

AI 赋能下的医学健康监测系统研究

黄伯韬 徐化坤 方童岩
指导教师：龚小谨

摘要：本项目旨在基于深度学习和预训练模型等相关技术，借鉴传统医学的脉诊与面诊过程，开发一款快速、准确、实时的心电信号与面部特征监测系统，以实现对人们身体和心理健康状况的实时监测和分析。我们使用修改后的基于 Transformer-Encoder 的深度学习模型提高了心电信号识别准确率，并使用基于 TransFER 架构的代码提取面部特征。我们还在本项目中引入了 CLIP 模型以增强特征提取能力。通过引入嵌入式开发板 RK3399，我们可以实现更低成本与更加便捷的心电信号的监测和处理。我们正在基于这一项目申请专利，并计划在学院团委指导下参加创新创业比赛，以推广该系统在医疗健康领域的积极影响，为人们提供更加便捷、准确、高效的监测服务。

关键词：深度学习；实时监测；心电图；面部特征；中医药

Medical Health Monitoring System Empowered by Artificial Intelligence

HUANG Botao, XYU Huakun, FANG Tongyan
Advisor: GONG Xiaojin

Abstract: This project aims to develop a fast, accurate, real-time system for monitoring and analyzing electrocardiogram (ECG) signals and facial features using deep learning and pre-trained models. We improved the accuracy of ECG signal recognition using a modified ResNet model, achieving higher efficiency. Facial features were extracted using a transformer-based code for expression recognition. We also explored the use of the CLIP model to enhance the feature extraction capabilities. An embedded development board was introduced to enable real-time ECG signal monitoring and processing. This system could significantly reduce operating costs and improve convenience. We plan to apply for a patent and participate in innovation and entrepreneurship competitions to promote this system's impact on healthcare.

Keywords: deep learning; real-time monitoring; ECG; facial feature; TCM

1 前言与项目背景

1.1 整体背景

中医 (TCM) 是一个历史悠久、已经使用了数千年的整体医疗系统。中医强调个性化诊断和治疗，需要根据患者独特的体质、症状和病史进行针对性的医疗方案。

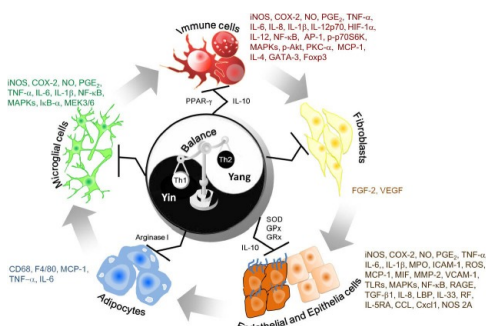


图 1 阴阳五行与中医诊断

中医诊断主要包括四种方法：望、闻、问、切。其中，触诊是最重要的诊断工具之一，包括脉诊和体格检查。脉诊是

中医诊断的核心之一，可以通过触诊脉搏来了解人体的脏腑功能、气血状况、阴阳平衡等情况。同时，体格检查也可以了解患者的体型、皮肤、舌苔等情况，为诊断提供更加全面的信息。

近年来，人工智能 (AI) 技术的发展为中医诊疗提供了新的机遇。利用机器学习、深度学习和自然语言处理等技术，可以对中医诊断的数据进行分析，挖掘潜在的规律和特征。例如，利用人工智能技术可以分析病人面部特征，辅助中医面诊；利用机器学习技术可以对中医脉诊数据进行分析，从而提高中医诊断的精确性和效率。

因此，结合中医传统的诊断方法和现代的人工智能技术，可以更加准确地诊断和治疗患者，为中医诊疗带来新的机遇和挑战。

1.2 脉诊在中医中的重要性

脉诊，又称脉诊或触诊，是中医的一项基本诊断技术。最早的脉诊记录可以追溯到《黄帝内经》，它是中医最古老的医学典籍之一。脉诊包括观察和解读脉搏质量，包括脉率、

强度、节奏和形状，以帮助确定患者的身体和情绪状态，以及他们的整体健康状况。中医的脉诊涉及对 12 个脉位的解读，这些脉位与不同的器官和身体系统相关。脉诊可以为中医诊断和治疗提供有价值的信息，包括识别人体气血失衡、检测是否存在致病因素以及确定最佳治疗方案。

1.3 中医面部诊断的重要性

面诊又称观诊，是中医的又一重要诊断方法。面部诊断包括观察患者的面部特征，例如颜色、纹理和形状，以帮助确定他们的整体健康状况。根据中医理论，面部的不同部位对应着不同的器官和身体系统。例如，鼻子与肺有关，眼睛与肝脏有关。面部特征的变化可以为中医诊断和治疗提供有价值的信息，包括识别人体气血失衡、检测是否存在致病因素以及确定最佳治疗方案。

1.4 AI 技术在中医诊断与医疗监测中的应用

人工智能技术的发展为中医诊疗提供了新的机遇。例如，OpenAI 开发的 CLIP 预训练模型已被证明可以有效地分析面部特征并为面部诊断提供有价值的信息。CLIP 可以分析患者的面部特征，协助中医师做出更准确的诊断。此外，人工智能技术可以帮助分析脉诊数据，识别人类医生难以发现的模式和特征，提高中医诊断的准确性和效率。

1.5 总结

综上所述，中医诊断包括视诊、听诊、问诊、触诊等多种方法，其中脉诊和面诊被认为是中医最重要的两种诊断工具。脉诊和面诊可以为中医诊疗提供有价值的信息，包括识别人体气血失衡、检测致病因素的存在以及确定最佳治疗方案。人工智能技术的发展，特别是人工智能在脉诊和面部诊断方面的应用，为中医诊疗提供了新的机遇和挑战。

2 项目过程与方法

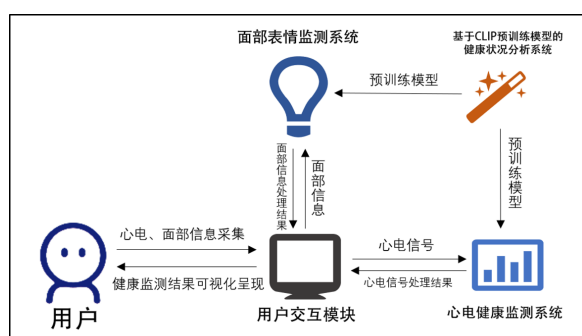


图 2 项目框架图

本项目基于深度学习以及预训练大模型等相关技术，旨在开发一款快速、准确、实时的心电信号与面部特征监测系统，通过对于中华民族传统医学的继承与创新，实现对于人们身体和心理健康状况的实时监测和分析。

2.1 深度学习模型开发与训练

首先，我们使用了基于 Transformer 模型的算法，对于心电信号和面部特征进行识别和提取。对于心电信号识别，我们使用了已有的单导联心电信号数据集 (PhysioNet/CinC Challenge 2017: Training Sets)，并在原有的 ResNet 模型上进行了改进。通过分析，我们发现原有的 ResNet 模型识别四种心电信号特征的准确率约为 87%，而我们通过修改后的模型提高了其识别准确率，实现了更高效的识别。对于面部特征提取，我们使用了基于 Transformer 模型的代码进行表情相关的面部特征提取。

2.2 新一代人工智能与预训练大模型

接着，我们研究了 CLIP 模型，思考是否可以将其用于提升心电与面部特征的提取能力。CLIP 是一个图片与文字的多模态模型。通过对于 CLIP 模型的学习，我们发现 CLIP 模型可以很好地处理图片与文字之间的对应关系，并通过一个 0-1 之间的数字来反映这二者之间的对应关系。我们认为，CLIP 模型也可以在面部状况的监测中发挥作用，例如输出表情、年龄、工作压力、或者直接的健康状况 (healthy) 等。

2.3 基于嵌入式开发板的系统开发

另外，我们改进了心电信号的输入方式。原来的单导联心电信号测量的过程中，我们要求被监测者使用 AD8232 录制一分钟的心电信号，并将音频信号以 wfdb 格式进行保存。但这个在实际应用中效果不会很好，因为不能实现实时监测。因此，我们还在项目中引入了嵌入式开发板 RK3399，通过 ADC 转换接口实现了实时心电信号的监测与处理，避免了格式转化等困扰，同时无需电脑也能进行心电监测，大大降低这一系统的运行成本。



图 3 嵌入式开发板实物图

2.4 专利申请与创新创业

最后，我们的项目计划在学院团委的指导下申请专利，并参加相应的创新创业比赛。我们相信这一项目的成果将会对医疗健康领域产生积极的影响，为人们的健康提供更加便捷、准确、高效的监测服务。

3 基于深度学习的模型开发

基于深度学习的模型开发流程可以大致分为以下几个步骤。首先是数据准备，也就是收集和整理数据，进行数据预处理、数据清洗和数据分割等工作，以便于模型训练和评估。其次是模型选择和设计，我们根据任务需求选择适合的模型架构，例如卷积神经网络（CNN）、循环神经网络（RNN）或 Transformer 等，然后设计模型结构，包括网络层数、激活函数、损失函数等。第三是模型训练。我们将准备好的数据输入模型中进行训练，调整模型参数使得损失函数最小化。常用的训练方法包括随机梯度下降（SGD）、自适应矩估计（Adam）等。第四是模型评估。我们使用测试数据集来评估模型的性能，通常使用准确率、精度、召回率和 F1 分数等指标来评估。第五是模型优化。我们对模型进行进一步优化，包括调整超参数、增加数据样本、数据增强、正则化等方法，以提高模型的性能。最后是部署应用：将训练好的模型部署到应用环境中，可以是服务器端、移动端或者嵌入式系统，以满足实际应用需求。在部署应用前需要进行模型压缩、加速、精度量化等处理。

3.1 基于深度学习的心电健康监测

使用 CINC2017 Challenge 作为训练数据集。在数据集中，输入一段一维的心电信号数据，并将其标注为四类：正常心电数据、异常心电数据、其它心电数据以及噪声数据。数据集的样例数据如下图所示。

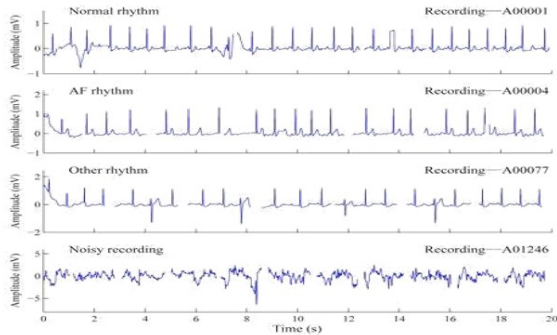


图 4 CINC17 示例数据

我们的深度学习网络模型基于 Kaiming He 在 2015 年提出的残差神经网络（ResNet）进行改进。ResNet 的核心思想是引入残差模块，通过跳过某些层，直接将一些输入信息传递到后面的层中，从而减轻了网络的训练难度和梯度消失的问题。这一改进极大地提高了深度神经网络的准确率。

针对本项目中的心电信号数据，我们引入了 Transformer-Encoder 结构，该结构在自然语言处理中已经取得了很好的效果。我们将心电信号数据转化为一维信号，并将其作为模型的输入。为了利用心电信号的频域特性，我们采用了傅里叶变换将时间域信号转化为频域信号。这样，我们可以更好

地提取信号的特征信息，并准确识别心电信号的异常。下图是我们所采用的网络模型架构图。

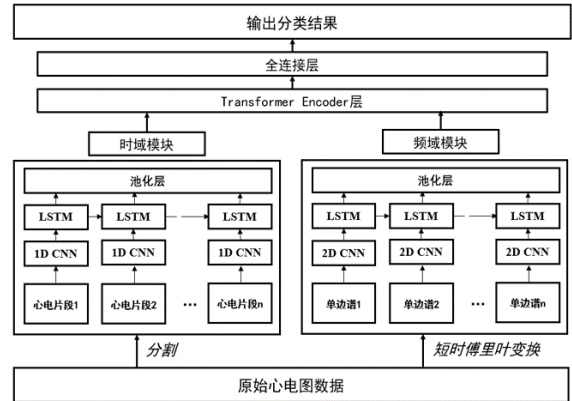


图 5 模型架构

在模型训练方面，我们使用了大量的心电信号数据进行模型的训练。最终，我们通过大量的实验，将模型的准确率提升到了 91%。

表 1 两种不同方式的测试集准确率

Model	ResNet	Transformer Based
Accuracy	0.871	0.913

3.2 基于深度学习的面部表情监测

3.2.1 面部表情数据集：RAF-DB

RAF-DB 数据集包含约 3 万张带有注释的面部图像，其中包括六种基本情感和一种中性表情。每张面部图像还有五个面部标志点、一个边界框以及种族、年龄范围和性别等属性的注释。本项目中使用了该数据集。

3.2.2 训练的模型：TransFER

TransFER 是一种已被发表的深度学习方法的架构，该方法用于表情分类等特定任务。该方法由三个主要组件组成：stem CNN、Local CNN 和 Multi-head Self-Attention Dropping。

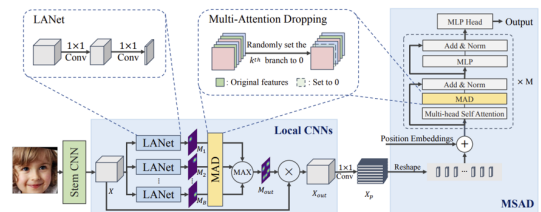


图 6 TransFER 模型架构

Stem CNN 是网络的初始部分，它的作用是从输入数据中提取特征。在这个方法中，TransFER 架构使用了一个已经训练好的卷积神经网络架构叫做 IR-50，它在各种任务和数据集上表现都很好。

Local CNN 用来从 Stem CNN 提取的特征图中提取更详细的信息。它们被称为“local”，因为它们只关注特征图的小区域，而不是整个特征图。这种方法可以帮助我们捕捉更细节的信息，而且这些信息可能被 Stem CNN 忽略掉了。

Multi-head Self-Attention Dropping (MSAD) 是一种用于提高网络性能的技术。它可以防止过拟合，避免网络过度学习训练数据，以致在新的数据上表现不佳。在训练过程中，MSAD 会随机删除一些连接，这样网络学习到的特征会更加稳健，不会过于依赖于某些特定的输入模式。

总之，TransFER 架构使用预训练的卷积神经网络和定制的模块来提取特征并对输入数据进行分类。这些特定的架构和技术的选择，是基于它们在类似任务中的已知性能。

通过训练，TransFER 架构在验证集上的表情识别准确率达到 90.91%。

4 基于预训练大模型的面部特征提取

4.1 CLIP 预训练模型的简介

预训练语言模型 (Pre-Trained Language Models) 采用“预训练-微调”两步走的训练流程，旨在学习通用的语言模式，以便在各种自然语言处理任务中快速应用。相比于传统的人工特征提取和模型设计，预训练语言模型能够利用更大规模的数据和更强大的模型进行训练，取得了很好的效果。

CLIP 是一种基于对比学习的多模态模型，它利用对文本-图像对的预训练方法，学习到文本-图像对之间的匹配关系。具体而言，CLIP 包括两个模型：Text Encoder 和 Image Encoder，Text Encoder 用来提取文本的特征，而 Image Encoder 用来提取图像的特征。这两个模型通过对比学习，学习到文本和图像的表示之间的相互关系。在 CLIP 的预训练阶段，通过对比学习的方法，让模型学习到文本-图像对之间的匹配关系，进而学习到语义表示。

与传统的预训练语言模型不同，CLIP 不仅仅是基于语言的预训练模型，而是同时考虑了图像和语言的信息。因此，CLIP 可以在多个领域中应用，如图像分类、目标检测、自然语言生成等。与其他预训练语言模型相比，CLIP 能够更好地捕捉语义信息，从而提升在多模态任务中的表现。

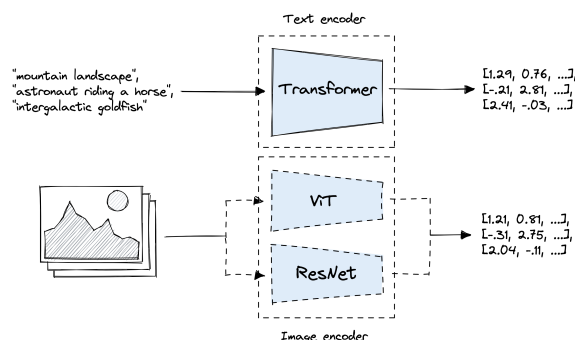


图 7 CLIP 预训练模型

4.2 CLIP 在本项目中的相关应用

4.2.1 身体健康状况的分析

在传统中医 (TCM) 中，面部诊断是用于评估一个人的整体健康状况和诊断潜在状况的重要诊断方法。这种诊断方法基于这样一种信念，即面部反映了身体的内部状态，包括器官、血液和能量。

中医从业者观察面部的颜色、形状、质地和其他特征，以识别不协调的模式并确定哪些器官系统可能受到影响。

例如，脸色苍白可能表示血虚，而脸色发红可能表示体内有热。肤色偏黄可能表示消化系统有问题，而肤色呈蓝紫色则可能表示血液循环不良或心血管问题。中医从业者还会观察眼睛、鼻子和嘴巴，以评估相应器官和系统的健康状况。

最近，随着人工智能 (AI) 技术的发展，中医面部诊断可以借助 CLIP（对比语言-图像预训练）等预训练大型模型得到进一步改进。CLIP 可以将自然语言描述与图像相关联，使其能够理解语言与视觉信息之间的关系。

在 TCM 中，CLIP 可用于通过分析人脸图像并将其与大型面部图像数据集进行比较以识别模式和相似性来辅助面部特征的诊断。这可以帮助中医师根据患者独特的面部特征和整体健康状况做出更准确的诊断并制定更有针对性的治疗计划。

总之，中医面相诊断是一种重要的诊断方法，可以提供有关一个人整体健康状况的有价值信息。借助像 CLIP 这样的预训练大型模型，中医师可以进一步提高面部诊断的准确性，并为患者提供更加个性化的治疗方案。

4.2.2 心理健康状况的分析

首先，我们可以将面部照片输入到 CLIP 模型中，利用其强大的图像分类功能来识别面部表情。由于 CLIP 已经在图像分类方面证明了其在预测方面的优越性，因此可以在此基础上实现表情识别。同时，我们可以使用传统的人脸检测技术，将面部图像切割出来，以便更准确地识别面部表情。

其次，我们可以将识别出的面部表情与已知的心理健康问题相关的特征进行交叉验证。例如，某些心理疾病可能在面部展现出某种特征，因此可以将这些特征与识别出的

面部表情进行比较。在这个过程中，我们可以利用前述的 Transformer 模型来对特征进行提取和匹配，以进一步提高预测的准确性。

最后，我们需要注意外部因素对脸部识别结果的影响，如光照、角度等。因此，我们需要进行进一步实验和讨论，以确保识别结果的准确性。可以在数据集上进行训练和测试，以评估模型的性能，并进行必要的调整和优化。

总之，利用 CLIP 预训练大模型可以帮助我们识别面部表情，进而监测人类的心理状况。这种方法可能存在一些限制，但仍然具有潜在的应用价值，并且可以在未来的研究中进一步发展和完善。

4.2.3 CLIP 的其它应用可能性

作为监控，用 CLIP 检测视频中的面部特征：CLIP 可以用于分析视频中的面部特征，通过抽帧分析和截取图片的方式，可以对面部在一段时间范围内的表情变化和特征进行判断和分析。这种应用可以用于情感识别、自动化监控等领域。

在总体健康状况的量化方面，CLIP 可以将照片与“Healthy”相关联，通过计算它们的相关系数来验证健康指数。这种应用可以用于评估个体的整体健康状况，从而提供健康管理和预防保健方面的指导。

需要注意的是，这些应用都需要临床数据支撑来验证其有效性。虽然 CLIP 在图像分类和自然语言处理等方面表现出色，但在医疗领域的应用需要进一步研究和探索。

5 健康监测系统的实际部署

5.1 系统的整体架构

首先，我们使用 BMD101 传感器收集人体的心电信号和一个外置摄像头捕捉人脸的信号，将这些信号通过蓝牙传输到嵌入式开发板上。在开发板上，我们使用滤波技术来净化信号，并通过实时可视化的方式来监测数据。

为了实时获取健康状况信息，我们可以利用已经训练完毕的预处理模型进行调用。这个模型能够分析收集到的数据，快速生成相关的健康状况结果，并将其在开发板上进行可视化。这项技术还可以部署在 Web 页面上，使用户能够随时随地获取最新的监测结果。

这种技术对于医疗领域和个人健康管理来说具有巨大的潜力，可以实时监测人们的健康状况，预测可能的疾病，提供早期干预措施，从而改善人们的生活质量。

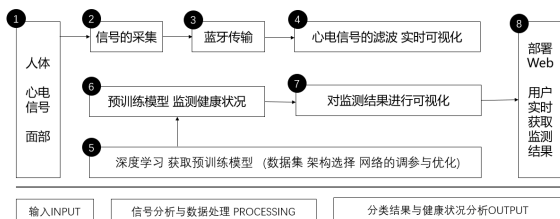


图 8 系统整体架构图

5.2 心电与面部信号的输入

输入部分是整个系统的数据采集和传输环节，主要包括 BMD101 传感器和蓝牙模块。

BMD101 传感器是一款高精度心电信号采集芯片，能够采集人体手指上的心电信号，同时具有抗干扰能力强、信噪比高等优点。通过采集到的心电信号，系统可以进行心率、心律等数据的分析和诊断。

蓝牙模块是将采集到的心电信号通过无线蓝牙技术传输到电脑或其他蓝牙设备上的关键部件。本项目中采用蓝牙 4.0 技术，具有低功耗、高速率、稳定性好等特点，能够实现高效可靠的数据传输。用户可以通过手机或其他蓝牙设备连接到系统，方便地进行心脏健康监测。

对于摄像头，我们使用 PZ-OV7670。这一摄像头便宜且方便使用，可以用来提取面部的相应特征。

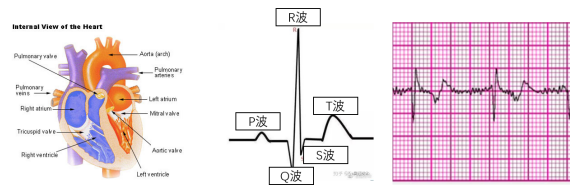


图 9 心电图的输入

5.3 信号的预处理

在输入部分，我们采集到的原始信号可能存在较多的噪声和低频分量，这会影响信号的准确性和稳定性。为了保证信号的准确性和稳定性，我们对原始信号进行预处理，包括去除噪声和增强面部特征等操作。

在本项目中，我们使用均值滤波器对低频分量进行平滑。基于采集的心电信号特性，我们采用去除均值漂移、移除低频高频分量的方法来进行信号预处理。根据面部信号的特性，其低频成分主要由照明条件和姿势等因素引起的变化所导致。因此，均值滤波器的截止频率被设置为 1Hz 以下，以保留面部信号中的高频特征成分。同时，通过对照片进行灰度化和直方图均衡化等处理，可以增强面部特征的对比度和细节，从而提高模型对于面部表情和特征的识别准确度。

经过预处理后的面部信号和心电信号，均能够更加准确地反映人体的健康状况。在嵌入式开发板上，我们可以将这些预处理后的信号进行实时处理，并调用相应的模型来提取有用的信息，最终展示给用户。

5.4 嵌入式开发板上的模型部署

首先，我们基于开发板的 Debian 系统安装和配置相关的库和环境，以支持预训练模型的加载和调用。其次，我们导入预训练模型：将已经训练好的预训练模型导入开发板的存储设备中。第三，我们调用预训练模型：在开发板上使用深度学习框架调用导入的预训练模型，加载到内存中，然后进行预测。第四，我们基于 RK3399 开发板提供的 Lite 模块

优化模型，在系统内进行量化、剪枝、模型压缩等方式来减少模型大小和计算量，提高模型在嵌入式设备上的速度和效率。最后，我们部署应用程序，将预训练模型集成到具体的应用程序中。

5.5 最终结果的呈现

基于此，我们开发了一种用于分析心电图 (ECG) 信号与面部信号，并将结果呈现给用户的新系统。该系统使用深度学习模型将心电信号分为正常、异常、噪声等波形类别，并在网页上以饼图形式呈现分类概率。此外，波形的时域和频域特征的动画显示呈现给用户，提供有关 ECG 信号的更深入的信息。五边形图用于呈现基于输入心电信号的详细分析结果，例如心脏活力、运动状态和压力水平。该系统为用户提供了一种方便、可靠的心脏健康监测工具，有望在医疗和健康管理领域发挥重要作用。结合面部表情分类模型和 CLIP 预训练大模型，可以将结果进行整合，形成更全面、更准确的健康状况评估。组合后的系统可以为用户提供更加个性化的健康管理方案，包括心电信号分类概率、面部表情分类结果、CLIP 预训练模型分析结果等相关健康指标和建议。

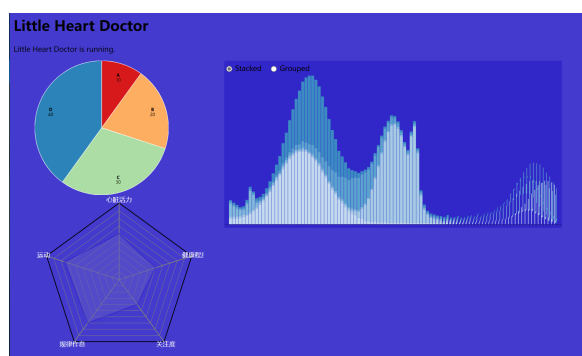


图 10 最终结果的呈现图

6 结束语

我们的项目旨在提供一个基于深度学习和预训练模型的快速、准确、实时的身心健康状况监测和分析系统。我们通过使用改进的 ResNet 与 Transformer-Encoder 模型成功地提高了 ECG 信号识别的准确性，并通过使用基于 Transformer 的代码提取了面部特征。此外，我们还探索了 CLIP 模型增强特征提取能力的潜力。通过集成嵌入式开发板 RK3399，我们实现了对心电信号的实时监测和处理，大大降低了系统的运行成本。

此外，我们的系统可以通过提供一种更方便、更有效的方式来监测一个人的健康状况，从而为中医的普及做出贡献。我们已经撰写新型发明专利申请表的初稿，并计划参加创新创业大赛，来推动我们的系统在医疗健康领域的积极影响。

总而言之，我们相信我们的系统将帮助人们更有效地监

测自己的健康状况，改进疾病的诊断和治疗，最终推动中医的发展和进步。

参考文献

- [1] ZHANG A, LIPTON Z C, LI M, et al. Dive into deep learning [A]. 2021.
- [2] LE M D, SINGH RATHOUR V, TRUONG Q S, et al. Multi-module recurrent convolutional neural network with transformer encoder for ecg arrhythmia classification[C/OL]// 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI). 2021: 1-5. DOI: [10.1109/BHI50953.2021.9508527](https://doi.org/10.1109/BHI50953.2021.9508527).
- [3] XUE F, WANG Q, GUO G. Transfer: Learning relation-aware facial expression representations with transformers [A]. 2021. arXiv: [2108.11116](https://arxiv.org/abs/2108.11116).
- [4] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision [A]. 2021. arXiv: [2103.00020](https://arxiv.org/abs/2103.00020).
- [5] LUO H, JI L, ZHONG M, et al. Clip4clip: An empirical study of clip for end to end video clip retrieval[A]. 2021. arXiv: [2104.08860](https://arxiv.org/abs/2104.08860).
- [6] 任慧霞, 马晓梅, 王锦玉, 等. 基于面部特征测量的中医五行人体质类型判别研究[J]. 中华中医药杂志, 2022(003): 037.