|  |
| --- |
|  |
| IRIS |
| 一个基于边缘计算和快速集成之卷积分析模型的呼吸系统疾病风险早期识别系统 |
|  |
| **陈威霖、刘锦熙、魏纞熹**  **上海中学国际部** |
| **2022/10/28** |

|  |
| --- |
|  |

# **摘要**

呼吸系统疾病是一种常见病、多发病，在中国有超过一亿的呼吸系统疾病患者，同时呼吸系统疾病也是不容忽视的疾病，如其中最常见的慢性阻塞性肺病（简称慢阻肺），已经是世界上第三大死亡原因，在中国的城市死亡率占第三位，在农村则占首位。我们计划开发一种创新的呼吸系统疾病风险早期识别与预警系统，可以居家连续监测，并且利用人工智能算法来有效分析与呼吸系统疾病相关的生理指标。

在概念设计上，本研究根据目前现行技术支持而且居家可监测的生理指标之中，选取与呼吸系统疾病最相关的血氧饱和度以及咳嗽音，设计并制作了一种可以连续录制咳嗽音、检测血氧饱和度的便携式设备，并将采得血氧数值，经过快速集成之卷积模型分析过的咳嗽音数据，结合指南规范的呼吸系统疾病风险问卷的数值、来分析判断用户的呼吸系统疾病风险。

在硬件设计上，首先考虑到用户的易用性，采取了便携式集成设备的方式，不但可以平时用于放置在居家场景收集信号，也可以在基层医疗场景如乡镇卫生院使用。其次为了保证数据采集的质量，没有采用直接在智能手机上收音采集的模式，而是采用将一个接触式脉搏血氧饱和度计与一组四声道麦克风模块加一个上位机的集成硬件模式进行，以确保数据的精确度，为将来医疗专业人士做诊疗参考的场景做准备。

在软件及人工智能算法上，本研究开发了一种快速集成卷积模型，主要针对咳嗽音来进行分析来判断用户肺部健康情况。数据方面采用开源的Covid 19-Cough [1]数据集，进行算法训练之后，在综合判断呼吸系统疾病风险结果显示，达到曲线下面积（AUC）82.5%。

*关键词：呼吸系统疾病、咳嗽音、脉搏血氧饱和度、生理指标、特征提取、综合特征参数、机器学习、卷积模型、远程监测、数字卫生、医疗健康*

# **目录**

摘要........................................................................................................................................................I

目录.......................................................................................................................................................II

第一章 绪论 1

1.1 研究背景与问题陈述 1

1.2 相关研究现状及本研究的思路 1

1.3 本文研究内容 2

第二章 硬件及功能设计 4

2.1 功能模块的选择 4

2.2硬件原型搭建 6

2.3采集功能的实现 7

2.4网页客户端的服务实现 8

第三章 咳嗽音信号数据与模型开发 10

3.1 数据收集 10

3.2数据预处理 10

3.3模型架构 10

3.4模型主干 11

3.5快速集成方法 12

3.6模型训练与评估 12

第四章 总结与展望 16

4.1 总结 16

4.2展望 16

参考文献 17

# **第一章 绪论**

**1.1 研究背景与问题陈述**

近年来呼吸系统疾病成了继心血管、肿瘤之后全球人类第三大死亡原因，且其发病率逐年上升，2017年全球慢性呼吸系统疾病的患病率约为7.1%，总患病人数达5.449亿［2］，其中，中国哮喘患病总人数已达到4570万人[3]，慢阻肺总患病人数约1亿人[4]，整体疾病患病率、发病率仍然处于高位且还有上升的趋势。在众多慢性呼吸系统疾病患者中，只有28.8%的哮喘患者曾在就医得到过明确诊断，慢阻肺漏诊高达70%[4]，更无从谈起得到及时、规范治疗。

中国的十五项公共卫生服务，主要覆盖了心血管、糖尿病、妇幼保健与重度精神病，呼吸系统疾病还不在其中，其将是心血管、糖尿病之外下一个急需关注且影响范围广泛的疾病领域。目前的呼吸系统疾病监测策略使用各种症状评分和基本生理指标的测量，包括脉搏率、脉搏血氧饱和度、肺活量、呼吸音和肺音等等。目前市面上的呼吸系统疾病检测设备包括肺活量计或是电子听诊器，都不适合非医疗专业人员独立操作，设备的成本以及可获得性都受限制。同时，监测得到的数据解释，也有赖于专业医疗人员的解读判断。但是由于在相关疾病领域医疗资源的不均衡以及新冠疫情流行期间就医不便的影响，亟需有简易的科技手段进行有效预警，帮助民众在早期连续监测并且识别呼吸系统疾病可能的风险，以提高民众的早诊早筛率，提升呼吸系统疾病的防控手段。如果能够做到慢性呼吸系统疾病的早期筛查与诊疗，可以显著提高病人的生活质量，减少急性加重的发生，对于民众个人以及社会整体的卫生经济效益巨大。

**1.2 相关研究现状及本研究的思路**

由于新冠疫情的流行，许多研究开始关注远程监测数据在医疗方面的使用与评估，呼吸系统疾病也在远程医疗的关注领域，但卫生组织没有明确认定何种方法最佳，所以大家都在探索当中。

肺活量测定是评估呼吸功能最常用的测试，主要用来识别和管理慢性阻塞性肺病 (COPD)、哮喘和其他影响呼吸系统的疾病[5]。然而，对肺活量测定的多项研究表明，初级保健机构在遵守质量标准方面成绩不佳[6]，在非医疗环境下居家使用肺活量计，因为用户的依从性差，无法证明可以增加额外的有用信息来监测呼吸系统疾病患者的预后或者预测急性加重，因此在考虑呼吸系统早期疾病识别时，要考虑选取的生理指标的医学解释性、用户的友善程度以及辅助决策的潜力。

脉搏血氧饱和度是评估哮喘与阻塞性肺疾病严重程度的有用诊断辅助手段[7]，血氧饱和度 (SpO2) 是衡量血液中携氧血红蛋白数量相对于不携氧血红蛋白数量的一种指标，即血液中血氧的浓度，正常的血氧饱和度水平在海平面为95-100%之间。血氧饱和度是监测组织氧合功能的一个重要指标并能一定程度上反应人体的呼吸系统疾病症状，血氧饱和度已被用作诊断COVID-19的指标[8]。有研究指出，血氧残差的标准偏差（每天的SpO2 变异性）和长期趋势的时间演变，在检测慢阻肺患者病情异常时非常准确，甚至可以检测到有长期恶化的病人的变化[9]。从脉搏血氧饱和度计获得的所有生命体征(脉搏率、血氧饱和度和呼吸率)都可以预测COPD加重事件，其中氧饱和度是最能预测的，其次是呼吸率和脉搏率 [10]。目前血氧饱和度测量的标准是使用光体积描记法 (PPG) 方法的脉搏血氧饱和度法，该方法通过用光照射皮肤后透射或反射的光量来测量血容量的变化。

咳嗽是一种非常常见的症状，患病率高达全球总人口的33% [11]。呼吸系统疾病的10-38%都伴随有慢性咳嗽，包括慢性鼻窦炎、哮喘、慢性阻塞性肺疾病(COPD)、肺炎、慢性支气管炎、肺癌等等[12]。目前，国内外对咳嗽音的研究主要集中在两方面，一方面是对咳嗽音的监测和记录，另一方面是对咳嗽音的分类识别。咳嗽音的分类到目前为止采取的特征主要是倒谱系数，其次是时域和与能量相关的特征。近年来多个研究希望藉由机器学习与算法提高呼吸系统疾病的诊断准确率，目前看到的研究领域主要包括卷积神经网络 (CNN)、隐马尔可夫模型等。

本研究选取了血氧饱和度以及咳嗽音作为主要的监测指标。

远程监测生理数据的趋势并且预测分析其进一步发展方向，使早期发现疾病或者避免病情加重成为可能，因为生命体征的相对变化可能比检测生命体征达到绝对阈值更为重要[13]。研究显示，大多数健康相关的监控系统需要用户的积极参与，因此必须考虑用户的易用性[14]，特别是老年人群的用户，研究发现他们甚至不激活基于手腕的监测设备[14]。因此，远程监控设备在初始设置、连接和操作程序方面必须是用户友好的，或者采取一种以用户被动的方式收集数据、不需要用户太多操作技巧的可靠的客观监测方式[15]。

本研究利用便携式设备方便用户在居家场景下实时收集咳嗽音及血氧饱和度值。

根据收集的数据进行分析并确定干预措施方面，许多研究发现软件和基于计算机的决策辅助或预警工具可以起到积极的作用[16]。对于用户来说，这些工具与支持系统的目的并不是做出明确的诊断或提出治疗方法，而是决定是否进一步寻求医疗协助或者提醒患者关于自身的健康风险。

本研究利用快速集成之卷积分析模型对咳嗽音进行分析，结合血氧饱和度指标和呼吸系统疾病道风险问卷，快速进行呼吸系统疾病的早期识别，目的就是提醒用户呼吸系统疾病的可能风险，提醒用户进一步寻求医疗协助。

**1.3 本文研究内容**

本文主要研究⽬标是设计和创建⼀个基于便携式集成设备和快速集成之卷积分析模型的呼吸系统疾病早期识别系统。主要组件包括：一个血氧咳嗽音分析仪的设计和制造，用以获得连续的血氧饱和度数值以及录制咳嗽音的生理监测数据，一个通过机器学习和信号处理技术建立的咳嗽音检测及分类的算法模型，一个网页软件来收集用户关键信息，包括用户的基础呼吸系统风险基线数据 (COPD-SQ问卷，COPD-SQ问卷是在中国人群中改良验证的筛查问卷 [17])、长期的血氧饱和度、咳嗽音的变化以及呼吸系统疾病的早期识别信息。

为了保证数据质量，我们没有采用直接在智能手机上收音采集的模式，而是另行设计便携式的集成设备，将一个接触式血氧饱和度仪与一组四声道麦克风模块加一个上位机的集成硬件模式进行，以确保数据的精确度，为将来医疗专业人士做诊疗参考的场景做准备。设备本身可支持对各种呼吸系统疾病的有效管理，用户通过收集和跟踪数据变化，可以更好地理解自己的健康情况，有助于提高预后及整体生活质量。数据与分析结果也可以为门诊或远程医疗就诊提供信息，方便与医疗保健提供者进行沟通，并进行实时有效的医疗干预，以降低发病率和住院风险。

本研究主要创新点：

(1) 连续监测血氧饱和度、咳嗽音等多维度数据，配比筛查问卷数值进行分析，精准判断呼吸系统疾病的早期风险，降低呼吸系统疾病早期漏判率；

(3) 本文采用了一种基于轻量级卷积网络MobileNet的模型，在基于咳嗽音的肺部健康监测任务上获得了良好的表现，且模型可在CPU设备上进行部署。

(4) 本文采用了一种创新的快速集成方法，加强了卷积网络在推理中的准确率，而不显著增加计算复杂度。

(5) 便携的集成设备使得连续监测血氧饱和度、咳嗽音成为可能，兼具用户友好度和可用性两方面优点。

# **第二章 硬件及功能设计**

考虑到客户的易用性，在有潜在呼吸系统系统疾病人群的居家场景中，需要通过一套便携的、紧凑的硬件设备，对血氧饱和度和咳嗽音数据进行实时采集和提取，并通过基于深度学习技术的咳嗽音呼吸系统疾病识别模型对咳嗽音进行分析，再结合用户的病史调查做风险比对，达到对呼吸系统疾病风险的初步识别和预判。

数据采集硬件设备的设计的目标：

(1)方便携带、容易组装，结构紧凑；

(2)数据采集精细，质量要求高；

(3)血氧饱和度和咳嗽音的整合数据的易用性。

经过对硬件结构原型的反复迭代。我们决定采用一套以Raspberry Pi为backbone的嵌入式的硬件框架，达到硬件系统的快速搭建和快速功能实现。

**2.1 功能模块的选择**

**2.1.1血氧数据采集模块**

我们调查并比较了各种带有血氧功能的应用模型，主要分为是三大类：传统的血氧传感器模块、智能穿戴设备上的血氧饱和度仪、以及医疗上夹在手指的小型血氧饱和度仪等。 由于智能穿戴设备上的血氧饱和度仪和医用小型血氧饱和度仪，均为封闭式设备，无法获得可供二次开发的实时血氧饱和度数据。所以我们决定采用单独的血氧传感器模块，并搭配一个简单易用的开发板的方案。

我们选择采用M5Stack厂商的Stick C+开发板，搭载M5Stack MAX30100血氧心率模块作为血氧饱和度数据采集模块。M5Stack 产品的主要特点，是能够像乐高一样，即插即用方式快速搭建自己的组件体，而且相对于传统开发板的简单功能验证，M5Stack更具备产品化特征。

M5Stack Max30100血氧传感器如图1所示。

目前无创脉搏血氧饱和度测量技术主要分为透射式双波长脉搏血氧饱和度检测和反射式双波长脉搏血氧饱和度检测，后者光斑强，易于观察。由于我们的系统需要通过夹在测量者手指上进行测量，我们选择反射式血氧传感器。M5Stack MAX30100 是一款集成有脉搏血氧饱和度仪和心率监测传感器的模块。该器件有两个LED发出红外光，一个光电探测器用来测量反射回来的光，可测量氧合血红蛋白（HBO2）和血红蛋白（HB）对红外光的吸收量，以检测血氧饱和度。其优化的光学器件和低噪声模拟信号处理器，让MAX30100 在血氧饱和度检测上，采集数据上表现优良。

M5Stack Stick C+是一款迷你的IoT开发板，集成 ESP32 芯片,具备Wi-Fi 功能，能够快速地搭建功能原型，简化的开发过程，并且有着丰富开源代码和活跃的论坛社区，有着丰富的开源资源，可以加速整个研究过程。

M5Stack Stick C+开发板如图2所示。

图1：M5Stack MAX30100 血氧传感器 图2： M5Stack Stick C+ 开发板

**2.1.2咳嗽音数据采集模块**

麦克风是用于记录咳嗽的声音。由于普通麦克风没有处理芯片，而我们的应用是一个智能语音识别的场景，需要较高的语音质量，所以，我们决定选用麦克风阵列。麦克风阵列，主要是由一组按一定几何结构（常用线形、环形）摆放的声学传感器组成，对采集的不同空间方向的声音信号进行空时处理，实现噪声抑制、混响去除、人声干扰抑制、声源测向、声源跟踪、阵列增益等功能，进而提高语音信号处理质量，以提高真实环境下的语音识别率。

我们选用基于Raspberry Pi的ReSpeaker 4-Mic阵列，它是一款适用于AI和语音应用的Raspberry Pi的四通道麦克风扩展板，有四个数字麦克风，支持片上语音算法，灵敏度为-26dBFS（全向，支持远场语音捕获，能够检测最远5米处的声音，即使在存在背景噪音的情况下也是如此）。在本项目系统中作为咳嗽音数据采集接口。

ReSpeaker 4-Mic麦克风阵列如图3所示。

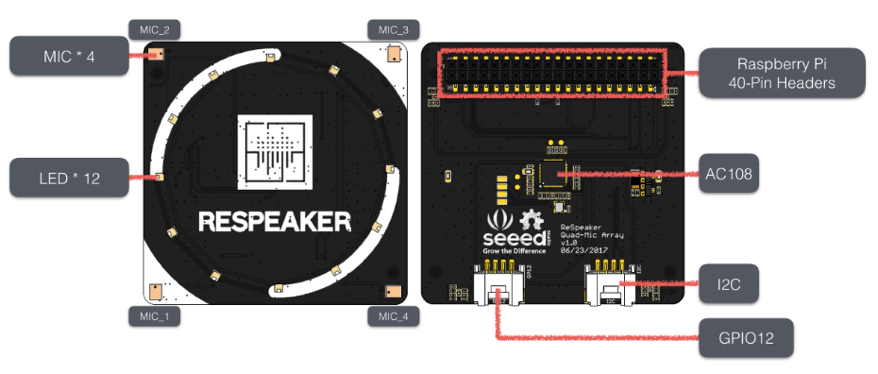


图3：ReSpeaker 4-Mic 麦克风阵列

**2.1.3主工作模块**

我们选用Raspberry Pi 4B 作为嵌入式硬件框架中的核心主工作模块。 其具备1.5Ghz运行的64位四核处理器，高达4GB RAM（可根据型号选择1GB、2GB、4GB、8GB），2.4/5.0 Ghz 双频无线LAN，蓝牙5.0/BLE，千兆以太网，USB3.0和PoE功能。其优秀的四核性能，能满足对数据采集处理、模型运算、网页后端支持的并行处理能力的需求。

Raspberry Pi 4B主工作模块如图4所示。

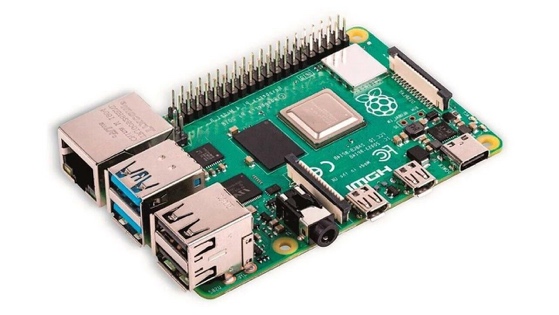


图4：Raspberry Pi 4B主工作模块

**2.2硬件原型搭建**

通过集成上述传感器、评估板和声音数据采集设备构建了一个快速紧凑的硬件原型。 M5Stack Stick C+ 与MAX30100传感器模块构成血氧饱和度数据采集模块，MAX30100通过I2C总线，将血氧饱和度数据实时传输到M5Stack Stick C+模块上。

ReSpeaker 4-Mic阵列通过排线（I2C）与主处理模块Raspberry Pi 4B直接连接，记录咳嗽音数据。同时，M5Stack C+ 通过WIFI发送血氧饱和度采集数据到Raspberry Pi 4B 上。

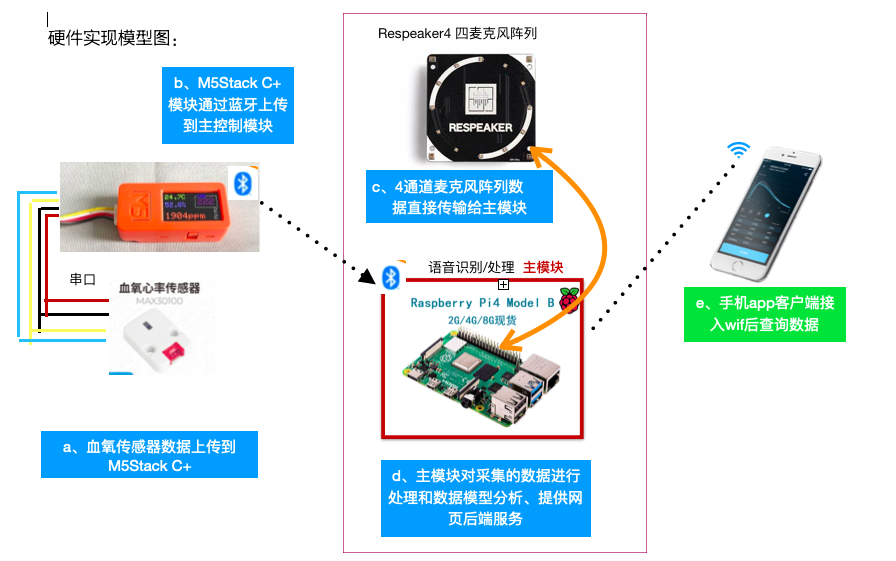


图5：硬件实现模型 - 概念图

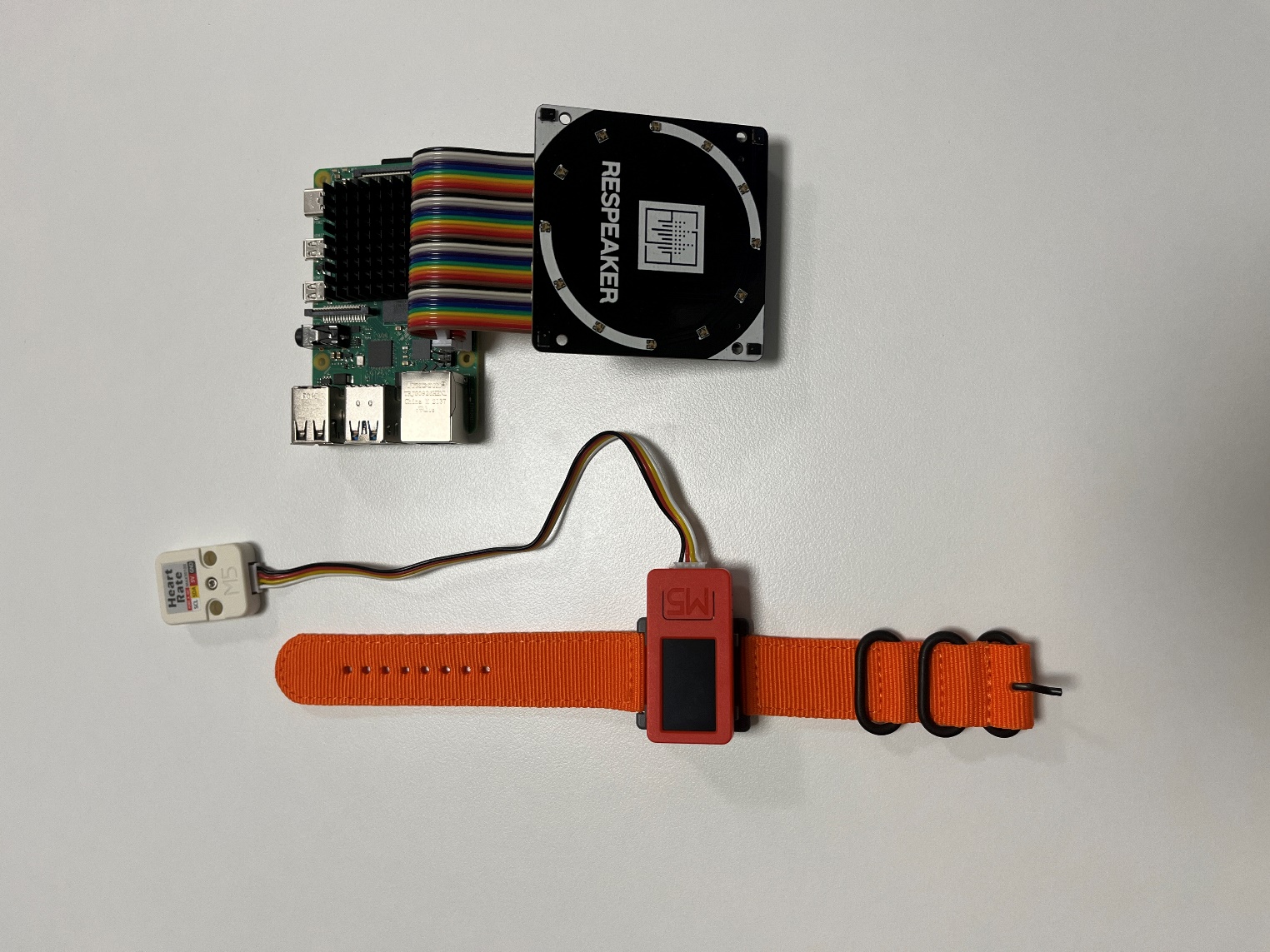


图6：硬件实现模型 - 实际图

**2.3采集功能的实现**

**2.3.1 血氧采集模块功能实现**

通过M5Stack MircoPython 快速实施定时血氧饱和度数据的采集功能，采集时间间隔为500ms，通过WIFI将数据上传到服务器端口。

图示

描述已自动生成

**2.3.2咳嗽音采集模块功能实现**

ReSpeaker 4-Mic麦克风阵列收集到的数据，通过40 pin head 排线 与 RaspBerry Pi主处理板相连。RaspBerry Pi 上通过三个线程，并行处理声音采集、4通道转2通道、血氧饱和度采集数据接收。 并按时间标签，将血氧饱和度采集数据和咳嗽音采集数据分别作为数据记录，插入到mongo数据库列表里，以备后期风险识别功能模块进行调用。线程功能描述如下：

图示

描述已自动生成



**2.4网页客户端的服务实现**

我们需要建立一个基于web网页形式的呼吸系统疾病风险识别的客户端应用框架。由于我们的主硬件功能模块，采用了python编程语言在RaspBerry Pi 上完成功能实现。而Django是Python体系下最成熟的web框架之一，因其能够快速开发网站应用的特性成为了中小型网站开发框架首选。 所以我们采用Django作为Web应用程序后端框架，其成熟的开源社区资源，可以使开发网站变得更简便、快速。

在客户端应用页面设计上，我们采用前后端完全分离的框架结构。我们采用了Vue.js 作为前端开发框架，代替Django本身较为孱弱的模板引擎。Vue.js 是一个优秀的前端界面开发 JavaScript 库，其聚焦在简化视图层的实现为目标，使其成为前端页面开发工具中的佼佼者。这种Django则作为服务端提供api接口，前后端实现完全分离，更适合我们的单页应用的开发构建。

画一个我们的网页的Django+Vue.js 结构大概实现示意图（类似下面的样子）

图示

描述已自动生成

图7：网页结构图

图8：网页客户端页面

补充一个客户端页面的样子。

# **第三章 咳嗽音信号数据与模型开发**

**3.1 数据收集**

在本研究中，我们采用开源且匿名化后的数据集Covid19-Cough [1] 训练咳嗽音分类模型。该数据集依托于一个呼叫中心和Telegram消息机器人，对1324名受试者进行咳嗽音的收集，且询问受试者是否目前是Covid19阳性患者。在这批受试者中，有682人报告为阳性，其中382人具有PCR检测结果, 295为有症状患者。该数据集总共收集了总时长为58分钟的咳嗽音录音，每个录音样本平均时长为2.6s. 因为我们的目标是根据咳嗽音判断患者肺部健康状况，我们将具有covid19症状的样本作为阳性样本，其他样本作为阴性样本。

同时，由于我们的应用在实际使用时可能会接收到各种声音干扰，我们需要使用一个模型来检测一段录音是否具有咳嗽声。由于Covid19-Cough数据集中的所有样本都是具有咳嗽音的，我们采用URBANSOUND8K数据集[18]作为没有咳嗽音的样本。该数据集包含8732个具有标注的音频片段，记录了城市中可能会出现的各种声音，如空调声，汽车鸣笛，音乐，施工声，小孩嬉戏等。我们将Covid19-Cough数据集与URBANSOUND8K数据集组成的新的数据集叫做Cough classification Dataset 1 (CoughCLS-1)数据集。

**3.2数据预处理**

当我们设备接收到一段录音之后，我们首先会对录音文件进行加载和重采样为32kHz的波形数据(numpy格式)，我们将波形数据统一padding为15 \* 32000 的numpy array。波形数据通过torchaudio工具包进行短时傅里叶变换(Short-Time Fourier Transform, STFT) 得到频谱表征(window size: 2048; stride: 302), 然后将频谱表征投影到20Hz 到24kHz的128组梅尔滤波器组上，得到梅尔频谱图。在训练时，我们对训练样本的音频采用添加噪声的方式(采用torchaudio实现)进行数据增强(data augmentation)。

**3.3模型架构**

如下图所示，我们主要采用模型框架进行咳嗽音分类。我们的模型框架主要有以下四个步骤：(a)首先将波形数据转化为梅尔频谱；(b)梅尔频谱作为单通道图输入预训练的卷积神经网络中进行特征提取；(c)卷积网络输出的特征图(feature map)经过集合平均池化，转化为特征向量；(d)特征向量经过一个两层的全连接层进行池化，输出标签预测。我们的模型采用深度学习框架PyTorch进行开发。

我们也对比了广泛在音频分析领域应用的特征提取加机器学习分类的方法[19]，具体步骤如下图：(a)首先对波形数据进行特征提取, 最后每个音频由一个特征向量表示；(b)特征向量通过机器学习模型得到分类结果。我们采用广泛使用的音频特征：(1)梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficients, MFCC)的均值和方差；(2)谱矩心；(3)过零率；(4)色度特征向量；(5)频谱截止特征。我们采用目前最强大的机器学习模型LightGBM [20]作为分类器。

相对于基于深度学习的模型，基于特征提取的机器学习模型依托于音频分析文献积累的经验，进行各种约定俗成的特征量的提取。这类方法在深度学习普及之前较为常见。其本身也具有推理效率高的优势。但是，其特征提取过程侧重于对波形的整体统计特性的挖掘，会损失一些细粒度的特性，且哪些特征对特定任务有效也是需要大量特征工程工作。对比之下，由于深度卷积网路具有强大的自动特征提取能力，所以其能够在各类音频任务上获得更优的表现。

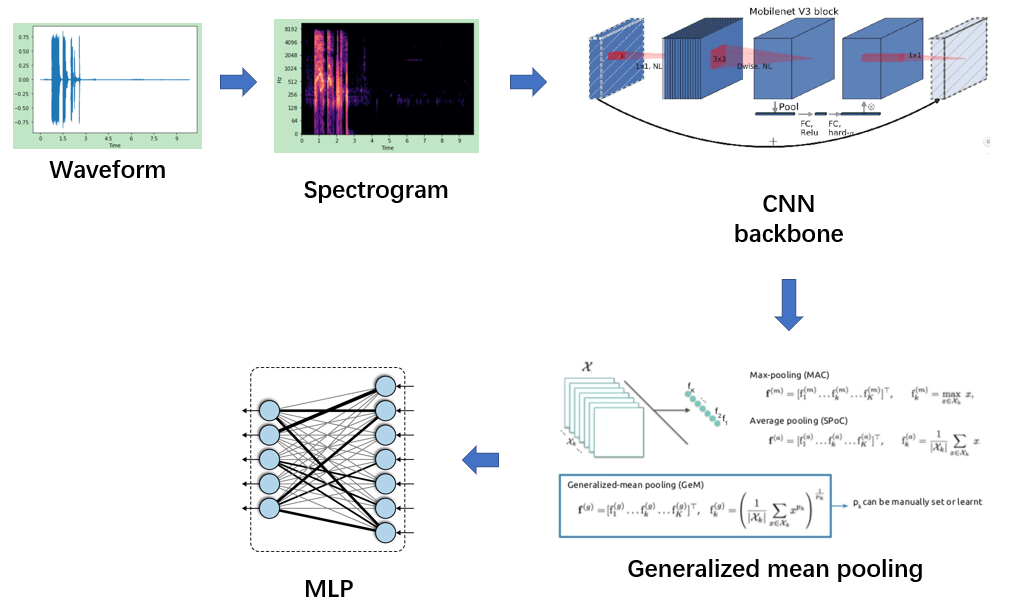


图9. 基于谱图与卷积神经网络的咳嗽音分类模型框架

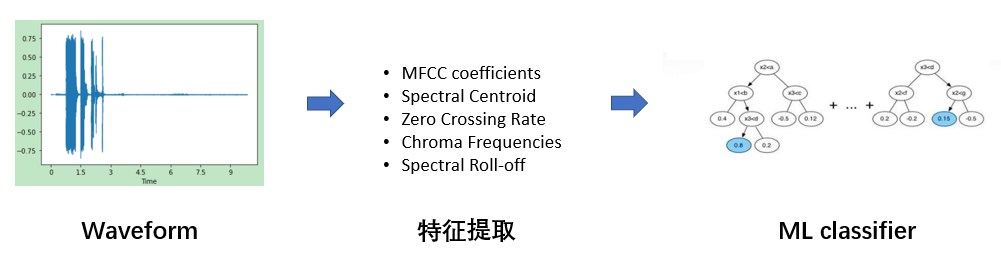


图10. 基于音频特征提取的机器学习模型

**3.4模型主干**

本研究中我们采用卷积神经网络[21]从梅尔频谱图中提取特征。卷积网络的主干由多个卷积层，激活函数和池化层组成，卷积层可以产生一组平行的特征图，它通过在输入图像上滑动不同的卷积核并执行一定的运算而组成。此外，在每一个滑动的位置上，卷积核与输入图像之间会执行一个元素对应乘积并求和的运算以将感受野内的信息投影到特征图中的一个元素。卷积神经网络在大型图像处理任务上获得了巨大成功[22]，并引领了第二次人工智能浪潮。

本研究采用预训练加微调的范式，这也是目前表现最为优异的方法。经过预训练的网络已经学习了大量图片中蕴含的语义知识，且具有良好的参数初始化，有利于整个模型在下游任务上的快速优化。我们采用在ImageNet上预训练的卷积神经网络作为主干。目前较为流行的网络主干有MobileNet[23], EfficientNet[24], ResNet[25], ResNeXt[26]. 其中ResNet[25]是最为广泛使用的网络结构，其提出的残差链接有利于更深的神经网络训练而不会有梯度消失问题，目前是各种神经网络必备的模块。MobileNet[23]和EfficientNet[24]在ResNet[26]基础上采用更为轻量化的网络设计，更加利于端侧的模型部署与应用。考虑到我们的应用需要部署在移动设备上，所以我们主要采用MobileNet这个轻量级的网络结构。最终我们采用Pytorch-Image-Models框架提供的预训练后的MobileNet模型权重作为我们的卷积主干。

**3.5快速集成方法**

一般采用卷积网络作为主干进行特征提取后，我们会使用一个多层全连接层进行分类结果预测。在众多研究中，模型集成[27]被大量使用。但是一般的模型集成需要采用多个不同的模型进行分类，这样不仅在训练过程中会消耗大量的显卡资源，而且在模型部署后推理过程中显著拖慢整个应用的速度。我们在本研究中开发了一种快速高效的模型集成方法，我们叫做快速集成方法（Fast-ensemble method）。快速集成方法（如图11所示）的实现如下：我们在一个卷积网络主干后接上K个不同的全连接层作为分类头。在训练时，每个分类头接收到的特征向量由于dropout的随机性而不同，这样训练后分类头可以学习到不同的参数，相当于通过不同的视角理解图像的特征向量，获得多种不同的logits结果。模型输出的logits是由这K个分类头的logits求平均后的结果。我们在实验中发现K=5可以有效提升模型效果而不至于过度降低模型的推理效率。

快速集成方法只在模型主干上增加了几个分类层，相比于模型主干，模型只增加了5k的额外参数，对模型显存占用不会产生显著影响，且由于分类层结构简单，在推理时也不会增加非常多的耗时。

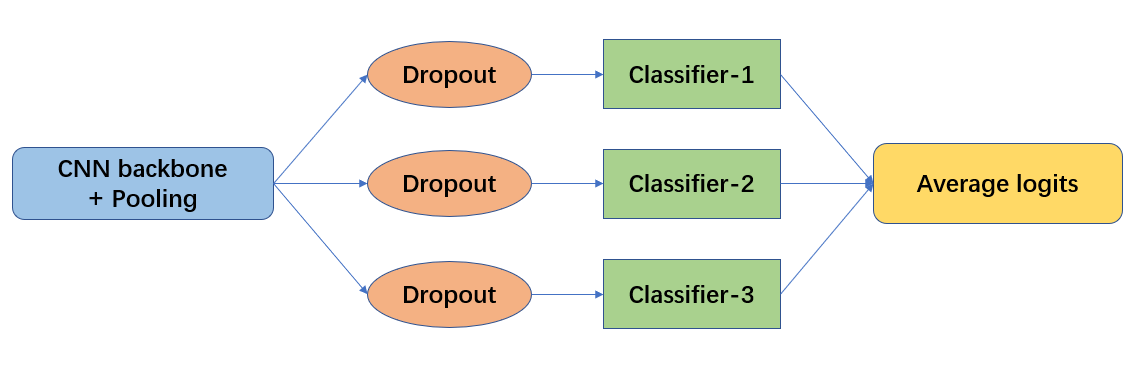


图11. 快速集成方法示意图.

**3.6模型训练与评估**

训练中我们采用batch size为16，预训练卷积主干使用学习率为2e-5，随机初始化的分类头采用2e-4的学习率，训练轮数最大为30。

**3.6.1咳嗽检测模型**

我们采用我们组合的CoughCLS-1数据集进行5折交叉验证的方式进行模型评估，评估的主要指标为5折交叉验证平均测试AUC分数。同时，我们还汇报以0.5为阈值时的平均真阳性率和平均真阴性率。模型的表现对比见下面的表1。我们可以看到，通过卷积神经网络，我们可以很高精度的检测一段音频中是否包含咳嗽音或是外界噪声。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 模型架构 | Auc(%) | TPR(%) | TNR(%) |
| 机器学习模型 | LightGBM | 78.5 | 73.2 | 80.1 |
| 深度学习模型 | MobileNet-v3  (not pretrained) | 86.8 | 80.6 | 85.7 |
|  | MobileNet-v3 | 97.1 | 91.5 | 94.2 |
|  | MobileNet-v3  + Fast Ensemble | 97.9 | 92.7 | 94.5 |

表1：各个不同的模型在CoughCLS-1数据集上进行5折交叉验证的表现

**3.6.2基于咳嗽音的肺部健康检测模型**

我们采用在Covid19-Cough数据集上进行5折交叉验证的方式进行模型评估，评估的主要指标为5折交叉验证平均测试AUC分数。同时，我们还汇报以0.5为阈值时的平均真阳性率和平均真阴性率。模型的表现对比见下面的表2. 在本研究中，我们默认采用预训练后的卷积网络。我们也对比了完全随机初始化的网络的表现，这些模型会标识“未预训练”。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 模型架构 | Auc(%) | TPR(%) | TNR(%) |
| 机器学习模型 | LightGBM | 65.3 | 56.7 | 61.1 |
| 深度学习模型 | MobileNet-v3  (not pretrained) | 73.1 | 58.7 | 72.7 |
|  | MobileNet-v3 | 81.1 | 70.6 | 80.3 |
|  | MobileNet-v3  + Fast Ensemble | 82.5 | 71.3 | 81.6 |

表2. 各个不同的模型在Covid19-Cough数据集上进行5折交叉验证的表现

我们可以看到，基于特征提取的机器学习模型在咳嗽音分类任务上表现明显差于基于深度卷积网络的模型，显示出深度卷积网络的表达能力和特征提取能力。在使用了我们的Fast Ensemble方法后，模型获得了1.4%的AUC提升, 这个提升经过Wilcoxon rank检验[28]后是显著的。同时，我们发现预训练加微调的迁移学习对我们咳嗽音分类是非常有益处的，未经过预训练的MobileNet-v3的表现非常差，这说明预训练阶段网络学习到了图像特征提取的能力，有利于下游任务的优化。

我们下面在图12中对比MobileNet-v3 (not pretrained),MobileNet-v3,MobileNet-v3 + Fast Ensemble这三个不同模型设定的训练损失函数收敛情况和在测试集上的泛化情况。我们可以看到是，未经过预训练的MobileNet-v3模型，其训练损失最开始有所下降后就一直处于震荡中，不再优化，其AUC分数也是没有达到一个比较高的水平。通过添加了Fast Ensemble机制，MobileNet-v3模型可以获得更加平稳的损失收敛，且最终得到更高的泛化性能。

|  |  |
| --- | --- |
| (a-1) MobileNet-v3 + Fast Ensemble模型的损失函数收敛情况 | (b-1) MobileNet-v3 + Fast Ensemble模型的auc分数收敛情况 |
| (a-2) MobileNet-v3模型的损失函数收敛情况 | (b-2) MobileNet-v3模型的auc分数收敛情况 |
| (a-3) MobileNet-v3 (not pretrained)模型的损失函数收敛情况 | (b-3) MobileNet-v3 (not pretrained)模型的auc分数收敛情况 |

图12. 各个卷积网络在Covid19-Cough数据集上的训练集损失和测试auc表现随着训练推进的变化情况

# **第四章 总结与展望**

**4.1 总结**

本研究的原型是一个概念验证，希望在没有专业医疗人员在场的情况下，为民众在居家场景下提供了一种容易获得的呼吸系统疾病的风险评估手段。

本研究中面临的挑战有许多。最初希望将血氧饱和度仪与录音设备进行进一步的整合，后来发现光学感应讯号与声音讯号之间容易互相干扰，因此为了确保数据收集的质量，最终还是将两个模块分开，再用上位机进行数据整合。另外的难点在算法迭加后如何维持模型的大小，以确保运行的流畅性，且不会损失细粒度。

最终本研究设计了一个完整的硬件加软件和算法模型的解决方案，它提供了一个准确、高效和可获得的端到端的方法，使得用户可以利用便携式的基础设备，对咳嗽音及脉搏血氧饱和度数据进行持续可靠的自我监测，并通过快速集成的卷积分析模型进行分析，以达成实时预警呼吸系统疾病风险的目的。其中，咳嗽音检测模型的AUC最高达到了97.9%，肺部健康诊断模型的AUC达到了82.5%，这个准确率（指哪个准确率？）超过了工程目标中设定的预期精度值（多少？）约10%，能够实现对呼吸系统疾病早期识别的目标。

**4.2展望**

本研究的呼吸系统疾病早期风险识别系统根据医学共识，同时监测了对呼吸系统疾病早筛预警较为关键的量表、血氧饱和度与咳嗽音，目前尚未有公开的数据库同时监测以上几种生理监测数据，本研究是此领域的一个突破。

未来基于这些数据，可以探索建立涵盖更多特征的呼吸系统疾病早筛模型，甚至基于这些特征融合，来判别不同种类的呼吸系统疾病，帮助用户及时判定风险，尽早寻求医疗协助。同时，长期的监测数据所揭示的生理数据的变化，也可以辅助医疗专业人员进行临床评估诊断，因此，未来也可以探索整合开发出更高的诊断准确性、临床可解释性、易用性的远程医疗诊断系统。

未来研究的其他想法包括与小区与乡镇卫生院合作，建立基于中国人群的咳嗽音、呼吸音数据库，一方面增加数据量以优化模型，另一方面可以在纳入中国人群特征后进行模型泛化，提升模型的准确率。

# **参考文献**

1. Project “fkthecovid,” “Dataset of recordings of induced cough,” Medical Investment Solutions, St Petersburg, Russia, Jan. 2021, license CC0 1.0 Universal. [Online]. Available: https://github.com/covid19-cough/dataset [Google Scholar]
2. GBD Chronic Respiratory Disease Collaborators. Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Respir Med. 2020 Jun;8(6):585-596. doi: 10.1016/S2213-2600(20)30105-3. PMID: 32526187; PMCID: PMC7284317.
3. Huang, Kewu et al. “Prevalence, risk factors, and management of asthma in China: a national cross-sectional study.” Lancet (London, England) vol. 394,10196 (2019): 407-418. doi:10.1016/S0140-6736(19)31147-X
4. Wang C, Xu J, Yang L, et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet. 2018;391(10131):1706-1717. doi:10.1016/S0140-6736(18)30841-9
5. Swanney MP, Odea CA, Ingram ER, Rodwell LT, Borg BM. Spirometry training courses: Content, delivery and assessment – a position statement from the Australian and New Zealand Society of Respiratory Science July 2017
6. Eaton T, Withy S, Garrett JE, Mercer J, Whitlock RM, Rea HH. Spirometry in primary care practice. The importance of quality assurance and the impact of spirometry workshops Chest 1999; 116: 416-23
7. Dalbak LG, Straand J, Melbye H. Should pulse oximetry be included in GPs' assessment of patients with obstructive lung disease? Scand J Prim Health Care. 2015;33(4):305-10. doi: 10.3109/02813432.2015.1117283. Epub 2015 Dec 11. PMID: 26654760; PMCID: PMC4750741.
8. Caputo, N.D.; Strayer, R.J.; Levitan, R. Early Self-Proning in Awake, Non-Intubated Patients in the Emergency Department: A Single ED’s Experience During the COVID-19 Pandemic. Acad. Emerg. Med. 2020, 27, 375–378. [CrossRef] [PubMed]
9. Clarke M, Gokalp H, Fursse J, Jones RW. Dynamic Threshold Analysis of Daily Oxygen Saturation for Improved Management of COPD Patients. IEEE J Biomed Health Inform. 2016 Sep;20(5):1352-60. doi: 10.1109/JBHI.2015.2464275.Epub 2015 Aug 4. PMID: 26259203.ProceedingsofTheFirstJointBMES/EMBSConferenceServingHumanity,AdvancingTechnology.pp.1017,Oct.13-16.Atlanta,GA,USA.
10. Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in Chronic Obstructive Pulmonary Disease: Identification and Prediction Using a Digital Health System. J Med Internet Res. 2017 Mar 7;19(3):e69. doi: 10.2196/jmir.7207. PMID: 28270380; PMCID: PMC5360891.
11. Chung, K.F.; Pavord, I.D. Prevalence, Pathogenesis, and Causes of Chronic Cough. Lancet 2008, 371, 1364–1374. [CrossRef]
12. Chung, K.F.; Pavord, I.D. Prevalence, Pathogenesis, and Causes of Chronic Cough. Lancet 2008, 371, 1364–1374. [CrossRef]
13. Hans Christian Riis, Morten H. Jensen, Simon LebechCichosz& Ole K. Hejlesen (2016) Prediction of exacerbation onset in chronic obstructive pulmonary disease patients, Journal of Medical Engineering & Technology, 40:1, 1-7, DOI: 10.3109/03091902.2015.1105317
14. M. Galarnyk, G. Quer, K. McLaughlin, L. Ariniello, and S. R. Steinhubl, “Usability of a wrist-worn smartwatch in a direct-to-participant randomized pragmatic clinical trial,” Digit. Biomarkers, vol. 3, no. 3, pp. 176–184, 2019.
15. Crooks MG, den Brinker A, Hayman Y, Williamson JD, Innes A, Wright CE, Hill P, Morice AH. Continuous Cough Monitoring Using Ambient Sound Recording During Convalescence from a COPD Exacerbation. Lung. 2017 Jun;195(3):289-294. doi: 10.1007/s00408-017-9996-2. Epub 2017 Mar 28. PMID: 28353117; PMCID: PMC5437191.
16. Eysenbach G. Consumer health informatics. BMJ (Clinical Research ed.). 2000 Jun;320(7251):1713-1716. DOI: 10.1136/bmj.320.7251.1713. PMID: 10864552; PMCID: PMC1127483.
17. Zhou J, Yu N, Li X, Wang W. Accuracy of Six Chronic Obstructive Pulmonary Disease Screening Questionnaires in the Chinese Population. Int J Chron Obstruct Pulmon Dis. 2022 Feb 10;17:317-327. doi: 10.2147/COPD.S341648. PMID: 35177900; PMCID: PMC8843418
18. URBANSOUND8K数据集见 https://urbansounddataset.weebly.com/urbansound8k.html
19. Ponomarchuk, A., Burenko, I., Malkin, E., Nazarov, I., Kokh, V., Avetisian, M., & Zhukov, L. (2022). Project Achoo: A Practical Model and Application for COVID-19 Detection From Recordings of Breath, Voice, and Cough. IEEE Journal of Selected Topics in Signal Processing, 16, 175-187.
20. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. NIPS.
21. Li, Z., Yang, W., Peng, S., & Liu, F. (2021). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE transactions on neural networks and learning systems, PP.
22. Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60, 84 - 90.
23. Howard, A.G., Sandler, M., Chu, G., Chen, L., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., & Adam, H. (2019). Searching for MobileNetV3. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 1314-1324.
24. Tan, M., & Le, Q.V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ArXiv, abs/1905.11946.
25. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.
26. Xie, S., Girshick, R.B., Dollár, P., Tu, Z., & He, K. (2017). Aggregated Residual Transformations for Deep Neural Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5987-5995.
27. Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2019). A survey on ensemble learning. Frontiers of Computer Science, 14, 241 - 258.
28. Frank Wilcoxon. 1945. Individual comparisons by ranking methods. Biometrics bulletin, 1(6):80–83.