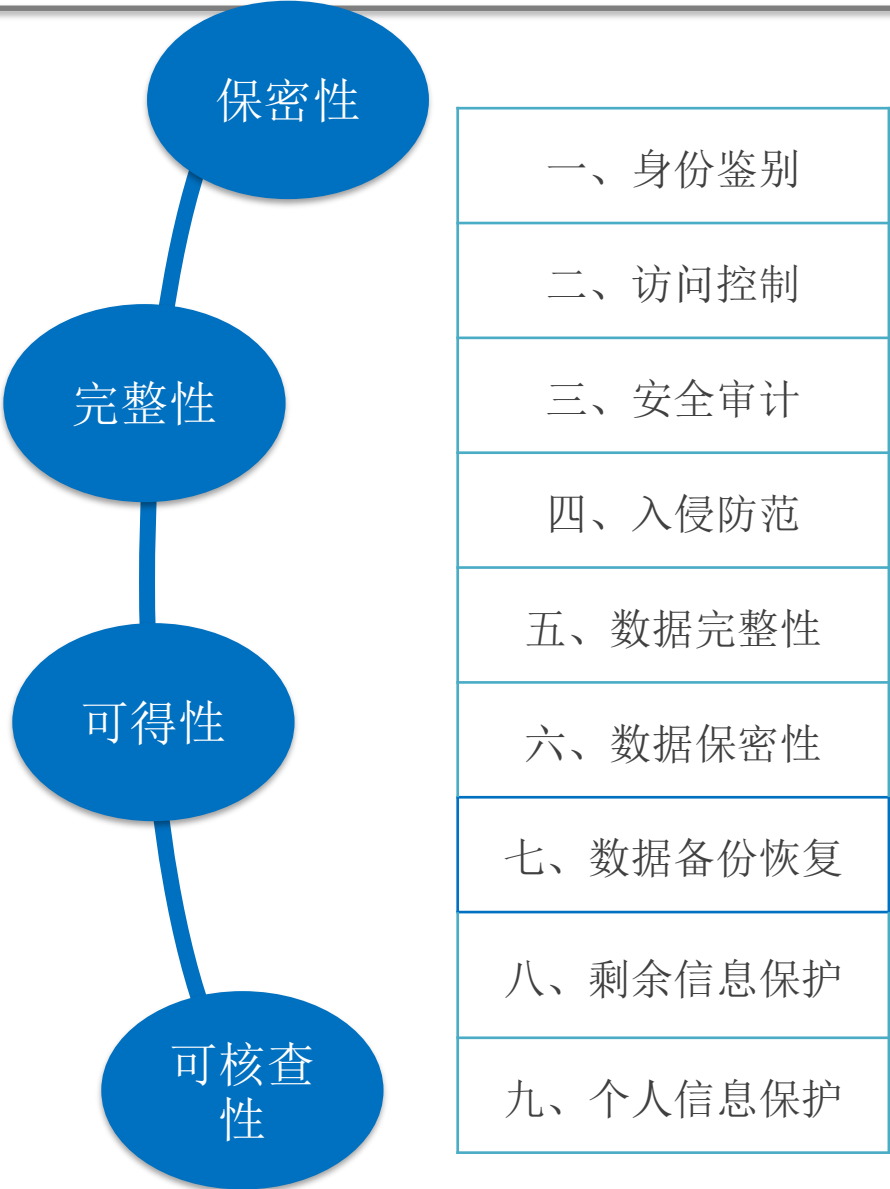
注：红圈部分将搭建环境并使用测试工具重点测试

构建测试能力3：加强软件安全测试能力

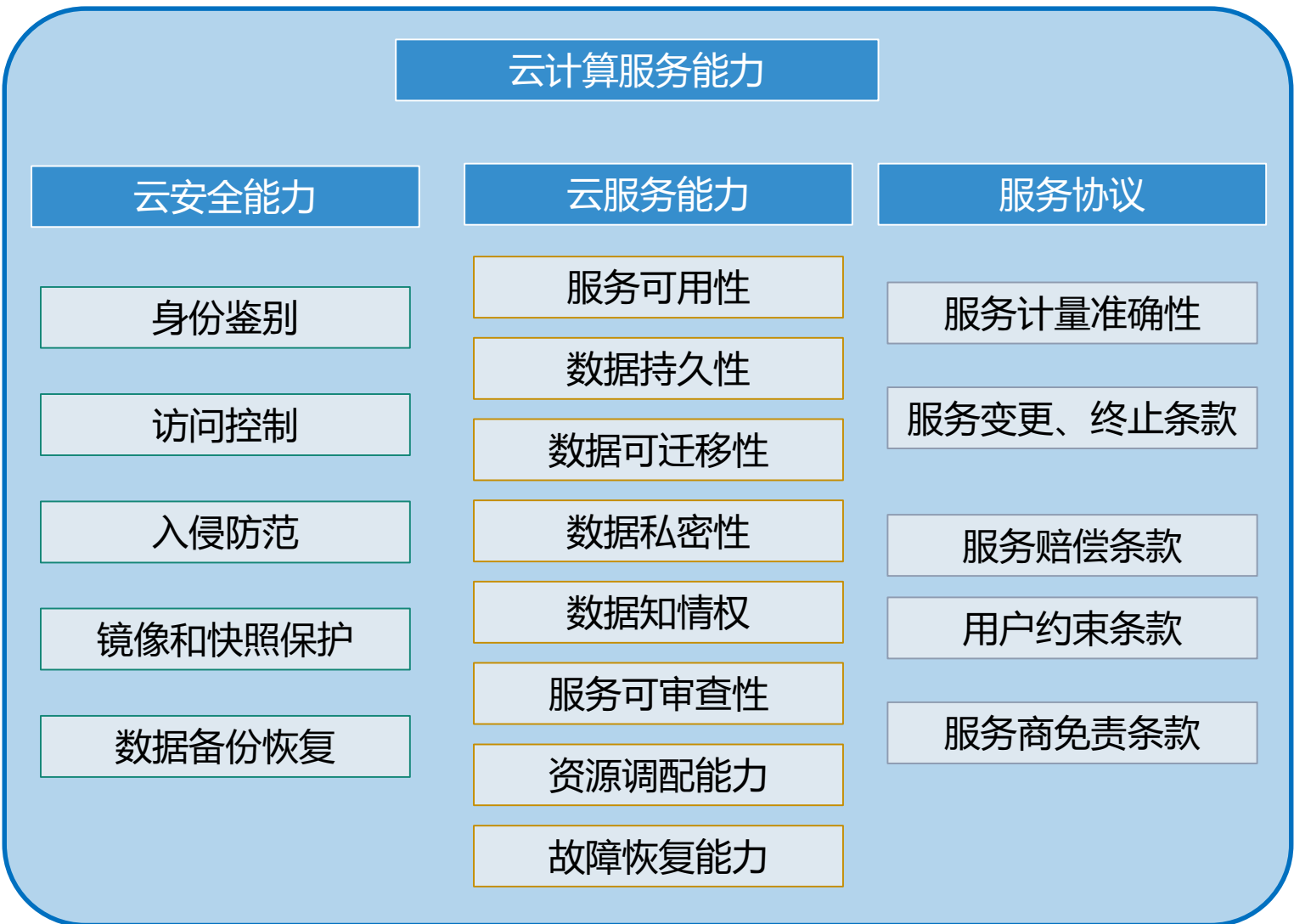
信通院安全
测试能力

- ❑ 国家信息安全等级保护测评机构
- ❑ 全面支撑工信部网安局安全保障工作
- ❑ 承建全国互联网基础资源与安全管理系统
- ❑ 联合承建**大数据协同安全技术**国家工程实验室

测评工具



构建测试能力4：打造云计算等新技术服务测试能力



信通院云计算评测能力

- 云计算标准与测试验证北京市重点实验室
- 开展可信云评测6年有余，累计近700项产品/服务通过评测；
- 已具备医疗云、政务云、金融云、电信云等行业云服务测试能力并开展测评。

2019国际医学人工智能论坛

第三批医疗云可信选型评估证书颁发仪式

影像云

浙江飞图影像科技有限公司
飞图影像云平台
浙江视联智慧医疗科技有限公司
视联网远程医疗影像诊断服务平台
杭州依图医疗技术有限公司
依图医疗智能影像云
上海蓝云网络科技有限公司
西门子 Samplay 云平台
中国电

远程会诊云

腾讯云计算（北京）有限责任公司
移动远程会诊平台
浙江飞图影像科技有限公司
飞图影像远程会诊云平台
浙江视联智慧医疗科技有限公司

云应用

平安国际智慧城市科技股份有限公司
平安智慧医疗云平台
通用环球医疗集团有限公司
互联网健康平台

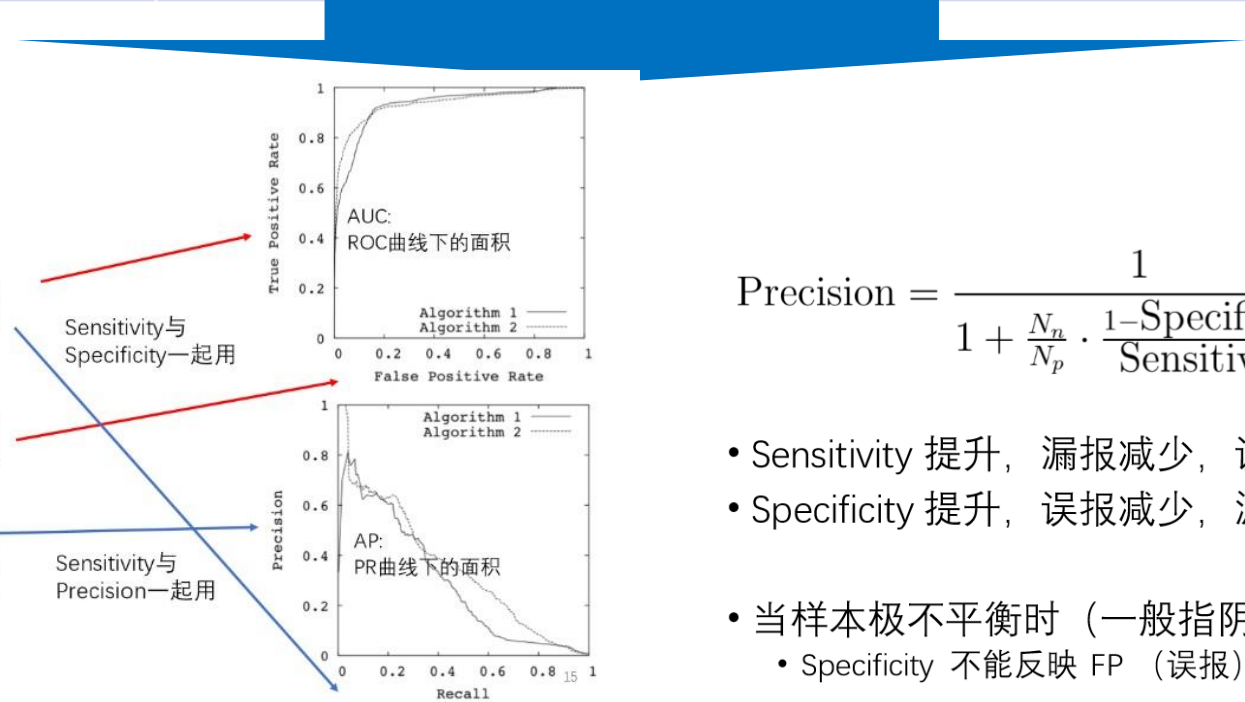
设备管理云

北京泽普科技有限公司
设备管理系统

构建测试能力5：厘清通用性能指标

指标	中文名称	计算公式	含义
True Positive Rate (TPR)	真阳性率	$TP / (TP + FN)$	阳性样本中，预测正确的样本比例
False Positive Rate (FPR)	假阳性率	$FP / (TN + FP)$	阴性样本中，预测错误的样本比例
Sensitivity	灵敏度	TPR	
Specificity	特异度	$1 - FPR$	阴性样本中，预测正确的样本比例
Precision	准确率	$TP / (TP + FP)$	预测为阳性的样本中，预测正确的比例
Recall	召回率	TPR	阳性样本的召回率
F-measure	F度量	$2 * P * R / (P + R)$	Precision和Recall的调和平均

- Sensitivity (Recall)
 - 与模型性能有关、反映漏报
 - 与测试样本的先验分布无关
- Specificity (1 - FPR)
 - 与模型性能有关、反映误报
 - 与测试样本的先验分布无关
- Precision
 - 与模型性能有关、反映误报
 - 与测试样本的先验分布有关
 - 更影响“观感”



$$\text{Precision} = \frac{1}{1 + \frac{N_n}{N_p} \cdot \frac{1 - \text{Specificity}}{\text{Sensitivity}}}$$

N_p : 测试集中的阳性样本数
 N_n : 测试集中的阴性样本数

- Sensitivity 提升，漏报减少，误报增多，Precision下降
- Specificity 提升，误报减少，漏报增多，Precision提升
- 当样本极不平衡时（一般指阴性远多于阳性，即 $N_n \gg N_p$ ）
 - Specificity 不能反映 FP（误报）明显增加，但Precision可以

构建测试能力6：分场景制定产品性能测试方案

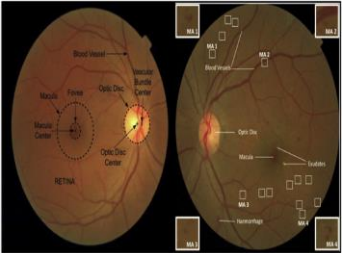
使用场景

产品用途

指标要求

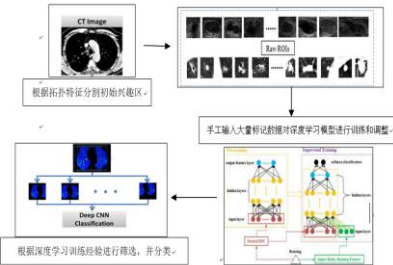
测试方法

眼底糖网筛查



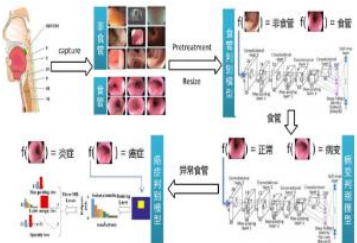
诊断结果：阴性/阳性；NO DR, MILD DR, MODERATE DR, SEVERE DR, PROLIFERATIVE DR
其他指标：诊断时长、诊断流程等服务指标

肺结节影像诊断



诊断结果：良性/恶性；恶性分级
诊断指标：肺结节检出精度
其他指标：诊断时长

食管癌筛查



诊断结果：阴性/阳性；恶性分级
其他指标：诊断时长、检出精度

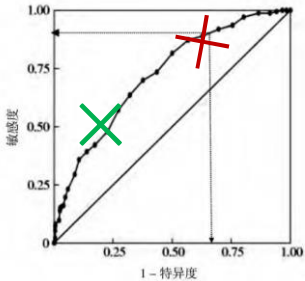
结果分类

二分类：非阴即阳
——特异性、灵敏性

多分类：特异性/灵敏性不再适用
①独立分类：KAPPA系数
②强相关分类：二次加权KAPPA系数

使用场景

疾病筛查：优先尽量保证少的漏诊



辅助诊断：优先保证尽量少的误诊率

① 基于测试集对算法设计结果进行评估，明确各场景及用途下假阴性与假阳性、重复性与再现性、鲁棒性/健壮性等评估要求，以证实算法性能满足算法设计要求。

② 分析算法性能影响因素及其影响程度，如采集设备、采集参数、疾病构成、病变特征等因素影响，以提升算法可解释性，并作为软件验证、软件确认的基础。

构建测试能力7：配合医疗机构开展临床试验方法研究

常见“主要评价指标”

- 1) 正确率 (accuracy), $\text{accuracy} = (\text{TP} + \text{TN}) / (\text{P} + \text{N})$, 被正确判断的样本数除以所有的样本数, 通常来说, 正确率越高, 效果越好;
- 2) 错误率 (error rate), 与正确率相反, $\text{error rate} = (\text{FP} + \text{FN}) / (\text{P} + \text{N})$, 描述被错误判断的比例, 正确与错误是互斥事件, 所以 $\text{accuracy} = 1 - \text{error rate}$;
- 3) 灵敏度 (sensitive), $\text{sensitive} = \text{TP} / \text{P}$, 表示的是所有患者中被判断为患者 (阳性) 的比例, 衡量了产品或方法对患者的识别能力;
- 4) 特效度 (specificity), $\text{specificity} = \text{TN} / \text{N}$, 表示的是所有正常人中被判断为正常人 (阴性) 的比例, 衡量了产品对正常人的识别能力;
- 5) 精度 (precision), 精度是精确性的度量, 表示被判断为患者的示例中实际为患者的比例, $\text{precision} = \text{TP} / (\text{TP} + \text{FP})$;
- 6) 召回率 (recall), 召回率是覆盖面的评估指标, 评估有多个患者被判断为患者, $\text{recall} = \text{TP} / (\text{TP} + \text{FN}) = \text{TP} / \text{P} = \text{sensitive}$, 可以看到召回率与灵敏度是一样的。

其它“评价指标”

- 1) 计算速度: 分类器训练和预测需要的时间;
- 2) 鲁棒性: 处理缺失值和异常值的能力, 所谓“鲁棒性”, 是指控制系统在一定 (结构, 大小) 的参数摄动下, 维持其它某些性能的特性。比如说, 计算机软件在输入错误、磁盘故障、网络过载或有意攻击情况下, 能否不死机、不崩溃, 就是该软件的鲁棒性。
- 3) 可扩展性: 处理大数据集的能力。
- 4) 可解释性: 产品的预测标准的可理解性。

推进国际标准：建设AI4H算法评估框架（数据集+平台）

The ITU-T Focus Group on Artificial Intelligence for Health (AI4H) is to establish a **standardized assessment framework for the evaluation** of AI-based methods for health, diagnosis, triage or treatment decisions.



推进国际标准：推动国内医疗标准数据集成为国际标准

Topic group	References	Leader
CVD risk prediction	FGAI4H-C-017-R1 WatIF Health / IEPH	Benjamin Muthambi
Classifying autism through analysis of brain imagery	B-018 Columbia University	Jongwoo Choi
Dermatology	FGAI4H-C-015 Fraunhofer Portugal, FGAI4H-C-021 Xiangya Hosp. Central S. Univ.	Maria Vasconcelos
Falls among the elderly	FGAI4H-C-014 Fraunhofer AICOS	Inês Sousa
Histopathology	FGAI4H-C-018 Charité Berlin et al	Frederick Klauschen
Neuro-cognitive diseases	FGAI4H-C-020 UNIL, CHUV (Alzheimers)	Marc Lecoultre
Ophthalmology	FGAI4H-C-026-R1, FGAI4H-C-022 Baidu	Arun Shroff
Psychiatry	FGAI4H-C-013 Zurich Univ., Charité Berlin (Paediatric)	Nicholas Langer
Snakebite and snake identification	FGAI4H-C-012-R1 University of Geneva	Rafael Ruiz de Castañeda
Symptom assessment	FGAI4H-C-019 Ada Health, FGAH4H-C-025 Your.MD, FGAI4H-D-022 Baidu	Henry Hoffmann
Tuberculosis	FGAI4H-C-023 ICMR (India), NICF (DoT, India)	Manjula Singh
Radiotherapy	FGAI4H-D-024 Infervision	Chen Kuan
Chest CT	FGAI4H-D-018 BioMind	Wu Zhenzhou

后续工作计划



建平台

配合卫健委医管中心构建**医学人工智能测评库公共服务平台**，在保证“数据不出医院”的前提下进行数据的接入汇总，以用于医学人工智能产品的性能测试。



定方案

根据医学人工智能产品的特点，召集产业界力量，共同探讨，以期达成行业共识、形成适用于**医学人工智能产品的测试方案**，并形成**测试方案的定期迭代更新机制**。



立标准

创新平台-标准化研究工作组
行业共识->行标->国标->国际标准。



出报告

医学人工智能产业发展白皮书；
医学人工智能产业调研报告；
医学人工智能产业动态监测报告。

感谢聆听！

