ICS 35.020
CCS L70

# T/

# 团 体 标 准

T/ XXXX—XXXX

# 医疗图文识别大模型技术要求

Technical requirements for medical image to text large model

（征求意见稿）

**在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。**

XXXX－XX－XX 发布 XXXX－XX－XX 实施

中国互联网协会 发 布

# 目 次

# 前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由××××提出。

本文件由××××归口。

本文件起草单位：

本文件主要起草人：

# 医疗图文识别大模型技术要求

## 1 范围

本文件规定了医疗图文识别大模型涉及的技术能力要求。

本文件适用于医疗图文识别大模型提供方评估自身技术与应用能力，管理方对医疗图文识别大模型服务提供方的服务能力进行要求，第三方评估医疗图文识别大模型提供方的能力

## 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 14396-2016 疾病分类与代码

YD/T 6488-2025 光学字符识别（OCR）服务技术要求和评估方法

IEEE P3394 Standard for Large Language Model Agent Interface

## 3 术语和定义

下列术语和定义适用于本文件。

3.1

**大模型 large model**

指可学习参数量大的深度学习模型，本文件中特指至少具备10亿参数（1B）的深度学习模型。

3.2

**图文识别大模型**

指可完成图像文本转化与理解的深度学习模型，包括OCR模型与大语言模型结合的模型、端到端完成图文转化与理解的模型。

3.3

**令牌 token**

指语言模型的数据最小单位，单个令牌可表示为一个单词、字符或符号，通过分词器进行标注。

3.4

**思维链 chain of thought**

指通过分步骤逻辑推理提升大语言模型解决复杂问题的能力。

## 4 缩略语

下列缩略语适用于本文件。

CPU：中央处理器（Central Processing Unit）

FN：假阴性（False Negative）

FP：假阳性（False Positive）

GPU：图形处理器（Graph Processing Unit）

ICD-10：国际疾病分类第十版（International Classification of Disease, 10th Revision）

IoU：交并比（Intersection over Union）

JSON：JavaScript对象表示法（JavaScript Object Notation）

ROC：接受者操作特性曲线（Receiver Operating Characteristic）

SNOMED CT：医学系统命名法——临床术语（Systematized Nomenclature of Medicine – Clinical Terms）

TN：真阴性（True Negative）

TP：真阳性（True Positive）

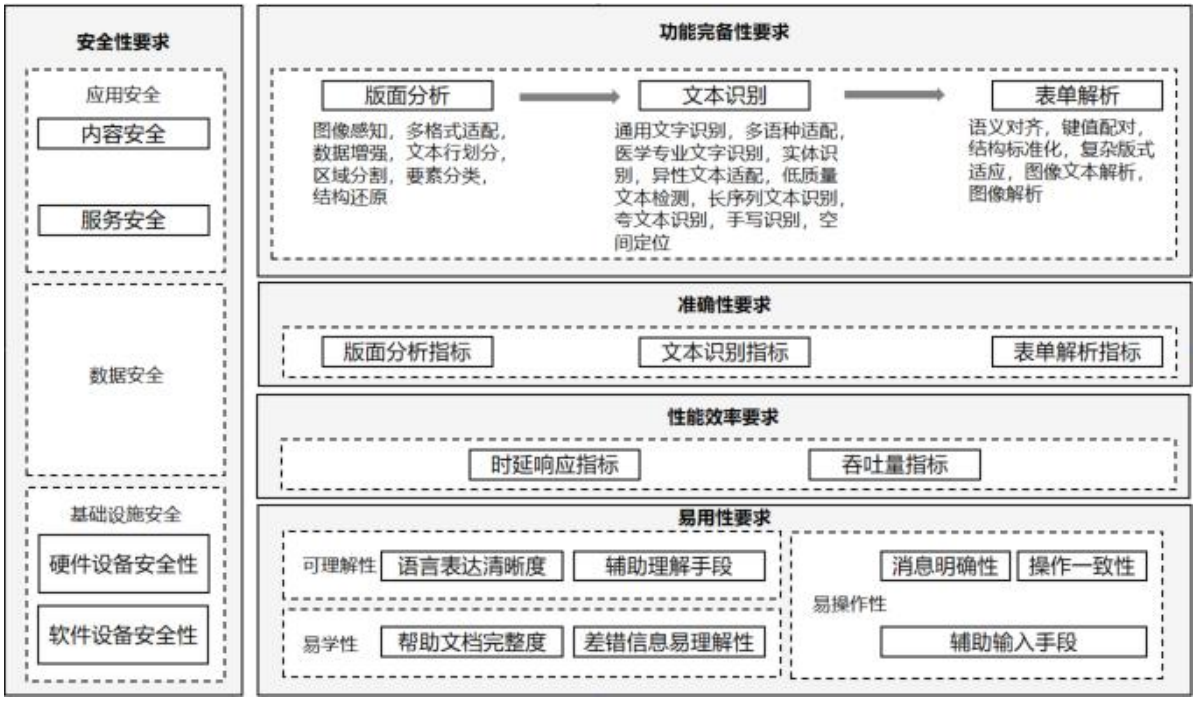
XML：可扩展标记语言（Extensible Markup Language）

## 5　总体要求

医疗图文识别大模型应支持版面分析、文本识别、表单解析功能以覆盖医疗场景图文识别全流程需求。通过规范大模型的功能，明确医疗图文识别大模型的应用范围，提升医学服务的效率及质量。

医疗图文识别大模型应在以下方面满足要求：功能完备性要求、准确性要求、性能效率要求、易用性要求、安全性要求。

图 1　医疗图文识别大模型技术要求架构图




## 6　通用功能需求

### 6.1　版面分析

医疗图文识别大模型应支持对医学文档的版面分析，捕捉医学文档图像基础特征并进行适配，解决稠密文字区分、文档要素识别、复杂版面理解等问题：

a）　图像感知能力：应支持对医学文档图像核心视觉特征提取，包括文字特征、版式结构特征、背景特征、色彩特征、亮度特征等；

b）　多格式适配能力：应支持对医学文档图像主流输入格式适配；

c）　数据增强能力：应支持对不同采集场景的医学文档图像进行分析并进行数据增强，基于图像采集场景特征优化图像精度；

d）　文本行划分能力：应支持对医学文档图像中文本区域划分为独立文本行实例，能区分稠密排列的相邻文本行、重叠度低的遮挡文本行，宜支持对重叠度高的相邻文本行生成划分并生成告警；

e）　区域分割能力：应支持对医学文档图像中不同文本区域（如标题区、正文区、表格区、备注区等）图像分割，区分功能不同的文本区域；

f）　要素分类能力：应支持对医学文档图像分割后区域进行分类，区分标题、正文、表格、备注、印章等要素；

g）　结构还原能力：宜支持对分类后医学文档图像要素进行排列，生成符合人类阅读习惯的要素顺序以降低图文识别后文本理解难度。

### 6.2　图文识别

医疗图文识别大模型应支持将图像化文字转化为字符化文字，解决医学文档图像中专业术语/符号识别、小字识别、长文本识别、跨文本识别等问题：

a) 通用文字识别能力：应支持对医学文档中通用汉字、数字、字母、标点的高精度识别；
b) 多语言适配能力：应支持对多语言医疗文档中文字的高精度识别；
c) 医学专业文字识别能力：应支持对医疗领域专业术语、特殊符号（如病理报告等级符号、检验结果符号等）、专用编码（如 ICD-10、SNOMED CT 等）的高精度识别，识别结果应符合医疗行业术语规范；
d) 实体识别能力：应支持对识别到的医疗领域专业术语基于编码规范（如 ICD-10、SNOMED CT 等）完成语义映射；
e) 异形文本适配能力：应支持对医学文档图像中倾斜、弯曲、纵向排版、不规则排版文本实例的轮廓识别能力；
f) 低质量文本检测能力：应支持对医学文档图像中遮挡文本、稠密文本、模糊文本的检测能力，对低质量、不完整文本目标进行识别、补全与告警；
g) 长序列文本识别能力：应支持对医学文档中的长序列文本实现连续识别，避免长序列导致的识别错误；
h) 跨文本识别能力：应支持对多医学文档（如全病案）处理中信息不一致情况进行提示与修改；
i) 自检纠错能力：应支持对医学文档识别结果进行自检，基于当前文档上下文信息、患者全病案信息、医学常识、原始医学文档图像纠正错误识别信息；
j) 手写识别能力：应支持对医学文档中手写体的高精度识别，包括病历医嘱、处方签名、手写备注等；
k) 空间定位能力：应支持识别结果的溯源，字符、文本行应与原始医学文档图像空间坐标对应。

### 6.3 表单解析

医疗图文识别大模型应支持将非结构化图表信息转化为结构化信息，解决医学文档图像中不规则表格排版识别复杂的问题：

a) 语义对齐能力：应支持基于原始医学文档图表的空间坐标与图文识别文本的语义信息进行对齐，判断语义关联度；
b) 键值配对能力：应支持对医学文档表格文本部分进行解析，将文本转化为键值配对结构；
c) 结构标准化能力：应支持将提取后键值配对、嵌套结构重构为标准结构化格式（如 JSON、XML 等）；
d) 复杂版式适配能力：应支持对复杂图表版式、非标准化图表版式进行解析；
e) 图像文本解析能力：应支持对图像包含的文本信息进行解析；
f) 图像解析能力：宜支持对医疗图像（如心电图、脑电图、折线图）进行分析并生成描述性文字。

## 7 准确性要求

### 7.1 版面分析指标

版面分析指标评估识别出的文本框实例是否为真实文本框：

a) 召回率：衡量真实文本框中，模型成功正确检测的比例，计算公式如下：

$$Recall = \frac{TP}{TP+FN}$$

式中：

Recall——召回率；

TP——真阳性；

FN——假阴性。

b) 精确率：衡量模型检测出的文本框中，为真实文本框的比例，计算公式如下：

$$Precision = \frac{TP}{TP+FP}$$

式中：

Precision——精确率；

TP——真阳性；
FP——假阳性。

c） 交并比（IoU）：模型预测的文本框区域先估算与真实文本框区域像素的交集相对于并集的比例，计算公式如下：

$$IoU = \frac{|X \cap Y|}{|X \cup Y|}$$

式中：
X——模型预测为文本框的区域像素集合；
Y——真实文本框像素集合。

d） Dice 系数：模型预测的分割区域与真实文本框区域之间的重叠程度，计算公式如下：

$$Dice = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

式中：
X——模型预测为文本框的区域像素集合；
Y——真实文本框像素集合。

## 7.2 文本识别指标

文本识别指标衡量识别出的文本与真实文本的一致程度：

a） 莱文斯坦距离（Levenshtein Distance）：模型识别的文本序列转换为人工标注序列所需的最小操作序列，一般采用动态规划计算，计算公式如下：

$$Lev(i,j) = \begin{cases} i \mid j \; if \; i = 0 \mid j = 0 \\ lev(i-1, j-1) \; if \; si = tj \\ 1 + \min\left(lev(i-1,j), lev(i,j-1), lev(i-1,j-1)\right) \; o.w. \end{cases}$$

式中：
s——字符串s；
t——字符串t；
si——字符串s中第i个字符；
tj——字符串t中第j个字符；

b） 字符准确率：识别正确的总字符数占总标注字符数的比例，计算公式如下

$$Acc = \frac{len(total) - Lev\ (i,\ j)}{len\ (annotated\_total)}$$

式中：
Lev（i，j）——莱文斯坦距离；
Len（total）——总字符数；
Len（annotated_total）——总标注字符数。

c） 单词准确率：将字符按医学分词器（tokenizer）分词后，识别正确的单词数占总标注单词数的比例，计算公式如下：

$$Acc_{tok} = \frac{N_{correct\_tok}}{N_{tok}}$$

式中：
$Acc_{tok}$——单词准确率；
$N_{correct\_tok}$——分词器分词后识别文本单词与标注文本单词一致的数量；
$N_{tok}$——标注文本分词器分词后单词总数。

d） 序列准确率：衡量整行无错识别能力，适用于票据、证件号等低错误要求场景，计算公式如下：

$$Acc_S = \frac{N_{correct\_S}}{N_S}$$

式中：
$Acc_S$——序列准确率；
$N_{correct\_S}$——完全正确识别的文本行数；

$N_S$——总文本行数。

### 7.3 表单解析指标

表单解析指标衡量模型能否正确回答关于表单内容的问题：

a) 回答准确率：针对表格内容进行提问（事实性提问与推理性提问），模型回答正确的比例，计算公式如下：

$$Acc_{tab\_crit} = \frac{N_{correct\_tab\_crit}}{N_{tab\_crit}}$$

式中：

$Acc_{tab\_crit}$——回答准确率；

$N_{correct\_tab\_crit}$——模型正确回答数量；

$N_{tab\_crit}$——模型总回答数量。

b) Rouge-L：针对生成类任务，衡量模型回答答案与正确答案序列匹配程度，计算公式如下：

$$ROUGE-N = \frac{\sum S\in\{ReferenceSummaries\}\sum gram_n\in SCount_{match}(gram_n)}{\sum S\in\{ReferenceSummaries\}\sum gram_n\in SCount(gram_n)}$$

式中：

N——即n-gram，文本内容滑动窗口字节数，参考值为2；

$Count_{match}(gram_n)$——参考摘要和生成摘要中共有的n-gram的数量；

$Count(gram_n)$——参考摘要中n-gram的数量。

c) BertScore：模型生成答案与正确答案语义匹配程度，计算公式如下：

$$sim(x_i, y_i) = \frac{Emb(x_i)\cdot Emb(y_i)}{\|Emb(x_i)\|\|Emb(y_i)\|}$$

$$Precision = \frac{1}{|x|}\sum_{x_i}\max_{y_i\in y} sim(x_i, y_i)$$

$$Recall = \frac{1}{|y|}\sum_{x_i}\max_{x_i\in x} sim(x_i, y_i)$$

$$BERTScore = \frac{2\times Precision\times Recall}{Precision+Recall}$$

式中：

Emb(xi)——句子x中词语在经过编码器后的嵌入向量；

Emb(yi)——句子y中词语在经过编码器后的嵌入向量；

sim(xi, yi)——两词语嵌入向量的余弦相似度；

Precision——精确率；

Recall——召回率。

## 8 性能效率要求

### 8.1 时延响应指标

时延响应指标衡量模型推理响应速度：

a) 首令牌输出时延：模型收到标准化输入请求后，流式输出第一个令牌的间隔时长，如具备推理能力，则计算输出第一个非思维链令牌的间隔时长；

b) 推理平均时延：具备推理能力的模型完成单次推理的平均耗时；

c) 99 分位推理时延：统计周期内位于 99%的首令牌输出时延，通过排除极端长尾时延，衡量模型时延稳定性。

### 8.2 吞吐量指标

吞吐量指标衡量大模型流式输出吞吐能力：

a) 每秒输出量：模型在稳定推理状态下每秒输出的令牌总数量；

b) 单日最大输出量：模型在运行状态下单日可处理与输出的令牌总量，衡量模型长期运行稳定性。

## 9 易用性要求

### 9.1 可理解性

#### 9.1.1 语言表达清晰程度

医疗图文识别大模型应用界面文字、提示及交互内容应简洁准确并以标准医学术语表示，避免口语化内容导致的歧义。

#### 9.1.2 辅助理解手段

医疗图文识别大模型涉及召回信息时应显式引用相关来源辅助医师理解。

### 9.2 易学性

#### 9.2.1 帮助文档完整性

医疗图文识别大模型应配备结构化帮助文档，含功能说明、操作指南及常见问题解答，支持关键词检索，内容随平台更新同步修订。

#### 9.2.2 差错信息易理解性

医疗图文识别大模型系统异常时，差错信息应明确原因并提供解决方案，不应以技术代码表述。

### 9.3 易操作性

#### 9.3.1 操作一致性

医疗图文识别大模型各功能模块操作逻辑、交互样式应保持统一，降低用户学习成本。

#### 9.3.2 消息明确性

医疗图文识别大模型进行图表理解与回答输出时，输出文本应简洁明了，不应使用冗长复杂的表述。

## 10 安全性要求

### 10.1 基础设备安全性

#### 10.1.1 硬件设备安全性

医疗图文识别大模型涉及的硬件设备（如网络设备、存储设备、计算设备等）的安全防护能力应包含：

a) 通用安全要求：
   1) 应满足物理安全保障要求，包含防火、防雷、防水、灾备等；
   2) 应满足功能安全保障要求，包含设备标签、硬件接口安全、固件安全、驱动程序安全等；
   3) 应满足管理安全保障要求，包含管理机制、管理人员、授权等；
b) 计算设备安全专用要求：
   1) 应具备保障人工智能加速芯片应具备通用安全保障能力，包含 AI 加速芯片信息窃取防护、架构安全漏洞防护等；
   2) 应具备保障人工智能加速芯片在异构场景下稳定运行的能力，包含 CPU 与 GPU 相结合的场景；
   3) 应具备保障人工智能加速芯片运行环境安全的能力。

#### 10.1.2 软件安全性

医疗图文识别大模型应支持多种设施如依赖库、AI框架、向量数据库、中间件、接口等具备安全防护能力，包含：

a) 漏洞管理：软件设施应定期进行漏洞扫描和修复，具备完善的漏洞响应机制；
b) 安全更新：软件设施应及时更新安全补丁，以防止新出现的安全威胁。

### 10.2 数据安全

除应符合 GB/T 35273-2020 外，医疗图文识别大模型还应满足以下数据安全要求：

a） 生物特征数据保护：

——大模型与系统采集的可用于识别特定自然人身份的生物特征数据，应参照 GB/T 35273-2020 中个人生物识别信息的保护要求进行同等防护；

——原始数据应在终端侧完成特征提取，应上传脱敏后的特征。

b） 数据留存与注销隔离：

——患者数据的存储期限应遵循国家关于病历与健康档案保存期限的相关规定，平台应在存储期满后对数据进行不可逆匿名化处理；

——患者注销账户后，因法律法规要求需留存的数据，应在完成去标识化处理后与日常运营数据库进行物理隔离，仅可用于法定的公共卫生统计或科研目的。

### 10.3 应用安全

#### 10.3.1 内容安全

医疗图文识别大模型输出内容（含生成内容、决策内容）应符合全人类普适的道德伦理及医学伦理要求。

a） 应支持尊重人权，包括医疗图文识别大模型输出内容（含生成内容、决策内容）应遵循人权的普遍性和不可侵犯性的原则，尊重人类平等、尊严和自由的权利；

b） 应支持无偏见歧视性，包括医疗图文识别大模型输出内容（含生成内容、决策内容）避免产生偏见及歧视性结果的程度；

c） 应符合科技伦理原则，包括增进人类福祉、坚持公平公正、推动透明可释、确保可控可信等；

d） 应遵循科技伦理指标，包括公平性、透明可释性、数据隐私、可控可靠性、内容向善、责任可追溯、可持续性等。

#### 10.3.2 服务安全

医疗图文识别大模型应支持服务安全可信、内容安全可信等应用相关内容具备安全防护能力，包含：

服务安全：医疗图文识别大模型涉及的模型安全性应满足模型安全保障要求，包含MTTF、服务安全性、服务合规性、反馈处置机制等。