同济大学大学生创新训练项目计划申请书

|  |  |  |  |
| --- | --- | --- | --- |
| 项目编号 |  | | |
| 项目名称 | 基于大模型的前列腺癌症肿瘤分期预测 | | |
| 项目负责人 | 朱俊泽 | 联系电话 | 18007721992 |
| 所在学院 | 计算机科学与技术学院 | | |
| 学号 | 2351114 | 专业班级 | 大数据1班 |
| 指导教师 | 倪张凯 | | |
| E-mail | 2715978522@qq.com | | |
| 申请日期 | 2024年 12月13日 | | |
| 项目期限 | 一年期 | | |

同济大学 创新创业学院

**填写说明**

1. 本申请书所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要。
2. 申请人可以是个人，也可为创新团队，首页只填负责人。“项目编号”一栏不填。
3. 本申请书为大16开本（A4），左侧装订成册。可网上下载、自行复印或加页，但格式、内容、大小均须与原件一致。
4. 负责人所在学院认真审核, 经初评和答辩，签署意见后，将申请书（一式两份）报送同济大学项目管理办公室。

#### 一、 基本情况

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 项目名称 | 基于大模型的前列腺癌症肿瘤分期预测 | | | | | | |
| 所属学科 | 学科一级门： 工学   学科二级类： 计算机类 | | | | | | |
| 项目来源 | C:\Users\asus\AppData\Local\Temp\ksohtml14632\wps1.wmfA、学生自主选题，来源于自己对课题的长期积累与兴趣  C:\Users\asus\AppData\Local\Temp\ksohtml14632\wps2.wmfB、导师推荐课题  C:\Users\asus\AppData\Local\Temp\ksohtml14632\wps3.wmfC、SITP直升课题 （原课题名称：） | | | | | | |
| 项目申请级别 | sitp | | | | | | |
| 申请金额 | 3000 | | 项目期限 | | 一年制 | | |
| 负责人 | 朱俊泽 | 性别 | 男 | 民族 | 汉族 | 出生年月 | 2005年 7月 |
| 学号 | 2351114 | 联系电话 | 宅： 手机：18007721992 | | | | |
| 指导教师 | 倪张凯 | 联系电话 | 宅： 手机：13646019080 | | | | |
| 项目简介 | | 为了解决医学影像数据质量参差不齐和标注过于稀疏导致的问题，本项目计划从视频数据的输入开始进行标签细化来为训练提供更精确指导信息，结合超声影像大模型的丰富基础知识，加入分割信息。在预处理过后的数据后，为了充分挖掘潜在的模式和特征的信息，标签输入到图像-文本预训练模型中，进行多模态学习。之后结合医疗基础大模型提出对诊断的辅助建议。帮助前列腺癌症分期诊断。 | | | | | |
| 负责人曾经参与科研的情况 | | 负责人曾多次参与sitp创新项目，目前在指导教师的课题组参与实习 | | | | | |
| 指导教师承担科研课题情况 | | 主持国家自然科学基金、青年基金；浦江人才计划项目和上海市领军人才（海外）青年人才项目 | | | | | |
| 指导教师对本项目的支持情况 | | 指导教师每周和项目组成员开会，定期单独和项目负责人交流进度 | | | | | |
| 项目组主要成员 | 姓名 | 学号 | 专业班级 | 所在学院 | 项目中的分工 | | |
| 朱俊泽 | 2351114 | 大数据1班 | 计算机科学与技术学院 | 项目负责人，核心算法设计，模型设计 | | |
| 王雪宸 | 2353597 | 计算机拔尖1班 | 计算机科学与技术学院 | 网页前端开发，后端链接设计 | | |
| 高晗博 | 2453585 | 计算机拔尖1班 | 计算机科学与技术学院 | 网页前端开发 | | |
| 季节 | 2353581 | 信息安全1班 | 计算机科学与技术学院 | 模型安全设计 | | |

#### 二、 立项依据（可加页）

|  |
| --- |
| **（1）研究目的**  前列腺癌是全球男性中第二常见的癌症类型，也是导致癌症相关死亡的主要原因之一。而检测前列腺癌症使用ct则成本较高，使用超声影像则过于依赖专业医生的评估内容，本项目研究目的就是通过深度学习的方法，结合大模型的丰富基础知识实现前列腺癌症分期诊断。  **（2）研究内容**  本项目主要以图像-文本预训练模型为核心结构，添加前处理和后处理模块。分别应用了相关医学大模型（分割、诊断对话），来解决训练数据标签过粗的问题和减小阳性粗标签中无关部分对模型的影响。并且结合后处理模块给医生提出相关建议。实现人机协同。并且致力于提出这样的训练范式：结合分割模型前处理-少样本预训练模型-结合基础医疗大模型后处理。   1. **国、内外研究现状和发展动态**   随着深度学习技术的普及，国内外不仅仅是前列腺癌症，很多很多的病情诊断也都渐渐有融合深度学习的方法进行自动诊断检测。问题是大多数基于深度学习的方法依赖医生提供的精标签，细到每一帧的分割。这样对医生提供的数据要求极高。视频质量参差不齐，标注过粗都能够影响到模型的精确性。并且相关工作缺乏一定可解释性，对医护人员的建议仅仅停留在做出一个最后的诊断。  目前结合深度学习的医工交叉方向越来越希望结合对话大模型的能力给医护人员提出诊断建议，或者是展示推理的推理链，让诊断更有逻辑性。  **（4）创新点与项目特色**  本研究项目致力于设计前处理模块，解决数据视频长短不一，标注标签过粗的问题。降低阳性数据中与病灶无关部分对图像-文本预训练模型的影响。  此外，本项目希望设计一个后处理模块，结合基础的医疗大模型为诊断流程提出诊断建议。为基础医护人员提供建议帮助其正确诊断。  最后我们希望把模型训练部署在应用平台上，根据使用和实时反馈、后续走访等方式不断优化模型，并且实现落地。  本项目的特色在于结合大模型的丰富基础知识，缓解数据质量问题。并且上线应用平台实现模型优化。  额外的，通过这样设计训练模块的思路，本项目希望能提出这样的前处理-预训练模型-后处理的方式来针对其他的病症的智能诊断。  **（5）技术路线、拟解决的问题及预期成果**  技术路线：   1. 结合基础的分割模型，在少样本训练下为输入的稀疏数据提供进一步的分割知识。挑选关键片段，计算其他片段和关键片段的相关性。 2. 结合前处理得到的关键片段和其他片段的相似性，输入图像-文本预训练模型，得到一个诊断结果。 3. 结合诊断结果和关键帧，通过医疗大模型为医护人员输出诊断建议和诊断结果。 4. 基于此模型开发出一个诊断平台，能通过输入输出数据不断优化模型。   拟解决的问题：   1. 压缩阳性样例中无关病灶部分对训练和诊断的影响。 2. 改善智能诊断的低解释性问题。   预期成果：   1. 经过前处理后能够通过粗标签的视频训练出准确性更高的模型 2. 能够提供给医护人员详细的诊断建议和结果 3. 能够设计出上线的应用平台，通过方便的操作来使用模型，同时通过医护人员使用的病例继续优化模型。   **（6）项目研究进度安排**  第一阶段：（2024.12-2025.2） 阅读论文，研究现有方法  第二阶段：（2025.2-2025.5） 前处理训练，数据集收集，模型复现  第三阶段：（2025.5-2025.9） 模型搭建，算法设计，微调模型  第四阶段：（2025.9-2025.12） 应用平台开发，落地  **（7）已有基础**  团队成员对项目建立起初步认知;在指导老师的引导下收集到足够的相关文献资料进行调研。前期调研已经了解到当前方向的国内外研究情况，对任务处理流程有清晰认知，包括前处理、分类、后处理模块的设计已经初有设计。并且已经调通了基础的图像-文本预训练的模型。已经收集到一部分初步的数据，锚定好了目标使用的基础大模型。在解决算力问题之后能够加速推进项目进行。 |

#### 三、 经费预算

|  |  |  |
| --- | --- | --- |
| **开支科目** | **预算经费（元）** | **主要用途** |
| 预算经费总额 | 3000 |  |
| 专用设备费(购置、维修、租赁) | 2000 | 服务器设备租赁 |
| 材料费 | 500 | 相关书籍、网盘、腾讯会议等平台会员 |
| 测试化验加工费 | 0 |  |
| 差旅费 | 500 | 校区交通 |
| 会议费 | 0 |  |
| 国际合作与交流费 | 0 |  |
| 出版/文献/信息传递/知识产权事务费 | 0 |  |
| 劳务费(含专家咨询费） | 0 |  |

#### 四、 项目组成员签名

|  |
| --- |
|  |

#### 五、 指导教师意见

|  |
| --- |
| **导师（签章）：**  **年        月        日** |

#### 六、 院系大学生创新创业训练计划专家组意见

|  |
| --- |
| **教学负责人（签章）：**  **年        月        日** |

#### 七、 学校大学生创新创业训练计划专家组意见

|  |
| --- |
| **负责人（签章）：**  **年        月        日** |

|  |  |
| --- | --- |
|  | 6 / 6 |