# 创意介绍

## 项目介绍

“康康肺吧”是一套基于多个病例的CT影像样本数据，学习训练样本的病灶纹理的全新AI算法模型。通过使用CNN卷积神经网络训练CT影像的识别网络，AI可以快速鉴别新冠肺炎影像与人体健康状态下肺部影像的区别，AI每识别一个病例平均只需要不到20秒，大大提高诊断效率，减轻医生压力。

## 项目可行性分析与应用前景

近年来，AI辅助诊断在精准医疗领域前景广阔，不少疾病在AI算法诊断下都能够达到极高的准确率。

本次新冠肺炎疫情来袭的背景之下，其防控难点之一就是医疗资源高度短缺，快速诊疗能力出现结构性缺失，尤其是核心疫区。由于短时间内的爆发式就诊需求，难以实现快速精准诊断及分诊。一位新冠肺炎病人的CT影像大概在300张左右，这给医生临床诊断带来巨大压力，医生对一个病例的CT影像肉眼分析耗时大约为5-15分钟，且CT识别对医生资历具有一定的要求，长时间高强度的工作带来疲劳也会影像诊断效率。由于医疗资源紧张，导致部分地区出现患者拥堵并伴随恐慌情绪，院内交叉感染风险增加，医疗资源无法精准匹配到不同病情程度的患者。

目前采用病灶定量评价的方法，涉及到病变累计的肺体积范围、密度等多因素，缺乏统一标准，以往采用传统手工勾画 ROI 的方法进行量化的评估，往往需要 5-6 小时，效率低，临床推广难，而利用这个新 AI 系统能够实现病变区域的自动检测，在 2-3 秒之内就能完成定量分析，极大提升了精准定量分析的效率。

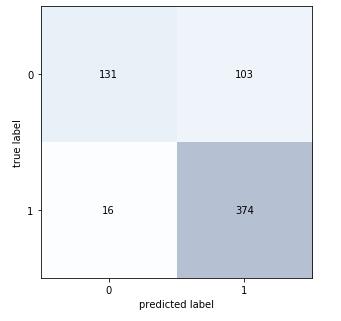
AI能够在其中发挥作用，对于医院和患者来说无疑都是一件好事。2月4日，国家工信部特别发布了《充分发挥人工智能赋能效用 协力抗击新型冠状病毒感染的肺炎疫情倡议书》，鼓励攻关并批量生产一批辅助诊断、快速测试、智能化设备、精准测温与目标识别等产品，助力疫病智能诊治。

伴随第六版新冠肺炎诊疗方案的发布，临床诊断无需依赖核酸检测结果，CT影像临床诊断结果可作为新冠肺炎病例判断的标准由此带来的CT阅片压也在急剧上升。。因此，使用AI算法对CT影响进行诊断完全合乎标准，且能够减缓临床医生诊断压力，提高确诊效率。

## 与同类产品比较

“康康肺吧”系统占存较小，操作简单，极易快速投入在日常医疗工作中。

在算法选择方面，常规的算法模型是以“精度 = 预测正确的测试样本数 / 总样本数”作为评估标准，其运算结果就容易忽略一些疑似病例甚至早期的轻微患者。而“康康肺吧”采用以“真阳性 /（真阳性 + 假阳性）”的计算结果作为评估标准。

根据测试结果得出比较数据：

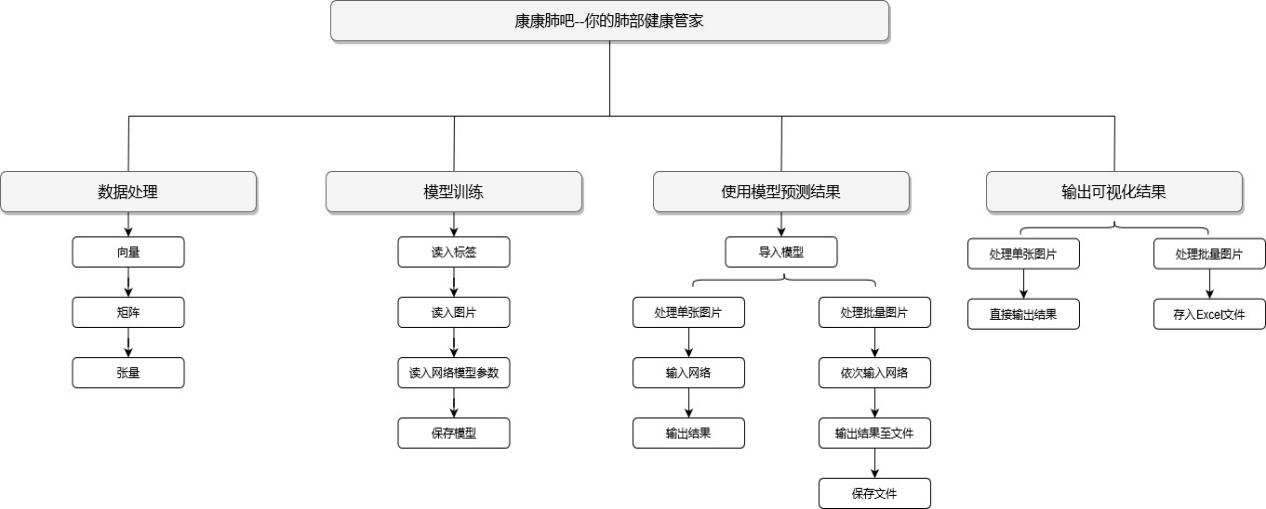
Precision = 374 / (374 + 103)= 78.40%

Recall = 374 / (374 + 16) = 95.89%

在此类突发性公共卫生事件中，面对高传染性疾病，要想最短时间内的实现早控制、早隔离，假阴性就必须直观地最小化，因为错误地诊断肺炎患者没有肺炎是比错误地诊断健康人作为肺炎患者更大的事情，“康康肺吧”能在最短时间内精确识别出确确诊病例和疑似病例，协助快速实现疫情控制隔离。

# 二、核心业务功能

## 2.1 总体功能结构图



## 2.2 功能模块介绍

#### 2.2.1 main.py功能模块

定义网络模型参数及函数

设定训练参数和即时输出

保存模型待用

#### 2.2.2 ****input\_set.py****功能模块

将训练图片从文件夹中提取

出来并统一化大小及冲量化

#### 2.2.3 ****solve.py****功能模块

定义Solve类 封装接口待用

导入训练好的模型

设计两种预测方式

#### 2.2.4 ****interface.py****功能模块

定义界面及各种控件

引用Solve中的方法并绑定到对应控件

预测所选择文件或文件夹

# 三、总体设计

## 3.1 数据集设计

test-NORMAL 234  
test-PNEUMONIA 390  
train-NORMAL 1341  
train-PNEUMONIA 3875  
val-NORMAL 8  
val-PNEUMONIA 8

## 3.2 界面设计

Image控件：预选单张图片

Folder控件：预选文件夹

Path：显示所选路径

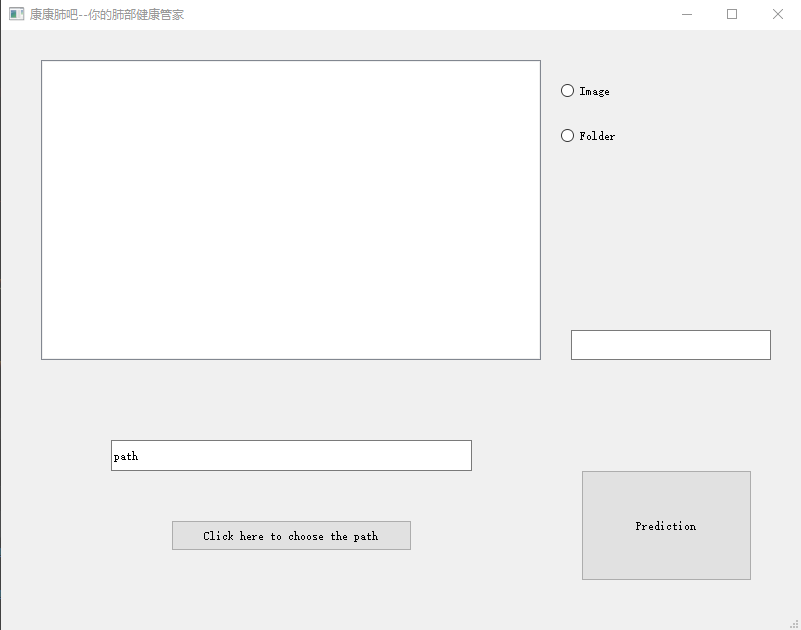
Click hero to choose the path：点击选择路径

View\_picture：显示所选图片

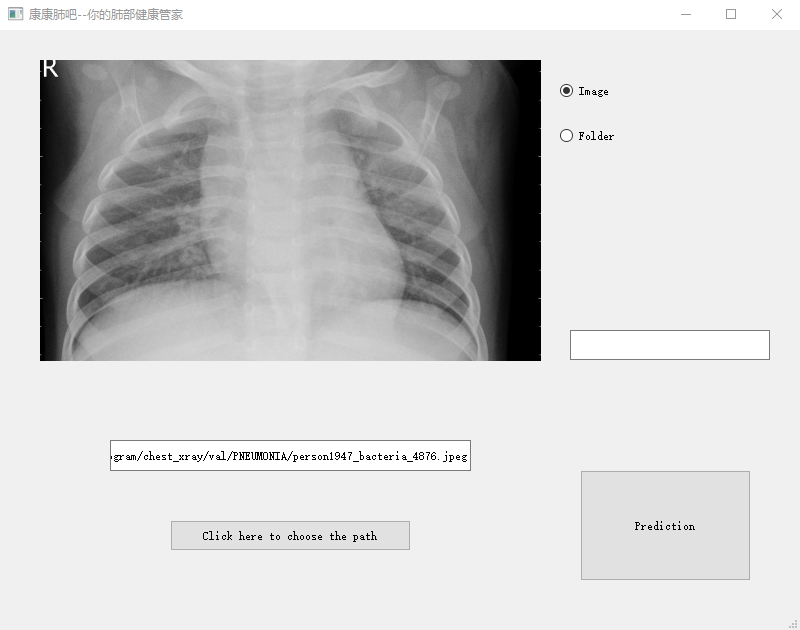
View\_result：显示结果

Prediction：运行程序

#### 3.2.1 待选界面



#### 3.2.2 单张图片预测界面



#### 3.2.3 文件夹分类界面



## 3.3 架构设计

#### 3.3.1 项目目录文件结构

folder:

│ folder.jpg

│ input\_set.py

│ interface.py

│ list.txt

│ main.py

│ solve.py

│ weights.06.hdf5

│

└───chest\_xray

├───test

│ ├───NORMAL

│ └───PNEUMONIA

├───train

│ ├───NORMAL

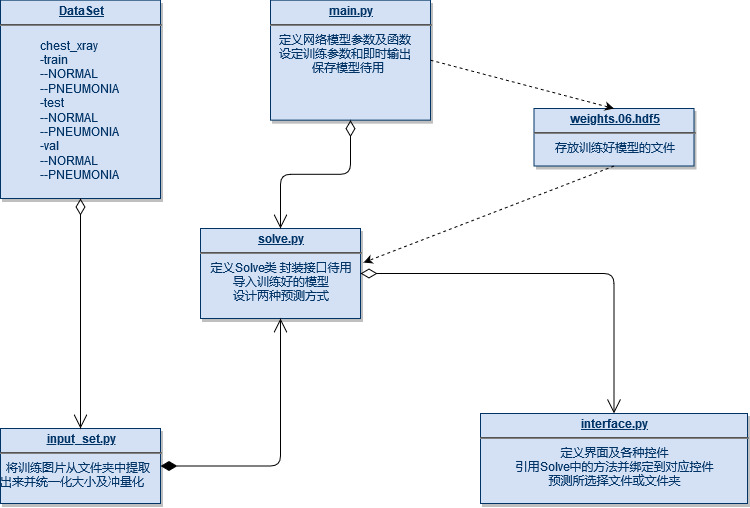
│ └───PNEUMONIA

└───val

├───NORMAL

└───PNEUMONIA

#### 3.3.2 项目层次结构



# 四、技术难点

本项目开发过程中，主要遇到2个技术难点，具体技术难点及解决方法如下：

## 4.1 数据集获取

建立项目的初步想法后进一步的研发，但相较于常规数据，公开医疗信息资源较少，训练模型的数据集搭建成为一大难点。

为了保障训练模型的有效搭建，小组成员分工后便在各大知名论坛（github.stackoverflow.kaggle...）开展了资源搜索。最终在Mendelay发现一项可有效应用于训练模型搭建的Dataset ：Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification（https://data.mendeley.com/datasets/rscbjbr9sj/2）查看License得出可用的结果后，整合了各平台网站的资源进一步完善训练集，最终解决训练集搭建难题。

## 4.2模型设计

在着手解决问题的初期，我们组决定先照搬之前的作品中的模型参数，也取得了不错的识别率，但预测时间稍有缓慢。而后参考一些相关的论文，发现医学上的AI模型的评判标准与普通的不一样，应为“宁可错杀一千，也不放过一个”的原则，在预测中应将阳性可能稍低的样本也标为阳性，并建议继续做多项检测。因此改变模型策略，重新开始设计与训练新的模型，将判断率上升到约96%，并将预测效率提高，以此来针对这次疫情前线的实际情况。

# 项目总结

参赛之初，我们便确立了AI识别的项目方向。数月里的不断学习，我们一点一滴地建立起认知框架，再通过大量的项目实践，为我们的知识体系增添砖瓦。从团队建立之初的探讨磨合，到中期的交流学习，再到后期的研发优化。我们收获的不止是对于机器学习认知，更可贵的是一种合作进步、探索求知的精神。这种积极的备赛氛围不断驱使着我们团队的每一个人，我们相信我们终将完成一部充满意义的作品。

2020年初，突如其来的新型冠状病毒肺炎扰乱了所有人的生活，我们开始疯狂地争抢起口罩，我们不得不放弃春节的团圆，我们被无形的可怕病毒锁死······我们能做点什么？我们好像可以做点什么！

近年来，AI在精准医疗领域前景广阔，不少疾病在AI算法诊断下都能够达到极高的准确率。面对着短时间内的爆发式就诊需求，我们开始着手查阅资料，并快速确立了项目框架。由于相关医疗信息的公开性较低，爬取的CT影像搭建的训练集具有一定狭隘性，但经更深入地搜集和多次测试优化后，成品系统已经具有较为成熟识别能力，在日常运行中可以有保障的提供相关医疗辅助。

在严峻的疫情面前，我们是渺小的，但我们认为如果我们能尽自己的能力为防疫献出一点帮助，那才是我们追求进步所具有的价值。近期看到阿里达摩院，腾讯云等企业团队相继研发退出了AI产品并迅速的投入到了一见的战役工作中去。相比起他们更专业更系统的研发运作，我们做出的产品像是一个学习奔跑的孩子，即使我们很小很弱，但我们永远不会放弃奔跑的权利和前进的机会，但我们总会追上他们的。能切实地帮助这个社会，便是每一种行业价值的光辉，就是迎着光的方向，我们的奔跑才赋有使命的价值。