|  |
| --- |
|  |
| IRIS |
| 基于边缘计算的AI 呼吸风险早期识别系统 |
|  |
| **陈威霖、刘锦熙、魏纞熹 Weilin Chen;Jinxi Liu;Lianxi Wei**  **上海中学国际部 Shanghai High School International Division** |
| **2022/12/2** |

|  |
| --- |
|  |

# **摘要Abstract**

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，呼吸系统疾病的早期筛查防治对减缓病程进展及减少肺功能的不可逆损伤起到决定性的作用，比如：其中最常见的慢性阻塞性肺病（简称慢阻肺，常见症状为咳、痰、喘，逐步演变为中重度肺损伤），已经是世界上第三大死亡原因。我们设计开发了一套基于边缘计算的AI算法系统，通过一套便携式设备对呼吸系统生理指标数据进行实时采集和分析，从而达到对呼吸系统疾病风险早期识别与预警。不仅可以对个体进行连续性监测，同时也可应用于社区或者基层卫生组织，进行早期呼吸系统疾病健康筛查。推而广之，它将为建立一套完备的呼吸系统疾病特征数据集，提供一种解决方案。

Respiratory diseases are common and frequently occurring diseases, and there are more than 100 million patients with respiratory diseases in China. Early screening and prevention of respiratory diseases are crucial for halting disease development and minimizing irreversible lung function impairment. Among the most common lung diseases, chronic obstructive pulmonary disease (also known as COPD, plagued by coughs, sputum, and wheezing, gradually progressing to moderate to severe lung damage) is already the third leading cause of death. We have designed and developed an artificial intelligence algorithm system based on edge computing. By using a set of portable devices, this algorithm system collects and analyzes respiratory physiological data in real time. This will allow early detection and early warning of respiratory diseases. In addition to providing continuous monitoring of individuals, the system can also be used for early screening of respiratory diseases in the community or primary care setting. Moreover, it will allow for the establishment of a comprehensive data set on respiratory diseases.

在概念设计上，本研究突出之处为：从目前现行技术支持而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边缘计算设备，并将采得血氧数值，以及通过快速集成卷积技术处理咳嗽音数据后的模型，结合一套符合指南规范的呼吸系统疾病风险问卷的数值，来多维度、综合分析判断使用者的呼吸系统疾病风险。从公共卫生预防角度，提高疾病早期诊断率以及提升卫生经济学效益。

The innovation points of the study are: Among the physiological parameters that can be monitored at home and are technically feasible, blood oxygen saturation and cough sounds, which are the most related to respiratory diseases, were selected. A portable edge computing device that can continuously record cough sounds and detect blood oxygen saturation was designed and developed. By combining the collected blood oxygen values, the fast integrated convolution model, and values from a guideline-compliant respiratory disease risk questionnaire, the user's respiratory disease risk is determined in a multidimensional and comprehensive manner. As far as public health prevention is concerned, it can improve early diagnosis and boost health economics.

在硬件设计特点上，本项目的突出之处为；首先考虑到使用者的易用性，采取了便携式集成设备的设计理念，不仅可以平时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。同时，在硬件上进行边缘计算，提高核心数据处理的效率。

在软件及人工智能算法上，本研究开发了一种基于轻量级卷积网络MobileNet的模型，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，且模型可在微处理器设备上进行部署。主要针对咳嗽音来进行分析来判断使用者肺部健康情况。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Table of Contents**

[摘要Abstract I](#_Toc119066524)

[目录Table of Contents II](#_Toc119066525)

[第一章绪论Introduction 1](#_Toc119066526)

[1.1 研究背景与问题陈述Research Background and Problem Statement 1](#_Toc119066527)

[1.2 相关研究现状及本研究的思路Prior Work 1](#_Toc119066528)

[1.3 本文研究内容Research content 2](#_Toc119066529)

[第二章硬件及功能设计Hardware and functional design 4](#_Toc119066530)

[2.1 功能模块的选择selection of functional modules 4](#_Toc119066531)

[2.2硬件原型搭建Hardware prototyping 6](#_Toc119066532)

[2.3采集功能的实现Realization of Data Collection 7](#_Toc119066533)

[2.4网页用户端的服务实现client-side Web application 9](#_Toc119066534)

[第三章咳嗽音信号数据与模型开发cough dataset and modeling 10](#_Toc119066535)

[3.1 数据收集Data Collection 10](#_Toc119066536)

[3.2数据预处理Data pre-processing 10](#_Toc119066537)

[3.3模型架构Model structure 10](#_Toc119066538)

[3.4模型主干Algorithmic skeleton 11](#_Toc119066539)

[3.5快速集成方法Fast-ensemble method 12](#_Toc119066540)

[3.6模型训练与评估Model Training and Evaluation 12](#_Toc119066541)

[第四章总结与展望Conclusions and future works 16](#_Toc119066542)

[4.1 总结Conclusions 16](#_Toc119066543)

[4.2展望Future Works 16](#_Toc119066544)

[参考文献References 17](#_Toc119066545)

# **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，但它是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是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。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帮助个人进行早期连续监测、识别呼吸系统疾病可能的风险，从而实现早诊早筛。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，对于个人以及社会整体的卫生经济效益巨大。

2020年1月30日，国际卫生组织（WHO）正式宣布COVID-19 (新型冠状病毒感染导致的呼吸系统疾病) 成为全球全球卫生紧急事件。此后的2年多时间，COVID-19在全球肆虐，截至目前已造成超过6.2亿人感染，并导致600多万人死亡。在新冠疫情期间，我们团队支持了小区的多项抗疫志愿者活动、协助小区普及患者教育并由学习新冠肺炎相关知识进一步了解到整体急性与慢性呼吸道疾病对社会带来的负担。因此，团队在校成立了第一个专注于呼吸系统疾病防治与管理的志愿者俱乐部，目标是倡导科学防治以及促进健康公平可及。结合医学与科创，设计呼吸系统疾病防治的创新手段。

**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，但目前该领域仍属于探索当中。

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、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，因为使用者的依从性差，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、使用者的友善程度以及辅助决策的潜力。

脉搏血氧饱和度(SpO2)是评估呼吸系统疾病严重程度的有用诊断辅助手段[7]，血氧饱和度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]。从脉搏血氧饱和度仪获得的所有生命体征(脉搏率、血氧饱和度和呼吸率)都可以预测慢性阻塞性肺病急性加重事件，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、慢性阻塞性肺病、肺炎、慢性支气管炎、肺癌等等[11]。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[12]。研究显示，大多数健康相关的监控系统需要使用者的积极参与。因此，必须考虑使用者的易用性[13]，特别是老年人群的使用者，而研究发现他们甚至不激活基于手腕的监测设备[13]。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使用者友好的，或者采取一种以使用者被动的方式收集数据、不需要使用者太多操作技巧的可靠的客观监测方式[14]。

本研究利用便携式设备方便使用者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[15]。对于使用者来说，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，目的就是提醒使用者呼吸系统疾病的可能风险，提醒使用者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边缘计算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：一套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使用者关键信息，包括使用者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[16])、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，使用者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。

本研究主要技术创新点：

(1)连续监测血氧饱和度、咳嗽音等多维度数据，配合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，且模型可在微处理器设备上进行部署；。

(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)一套基于边缘计算的便携集成设备使得连续监测血氧饱和度、咳嗽音成为可能，兼具使用者友好度和可用性两方面优点。

# **第二章硬件及功能设计HARDWARE AND FUNCTIONAL DESIGN**

考虑到使用者的易用性，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。

Considering the ease of use for users with potential respiratory diseases at home, it is necessary to collect and extract blood oxygen saturation and cough sound data in real time through a set of portable and compact hardware devices, and to analyze the cough sound by a cough sound respiratory disease recognition model based on deep learning technology. Combined with the user's medical history, it can identify and predict the risk of respiratory diseases.

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；

(4)设备端的数据边缘计算，提升核心处理效能。

Data acquisition hardware device design objectives:

(1) Portability, ease of assembly, and compactness.

(2) Fine data collection and high quality requirements;

(3) Ease of use of integrated data for blood oxygen saturation and cough sounds.

(4) Data edge computing on the device side to improve core processing performance.

经过对从硬件原型快速搭建，以及对硬件结构的反复迭代。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。

After rapid prototyping and repeated iterations of the hardware structure, we decided to use embedded hardware frameworks with a Raspberry Pi as the backbone in order to achieve rapid hardware construction and function implementation.

**2.1 功能模块的选择SELECTION OF FUNCTIONAL MODULES**

**2.1.1血氧数据采集模块Blood Oxygen Module**

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、以及医疗用指尖或耳垂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。

We investigated and compared a variety of oximetry application models, which were divided into three main categories: traditional oximetry sensor modules, oximetry on smart wearable devices, and medical fingertips or ear-clip oximetry devices. Because oximetry on smart wearable devices and the fingertips or ear-clip medical oximetry devices are closed devices, real-time oxygen saturation data for secondary development cannot be obtained. Therefore, we decided to use a separate blood oxygen sensor module with a simple and easy-to-use development board.

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。

We chose the M5Stack Stick C+ development board with the M5Stack MAX30100 Oximetry Heart Rate module as the oximetry data acquisition module. The main feature of the M5Stack product is the ability to quickly build components in a plug-and-play manner, just like Lego, and it is more product-oriented than the simple functional verification of traditional development boards.

M5Stack Max30100血氧传感器如图1所示。

Figure 1: M5Stack Max30100 pulse oximetry and heart-rate sensor

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。

Current non-invasive pulse oximetry technologies are divided into transmissive dual-wavelength pulse oximetry and reflective dual-wavelength pulse oximetry, with the latter having a strong light spot and being easy to observe. Since our system requires measurements to be taken by clipping on the finger of the measurer, we chose a reflective oxygen sensor.

The M5Stack MAX30100 is a module with an integrated pulse oximeter and heart rate monitoring sensor. The device has two LEDs emitting infrared light and a photodetector to detect the reflected light, which measures the amount of infrared light absorbed by oxyhemoglobin (HBO2) and hemoglobin (HB) to detect blood oxygen saturation. The MAX30100's optimized optics and low-noise analog signal processor make it an ideal device for measuring blood oxygen saturation and collecting data.

M5Stack StickC+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。

The M5Stack StickC+ is a mini IoT development board with an integrated ESP32 chip, Wi-Fi capability to quickly build functional prototypes and simplify the development process. With rich open source code and an active forum community, it provides a wealth of open source resources to accelerate research and development.

M5Stack StickC+开发板如图2所示。

Figure 2: M5Stack StickC+ Development Board

图1：M5Stack MAX30100 血氧传感器Figure 1: M5Stack Max30100 pulse oximetry and heart-rate sensor 图2： M5Stack StickC+ 开发板 Figure 2: M5Stack StickC+ Development Board

**2.1.2咳嗽音数据采集模块Cough sound data acquisition module**

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。

The microphone is used to record the sound of coughing. Since ordinary microphones do not have processing chips, and our application is an intelligent speech recognition scenario that requires high voice quality, we decided to use microphone arrays. The microphone array is mainly composed of a group of acoustic sensors arranged according to a certain geometric structure (commonly linear, ring-shaped). It carries out space-time processing of the collected sound signals in different spatial directions to achieve noise suppression, reverberation removal, human voice interference suppression, sound source direction finding, sound source tracking, array gain and other functions, and then improves the quality of speech signal processing to improve the speech recognition rate in real environments.

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。

We chose the Raspberry Pi-based ReSpeaker 4-Mic array, a four-channel microphone expansion board for Raspberry Pi for AI and speech applications, with four digital microphones, support for on-chip speech algorithms, and a sensitivity of -26 dBFS (omnidirectional, supports far-field speech capture, Able to detect sounds up to 5 meters away, even in the presence of background noise) It is used as a cough sound data acquisition interface in the system of this project.

ReSpeaker 4-Mic麦克风阵列如图3所示。

Figure 3: ReSpeaker 4-Mic Array

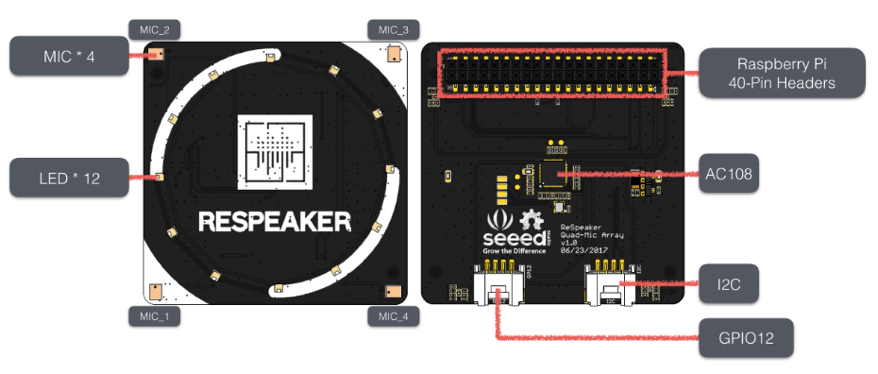


图3：ReSpeaker 4-Mic 麦克风阵列 Figure 3: ReSpeaker 4-Mic Array

**2.1.3主工作模块 Main module**

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。

We chose the Raspberry Pi 4B as the core working module in our embedded hardware framework. It has a 64-bit quad-core processor running at 1.5Ghz, up to 4GB RAM (1GB, 2GB, 4GB, 8GB depending on the model), 2.4/5.0 Ghz dual-band wireless LAN, Bluetooth 5.0/BLE, Gigabit Ethernet, USB3.0 and PoE capabilities. Its excellent quad-core performance can meet the parallel processing needs of data collection and processing, model computing, and web backend support.

Raspberry Pi 4B主工作模块如图4所示。

Figure 4: Raspberry Pi 4B core working module

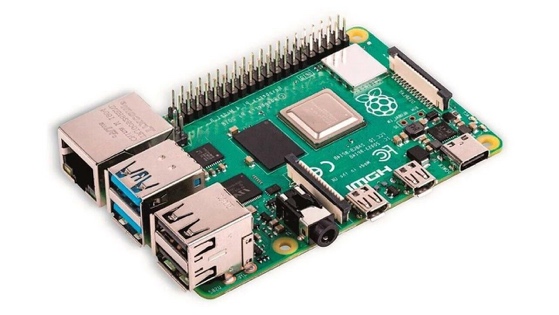


图4：Raspberry Pi 4B主工作模块Figure 4: Raspberry Pi 4B core working module

**2.2硬件原型搭建HARDWARE PROTOTYPING**

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结构。 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 StickC+模块上。

A fast and compact hardware architecture is constructed by integrating the above sensors, evaluation boards, and sound data acquisition devices. The M5Stack Stick C+ and MAX30100 sensor module form the oximetry data acquisition module. The MAX30100 transmits oximetry data to the M5Stack Stick C+ module in real time via the I2C bus.

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 Stick C+ 通过WIFI发送血氧饱和度采集数据到Raspberry Pi 4B 上。

The ReSpeaker 4-Mic array is directly connected to the main processing module, Raspberry Pi 4B, via a cable (I2C) to record cough sound data. Meanwhile, the M5Stack Stick C+ sends oxygen saturation data to the Raspberry Pi 4B via WIFI.

为了产品的易用性、收音清晰度并且兼顾散热功能，我们参考了水果常见的包装篮设计，设计了一款3D打印的外盒来装置传感器、评估板和声音数据采集设备，为了后续维修方便也设计了易于开启的卡扣。

Considering ease of use, radio clarity and heat dissipation, we referenced a supermarket basket and designed a 3D-printed outer box to house the sensors, evaluation board and sound data acquisition device. We designed an easy-open clip for subsequent maintenance.

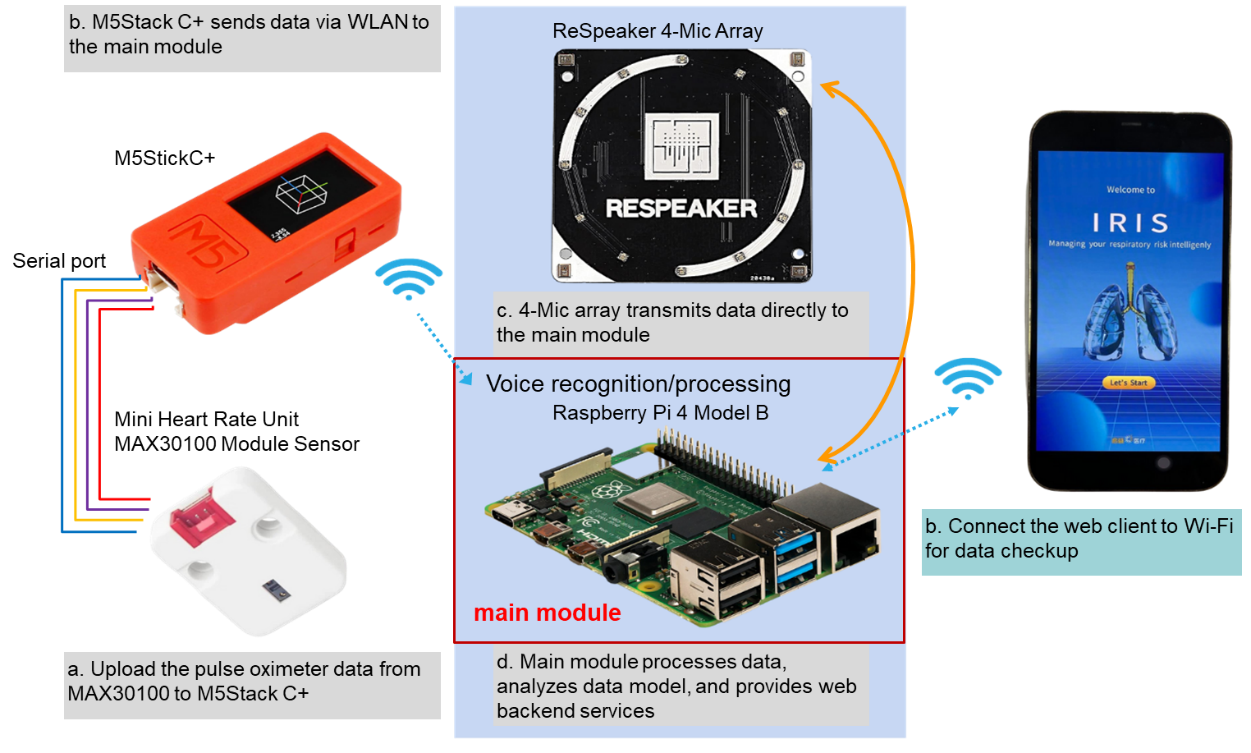


图5：硬件实现模型 - 概念图Figure 5: Hardware implementation model-Concept diagram

图片包含 游戏机

描述已自动生成图片包含 室内, 小, 桌子, 游戏机

描述已自动生成

图6：硬件实现模型 - 实际图 Figure 6: Hardware implementation model - actual view



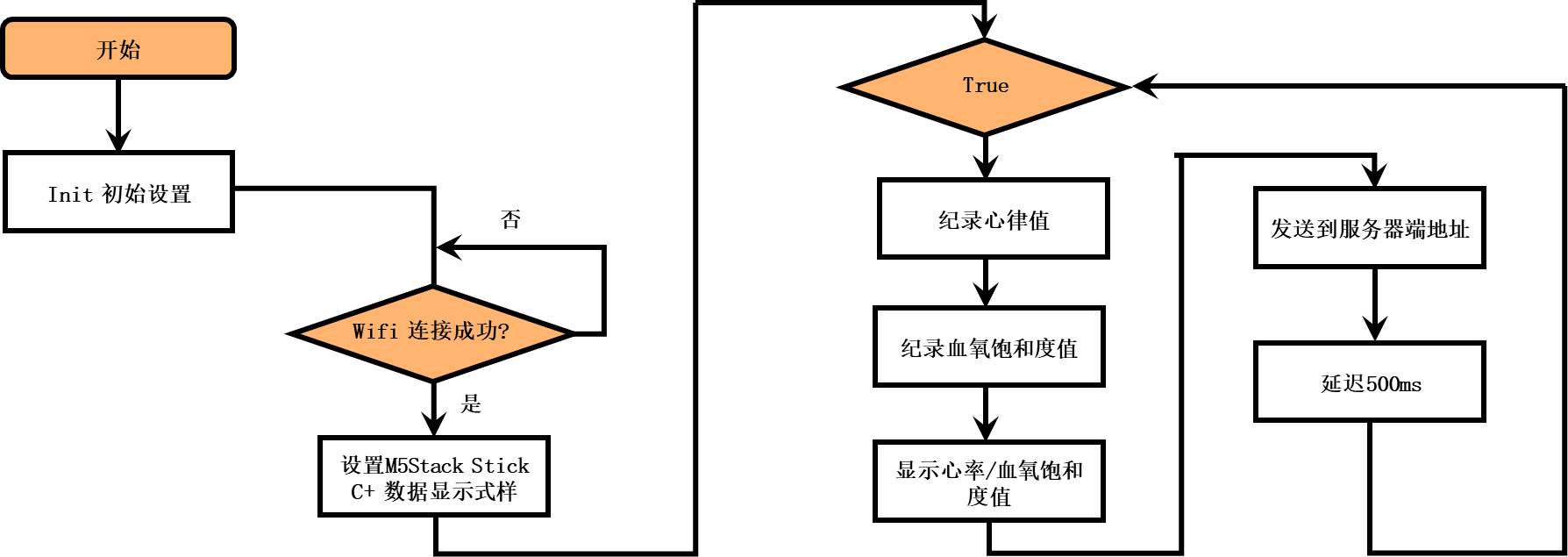
图7：设备3D外盒 - 实际图Figure 7: 3D outer box of the device - actual view

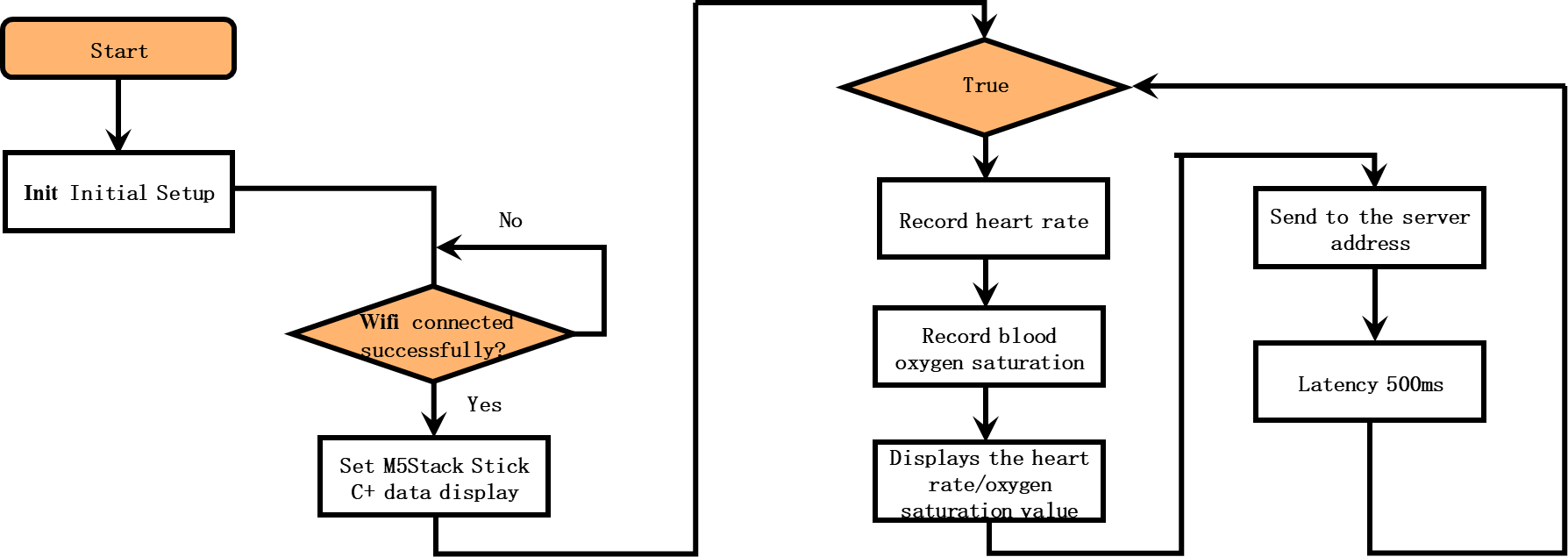
**2.3采集功能的实现REALIZATION OF DATA COLLECTION**

**2.3.1 血氧采集模块功能实现Blood oxygen acquisition module function implementation**

通过M5Stack MicroPython 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。

Rapid implementation of timed oximetry data acquisition via M5Stack MicroPython with 500ms acquisition interval and data upload to the server port via WIFI.

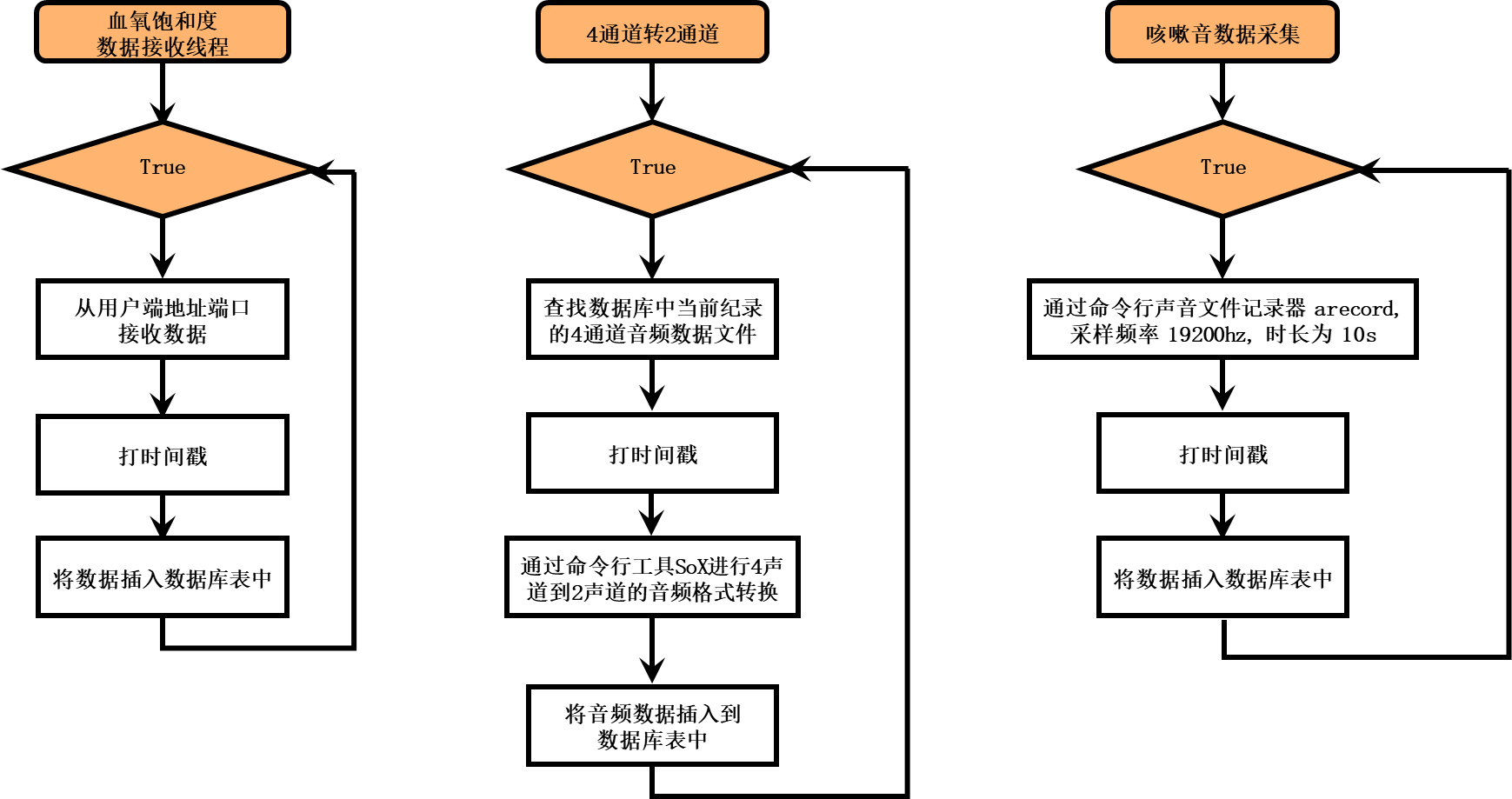


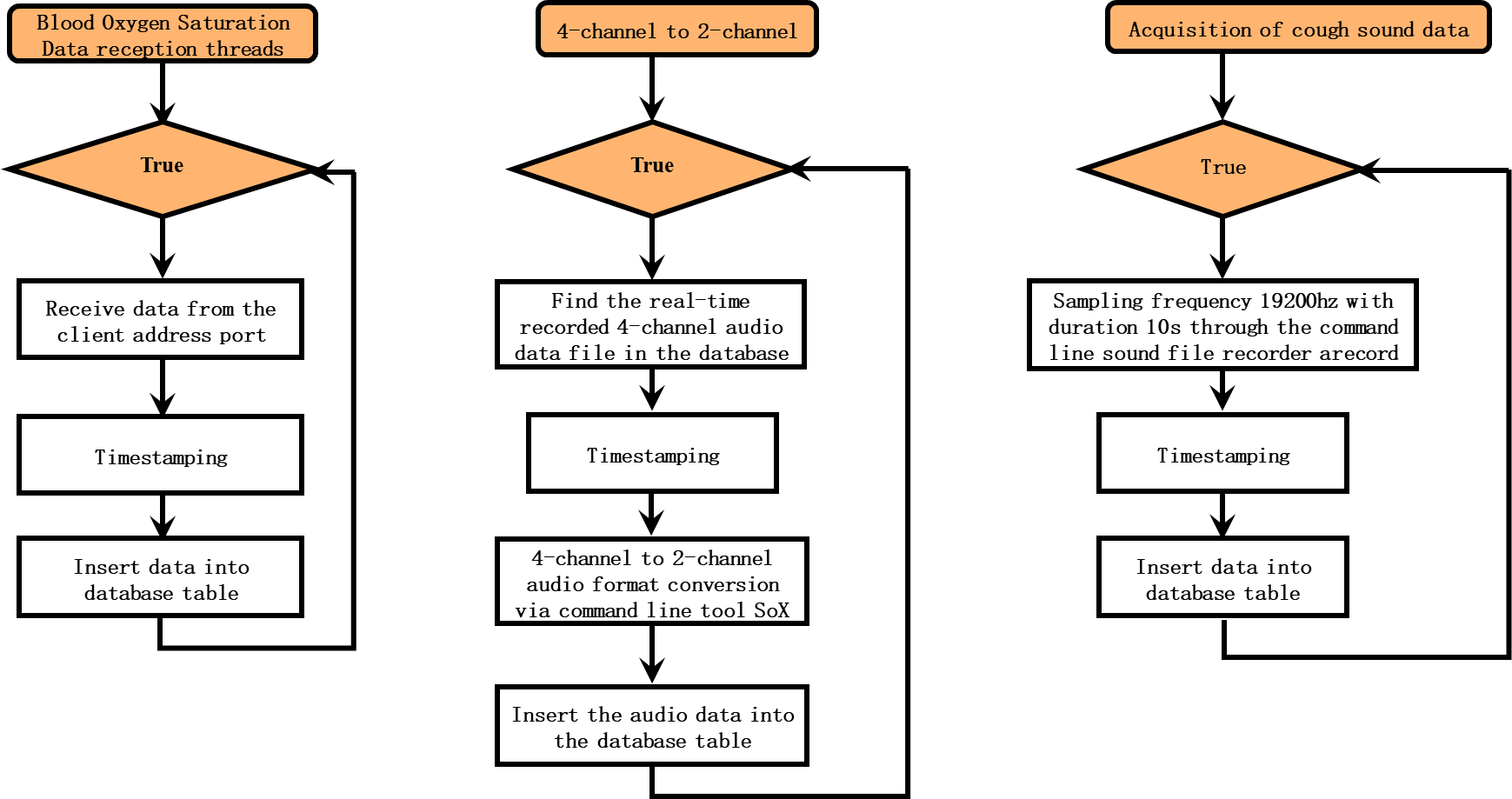


**2.3.2咳嗽音采集模块功能实现Cough sound acquisition module function implementation**

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：

The data collected by the ReSpeaker 4-Mic microphone array is connected to the main processing board of the RaspBerry Pi via a 40-pin head cable, which processes sound acquisition, 4-channel to 2-channel, and oximetry data reception in parallel through three threads. The oximetry data and the cough sound data are recorded and inserted into the mongo database list by time tag for later call by the risk identification module. The thread function is described as follows.





**2.4网页用户端的服务实现CLIENT-SIDE WEB APPLICATION**

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用户端应用框架。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。

We need to establish a user-end application framework for respiratory disease risk identification based on web pages. Due to our main hardware function module, we use Python to implement the function on the Raspberry Pi. Django is one of the most mature web frameworks under the Python system, and it has become the first choice for small and medium-sized web development frameworks because of its ability to develop web applications quickly. The mature open-source community resources of Django make it an attractive backend web application framework.

在用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。

For the design of the user-side application pages, we used a completely separate framework structure for the front-end and back-end. We used Vue.js as the front-end development framework instead of Django's weak templating engine. Vue.js is an excellent JavaScript library for front-end interface development with a focus on simplifying the implementation of the view layer, making it the leading front-end page development tool. Django, on the other hand, provides a server-side api, with a complete separation of front-end and back-end implementation, which is more suitable for the development and construction of single-page applications.

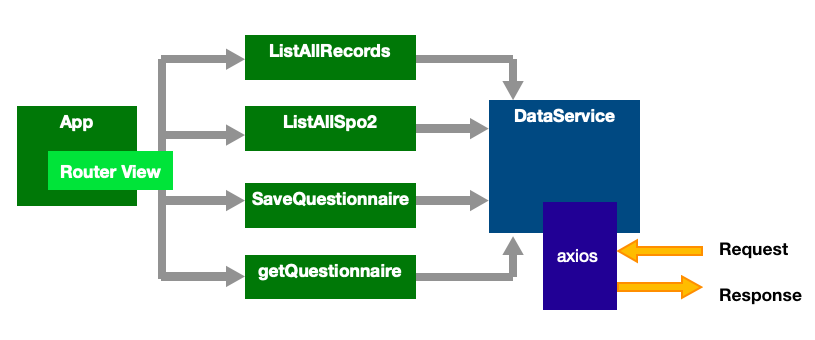


图7：网页Django-Vue结构图Figure 7: Django-Vue structure



图8：网页用户端页面Figure 8: Client side web page

# **第三章咳嗽音信号数据与模型开发COUGH DATASET AND MODELING**

**3.1 数据收集DATA COLLECTION**

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，同时也收取是否有呼吸道症状。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，我们将具有症状的样本作为阳性样本，其他样本作为阴性样本。

In this study, we used the open-source and anonymized dataset Covid19-Cough [1] to train the cough sound classification model. Using a Telegram messaging bot and call center, 1,324 subjects were asked whether they were COVID-19 positive and if they had respiratory symptoms. Of this cohort, 682 subjects were reported positive, of which 382 had PCR test results and 295 were symptomatic. A total of 58 minutes of cough sound recordings were collected in this dataset, with an average duration of 2.6 s per recording sample. Because our goal is to determine the user's lung health based on the cough sound, we treat the symptomatic samples as positive and the other samples as negative

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[17]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。

Meanwhile, since our application may pick up various sound disturbances during practical use, we need to use a model to detect whether a recording has a cough sound or not. Since all samples in the Covid19-Cough dataset are with cough sounds, we use the URBANSOUND8K dataset [17] as samples without cough sounds, that is, negative samples. The dataset contains 8732 annotated audio clips of various sounds that may appear in the city, such as air conditioning, car horns, music, construction sounds, children playing, etc. We combine the Covid19-Cough Dataset with the URBANSOUND8K dataset to create a new dataset called the Cough classification Dataset 1 (CoughCLS-1).

**3.2数据预处理DATA PRE-PROCESSING**

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)。

When a recording is received by our device, we first load and resample the recording file into a 32kHz waveform (numpy format), which is padded to a numpy array of 15 \* 32000. The waveform data were subjected to Short-Time Fourier Transform (STFT) using the torchaudio toolkit to obtain the spectral representation (window size: 2048; stride: 302), which was then projected onto a 128 MEL filter bank from 20 Hz to 24 kHz to obtain a MEL spectrogram. During training, we add noise (implemented by torchaudio) to the audio of training samples for data augmentation.

**3.3模型架构MODEL STRUCTURE**

我们主要采用如下图所示的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)经过幂平均池化，转化为特征向量；(d)特征向量经过一个两层的全连接层输出标签预测。我们的模型采用深度学习框架PyTorch进行开发。

We mainly use the model framework presented in the figure below for cough sound classification. Our model framework has the following four steps: (a) Firstly, the waveform data is transformed into a MEL spectrum; (b) The MEL spectrum is used as a single channel image into a pre-trained convolutional neural network for feature extraction; (c) The feature map output from the convolutional network is pooled by generalized averaging and transformed into feature vectors; (d) The feature vectors will go through a two-layer fully connected layer to produce a probability prediction (i.e., logits) for each class. Our model was developed using the deep learning framework PyTorch.

在本研究中，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[18]，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 [19]作为分类器。

In this study, we also compared the feature extraction and machine learning classification method widely applied in the field of audio analysis [18]. The specific steps are as follows: (a) Firstly, feature extraction is performed on the waveform data, and finally each audio is represented by a feature vector; (b) The feature vectors are passed through a machine learning model to obtain the classification result. We employ widely used audio features: (1) mean and variance of Mel-Frequency Cepstral Coefficients (MFCC); (2) spectral centroid; (3) zero-crossing rate; (4) chrominance feature vector; (5) spectrum cutoff feature. We use LightGBM [19], one of the most powerful machine learning model, as a classifier.

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。

In contrast to deep learning-based models, feature extraction-based machine learning models rely on the accumulated experience of audio analysis literature to extract various conventional features. Prior to the advent of deep learning, this type of approach was common. Additionally, it has high inference efficiency. While the feature extraction process is focused on mining the overall statistical characteristics of the waveform, it will lose some fine-grained features, and it also requires a lot of engineering to identify which features are effective for specific tasks. Deep convolution networks, on the other hand, achieve superior performance on audio tasks due to their ability to automatically extract features.

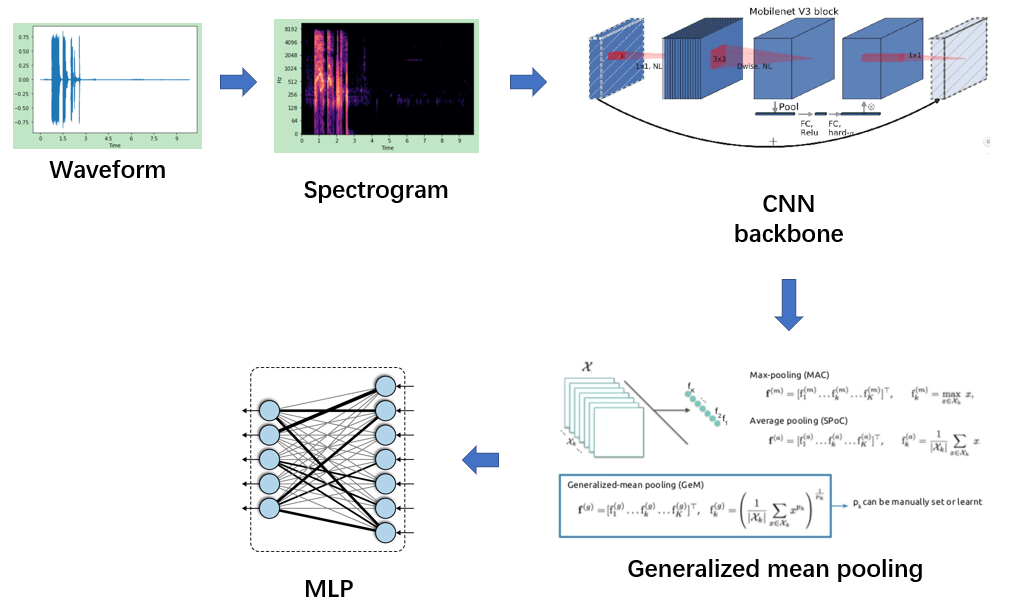
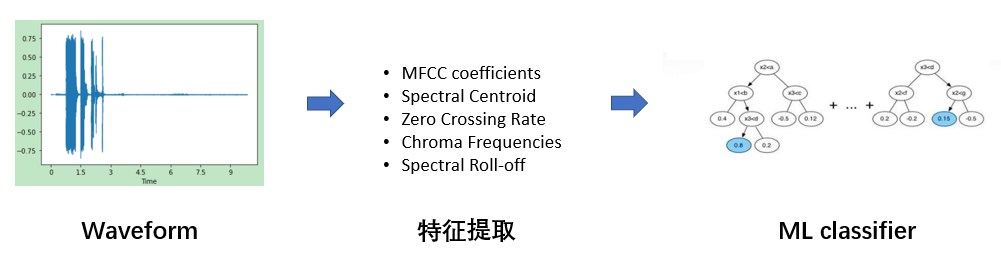


图9. 基于谱图与卷积神经网络的咳嗽音分类模型框架Figure 9. framework of cough sound classification model based on spectrogram and convolutional neural network



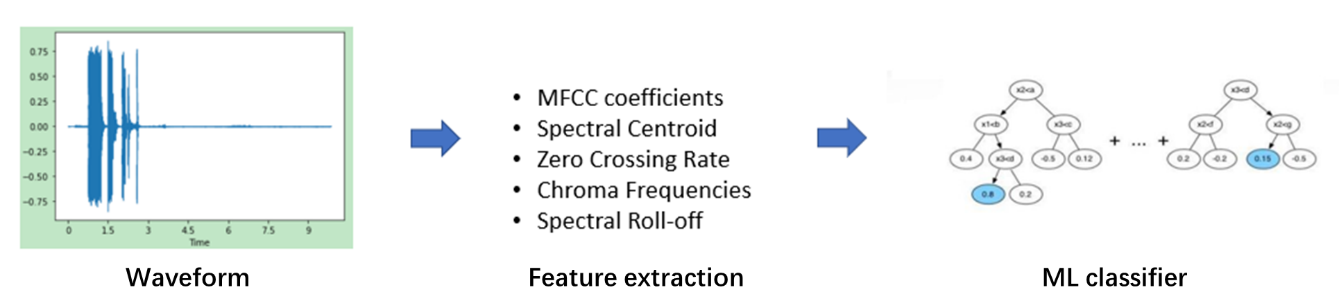


图10. 基于音频特征提取的机器学习模型

Figure 10. Machine learning model based on audio feature extraction

**3.4模型主干ALGORITHMIC SKELETON**

在本研究中，我们采用卷积神经网络[20]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进行不同的卷积操作得到。。卷积操作具有有效的特征提取能力，其组成的卷积网络可以在不同的深度抽取图像的各个不同粒度的特征。卷积神经网络在大型图像处理任务上获得了巨大成功[21]，并引领了第二次人工智能浪潮。

In this study, we employed a convolutional neural network [20] to extract features from the MEL spectrogram. The backbone of a convolutional network consists of multiple convolutional layers, activation functions, and pooling layers. Convolutional layers can produce a set of parallel feature maps by using different convolutional kernels. Thus, it can extract meaningful features for the task at hand. Convolutional layers at different depth can extract features of different gradularity. Convolutional neural networks have been highly successful at large image processing tasks [21] and have led the second wave of artificial intelligence.

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[22], EfficientNet[23], ResNet[24], ResNeXt[25]. 其中ResNet[26]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和EfficientNet在ResNet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。

This study uses the paradigm of pre-training and fine-tuning, which is the most efficient so far. The pre-trained network has learned a large amount of semantic knowledge embedded in the images and has robust parameter initialization, which facilitates fast optimization of the whole model for downstream tasks. We employ a convolutional neural network pre-trained on ImageNet as the backbone. The most popular network backbones are MobileNet [22], EfficientNet [23], ResNet [24], ResNeXt [25]. Among them, ResNet [26] is the most widely used network structure. Its proposed residual link is conducive to deeper neural network training without gradient disappearance, and is currently an essential module of various neural networks. Based on ResNet, MobileNet and EfficientNet adopt a more lightweight network design, which is more suitable for deployment on the edge devices. Considering that our application needs to be deployed on mobile devices, we mainly use MobileNet as a lightweight network structure. Finally, we adopt the pre-trained MobileNet model weights provided by the Pytorch-Image-Models framework as our convolution backbone.

**3.5快速集成方法FAST-ENSEMBLE METHOD**

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[26]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）。

Normally, we use a multi-layer fully connected layer to predict the classification result after using the convolutional network as the backbone for feature extraction. Model ensemble [26] has been heavily employed in numerous studies. However, general model ensemble requires the use of several different models for classification. This not only consumes a large amount of GPU resources during training, but also significantly slows down the whole application during inference after model deployment. In this study, we develop a fast and efficient model integration method, which we call the Fast-ensemble method (FEM).

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。

The fast ensemble method (as shown in Figure 11) is implemented as follows: we connect K different fully connected layers as classification heads after a convolutional network backbone. During training, each classification head receives different feature vectors due to the randomness of dropout, so that the classification heads can learn different parameters after training. This is equivalent to understanding the feature vectors of the image from various perspectives and obtaining multiple distinct logit results. The logits output by the model are the result of averaging the logits of these K classification heads. In our experiments, we found that K=5 can effectively improve the performance of the model without excessively reducing inference efficiency.

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。

The fast ensemble method only adds several classification layers to the backbone of the model. As compared to the model backbone, the model adds only 5k additional parameters, which does not significantly impact the memory footprint. In addition, the classification layer doesn't add much time to inference due to its simple structure.

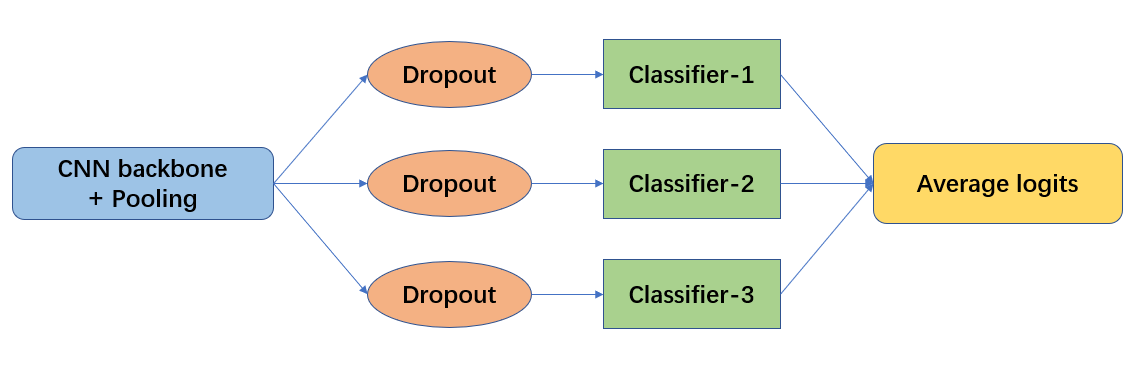


图11. 快速集成方法示意图.

Figure 11. Schematic diagram of the fast ensemble method.

**3.6模型训练与评估 MODEL TRAINING AND EVALUATION**

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。

For training, we use a batch size of 16, a learning rate of 2e-5 for the pre-trained convolutional backbone, a learning rate of 2e-4 for the randomly initialized classification head, and a maximum number of training epochs of 30.

**3.6.1咳嗽检测模型Cough detection model**

我们采用组合后的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，我们还报告并列出了以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。

We used the combined CoughCLS-1 dataset to perform 5-fold cross validation for model evaluation, and the main metric evaluated was the 5-fold cross validation average test AUC score. At the same time, we also report and list the average true positive rate and the average true negative rate with a threshold of 0.5. The performance comparison of the models is shown in Table 1 below. We can see that with convolutional neural networks, we can detect whether an audio contains cough sounds ornot with high accuracy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 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 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Model** | **AUC(%)** | **TPR(%)** | **TNR(%)** |
| **Machine learning** | LightGBM | 78.5 | 73.2 | 80.1 |
| **Deep learning** | 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

Table 1: Performance of various models with five-fold cross-validation on the CoughCLS-1 dataset

**3.6.2基于咳嗽音的肺部健康检测模型Lung health detection Model based on cough sounds**

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，我们还报告并列出了以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”。

We used 5-fold cross-validation on the Covid19-Cough dataset for model evaluation, and the main metric evaluated was the 5-fold cross-validation average test AUC score. Also, we report and list the average true positive and the average true negative with a 5-fold threshold of 0.5. The performance of the models is compared in Table 2 below. In this study, we used the pre-trained convolutional network by default. We also compare the performance of completely randomly initialized networks, which are marked as "not pre-trained".

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 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 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Model** | **AUC(%)** | **TPR(%)** | **TNR(%)** |
| **Machine learning** | LightGBM | 65.3 | 56.7 | 61.1 |
| **Deep learning** | 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 |

Table 2: Performance of different models with five-fold cross-validation on the Covid19-Coug dataset

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[27]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。

In the cough sound classification task, we can see that a machine learning model based on feature extraction performs significantly worse than a model based on deep convolutional networks. This demonstrates the expressive power and feature extraction abilities of deep convolutional networks. After using our Fast Ensemble method, the model achieved an AUC improvement of 1.4%, which is significant after the Wilcoxon rank test [27]. At the same time, we found that the pre-training then fine-tuning paradigm is very beneficial for cough sound classification. The performance of MobileNet-v3 without pre-training is very poor, which indicates that the network has learned the ability to extract image features in the pre-training stage, which is conducive to the optimization of downstream tasks.

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，以及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。

We compare the convergence of the training loss function and the generalization on the test set for MobileNet-v3 (not pretrained), MobileNet-v3, and MobileNet-v3 + Fast Ensemble in Figures 12. We can see that the randomly initialized MobileNet-v3 model has been in oscillation after the initial decrease in training loss and is no longer optimized, and its AUC score does not reach a relatively high level. By adding the Fast Ensemble mechanism, the MobileNet-v3 model can obtain smoother loss convergence and eventually higher generalization performance.

|  |  |
| --- | --- |
| (a-1) MobileNet-v3 + Fast Ensemble模型的损失函数收敛情况  (a-1) Convergence of loss function for MobileNet-v3 + Fast Ensemble model | (b-1)MobileNet-v3 + Fast Ensemble模型的AUC分数收敛情况  (b-1) Convergence of AUC scores for MobileNet-v3 + Fast Ensemble model |
| (a-2) MobileNet-v3模型的损失函数收敛情况  (a-2) Convergence of loss function for MobileNet-v3 model | (b-2) MobileNet-v3模型的AUC分数收敛情况  (b-2) Convergence of AUC scores for MobileNet-v3 models |
| (a-3) MobileNet-v3 (not pretrained)模型的损失函数收敛情况  (a-3) Loss function convergence for MobileNet-v3 (not pretrained) models | (b-3) MobileNet-v3 (not pretrained)模型的AUC分数收敛情况  (b-3) AUC score convergence for MobileNet-v3 (not pretrained) models |

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

Figure 12. Training set loss and test AUC performance of each convolutional network on the Covid19-Cough dataset as training advances

# **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，为普通人在居家场景下，提供了一种容易获得的呼吸系统疾病的风险评估手段。同时，也希望对社区或者基层医疗中心，提供一套呼吸系统疾病早筛手段。由于本项目采用的开源且匿名化后的数据集Covid19-Cough [1] 训练咳嗽音分类模型。该数据集是以国外患者数据为主，数据集中数据量有限，分类颗粒度不够高。那么，通过本项目的设备，也为建立一套以我国患者为主的呼吸系统病征生理指标数据集，提供了可靠的解决方案。

本研究过程中，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，由于在项目实践中，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，最终经过分析和比较硬件模型结构，决定将两个模块分开，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较为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，使得使用者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%，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，帮助使用者及时判定风险，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(2021). Dataset of recordings of induced cough.Medical Investment Solutions.(St Petersburg, Russia) license CC0 1.0 Universal. Available: https://github.com/covid19-cough/dataset
2. GBD Chronic Respiratory Disease Collaborators (2020). Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet. Respiratory medicine, 8(6), 585–596. https://doi.org/10.1016/S2213-2600(20)30105-3
3. Huang, K., Yang, T., Xu, J., Yang, L., Zhao, J., Zhang, X., Bai, C., Kang, J., Ran, P., Shen, H., Wen, F., Chen, Y., Sun, T., Shan, G., Lin, Y., Xu, G., Wu, S., Wang, C., Wang, R., Shi, Z., … China Pulmonary Health (CPH) Study Group (2019). Prevalence, risk factors, and management of asthma in China: a national cross-sectional study. Lancet (London, England), 394(10196), 407–418. https://doi.org/10.1016/S0140-6736(19)31147-X
4. Wang, C., Xu, J., Yang, L., Xu, Y., Zhang, X., Bai, C., Kang, J., Ran, P., Shen, H., Wen, F., Huang, K., Yao, W., Sun, T., Shan, G., Yang, T., Lin, Y., Wu, S., Zhu, J., Wang, R., Shi, Z., … China Pulmonary Health Study Group (2018). Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet (London, England), 391(10131), 1706–1717. https://doi.org/10.1016/S0140-6736(18)30841-9
5. Swanney, M. P., O'Dea, C. A., Ingram, E. R., Rodwell, L. T., Borg, B. M., & ANZSRS Spirometry Training Course Working Group (2017). Spirometry training courses: Content, delivery and assessment - a position statement from the Australian and New Zealand Society of Respiratory Science. Respirology (Carlton, Vic.), 22(7), 1430–1435. https://doi.org/10.1111/resp.13133
6. Eaton, T., Withy, S., Garrett, J. E., Mercer, J., Whitlock, R. M., & Rea, H. H. (1999). Spirometry in primary care practice: the importance of quality assurance and the impact of spirometry workshops. Chest, 116(2), 416–423. https://doi.org/10.1378/chest.116.2.416
7. Dalbak, L. G., Straand, J., &Melbye, H. (2015). Should pulse oximetry be included in GPs' assessment of patients with obstructive lung disease?. Scandinavian journal of primary health care, 33(4), 305–310. https://doi.org/10.3109/02813432.2015.1117283
8. Caputo, N. D., Strayer, R. J., & Levitan, R. (2020). Early Self-Proning in Awake, Non-intubated Patients in the Emergency Department: A Single ED's Experience During the COVID-19 Pandemic. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine, 27(5), 375–378. https://doi.org/10.1111/acem.13994
9. Clarke, M., Gokalp, H., Fursse, J., & Jones, R. W. (2016). Dynamic Threshold Analysis of Daily Oxygen Saturation for Improved Management of COPD Patients. IEEE journal of biomedical and health informatics, 20(5), 1352–1360. https://doi.org/10.1109/JBHI.2015.2464275
10. Shah, S. A., Velardo, C., Farmer, A., &Tarassenko, L. (2017). Exacerbations in Chronic Obstructive Pulmonary Disease: Identification and Prediction Using a Digital Health System. Journal of medical Internet research, 19(3), e69. https://doi.org/10.2196/jmir.7207
11. Chung, K. F., &Pavord, I. D. (2008). Prevalence, pathogenesis, and causes of chronic cough. Lancet (London, England), 371(9621), 1364–1374. https://doi.org/10.1016/S0140-6736(08)60595-4
12. Riis, H. C., Jensen, M. H., Cichosz, S. L., & Hejlesen, O. K. (2016). Prediction of exacerbation onset in chronic obstructive pulmonary disease patients. Journal of medical engineering & technology, 40(1), 1–7. https://doi.org/10.3109/03091902.2015.1105317
13. Galarnyk, M., Quer, G., McLaughlin, K., Ariniello, L., &Steinhubl, S. R. (2019). Usability of a Wrist-Worn Smartwatch in a Direct-to-Participant Randomized Pragmatic Clinical Trial. Digital biomarkers, 3(3), 176–184. https://doi.org/10.1159/000504838
14. Crooks, M. G., den Brinker, A., Hayman, Y., Williamson, J. D., Innes, A., Wright, C. E., Hill, P., &Morice, A. H. (2017). Continuous Cough Monitoring Using Ambient Sound Recording During Convalescence from a COPD Exacerbation. Lung, 195(3), 289–294. https://doi.org/10.1007/s00408-017-9996-2
15. Eysenbach G. (2000). Consumer health informatics. BMJ (Clinical research ed.), 320(7251), 1713–1716. https://doi.org/10.1136/bmj.320.7251.1713
16. Zhou, J., Yu, N., Li, X., & Wang, W. (2022). Accuracy of Six Chronic Obstructive Pulmonary Disease Screening Questionnaires in the Chinese Population. International journal of chronic obstructive pulmonary disease, 17, 317–327. https://doi.org/10.2147/COPD.S341648
17. Salamon, J., Jacoby, C., & Bello, J. P. (2014, November). A dataset and taxonomy for urban sound research. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 1041-1044).Available: https://urbansounddataset.weebly.com/urbansound8k.html
18. 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(2), 175–187. https://doi.org/10.1109/JSTSP.2022.3142514
19. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.
20. Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems.
21. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.
22. Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1314-1324
23. Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.
24. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
25. Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500).
26. Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. A survey on ensemble learning, Frontiers of Computer Science 14 (2)(2019) 241–258. doi: 10.1007/s11704-019-8208-z. URL https://doi. org/10.1007/s11704-019-8208-z.
27. Wilcoxon, F. (1946). Individual comparisons of grouped data by ranking methods. Journal of economic entomology, 39(2), 269-270.