

# 硕士学位论文

## 基于癫痫监测手环的数据处理算法与软件 开发

**DATA PROCESSING ALGORITHM AND  
SOFTWARE DEVELOPMENT BASED ON  
EPILEPSY MONITORING BRACELET**

余锡彪

哈尔滨工业大学

2021 年 6 月

国内图书分类号：TP181  
国际图书分类号：004

学校代码：10213  
密级：公开

## 工程硕士学位论文

# 基于癫痫监测手环的数据处理算法与软件 开发

硕 士 研 究 生：余锡彪

导 师：梁廷伟教授

申 请 学 位：工程硕士

学 科：航天工程

所 在 单 位：航天学院

答 辩 日 期：2021 年 6 月

授予学位单位：哈尔滨工业大学

Classified Index: TP181

U.D.C: 004

Dissertation for the Master Degree in Engineering

**DATA PROCESSING ALGORITHM AND  
SOFTWARE DEVELOPMENT BASED ON  
EPILEPSY MONITORING BRACELET**

|                                       |                                |
|---------------------------------------|--------------------------------|
| <b>Candidate:</b>                     | Yu Xibiao                      |
| <b>Supervisor:</b>                    | Prof. Liang Tingwei            |
| <b>Academic Degree Applied for:</b>   | Master of Engineering          |
| <b>Speciality:</b>                    | Astronautical Engineering      |
| <b>Affiliation:</b>                   | School of Astronautics         |
| <b>Date of Defence:</b>               | June, 2021                     |
| <b>Degree-Conferring-Institution:</b> | Harbin Institute of Technology |

## 摘 要

癫痫是一种常见的精神类疾病,我国患者基数大,而且每年新增很多患者,这种疾病给患者及其家属带来了很大的精神压力,更主要的是严重威胁了患者的生命健康安全。目前癫痫检测大多采用脑电信号,虽然这种方法成熟、准确率高,但是采用脑电信号检测癫痫的设备体积比较大,不利于便携,因此研究出搭载癫痫检测算法的可穿戴设备非常有必要,对癫痫患者及其家属来说是福音。基于此背景,本文提出了基于癫痫监测手环的数据处理算法,并且研究开发了配套癫痫监测系统软件。

课题中采集患者数据的设备为癫痫手环,手环搭载皮肤电传感器、心率传感器和三轴加速度传感器,采集这三种信号进行后续分析。对初始数据集进行预处理、特征提取与选择、标记与分割、标准化这四步操作,其中提取了 47 个特征,包含时域特征、频域特征,并且引入了近似熵、样本熵以及模糊熵。

经过预处理和特征提取后的数据集输入机器学习算法模型中,本文对比研究了逻辑回归算法、支持向量机算法、AdaBoost 算法以及 XGBoost 算法,经过仿真对比实验得出使用 XGBoost 算法进行癫痫检测的效果最好,检测发病的准确率为 98%。由于嵌入式设备硬件资源有限,无法支撑复杂的自学习算法,考虑使用边界函数监测癫痫发作,由此进行测试,得到线性核支持向量机算法边界函数效果最好,准确率为 88.50%,误报率为 5.14%,漏报率为 6.36%。

并且对预处理过程中使用的插值方式、数字滤波算法和窗口重复率进行了对比实验,得出以下结论:使用三阶 B 样条曲线插值、五点中值滤波算法、窗口重复率为 80%时的效果更优。

本文最后还研究了癫痫监测系统的数据流以及软件涉及到的接口技术,并由此开发了云平台软件和手机端软件,开发的软件均可正常运行,并实现了预期的效果。

**关键词:** 癫痫; 皮肤电导; 心率测量; 三轴加速度信号; 机器学习算法

## Abstract

Epilepsy is a common mental disease, the patient base in our country is large, and many new patients every year, this disease has brought great mental pressure to patients and their families, more importantly, is a serious threat to the patient's life and health safety. At present, EEG signals are mostly used in epilepsy detection. Although this method is mature and accurate, the equipment used to detect epilepsy by EEG signals is relatively large, which is not conducive to portability. Therefore, it is very necessary to develop a wearable device equipped with epilepsy detection algorithm, which is a good news for epilepsy patients and their families. Based on this background, this paper proposes a data processing algorithm based on epilepsy monitoring bracelet, and studies and develops a matching epilepsy monitoring system software.

In this project, the device for collecting patient data is the epilepsy bracelet, which is equipped with skin electrical sensor, heart rate sensor and triaxial acceleration sensor to collect these three signals for subsequent analysis. The initial data set is preprocessed, feature extraction and selection, marking and segmentation, and standardization. 47 features are extracted, including time-domain features and frequency-domain features, and approximate entropy, sample entropy and fuzzy entropy are introduced.

After preprocessing and feature extraction, the data set is input into the machine learning algorithm model. In this paper, the logistic regression algorithm, support vector machine algorithm, AdaBoost algorithm and XGBoost algorithm are compared and studied. Through simulation and comparison experiments, it is concluded that the XGBoost algorithm has the best effect in epilepsy detection, and the accuracy rate of detection of disease is 98%. Due to the limited hardware resources of the embedded device, it cannot support the complex self-learning algorithm. Considering the use of boundary function to monitor epileptic seizures, the test results show that the boundary function of the linear kernel support vector machine algorithm has the best effect, with an accuracy rate of 88.50%, a false alarm rate of 5.14% and a missing alarm rate of 6.36%.

In addition, the interpolation method, digital filtering algorithm and window

repetition rate used in the preprocessing are compared and the following conclusions are drawn: the third-order B-spline curve interpolation, five-point median filtering algorithm and window repetition rate of 80% are better.

In the end, this paper also studies the data flow of epilepsy monitoring system and the interface technology involved in the software, and thus develops the cloud platform software and mobile terminal software. The developed software can run normally, and achieves the expected effect.

**Keywords:** Epilepsy, Skin conductance, Heart rate measurement, Triaxial acceleration signal, Machine learning algorithm

# 目 录

|                           |    |
|---------------------------|----|
| 摘 要 .....                 | I  |
| Abstract.....             | II |
| 第 1 章 绪 论 .....           | 1  |
| 1.1 课题研究背景与意义 .....       | 1  |
| 1.2 癫痫监测算法国内外研究现状 .....   | 2  |
| 1.2.1 国内研究现状 .....        | 2  |
| 1.2.2 国外研究现状 .....        | 2  |
| 1.2.3 文献综述与分析 .....       | 3  |
| 1.3 癫痫监测设备国内外研究现状 .....   | 4  |
| 1.3.1 国内研究现状 .....        | 4  |
| 1.3.2 国外研究现状 .....        | 4  |
| 1.3.3 文献综述与分析 .....       | 5  |
| 1.4 主要研究内容与规划 .....       | 6  |
| 1.4.1 论文主要研究内容 .....      | 6  |
| 1.4.2 论文结构规划 .....        | 7  |
| 第 2 章 数据采集与特征提取 .....     | 9  |
| 2.1 基本原理 .....            | 9  |
| 2.2 数据采集 .....            | 9  |
| 2.3 数据预处理 .....           | 12 |
| 2.3.1 同步采样处理 .....        | 12 |
| 2.3.2 数字滤波去噪 .....        | 14 |
| 2.3.3 数据分割 .....          | 16 |
| 2.4 特征提取 .....            | 17 |
| 2.4.1 特征提取 .....          | 17 |
| 2.4.2 特征选择 .....          | 20 |
| 2.5 数据集处理 .....           | 22 |
| 2.5.1 数据集标记与分割 .....      | 22 |
| 2.5.2 数据集标准化 .....        | 22 |
| 2.6 本章小结 .....            | 23 |
| 第 3 章 癫痫监测的数据处理算法研究 ..... | 24 |
| 3.1 逻辑回归算法 .....          | 24 |

|                                    |           |
|------------------------------------|-----------|
| 3.1.1 逻辑回归算法原理 .....               | 24        |
| 3.1.2 算法模型训练 .....                 | 26        |
| 3.2 支持向量机算法 .....                  | 27        |
| 3.2.1 支持向量机算法原理 .....              | 27        |
| 3.2.2 算法模型训练 .....                 | 29        |
| 3.3 AdaBoost 算法.....               | 30        |
| 3.3.1 AdaBoost 算法原理 .....          | 30        |
| 3.3.2 算法模型训练 .....                 | 31        |
| 3.4 XGBoost 算法 .....               | 32        |
| 3.4.1 XGBoost 算法原理 .....           | 32        |
| 3.4.2 算法模型训练 .....                 | 33        |
| 3.5 算法模型对比实验 .....                 | 33        |
| 3.6 测试实验 .....                     | 34        |
| 3.6.1 边界函数测试实验 .....               | 34        |
| 3.6.2 临床数据验证 .....                 | 35        |
| 3.7 本章小结 .....                     | 37        |
| <b>第 4 章 癫痫监测系统软件设计与开发 .....</b>   | <b>38</b> |
| 4.1 癫痫监测软件功能模块设计 .....             | 38        |
| 4.2 数据库设计 .....                    | 40        |
| 4.2.1 数据库的 E-R 模型.....             | 40        |
| 4.2.2 数据结构设计 .....                 | 41        |
| 4.3 数据流设计 .....                    | 43        |
| 4.4 软件数据接口设计 .....                 | 44        |
| 4.4.1 云平台数据接口设计 .....              | 45        |
| 4.4.2 APP 端数据接口设计 .....            | 47        |
| 4.5 云平台主要模块界面 .....                | 49        |
| 4.6 软件平台测试 .....                   | 50        |
| 4.7 本章小结 .....                     | 51        |
| <b>结 论 .....</b>                   | <b>53</b> |
| <b>参考文献 .....</b>                  | <b>55</b> |
| <b>哈尔滨工业大学学位论文原创性声明和使用权限 .....</b> | <b>59</b> |
| <b>致 谢 .....</b>                   | <b>60</b> |



## 第 1 章 绪 论

### 1.1 课题研究背景与意义

本课题是预研项目的其中一个研究部分，瞄准了社会需求。

癫痫属于神经系统的一种慢性疾病，癫痫发作的最主要机理是大脑神经元放电不正常，患者通常是症状突然发作，发作的时间比较短暂，并且具有复发性<sup>[1]</sup>。2018 年世界卫生组织公布了一项统计数据，其中显示全球已经超过 5000 万人患有癫痫病，其中不发达国家的癫痫患者占很大比例。我国癫痫患者的人口基数大，数量目前超过 900 万，并且每年新增 60 万，给患者的家庭带来了很大的精神负担和经济负担，更重要的是，癫痫患者的生命健康安全受到了严重的威胁。

目前癫痫的检测主要是在医院中进行，通过专业的脑电仪器得到患者的脑电图，依此检测癫痫并诊断癫痫的类型，从而确定接下来的治疗方案，但是这个过程比较耗时间，要求患者长期住院监测，有时甚至需要持续数月的时间，长时间的住院环境会给患者带来精神压力，影响患者情绪，给诊断带来影响，因此亟需拥有一种更快速高效的癫痫识别手段<sup>[2]</sup>。其次，癫痫发作间隔时间不定，部分癫痫患者即使在长期的监测过程里，仍然可能没有发作，造成医护资源的占用浪费。

市场上已然存在一些 EEG 便携式监测设备，但是据 Bruno 等人的研究，在监测设备外形因素偏好方面，智能手表或戒指普遍被接受，因为这些外形外观较为常见，相反需要佩戴头戴式设备的脑电装置在患者中不受欢迎<sup>[3]</sup>，这类装置不仅不太美观，而且被认为暴露隐私，在睡眠期间佩戴头皮来记录 EEG 信号是不切实际且不舒适，因此，目前不足以用于可穿戴和自动癫痫发作检测系统<sup>[4]</sup>。

本课题研究的算法与软件开发基于课题组自研的以 STM32 芯片为核心的癫痫监测手环，该癫痫手环内嵌多种传感器，其中包括皮肤电传感器、心率传感器与三轴加速度传感器，通过采集到的皮肤电信号、心率信号与合成的加速度信号生理数据进行算法训练。用这三种信号代替脑电信号检测癫痫发作可以有效减小可穿戴设备的体积，从而将癫痫检测设备研制成大多数患者可以接受的手环形式，从而保障癫痫患者的生命健康安全，使得他们的家人安心。当手环有了软件平台的支撑，在手机上随时查看生理数据分析图得以实现，并且可以即时上传数据到云端，保存生理数据以便不断优化算法提高准确率，还可以

给患者提供其他便捷服务，比如在线问诊功能，患者在任何地方都能联系到医生咨询健康问题，从而极大的方便患者。因此，本课题对癫痫患者这一特定人群非常有意义。

## 1.2 癫痫监测算法国内外研究现状

### 1.2.1 国内研究现状

本节主要分析近几年国内学者在癫痫检测算法研究上的成果。

2016 年，山东大学的耿东云在小波神经网络算法的基础上改进，融合了小波变化与前馈神经网络算法，取得了较高的准确率和较低的误检率<sup>[5]</sup>。2017 年，李明阳、陈万忠等人提出了将双密度离散小波变换(DDDWT)以及 Log-Logistic 参数回归算法应用于特征工程中，之后使用遗传算法改进过的支持向量机算法对处理过的脑电信号进行分类，效果良好<sup>[6]</sup>。2018 年，兰州大学的吴浩博提出了一种对脑电信号数据集进行离散小波处理并提取样本熵特征和其他时频特征，其中对样本熵算法进行了优化，运算效率得到提高，基于此使用支持向量机和朴素贝叶斯算法对信号进行分类<sup>[7]</sup>。2020 年，兰州理工大学的尹倡隆对脑电信号数据进行小波分析，并提取变换后数据集的统计时域特征以及非线性特征，对比了随机森林和 AdaBoost、XGBoost 算法的癫痫发作数据分类效果，并提出了一维卷积神经网络和 XGBoost 算法相结合的模式识别算法，分类效果较前者更优<sup>[8]</sup>。

### 1.2.2 国外研究现状

自上个世纪八十年代国外就开始研究癫痫监测方法了，Gotman 是这一时期首位提出相关方法的学者，他提出了一种不依赖于临床症状的脑电图癫痫发作自动检测方法，基础是将脑电图分解为基本波，对此进行运算得到特征，识别多种类型的癫痫发作<sup>[9]</sup>。

在最近几年，对癫痫监测算法的研究也越来越多，成果也越来越丰富，以下是近几年的一些研究成果。

2018 年，Erdem 和 Mustafa 等人在特征提取阶段计算倒谱系数(MFCCs)，并将其应用于基于神经网络算法的癫痫发作识别中，由于该方法所采用的神经网络具有非迭代性，因此该系统具有应用于实时癫痫检测系统的潜力<sup>[10]</sup>。Hussain 采用了基于时频域特征、非线性特征、基于小波熵和少量统计特征的变化特征提取策略，模式识别方法选用了 k 近邻(KNN)算法、不同核的支持向

量机以及集成分类器进行分析对比<sup>[11]</sup>。2019年,Raghu和Sriraam等人提出对原始数据集进行离散小波变换,由分解后的每个子带的小波系数来估计 sigmoid 熵,在此基础上提取特征,并使用非线性内核的支持向量机分类算法进行分类,由此可以实现癫痫的自动检测<sup>[12]</sup>。2020年,Nabil和Benali等人利用离散小波变换对脑电信号进行子信号分解,然后从得到的子信号中确定不同的非线性参数,如近似熵(ApEn)、最大李雅普诺夫指数(LLE)和统计参数,并将这些特征数据集作为支持向量机分类器的输入进行癫痫发作识别<sup>[13]</sup>。Sharmila和Geethanjali研究了特征,首次尝试利用波形长度等时域特征,利用经滤波和离散小波变换(DWT)后的脑电图数据直接得到癫痫发作的零点交叉数(ZC)和斜率符号变化数(SSC),此外还研究了其他时域特征如平均绝对值、标准差、平均功率,采用的是朴素贝叶斯和支持向量机分类器研究特征的性能<sup>[14]</sup>。Raghu和Sriraam等人提出了一种计算效率高的自动癫痫检测模型,该模型使用一种新的特征称为连续分解指数,在计算效率上明显优于以往的小波等特征提取方法,模式识别算法选用的是支持向量机分类器<sup>[15]</sup>。Glory与Vigneswaran等人提出了一种基于自适应 Haar 小波的二值 Grasshopper 优化算法与深度神经网络相结合的癫痫发作检测深度学习模型,使用的 EEG 数据集来自波恩大学的伯尔尼大学和 CHB-MIT 脑电图数据库<sup>[16]</sup>。Hanan与Shahab等人提出了一种基于协方差矩阵和自适应 Boosting 最小二乘支持向量机框架的两阶段脑电图癫痫发作检测分类方法<sup>[17]</sup>。2021年,Rout与Susant等人利用脑电信号的变分模式分解(VMD)和近似熵进行特征提取,并与多层多核随机向量函数链接网络分类器相结合,有效地识别癫痫发作<sup>[18]</sup>。

### 1.2.3 文献综述与分析

分析癫痫监测算法相关的文献,可以看出目前研究的算法大都是基于脑电信号的,因此目前由脑电信号进行癫痫发作监测比较成熟,而用皮肤电信号和心率信号来分析癫痫发作的少之又少,但是改用皮肤电信号和心率信号进行监测会减小监测设备的体积增加便携性。提取的特征有时域特征、频域特征,时域特征主要是统计特征,频域特征有进行小波变换分解运算的,也有进行频谱分析的,非线性特征主要是近似熵、样本熵。使用的监测算法包括机器学习算法和深度学习算法,具体比如 k 近邻、支持向量机、随机森林、AdaBoost 算法、XGBoost 算法、神经网络算法,并且都取得了不错的效果。本文将会基于三轴加速度信号、皮肤电信号和心率信号进行癫痫算法的研究,并且提取时域、频域,包含非线性特征,研究机器学习算法对癫痫监测的效果。

## 1.3 癫痫监测设备国内外研究现状

### 1.3.1 国内研究现状

2016 年,张根选团队研制了一种可穿戴贴片式心电传感器设备,核心模块有心电采集模块、两个电极以及蓝牙模块,通过电极连接人体和心电传感器来采集数据,并由此计算得到 RRI 间期数据,内部蓝牙模块发送数据到手机安卓端,安卓端或者云端算法对癫痫发作进行识别与预测<sup>[19]</sup>。燕山大学的路承彪教授团队致力于研究便携式癫痫监测系统的设计与研究,他们团队提出一种基于嵌入式系统技术和 Android 智能操作系统平台,结合改进的递归熵(Entropy of recurrence quantification, ENTR)为核心算法的便携式癫痫发作监测系统设计方案,采集 EEG 信号,让癫痫监测算法在 Android 操作系统上的移植和实现<sup>[20]</sup>。

国内的瑞尔脑康公司曾研制了一款便携式癫痫手环并开发了配套软件管理系统,通过软件系统可以对癫痫数据统一管理和分析,这家公司于 2017 年在中国抗癫痫协会举办的“第七届 CAAE 国际癫痫论坛”中发布该产品。遗憾的是目前市场已经购买不到该公司的产品,也没有该公司的消息。

### 1.3.2 国外研究现状

2014 年, A Bideaux 等人将一种采集多参数的系统应用到癫痫病人的移动监测,采集到的监测参数主要包括 ECG、GSR、多通道 EEG 以及运动传感器,通过蓝牙通讯将多参数数据传输到家庭网关,家庭网关利用框架 x Affect 对数据进行整合,进行发作事件监测,数据存储,也可将数据以及分析结果上传至服务器,用户也可以通过家庭网关下载服务器中数据文件等,与服务器进行互动<sup>[21]</sup>,在此项研究中,仅仅介绍了系统的实现框架,并没有对系统的具体实现形式和参数分析方法进行说明,而且系统不能实现癫痫的病灶定位。2017 年, Abdullah Ahmed 提出了一种无线,低功耗,低成本的可穿戴式两部分,用于实时癫痫发作检测。利用心电图仪 (ECG),皮肤电活动 (EDA),身体运动和呼吸速率 (BR) 的参数,提出了一种多标准决策系统 (MCDS),该系统可减少错误警报和真实阴性。胸部和手穿戴式可穿戴设备的组合可连续检测这些参数,并通过 BLE 4.0 将数据传输到智能手机应用程序,其中基于长短期记忆 (LSTM) 的异常检测算法和逻辑分类器确定癫痫发作的发生情况。通过对综合数据进行测试,可以达到 96% 的精度<sup>[22]</sup>。

目前最先进的成熟商用产品是来自美国的 Empatica 公司研究的可穿戴癫痫手环,这款手环被命名为“Embrace”,于 2018 年 2 月获得美国 FDA 批准<sup>[23]</sup>。

“Embrace”可以测量癫痫发作的多重指标，在真实环境中实现了很高的精度<sup>[23]</sup>。其拥有一套完整的健康监控平台，包括手环通过蓝牙连接到手机，手机通过网络将数据传输到云端，再在 dashboard 中进行数据的分析，可以使用 Web 浏览器访问。除此之外，还有几种可商购的可穿戴式癫痫发作检测系统，例如 EpiLert（以色列的 BioLert 研发），SmartWatch（美国的 Smart Monitor 公司研发），Epi-Care Free（丹麦的 Danish Care 研发）以及目前正在临床评估中的设备<sup>[24]</sup>。

2018 年，Steven Boeckx 等人展示了一种小型可穿戴多传感器设备，用于监测患者日常生活中的癫痫活动和癫痫发作。设备通过三通道脑电图（EEG）测量大脑的电活动，并通过光电容积描记法（PPG）和使用六轴加速度计和陀螺仪的运动来测量心率<sup>[25]</sup>，传感器收集与癫痫发作检测有关的生理信号，然后将数据同步到云中，以进行高级数据分析并生成全面的患者报告，从而为医生提供了新的工具来监视患者日常生活中的状况<sup>[26]</sup>。同年，Dionisije Sopic 提出了 e-Glass，这是一种基于四个脑电图（EEG）电极的可穿戴系统（眼镜式），使用离散小波变换、熵度量进行非线性特征提取，基于随机森林进行分类，用于检测癫痫发作。根据 e-Glass 的早期预警，可以通知护理人员进行抢救，以避免由于潜在的神经系统疾病，癫痫突然死亡，或癫痫发作期间的事故，该系统可达到 93.80% 的灵敏度和 93.37% 的特异性<sup>[27]</sup>。2020 年，Mostafa Gheryani 等人提出了一种用于从生理和惯性数据中实时检测夜间癫痫发作的新型模型，该模型是通过集成了肌肉活动传感器，3D 加速度计和 3D 陀螺仪的无线腕带收集的。腕带将数据传输到便携式设备进行处理和癫痫发作检测<sup>[3]</sup>。

### 1.3.3 文献综述与分析

无论是国内还是国外的研究成果，都主要是采集 EEG 信号，然后对得到的数据进行分析，研究相应的算法，有一部分配套了相应的手机 APP 软件和云平台辅助软件，使得无论是用户还是主治医生、算法研究人员使用都十分方便。但是如前面所说，测量 EEG 信号需要用一些比较复杂的仪器，使用电极测量不可避免。即便是有些研究团队研制了便携式采集 EEG 信号的工具，比如头戴式、眼镜式便携癫痫监测系统，但是对于用户来说也存在不方便，比如头戴式不美观，有可能会被歧视，而眼镜式对于有些近视或者老花眼的患者来说就不适用了。有些研究成果采取的是采集心电等信号，采集仪器一般是胸部贴片之类，这会使得用户不舒适，所以也有待改进。就针对云平台软件和 APP 来说，有些研究成果两者都不配套，而是使用单独的嵌入式设备显示，大多数研究成果配套了 APP 软件，将研究的算法写入 APP 中，对生理信号数据进行处理后记录在

APP 中，方便患者或主治医师查看，少数研究成果配套了 APP 软件和云平台软件。

虽然现在国外的一些商用产品，如 Empatica 研究的智能手环“Embrace”等是腕带式，测量的是皮肤电、三轴加速度等信号，患者使用起来方便舒适，也配备了相应的 APP 和云平台软件，方便患者和主治医生查看，但是这些产品由于某些原因（比如法律法规等）不能在国内使用，国内的医生也没有权限查看软件平台的数据。而相类似的产品在国内属于空白，相类似的研究成果在国内监测准确率达不到国外水平，生理数据分析算法需要深入研究。本课题中将会研究开发项目配套的云平台软件与 APP 后台数据接口，深入研究动态数据算法。

## 1.4 主要研究内容与规划

### 1.4.1 论文主要研究内容

本课题的研究目标是在癫痫监测手环采集人体生理信号的基础上，研究准确率尽可能高的癫痫检测算法，并研发一套功能全面、稳定性高的癫痫监测软件，用于收集、处理、分析用户的生理信号，更好的监测癫痫患者的发作情况及实现其他一些功能。

整个癫痫监测系统的数据流是：首先通过蓝牙连接到手机，将传感器采集到的数据传到手机上，手机再通过移动网络以 json 的格式将生理数据传送到服务器端。服务器得到生理数据后，将其存储到 MySQL 数据库中。当需要分析数据时，可以登录云平台管理系统，服务器中先将所需要的数据简单处理，以表格和折线图的形式呈现给管理员，并且算法工程师或者医生等相关人员如果需要拿到这些数据，可以在云平台导出包含生理数据的 Excel 文件，十分方便简洁。除此之外，云平台 and APP 还包含其他一些额外辅助功能，比如健康推送模块、专家咨询模块，等等。

本课题的主要工作是研究、实现能够检测出癫痫发作的算法，并且设计出配套 APP 端的数据接口以及 Web 端的前台界面与数据接口。

癫痫数据处理算法的实现：结合取得的实验数据，这些数据包括患者的心率信号、皮肤电信号、三轴加速度信号，用这些数据来训练算法模型，将新数据代入得到的模型用来判断是否发病，使之能在一定程度上可以有效判断患者的发病情况。

整个算法研究过程包括以下几个步骤：数据采集、数据预处理（包括统一

频率、数字滤波去噪、数据分割步骤)、特征提取(包括特征提取、特征选择步骤)、数据集处理(包括数据集标记与分割、数据集标准化步骤)、算法模型训练、测试算法效果。

关于云平台与 APP 端的数据接口需要实现如下功能:管理员(比如算法开发人员、医生)的登录验证,普通用户的登录验证,数据上传,数据下载(导出 Excel 文件的形式),数据分析(在后台对数据进行简单处理得到前台可视化所需要的数据),合作医生信息介绍、健康推送与新闻推送、产品介绍、APP 操作介绍等功能所需要的增删改查操作。考虑到后期存储的数据量可能非常大,因此在这一阶段将会从多方面考虑设计数据库结构,在开发接口时注意数据库事务的管理。

关于云平台可视化界面(前端)大致包括两个部分,分别是普通用户可视部分和管理系统界面。在普通用户可视界面,大体包括如下模块:产品介绍、APP 操作介绍、健康推送、新闻推送、项目简介,在管理系统界面则大体包括如下模块:用户数量统计、用户可视化界面信息编辑、APP 和 DFU 的安装包上传模块、数据图表可视化模块、数据下载模块。

#### 1.4.2 论文结构规划

基于上一小节所述的研究内容,本论文共分为四个章节和最后结论部分,每一部分的内容将从以下方面进行展开。

第一章为绪论部分。这部分首先讨论课题的研究背景,并分析研究意义,表明本课题研究的必要性;接着介绍癫痫监测算法的国内外研究现状,并对参考文献进行了综述总结;紧接着对癫痫监测设备的国内外研究现状进行介绍,通过这两个小部分的分析,展现了本课题相关的研究成果与研究方向;最后阐述本论文的研究内容与规划。

第二章为数据采集与特征提取部分。这部分首先介绍本课题的基本原理,即为什么采用皮肤电信号、心率信号、三轴加速度信号进行癫痫检测;接着阐述数据采集过程,即采集使用的硬件与采集对象,以及采集到的数据格式;其次介绍数据预处理,主要包括统一频率、数字滤波降噪与数据分割,其中还介绍仿真对比实验,包含不同插值方式的对比实验、不同数字滤波方式的对比实验、不同窗口重复率的对比实验,介绍实验的条件以及对比结果;紧接着阐述特征提取过程,具体分为两个步骤,分别为特征提取与特征选择,特征选取了频域特征、时域的统计特征与非线性特征;最后讨论对数据集处理,其中包括数据集的标记与分割以及数据集的标准化。

第三章为癫痫监测的动态数据算法研究部分。这一部分将使用四种机器学习算法对癫痫监测进行研究，这四种算法分别是逻辑回归、支持向量机、AdaBoost 算法与 XGBoost 算法。首先对每一种算法的基本理论进行详细介绍，其次阐述相应算法的训练过程，并将这四种算法的训练情况作仿真对比实验。最后将介绍测试实验，包含边界函数测试实验与由 5 秒钟窗口数据判断发病测试实验，并展示测试结果。

第四章为癫痫监测系统软件设计与开发部分。这部分首先介绍癫痫监测软件功能模块，将详细阐述云平台软件与安卓端软件的 6 个功能模块。紧接着展示数据库设计的 E-R 模型与主要的 8 个数据库表，并详细描述系统的数据流，揭示癫痫手环、手机 APP 与云平台三者之间的通信关系。其次重点介绍云平台软件与安卓端软件的数据接口，这些数据接口支撑软件实现主要的功能。最后展示云平台的主要模块界面，将会具体展示查询获取生理信号数据界面和训练算法模型界面，并介绍软件的使用测试结果。

最后一部分是本文的结论。这部分对基于癫痫手环的动态数据算法研究工作和癫痫检测系统配套软件接口技术研究工作进行了总结，总结了算法研究中仿真实验的结果和软件开发的接口种类，并针对研究工作中的不足之处给出了对未来工作的展望。



## 第 2 章 数据采集与特征提取

### 2.1 基本原理

癫痫的发病原理是大脑中的神经元产生突发性的不正常放电现象，与交感神经的活跃程度有密切联系。身体中唯一仅仅只受交感神经影响并且完全不受副交感神经影响的器官就是皮肤，皮肤电信号可以反映皮肤的特征，因此要想对交感神经的活跃神经进行监测，可以依据测量皮肤电的微小变化来实现。

心率的变化有可能会与交感神经系统与副交感神经系统的活动存在联系，因此可以通过提取 HRV（心率变异性，heart rate variability）中的高频分量来评估副交感神经或迷走神经系统的活跃程度。并且有研究表明癫痫的发作可以在一定程度上通过心率的变化来反映，仍然是通过测量心率的变异性参数 HRV<sup>[28]</sup>。

癫痫发作时患者大多会发生全身强直或者是痉挛抽搐的现象，手部会表现出比较明显的振动，因此可以通过测量手部的加速度信号来监测癫痫发作。

本课题将通过综合监测癫痫患者的皮肤电信号、心率信号、三轴加速度信号的异常变化来对癫痫发作进行实时检测。

### 2.2 数据采集

本课题使用基于 STM32F103ZET6 嵌入式芯片为控制器的癫痫监测手环来采集生理信号，该癫痫监测手环包括多个外设单元，比如动作检测单元、皮肤电检测单元、心率检测单元、蓝牙通信单元等，这些检测单元内含精确性高的传感器，测量准确。

加速度传感器选用 ATK-MPU6050-V1.1 传感器模块，这个模块的核心采用了 InvenSense 公司生产的 MPU6050 芯片，这块芯片内含三轴加速度传感器和三轴陀螺仪传感器，并将其完美地进行整合，此外包含数字运动处理器（Digital Motion Processor, DMP），通过 DMP 可以提高运算运动姿态的效率，从而可以使运动处理运算在操作系统上的压力减小，再通过 IIC 接口进行通信来输出运动姿态运算处理后的数据<sup>[29]</sup>。

皮肤电传感器采用皮肤电反应 SC 传感器。皮肤电信号具体来说就是反映皮肤的导电特性，当人的情绪发生变化时，皮肤电信号产生相应的变化，其中的机理主要是与人体汗腺分泌有关，并且皮肤电信号可以在一定程度上反映交感神经的活跃程度。课题中采用的皮肤电反应 SC 传感器可以精确的采集皮肤

电信号。

心率检测传感器选用 MAX30102 芯片，其集成了光电检测器、光器件、红外光和红光的 LED 和低噪声电子电路，关断电源可以通过软件方式来控制，当模块保持待机状态时电流几乎为零，电源可以持续进行供电，因此非常适合在可穿戴装置中应用采集心率这项生理数据，可以将采集到的心率数据通过标准的 I2C 兼容的通信接口参数到微控制器中进行运算操作。

癫痫手环硬件结构如图 2-1 所示，癫痫手环外观图如图 2-2 所示。

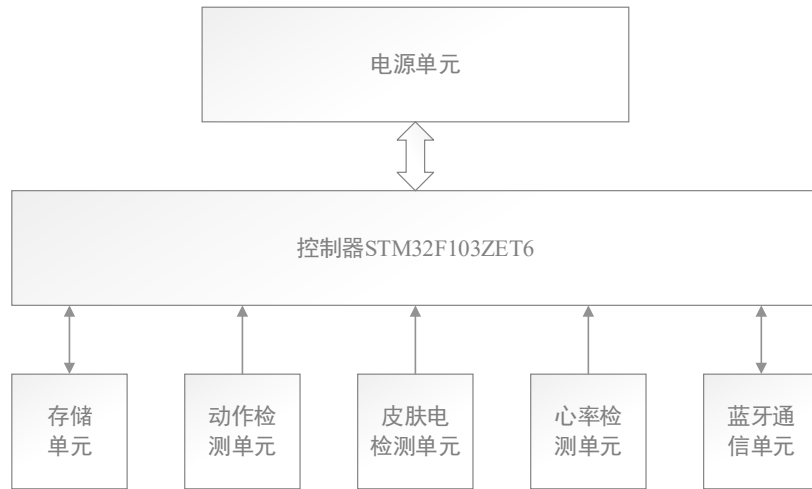


图 2-1 癫痫手环硬件结构图

Figure 2-1 Hardware Structure Diagram of Epilepsy bracelet



图 2-2 癫痫手环外观图

Figure 2-2 Appearance diagram of epilepsy bracelet

本课题的实验数据均是于 2020 年 8 月到 2021 年 3 月期间在首都医科大学三博脑科医院测癫痫患者临床试验采集，采集现场如图 2-3 所示，考虑到隐私问题，图片已进行打码处理。癫痫患者的年龄分布范围广，各个年龄段的都有

涉及，并且性别有男有女，癫痫症状有大发作也有小发作，因此可以保证后续研究的算法的普适性。



图 2-3 采集现场的照片

Fig. 2-3 Photos of the collection site

本次共采集到了 12 名癫痫患者的有效数据，在这 12 名患者腕部戴癫痫手环监测的过程中均有癫痫发作，包括大发作和小发作，共有 20 次癫痫发作的过程，数据量以万为单位计。

癫痫监测手环首先通过蓝牙通信模块采集好的生理信号数据发送给手机，手机通过移动网络以 HTTP 协议的方式将生理数据上传到服务器的数据库中。这样生理信号数据就上传到云端了。

本课题的动态数据算法研究部分使用 python 语言编程，这里是使用 pymysql 包中的函数来连接数据库使用特定的数据库表，从而获取指定患者和时间范围内的数据。获取到的数据的格式是（患者 ID，皮肤电信号，心率信号，腕动次数，加速度信号 00，加速度信号 01，…，加速度信号 19，采集时间），是以元组的形式呈现。需要说明的是，其中单条数据中有 20 个加速度信号数值，这是因为加速度信号的采集频率是 20Hz，而皮肤电信号和心率信号的采集频率均为 1Hz，这里的加速度信号数值是三轴加速度信号合成之后的数值，包含正负方向。腕动次数并不是采用单独的传感器采集得到的信号，是通过芯片的嵌入式算法运算而得，每 60 秒钟清零一次，采用累加方式，当腕动速度比较快时，累加的速度也随之比较快。数据格式如表 2-1 所示。

上表中的 account 表示患者的 ID，gsr 表示皮肤电信号数值，hrt 表示心率信号数值，wrist 表示腕动次数，acc00 表示加速度信号 01 数值，acc19 表示加速度信号 19 数值，dateTime 表示采集时间。

表 2-1 数据格式示例

| account     | gsr  | hrt | wrist | acc00     | ... | acc19     | dateTime |
|-------------|------|-----|-------|-----------|-----|-----------|----------|
| 176****7651 | 3930 | 130 | 3     | -0.125485 | ... | -0.121409 | 03:11:48 |
| 176****7651 | 3835 | 130 | 3     | -0.116746 | ... | -0.116757 | 03:11:49 |
| 176****7651 | 3861 | 131 | 6     | -0.121256 | ... | 0.1221    | 03:11:50 |
| 176****7651 | 3792 | 132 | 7     | -0.122637 | ... | 0.117563  | 03:11:51 |
| 176****7651 | 3943 | 132 | 8     | 0.124138  | ... | -0.120551 | 03:11:52 |

## 2.3 数据预处理

### 2.3.1 同步采样处理

#### (1) 基本介绍

在本课题中，加速度信号的采样频率为 20Hz，而皮肤电信号和心率信号的采样频率为 1Hz，因此加速度信号数据量是皮肤电信号和心率信号数据量的 20 倍，后续需要对三种信号进行特征融合，特征融合有多种方法，这里采用的是将皮肤电信号和心率信号的生理数据进行插值处理，使其数据量扩充 20 倍，与加速度信号的数据量保持一致。

本文研究的是动态数据，即在使用癫痫手环采集患者生理数据的同时需要对发病情况进行分析并且判断是否发病，而插值不能仅仅对一条数据进行插值，对多条数据进行插值会使得监测产生延迟。考虑到实际情况，允许延迟 5 秒钟判断发病情况，因此一个窗口的数据量取 5 秒钟采集的数据量，即 5 条皮肤电信号数据和心率信号数据和 100 条加速度信号数据，在窗口内对皮肤电信号数据和心率信号数据进行插值处理，从而可以实现皮肤电信号和心率信号的采集频率与加速度信号的采集频率同步，这个过程也称为同步采样处理。

信号处理中常用的插值方式有二阶 B 样条曲线插值和三阶 B 样条曲线插值，这里插值使用的函数是 scipy 库中 interpolate 提供的 interp1d 函数，函数中的“kind”参数选择“quadratic”，代表使用的是二阶 B 样条曲线插值，如果参数选择的是“cubic”，则代表使用的是三阶 B 样条插值。

#### (2) 仿真实验基础理论

仿真实验中使用的电脑的 CPU 为 Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz，运行内存为 16.0 GB，操作系统版本为 windows 10。算法程序使用 Python 语言实现，运行的环境是 Anaconda3。

算法模型的输出结果通过混淆矩阵来反映，通过混淆矩阵可以得到准确率、精确率和召回率。在本课题中，发病的这一类定义为正类，未发病的这一类定义为负类，算法模型训练好之后在测试集上判断的结果可分为 4 种情况，分别是：将正类样本划分到正类的数目，可记为  $TP$ ；将正类样本划分到负类的数目，可记为  $FN$ ；将负类样本划分到正类的数目，可记为  $FP$ ；将负类样本划分到负类的数目，可记为  $TN$ 。

准确率，表示判断正确的样本在所有样本中的占比，可以用如下公式表示：

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (2-1)$$

精确率，表示真正类样本（即将正类样本划分到正类这种情况）在所有预测为正类的样本中的占比，可以用如下公式表示：

$$P = \frac{TP}{TP + FP} \quad (2-2)$$

召回率，表示真正类样本在所有正类样本中的占比，可以用如下公式表示：

$$R = \frac{TP}{TP + FN} \quad (2-3)$$

准确率、精确率和召回率这三个指标均是越高越好，可以用这些指标来评价模型学习的结果。

### （3）实验输入数据集

同步采样处理部分的仿真对比实验数据源是通过 10 位患者戴癫痫手环进行临床试验采集得到的，共监测到了 19 次癫痫发作，由于经过了精密的癫痫脑电监测设备与专业医生共同判断标记，因此这些数据的发病标记准确。数据采集过后首先进行简单筛选，这里的筛选主要针对发病数据，因为发病数据集中有一些的三轴加速度信号、皮肤电信号与心率信号的波形基本上都是平缓的，几乎没有任何波动，这些数据集则可以备筛除掉，这些波形可以在 Excel 中绘制而成，并且属于时域波形。除此之外，由于癫痫手环在佩戴之初一般来说皮肤电的数据会特别高，因此这部分的数据集不稳定，也是需要筛除掉的，类似不正常的数据集也都需要筛除掉。

将经过筛选过后的数据集进行预处理与特征工程，具体包含统一频率、数字滤波去噪、数据分割、特征提取与选择、数据集标记与分割、数据集标准化这些过程，经过这些处理过后得到的正负样本数据量均为 8169 条。

需要说明的是，本章后续的数字滤波去噪部分、数据分割部分中的仿真对比实验以及下一章的算法模型对比实验、边界函数测试实验中使用的输入数据

均与本部分的输入数据相同。

#### (4) 不同插值方式的仿真结果对比

仿真实验中分别采用二阶 B 样条曲线插值和三阶 B 样条曲线插值对皮肤电信号和心率信号进行插值处理,对比验证两种插值方式对算法训练结果的影响。

进行不同插值方式的对比实验时,控制其他条件相同,包括对数据集滤波去噪的方式均选用五点中值滤波,对数据集进行滑窗分割时选用的窗口重复率均为 80%,选用的机器学习算法模型均为逻辑回归算法。两种插值方式处理数据集的对比结果如表 2-2 所示:

表 2-2 不同插值方式处理数据的结果对比表

| 不同的插值方式     | 准确率  | 判断为发病的精确率 | 判断为未发病的精确率 |
|-------------|------|-----------|------------|
| 二阶 B 样条曲线插值 | 0.87 | 0.89      | 0.85       |
| 三阶 B 样条曲线插值 | 0.89 | 0.90      | 0.87       |

由此可以看出,无论是准确率、判断为发病的精确率还是判断为未发病的精确率,对数据集进行三阶 B 样条插值的结果都要比进行二阶 B 样条插值高,因此三阶 B 样条插值的效果更优。

### 2.3.2 数字滤波去噪

#### (1) 基本介绍

在使用癫痫监测装置采集生理信号数据的过程中,不可避免会掺入环境噪声,文献[30]中介绍了两种去除噪声的方式,一种是小波阈值去噪<sup>[30]</sup>,这种方式中最重要的一步是对原始数据小波分解,根据计算得到的阈值对各尺度分量过滤从而达到去噪效果,再进行重构得到需要的数据;另一种方式则是使用滤波算法对信号数据进行滤波。本课题中使用的是通过滤波算法滤波。

滤波过程是对信号其中部分频段信号进行抑制,通过另一部分频段信号。滤波算法按照信号保留的频段可以分为五类,分别是低通滤波算法、高通滤波算法、带通滤波算法、带阻滤波算法以及全通滤波算法<sup>[31]</sup>。低通滤波算法根据设定的截止频率会保留信号中低于此频率的部分或直流部分,抑制信号中高于此频率的部分,与之相反的是高通滤波算法。带通滤波算法有选择性的保留信号中的部分频段,当频率高于该频段或者低于该频段均会被抑制衰减,与之相反的是带阻滤波算法。全通滤波算法允许信号的全频段通过,一般作延时系统用。

在信号处理中常用的滤波方式有巴特沃斯滤波、切比雪夫滤波以及中值滤波。前两种滤波算法的幅频特性曲线各有特点，其中巴特沃斯滤波算法无论是在通频带还是在阻带上的曲线都是平坦的；与之不同的是，切比雪夫 I 型滤波算法在通频带上的曲线存在波动，而在阻带上是平坦的；切比雪夫 II 型滤波算法在通频带上的曲线是平坦的，在阻带上则存在波动，这些波动都是等波纹的。巴特沃斯滤波算法与切比雪夫滤波算法相比，过渡带衰减会比前者慢。本论文中中值滤波选用的是五点中值滤波，也就是在每 5 个点之间进行中值滤波处理，中值滤波可以很好的消除椒盐噪声。

由于本文中的原始数据中有皮肤电信号、心率信号和加速度信号，其中皮肤电信号和心率信号已经经过插值处理，所以是波形是比较光滑的，噪声信号很小以至于可忽略不计，因此省去滤波过程，后续只需要对加速度信号进行滤波去噪处理，查阅相关资料可知癫痫发作时频谱基本上会集中在 0-20Hz 范围内<sup>[32]</sup>，低通滤波满足要求。三种滤波算法处理一段连续 30 秒钟的加速度信号数据（共 600 条数据）的效果示例如图 2-4 所示，包含原始数据的波形以及三种滤波算法处理过的数据波形（左上图为原始信号的波形，右上图为巴特沃斯滤波效果图，左下的图为切比雪夫滤波效果图，右下的图为五点中值滤波效果图）。

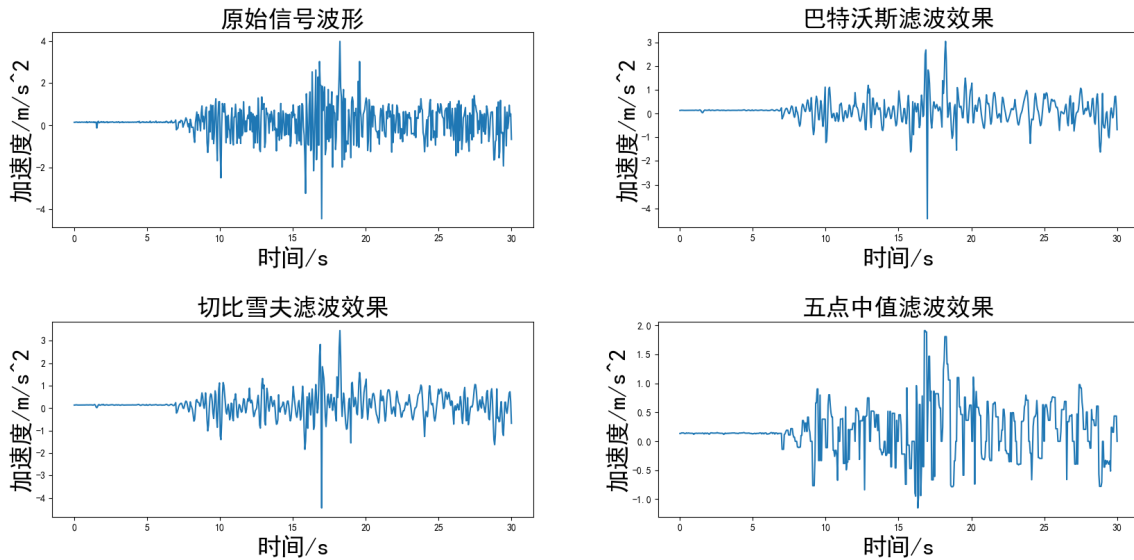


图 2-4 原始信号与滤波前后的波形对比图

Fig. 2-4 Waveform comparison between original signal and before and after filtering

## (2) 不同数字滤波方式的仿真结果对比

在仿真实验中分别使用巴特沃斯滤波、切比雪夫滤波以及五点中值滤波对加速度信号数据进行处理，通过对比算法训练模型的最终结果来选取最优的滤波器算法。

具体过程是：首先在这个对比实验中分别使用这三种滤波算法对数据集中的加速度信号进行处理，将处理过后的数据进行特征工程过后送达机器学习算法模型中训练，从而比较训练结果。同样地，在进行不同滤波算法的对比实验中时，需要控制其他条件相同，包括统一频率均采用三阶 B 样条插值，对数据集进行滑窗分割时选用的窗口重复率均为 80%，选用的机器学习算法模型均为逻辑回归算法。以上三种滤波算法处理数据集的对比结果如表 2-3 所示：

表 2-3 不同滤波算法处理数据的结果对比表

| 不同的滤波算法  | 准确率  | 判断为发病的精确率 | 判断为未发病的精确率 |
|----------|------|-----------|------------|
| 巴特沃斯滤波算法 | 0.87 | 0.89      | 0.86       |
| 切比雪夫滤波算法 | 0.86 | 0.89      | 0.85       |
| 五点中值滤波算法 | 0.89 | 0.90      | 0.87       |

由表 2-3 中的结果可知，无论是从准确率、判断为发病的精确率还是从判断为未发病的精确率来看，对数据集中的加速度信号进行五点中值滤波处理的效果是最好的，其次是巴特沃斯滤波算法，最后再是切比雪夫滤波算法，因此最终选择对数据集中的加速度信号进行五点中值滤波处理。

### 2.3.3 数据分割

#### (1) 基本介绍

为了便于后续的特征提取，需要对数据分割处理，本课题中使用滑窗的方式实现数据的分割。

癫痫监测手环采集加速度信号的频率为 20Hz，采集皮肤电信号和心率信号的频率虽然为 1Hz，但是已经进行了插值处理，所以也可以等同于 20Hz。由于后续需要依据 5 秒钟的生理数据进行发病判断，所以选取滑动窗口的时间为 5 秒钟，也就是说每个窗口有 100 个数据点。

还有另外一个参数——“窗口重复率”需要确定。“窗口重复率”表示在一个窗口中数据重复量占数据总量的比重，这是一个十分重要的参数，对后续模型训练的结果有影响。

#### (2) 不同窗口重复率的仿真结果对比

在仿真实验环节对“窗口重复率”进行了对比实验，对数据采用不同的“窗口重复率”进行滑窗分割处理，分别设置“窗口重复率”为 50%、80%，由最终模型训练结果来选择最合适的“窗口重复率”。



经过这两种滑窗处理后最终得到的数据集的容量大小不相等，窗口重复率为 80%时的数据量要更大。在这个对比实验中分别使用这两种窗口重复率对数据集进行滑窗，将最终得到的数据集送到算法模型中训练，比较模型的训练结果。与前两个对比实验类似，控制其他条件相同，即对数据统一频率使用的插值方式统一为三阶 B 样条插值，数字滤波处理使用的方式统一为五点中值滤波，选用的机器学习算法模型均为逻辑回归算法。使用以上两种窗口重复率对数据集滑窗处理的对比结果如表 2-4 所示：

表 2-4 不同窗口重复率对数据集滑窗的结果对比表

| 不同的窗口重复率 | 准确率  | 判断为发病的精确率 | 判断为未发病的精确率 |
|----------|------|-----------|------------|
| 80%      | 0.88 | 0.90      | 0.87       |
| 50%      | 0.81 | 0.82      | 0.80       |

通过表 2-4 中所示对比结果可以看出，当滑窗分割的窗口重复率为 80%时，模型预测的准确率、判断为发病的精确率和判断为未发病的精确率都要比窗口重复率为 50%时的要高，因此可得窗口重复率选择 80%的效果更优。

## 2.4 特征提取

### 2.4.1 特征提取

特征提取这个过程是模式识别问题中的重要一环，特征选取的合适与否决定了分类的准确率，因此是十分重要的。有些特征是可以直接测量出来的，这种特征属于原始特征，而有些特征则需要进行一些的计算才能得到。本课题中的数据是生理信号数据，这些数据本身没有代表性，其特征就需要经过一系列的运算得到。

特征可以分为时域特征、频域特征。时域特征是直接对时间序列进行处理运算得到的特征，一般来说比较直观。频域特征中最常见的是对时间序列进行傅里叶变化得到频域下的序列，再对新的序列进行处理运算得到。频域下的特征往往能更深刻地反映现象。

本课题主要参考文献[30]和文献[33]，其中时域特征大多是与数学统计运算相关的特征，包括平均值、方差、四分位距等等，共选取了 13 个时域特征。频域特征选取了一个，为峰值频点。频谱分析首先是对信号序列进行离散傅里叶变换，得到频域下的信号序列，然后选取其中的峰值点。对时间序列进行离散

傅里叶变换的公式如下：

$$DFT[h(nT_s)] = \sum_{n=0}^{N-1} h(nT_s) e^{-j \frac{2\pi mk}{N}} \quad (2-4)$$

其中， $h(nT_s)$  是函数  $h(t)$  的  $N$  个离散点， $T_s$  是时间间隔， $n = 0, 1, \dots, N-1$ ，这  $N$  个点可以看成是一个时间序列，且  $k = 0, 1, \dots, N-1$ 。

反映癫痫发作的加速度信号、皮肤电信号与心率信号既属于时序信号，也属于生物信号。生物信号具有不平稳的特征，反映在数值上则具有非线性的特性，因此加上非线性特征很有必要，能更好的反映生理类信号的特点。一些熵值是常见的非线性特征，熵的数值能评判一段信号序列的混乱程度，该序列的混乱程度越高，熵的数值随之越大，反之序列越不混乱熵的数值则随之越小。在本论文中使用的熵值有三种，分别是近似熵（Approximate Entropy, ApEn）、样本熵（Sample Entropy, SampEn）以及模糊熵（Fuzzy Entropy, FuzzyEn）。

已知一个时间序列  $\{x(i), i = 1, 2, \dots, n\}$ ，序列的标准差记为  $SD$ ，近似熵的计算步骤如下所示：

（1）划分序列，以  $m$  为窗使之分成  $n-m+1$  个，每一个划分后的序列为  $X_i(t) = (x_i(t), x_{i+1}(t), \dots, x_{i+m-1}(t))$ ；

（2）计算所有序列之间的距离，公式为： $d_{ij} = \max |x_{i+k}(t) - x_{j+k}(t)|$ ，这些距离可以形成一个矩阵  $D$ ，其中  $k = 0, 1, \dots, m-1$ ；

（3）首先确定系数  $r$ ，一般取  $0.1-0.25$ ，则阈值  $F = r \times SD$ ，把每一行中元素大于阈值  $F$  的个数占  $n-m+1$  的比例记为  $C_i^m(t)$ ，并计算这  $n-m+1$  个数的对数平均值，记为  $\Phi^m(t) = \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \ln C_i^m(t)$ ；

（4）令  $m = m+1$ ，重复步骤（1）到（3）；

（5）近似熵的计算公式为：

$$ApEn(t) = \Phi^m(t) - \Phi^{m+1}(t) \quad (2-5)$$

样本熵的计算步骤与近似熵类似，前两个步骤完全相同，故下面从第三步开始：

（3）确定阈值  $F = r \times SD$ ，把每一行中元素大于阈值  $F$  的个数占  $n-m$  的比例记为  $C_i^m(t)$ ，并计算这  $n-m$  个数的平均值，记为  $\Phi^m(t) = \frac{1}{n-m} \sum_{i=1}^{n-m} C_i^m(t)$ ；

（4）令  $m = m+1$ ，重复步骤（1）到（3）；

（5）样本熵的计算公式为：

$$SampEn(t) = \ln \Phi^m(t) - \ln \Phi^{m+1}(t) \quad (2-6)$$

同理，模糊熵的计算步骤与近似熵类似，前两个步骤完全相同，故下面从第三步开始<sup>[34]</sup>：

(3) 模糊隶属度的计算公式为： $D_{ij}^m = \mu(d_{ij}^m, n, r) = \exp(-\frac{(d_{ij}^m)^n}{r})$ ，再计算自身以外的全部隶属度的平均值： $\Phi^m(n, r) = \frac{1}{n-m} \sum_{j=1}^{n-m} (\frac{1}{n-m-1} \sum_{j=1, j \neq i}^{n-m} D_{ij}^m)$ ；

(4) 令  $m = m+1$ ，重复步骤 (1) 到 (3)；

(5) 模糊熵的计算公式为：

$$FuzzyEn(t) = \ln \Phi^m(t) - \ln \Phi^{m+1}(t) \quad (2-7)$$

本课题提取了与加速度信号、皮肤电信号、心率信号相关的共 51 个特征，其中三种信号各 17 个特征。除此之外还融合了腕动次数的特征，腕动次数包含 1 个特征，此特征为峰值频点，属于频域特征。17 个特征如表 2-5 所示。

表 2-5 17 个特征的介绍

| 特征缩写         | 特征名称       | 相关描述            |
|--------------|------------|-----------------|
| peak_to_peak | 峰峰值        | 数组元素中的最大值减最小值   |
| std          | 方差         | 数组元素的方差         |
| mean         | 平均值        | 数组元素的平均值        |
| energy       | 能量         | 见公式(2-8)        |
| iqr          | 四分位距       | 见公式(2-9)        |
| d1iqr        | 一阶差分的四分位距  | 一阶差分见公式(2-10)   |
| d2iqr        | 二阶差分的四分位距  | 二阶差分见公式(2-11)   |
| d1std        | 一阶差分的标准差   | 数组元素间一阶差分的标准差   |
| d2std        | 二阶差分的标准差   | 数组元素间二阶差分的标准差   |
| d1peak       | 一阶差分的峰峰值   | 数组元素间一阶差分的峰峰值   |
| d2peak       | 二阶差分的峰峰值   | 数组元素间二阶差分的峰峰值   |
| d1absmean    | 一阶差分的绝对值均值 | 数组元素间一阶差分绝对值的均值 |
| d2absmean    | 二阶差分的绝对值均值 | 数组元素间二阶差分绝对值的均值 |
| fpeak        | 峰值频点       | 相关的公式见公式(2-4)   |
| apEn         | 近似熵        | 见公式(2-5)        |
| sampEn       | 样本熵        | 见公式(2-6)        |
| fuzzyEn      | 模糊熵        | 见公式(2-7)        |

能量的公式如下所示：

$$energy = |mean|^2 \quad (2-8)$$

其中，mean 表示平均值。

四分位距的公式如下所示：

$$iqr = Q1 - Q3 \quad (2-9)$$

其中，Q1 表示第一四分位数，也就是位于数据中四分之一处的值，同理，Q3 表示位于数据中四分之三处的值。

一阶/二阶差分的公式如下所示：

$$1d_n = x_{n+1} - x_n \quad (2-10)$$

$$2d_n = x_{n+2} - x_n \quad (2-11)$$

其中， $x_n$  表示数据中的第  $n$  个样本。

## 2.4.2 特征选择

由于在特征提取过程中会存在选择的不确定性，有些提取的特征可能会是冗余特征，对算法模型的性能不会起到提升作用，甚至有些特征可能会对算法模型有负面作用，即可能会降低识别准确率，因此进行进一步的特征选择十分有必要。把冗余特征和对模型有负面作用的特征剔除可以增强模型的鲁棒性和泛化能力，并且能够减少算法模型的运算时间。

对特征进行选择往往选用以下方法：第一种方法是特征过滤，就是以特征的发散程度或者相关性为评判指标来评分，最终筛选出评分结果高的特征；第二种方法是特征筛选，即多次增加新特征或减少已有特征并运行模型，比较增减特征前后的模型性能从而来判断特征对模型的作用；第三种方法是嵌入法，即通过运行特定系列的算法模型来对特征的重要性进行综合评分，此时模型学习训练过程与特征选择是一起进行的。本论文选用嵌入法对特征进行选择，使用的算法模型是随机森林算法，最终可以得到所有特征的重要性评分，据此筛选评分靠后的特征。下面对随机森林的基础理论进行简要介绍。

随机森林应用非常广泛，是由多个基学习器组合而成的集成学习分类器，其中基学习器是决策树算法，最终分类结果取各个基学习器分类结果的众数，除此之外也可以对分类过程中的特征进行重要性评价。这里的“随机”包含两层含义，首先是训练集是随机的，即每一个基学习器的训练集使用 bootstrap 抽样方法得到，bootstrap 抽样是随机可放回抽样，因此可以保证每个决策树的训练集不尽相同；其次是特征是随机的，每一棵决策树的特征都是从所有的  $M$  个

特征中随机选出的  $m$  个特征，要求是  $m$  的值远远小于  $M$ 。经此操作构建的每一棵决策树都是有差异的，集成的学习器分类结果由各基学习器的分类结果投票表决<sup>[35]</sup>。

当特征集是多维时，即有多个特征，随机森林可对特征进行重要性评价，介绍原理之前首先引入基尼不纯度指数  $I$ ，公式如(2-12)所示。

$$I(j) = \sum_i p(y=i|\text{节点}=j)(1-p(y=i|\text{节点}=j)) \quad (2-12)$$

对于单个样本来说，这个指数表示了使用分支的概率分布随机选择选择标签时分类错误的概率大小。

对于特征  $x_i$  来说，其重要性评价值  $Z$  可以表示为式(2-13)。

$$Z(x_i) = \frac{1}{N_{\text{树}}} \sum_i \sum_k \frac{N_k}{N} \Delta I_{x_i} \quad (2-13)$$

其中， $N_{\text{树}}$  表示森林中决策树的个数， $N_k$  表示单个决策树中到达节点  $k$  的样本数， $N$  表示单棵树的样本数。

由此，每个特征的重要性均可以以数值形式表示，所以可以对所有特征的重要性进行排序进而对特征进行选择，以上就是随机森林算法的基本原理。

表 2-6 加速度信号的特征

| 加速度信号的特征（17 个） |       |       |        |         |           |
|----------------|-------|-------|--------|---------|-----------|
| peak_to_peak   | std   | mean  | energy | iqr     | d1iqr     |
| d2iqr          | d1std | d2std | d1peak | d2peak  | d1absmean |
| d2absmean      | fpeak | apEn  | sampEn | fuzzyEn |           |

表 2-7 皮肤电信号与心率信号的特征

| 皮肤电信号的特征（13 个） |        |           | 心率信号的特征（16 个） |        |           |
|----------------|--------|-----------|---------------|--------|-----------|
| peak_to_peak   | std    | mean      | peak_to_peak  | std    | mean      |
| energy         | iqr    | d1iqr     | energy        | iqr    | d1iqr     |
| d2iqr          | d1std  | d2std     | d2iqr         | d1std  | d2std     |
| d1peak         | d2peak | d1absmean | d1peak        | d2peak | d1absmean |
| d2absmean      |        |           | d2absmean     | apEn   | sampEn    |
|                |        |           | fuzzyEn       |        |           |

通过随机森林算法的训练，结果显示皮肤电的频域特征（**fpeak**）和三种熵特征（**apEn**、**sampEn**、**fuzzyEn**）的重要性比较低，分别为：0.0024、0.0026、0.0022、0.0035，并且心率的频域特征（**fpeak**）重要性比较低，为：0.0024，所以筛除掉这五个特征，剩下共 47 个特征，其中包括腕动次数的特征 **fpeak**，加速度信号的特征如表 2-6 所示，皮肤电信号的特征、心率信号的特征如表 2-7 所示。

综上所述，选取以上这 47 个特征进行后续的计算模型训练。

## 2.5 数据集处理

### 2.5.1 数据集标记与分割

要想让经过前面处理得到的特征集可以输入到算法模型函数中，还需进一步处理，首先需要对数据做标记，在这里对发病的特征集标记为 1，对未发病的特征集标记为 0。

标记好数据之后需要再对数据集进行分割处理，这是因为机器学习算法模型同时需要训练集与测试集，从而才可以输出机器学习的训练效果。在这里进行“三七分”，即将 70%的数据划入训练集，剩下的 30%的数据划入测试集。

无论是训练集还是测试集，发病样本的数据都会要比未发病样本的数据少不少，甚至不是同一个数量级，即产生了数据不平衡。数据不平衡会对算法训练结果产生非常大的影响，由于在这里发病的样本比未发病的少很多，如果不作处理那么结果显示的判断发病准确率会比实际的要低。因此需要处理数据不平衡的问题，可以在数据集处理的函数中需要控制发病样本的数据量与未发病样本的数据量保持一致。

### 2.5.2 数据集标准化

在本课题中，需要用到加速度信号、皮肤电信号与心率信号的特征来监测癫痫的发作，可是还存在一个问题就是这三种信号值的数量级不一样，大小相差非常大。正是因为如此，三种信号的特征值的数量级也会不一样，如果不做处理将会严重影响算法模型的训练结果，因此需要做标准化处理。经过标准化处理之后，可以使得数据集中的数据符合正态分布，也就是使得均值为 0，标准差为 1。标准化的公式如(2-14)所示。

$$z = \frac{x - \mu}{\sigma} \quad (2-14)$$

其中,  $z$  为进行标准化之后的数值,  $x$  为初始数值,  $\mu$  为初始数组的均值,  $\sigma$  为初始数组的标准差。

在实际编程中标准化使用的方法是 `sklearn` 中 `preprocessing` 提供的 `StandardScaler` 函数。

特征集在经过以上处理后, 就可以直接用于算法模型中。

## 2.6 本章小结

本章一开始介绍了课题的基本原理, 解释了选择用三轴加速度信号、皮肤电信号与心率信号来进行癫痫发作识别的原因。

在数据采集部分, 首先介绍了所使用的硬件, 即癫痫监测手环, 具体介绍了其中的传感器模块与硬件结构。其次描述了采集的数据来源与数据格式。

在数据预处理部分, 由于加速度信号的采集频率与皮肤电信号和心率信号不一致, 本文选用插值方式进行统一采样频率操作, 并且进行了仿真实验, 对比了二阶 B 样条曲线插值和三阶 B 样条曲线插值处理数据对最终结果的影响, 最终得出三阶 B 样条曲线插值效果最好。采集的数据中不可避免会存在噪声干扰, 本文选用滤波算法进行去噪, 在仿真实验中对比了巴特沃斯滤波、切比雪夫滤波以及中值滤波算法处理数据对最终结果的影响, 最终得出巴特沃斯滤波的效果最好。此时的数据并不能直接进行特征提取, 还需要对数据进行分割, 这里使用滑窗方式, 每个窗口包含 5 秒钟的数据, 即 100 个点, 在仿真实验中, 对比了窗口重复率为 80% 和 50% 两种情况下滑窗分割数据对最终结果的影响, 最终得出以 80% 作为窗口重复率的效果最好。

在特征提取过程中, 提取了时域特征和频域特征, 考虑到信号属于生物信号, 存在不平稳性和非线性的特点, 因此, 提取了包括近似熵、样本熵和模糊熵在内的非线性特征。之后使用了随机森林算法进行特征选择, 筛除了 5 个特征, 最终共选择了 47 个特征。

为了将数据处理成可以输入算法模型的形式, 需要对数据集进行标记与分割, 之后对数据集进行标准化操作, 标准化不仅可以统一量纲, 还可以提高算法模型的运行速度。至此, 数据预处理过程结束。

## 第 3 章 癫痫监测的数据处理算法研究

本课题在算法部分研究的目标是根据癫痫手环实时采集到的生理信号数据来识别癫痫是否发作，采用的实现实时监测的方法是：首先根据已经获取到的实验数据进行算法模型的学习训练，然后将实时采集的生理信号数据输入到训练好的模型中进行判断癫痫发作。这里的算法模型使用的是机器学习算法，本论文研究了逻辑回归算法、支持向量机算法、AdaBoost 算法以及 XGBoost 算法对癫痫监测的效果，本章将介绍这些算法的基础理论，并阐述算法模型训练的过程。

### 3.1 逻辑回归算法

#### 3.1.1 逻辑回归算法原理

逻辑回归（Logistic Regression），从字面上看有“回归”二字，但是该算法并不是回归算法，而是属于分类算法的范畴，在处理二分类问题中应用广泛，也可以推广到多分类问题的处理中。

逻辑回归算法的基础是线性回归算法，线性回归算法的模型可记为：

$$y = \omega^T x + b \quad (3-1)$$

在此基础上可引申出“广义线性回归模型”，可记为：

$$y = \varphi(\omega^T x + b) \quad (3-2)$$

其中  $\varphi(t)$  为单调可微的函数。

如果能找到这样一个函数  $\varphi(t)$  可以使模型的预测值  $y$  与分类问题中的标记值产生对应关系，则可以据此解决分类问题。通过单位阶跃函数可以使预测值与标记值产生对应关系，但是该函数不是可微的，因此不符合要求。于是选择了对数几率函数（Logistic Function），该函数的函数值只会接近 0 或接近 1，可以与标记值对应起来，曲线图如图 3-1 所示，通过曲线图可以看出趋势与单位阶跃函数非常接近，因此可以令此函数为  $\varphi(t)$ ，函数表达式如(3-3)所示<sup>[36]</sup>。

$$\varphi(t) = \frac{1}{1 + e^{-t}} \quad (3-3)$$



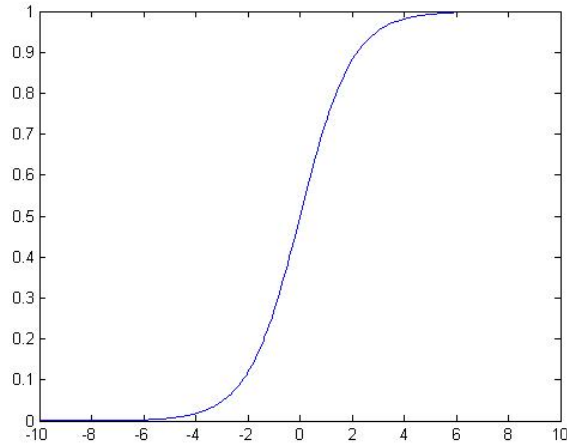


图 3-1 对数几率函数曲线图

Fig. 3-1 Log probability function curve

将式(3-3)代入(3-2)可得：  $y = \frac{1}{1 + e^{-(\omega^T x + b)}}$ ，由此可得：

$$\frac{y}{1-y} = e^{\omega^T x + b} \quad (3-4)$$

这个式中的  $y$  也可以看成是  $x$  为正样本的概率值，若  $y > 0.5$ ，可视为正样本，同理，若  $y < 0.5$  则视为负样本。概率形式如(3-5)、(3-6)所示。

$$P(Y=1|x) = \frac{e^{\omega^T x + b}}{1 + e^{\omega^T x + b}} \quad (3-5)$$

$$P(Y=0|x) = \frac{1}{1 + e^{\omega^T x + b}} \quad (3-6)$$

有了这种概率形式则可以很方便的参数  $\omega$  和  $b$  进行估计，在这里使用的是极大似然法求解。样本总数为  $n$ ，每一个样本点记为  $(x_k, y_k)$ ，这里  $k=1, 2, \dots, n$ ，令  $P(Y=1|x_k) = p_k$ ，则  $P(Y=0|x_k) = 1 - p_k$ ，并且令  $\beta = (\omega; b)$ ，令  $\bar{x} = (x; 1)$ ，所以对数似然函数如式(3-7)所示。

$$l(\beta) = \ln \prod_{k=1}^n [p_k^{y_k} (1-p_k)^{1-y_k}] = \sum_{k=1}^n [y_k \ln p_k + (1-y_k) \ln(1-p_k)] \quad (3-7)$$

化简可得：

$$l(\beta) = \sum_{k=1}^n [y_k (\beta^T \bar{x}_k) - \ln(1 + e^{\beta^T \bar{x}_k})] \quad (3-8)$$

接下来求  $l(\beta)$  最大时  $\beta$  的值，其中  $\beta$  中包括参数  $\omega$  和参数  $b$  的值。这时可

使用数值方法求解，往往使用梯度下降法或者牛顿法这样的迭代算法进行求解，由此可以求得参数的值。

### 3.1.2 算法模型训练

本课题中逻辑回归算法模型直接调用 `sklearn` 库中 `linear_model` 提供的 `LogisticRegression` 函数来实现。首先需要设定 `LogisticRegression` 函数中的参数，其中最主要的有两个参数，分别是 `C` 和 `solver`。`C` 参数是正则化系数的倒数，主要表征对模型的惩罚，该数值越大，对模型的约束程度越弱，在此设置为 10；`solver` 参数可以用来对模型进行提升优化，在此选择“`liblinear`”，并且这也是默认的参数，表示选择坐标轴下降法迭代方法对损失函数进行优化。

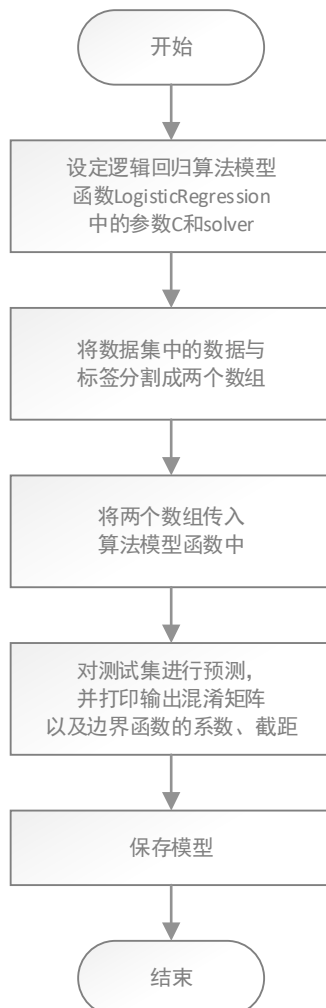


图 3-2 逻辑回归程序流程图

Figure 3-2 Logical Regression Program Flow Chart

具体步骤是将数据集中的数据与标签分割成两个数组，将含有数据的数组

与含有标签的数组分别传进模型中进行训练，并对测试集进行预测输出，使用模型对应的方法打印输出混淆矩阵，混淆矩阵可以很好的表示模型训练的效果。由于最终需要测试边界函数，所以还需要打印输出边界函数的系数值与截距。最后保存模型。逻辑回归算法的程序流程图如图 3-2 所示。

## 3.2 支持向量机算法

### 3.2.1 支持向量机算法原理

本节介绍的支持向量机（Support Vector Machines, SVM）不仅能处理分类问题，还能处理回归问题。与逻辑回归不同的是，支持向量机不仅在样本线性可分情况下可以使用，而且非线性可分的问题同样能够处理，应用场景更加广泛。需要注意的是，此算法在大样本的情况下，无论从时间复杂度还是空间复杂度来看效果都不是非常理想，所以此时不适合使用。

训练数据集是由原始输入数据映射到特征空间上的数据集，设有  $m$  个训练样本，记为  $X = \{x_1, x_2, \dots, x_n\}$ ，其中  $x_i$  是  $m$  维向量，且  $x_i \in R^m$ ，分类标签集为  $Y = \{y_1, y_2, \dots, y_n\}$ ，其中  $y_i$  只有两种取值，即 -1 和 1。这里假设数据集线性可分，需要找到一个超平面将两类正样本集（标签为 1 的样本）和负样本集（标签为 -1）的样本集进行分隔开，超平面的表达式记为  $\omega^T \cdot x + b = 0$ ，其中  $\omega$  为  $m$  维向量，即  $\omega = (\omega_1, \omega_2, \dots, \omega_m)$ 。原理图如图 3-3 所示<sup>[37]</sup>。

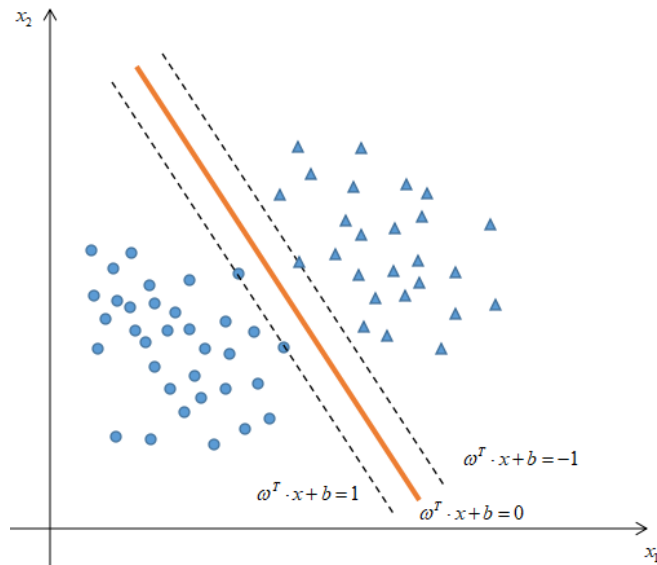


图 3-3 线性支持向量机原理图

Figure 3-3 Schematic diagram of linear support vector machine

图中圆形点表示正样本，三角形点表示负样本点。两个边界线（图中虚线

所示) 会包含若干个样本点, 这些样本点则称为支持向量, 这两条线称为间隔边界。为了保证分类预测尽可能准确, 需要使得两条间隔边界之间的距离最大。

由边界的方程可知, 两条边界的距离为:  $d = \frac{2}{\|\omega\|}$ , 为了使  $d$  最大, 就是需要保证  $\frac{1}{\|\omega\|}$  最大, 则需要使得  $\|\omega\|^2$  最小, 于是可以得到以下式子:

$$\begin{cases} \min & \frac{1}{2}\|\omega\|^2 \\ \text{s.t.} & y_i(\omega^T \cdot x_i + b) \geq 1 \end{cases} \quad (3-9)$$

其中,  $i=1, 2, \dots, n$ .

上面描述的间隔超平面属于硬间隔, 有些时候样本数据集不能线性可分, 即式(3-9)中的约束条件不能对所有样本点均满足, 需要去除一部分样本点才能够满足线性可分条件, 因此不能很好地分隔正负样本, 此时则有必要引入软间隔, 也就是引入松弛变量 (Slack Variable), 式 (3-9) 则变为:

$$\min \quad \frac{1}{2}\|\omega\|^2 + C \sum_i \zeta_i \quad (3-10)$$

其中  $C$  称为惩罚参数, 约束条件相应的转变为:

$$y_i(\omega^T \cdot x + b) \geq 1 - \zeta_i \quad (3-11)$$

其中,  $i=1, 2, \dots, n$ ,  $\zeta_i \geq 0$ .

软间隔的引入使得模型具有更好的灵活性和鲁棒性, 可以将能处理的数据集扩展到线性不可分的数据集。于是式(3-10)和(3-11)则变成了凸二次规划 (Convex Quadratic Programming) 问题, 这个问题中  $\omega$  的解是唯一可确定的, 而  $b$  的解存在一个范围, 并不是唯一可确定的。接下来使用拉格朗日乘子法求解, 于是可转化成对偶问题。拉格朗日形式如式(3-12)所示:

$$L(\omega, b, \zeta, \alpha, \mu) = \frac{1}{2}\|\omega\|^2 + C \sum_{i=1}^n \zeta_i - \sum_{i=1}^n \alpha_i (y_i(\omega \cdot x_i + b) - 1 + \zeta_i) - \sum_{i=1}^n \mu_i \zeta_i \quad (3-12)$$

其中,  $\alpha_i \geq 0$  且  $\mu_i \geq 0$ .

由(3-12)求解  $\omega$  以及  $b$  和  $\zeta$  的极小值并转化代入原式可得式(3-13):

$$\min_{\omega, b, \zeta} L(\omega, b, \zeta, \alpha, \mu) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (3-13)$$

由此可得对偶问题如式(3-14)所示:

$$\max_{\alpha} \quad \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (3-14)$$

约束条件为：  $\sum_{i=1}^n \alpha_i y_i = 0$ ，  $0 \leq \alpha_i \leq C$ ， 其中，  $i = 1, 2, \dots, n$ 。

转化为对偶问题过后就非常适合使用核技巧（Kernel Trick），即引入核函数，由此支持向量机能够方便地处理样本线性不可分的问题。常用的核函数有径向基核、多项式核和 Sigmoid 核，除此之外还可以自定义核函数，这些核函数必须为正定核，以上介绍的三种核函数如下所示。

径向基核函数：  $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$ ， 其中  $\gamma$  为参数。

多项式核函数：  $K(x_i, x_j) = (x_i^T \cdot x_j + r)^c$ ， 其中  $r$  与  $c$  均为参数。

Sigmoid 核函数：  $K(x_i, x_j) = \frac{1 - \exp(-2(\gamma x_i^T \cdot x_j + r))}{1 + \exp(-2(\gamma x_i^T \cdot x_j + r))} = \tanh(\gamma x_i^T \cdot x_j + r)$ ， 其中  $\gamma$

与  $r$  均为参数。

这时，可以将对偶形式中的内积替换成核函数，目标函数即为式(3-15)所示。

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3-15)$$

由此可以方便地求解其中的参数和构造决策函数，这里面求解的参数为  $\alpha^*$  和  $b^*$ ，决策函数为  $\varphi(x) = \text{sign}(\sum_{i=1}^n \alpha_i^* y_i K(x_i, x_j) + b^*)$ 。

### 3.2.2 算法模型训练

本课题中支持向量机算法模型直接调用 sklearn 库中 svm 提供的 SVC 函数来实现。首先使用网格搜索法来选取局部最优的函数参数，然后将选取的最优参数回代入 SVC 函数中，在这里还有一个比较重要的参数需要手动指定，那就是“kernel”参数，在这里设置为“linear”，表示选取的是线性核函数，这样最终可以得到一个线性超平面来进行分类。

剩下的步骤与逻辑回归算法模型类似，即先将数据集进行分割，再将分割好的数据数组与标签数组传入模型中进行训练，对测试集进行预测输出，最终打印输出混淆矩阵、边界函数的系数与截距，并保存模型<sup>[38]</sup>。支持向量机算法的程序流程图如图 3-4 所示。

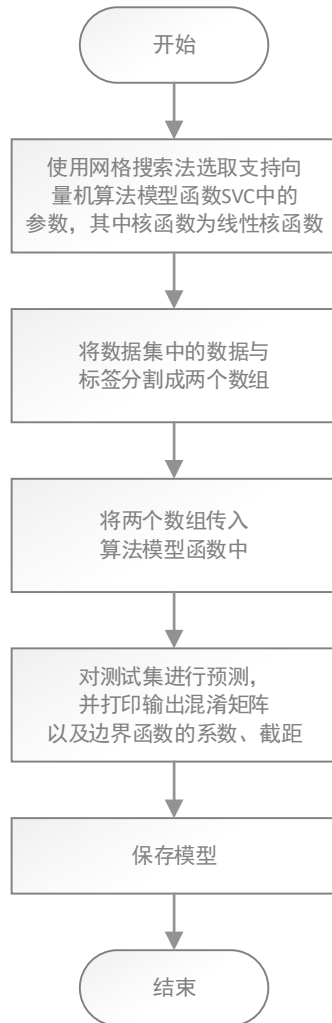


图 3-4 支持向量机程序流程图

Figure 3-4 support vector machine program flow chart

### 3.3 AdaBoost 算法

#### 3.3.1 AdaBoost 算法原理

AdaBoost 算法 (Adaptive Boosting) 属于集成学习的范畴，这种算法的工作原理是由数个基学习器 (弱学习器) 来组合成强学习器。数据集放进单个弱学习器中进行训练得到的结果可能不理想，但根据该算法组合成的强学习器处理这种数据集往往会得到不错的效果。

AdaBoost 算法大致分为以下三个步骤：

第一步，假设数据集包含  $n$  个样本，特征数据集为  $\{x_1, x_2, \dots, x_n\}$ ，每一个样本的标签  $y_i$  取 +1 或 -1，其中  $i=1, 2, \dots, n$ 。首先对每个样本的权值进行初始化，

记为  $W_1 = (\omega_{11}, \omega_{12}, \dots, \omega_{1n})$ ，其中每个分量均为  $\frac{1}{n}$ ，即在第一轮中保证每个样本起

的作用都一样，在此基础上学习第一个基学习器  $G_1(x)$ 。

第二步，确定基学习器的个数  $m$ ，总共需要学习  $m$  次基学习器，其中在第一步中已经学习过第一个基学习器，因此接下来需要反复学习  $m-1$  次基学习器，每次学习的步骤都相同，假设是第  $j$  次学习，则需要进行以下操作：首先计算第  $j$  个基学习器  $G_j(x)$  在使用样本权值  $W_j$  加权过后的数据集上的分类误差率

$e_j = \sum_{G_j(x_i) \neq y_i} \omega_{ji}$ ，该式表明了分类误差率与样本权值之间的关系；接下来计算基学

习器的系数  $\alpha_j$ ，该系数满足  $\alpha_j = \frac{1}{2} \log(\frac{1}{e_j} - 1)$ ，由此式可以看出，分类误差率越

小，基学习器的系数越大，这也就表明该基学习器的重要性越大；最后需要及

时更新样本的权值，权值更新公式为  $\omega_{j+1,i} = \frac{\omega_{ji} e^{-\alpha_j y_i G_j(x_i)}}{\sum_{i=1}^n \omega_{ji} e^{-\alpha_j y_i G_j(x_i)}}$ ，通过此式可知，分

类误差率越大，则权值越大，也就是若分类错误，权值会随之增大，算法更重视处理错误分类的情况<sup>[39]</sup>。

第三步， $m$  个权值  $\alpha_j$  与  $m$  个基学习器  $G_j(x)$  进行线性组合，将需要预测的数据集放进该线性组合式中进行运算从而可以得到判断结果，这也是由  $m$  个弱学习器转换成一个强学习器的过程。

### 3.3.2 算法模型训练

本课题中支持向量机算法模型直接调用 `sklearn` 库中 `ensemble` 提供的 `AdaBoostClassifier` 函数来实现的。该函数中有三个比较重要的参数，第一个重要的参数是“`base_estimator`”，该参数用来设置使用哪种基学习器，在 `sklearn` 中的分类器均可以被选择，在这里常用的学习器是 `CART` 决策树，该参数的默认值也是这种学习器，即默认选择“`DecisionTreeClassifier`”；第二个重要的参数是“`n_estimators`”，该参数用来设置基学习器的迭代次数，这个值设置大了可能会出现过拟合，如果设置小了可能会出现欠拟合，因此需要选择合适的迭代次数，该参数的默认值为 50；第三个重要的参数是“`learning_rate`”，这个参数表示学习率，可以表示基学习器的所占权重，也反映了梯度的收敛速度，如果设置小了会让收敛速度比较慢，设置大了有可能会与最优解错过，该参数的默认值为 1.0，表示每个学习器被认为具有相等的重要性。

在这里均使用函数的默认值，即基学习器选用 CART 决策树，学习器的选择迭代 50 次，学习率为 1.0。将经过特征工程处理好的数据集中 70% 的数据输入到算法模型中进行学习训练，并预测其中 30% 的数据，打印混淆矩阵，得到算法模型预测的准确率与召回率。

## 3.4 XGBoost 算法

### 3.4.1 XGBoost 算法原理

XGBoost 算法 (Extreme Gradient Boosting) 也属于集成学习中的 boosting 算法，是由陈天奇等人研究开发出的一种提升树模型，该算法的性能较 GBDT 算法而言有了很大的提升，在机器学习相关的算法竞赛以及工业中应用很多并且使用效果非常好。

XGBoost 算法由很多决策树集成，第  $k$  棵决策树的函数为  $f_k(x)$ ，则样本  $x_i$  的预测值表达式如式(3-16)所示：

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (3-16)$$

其中， $K$  表示决策树分类器的个数。

目标函数由损失函数和正则化函数组成，如式(3-17)所示：

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) + Const \quad (3-17)$$

其中， $Const$  表示常数项，正则化函数项减少了模型的复杂度，防止过拟合。

前一次预测的结果结合新加入的树可以得到加入新决策树过后的预测结果，如式(3-18)所示，因此不断加新决策树的过程是不断减少损失函数的过程。

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (3-18)$$

因此目标函数可表示为：

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) + Const \quad (3-19)$$

将第一项用泰勒展开式展开到第二项，由此可得：

$$Obj^{(t)} \approx \sum_{i=1}^n \left[ l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) + Const \quad (3-20)$$

其中， $g_i$  和  $h_i$  分别为  $l(y_i, \hat{y}_i^{(t-1)})$  对  $\hat{y}_i^{(t-1)}$  的一阶导和二阶导。



由此不断加入新的决策树的过程中尽可能减少目标函数的值，控制损失函数值在较低水平从而提高模型性能。并且算法支持并行化，训练效率高<sup>[40]</sup>。

### 3.4.2 算法模型训练

本课题中支持向量机算法模型直接调用 `xgboost` 库中提供的 `XGBClassifier` 函数来实现的。该函数有三个重要的参数需要设置，首先是“`objective`”参数，这个参数用来选择损失函数，默认值为“`linear`”；其次是“`seed`”参数，通过设置这个参数来选择随机数的种子，默认值为 0；还有“`eta`”参数表示学习率，用来表示每一步的权重，默认值为 0.3。

本文仿真实验中使用的参数均为默认参数，原始数据集经预处理和特征工程过后输入到 `XGBoost` 学习器，并打印混淆矩阵，得到算法模型预测的准确率与召回率<sup>[41]</sup>。

### 3.5 算法模型对比实验

对数据统一频率使用的插值方式统一为三阶 B 样条插值，数字滤波处理使用的方式统一为五点中值滤波，分割滑窗中使用的窗口重复率统一为 80%，分别对经过预处理和特征工程的数据集进行四种机器学习算法模型训练。

逻辑回归算法、支持向量机算法、`AdaBoost` 算法与 `XGBoost` 算法训练的结果对比如 3-1 表所示。

表 3-1 四种算法运行结果对比表

| 算法运行结果                        |        | 精确率  | 召回率  | 准确率  |
|-------------------------------|--------|------|------|------|
| 逻辑回归算法                        | 判断为发病  | 0.90 | 0.86 | 0.88 |
|                               | 判断为未发病 | 0.87 | 0.90 |      |
| 支持向量机（ <code>linear</code> 核） | 判断为发病  | 0.89 | 0.87 | 0.88 |
|                               | 判断为未发病 | 0.87 | 0.90 |      |
| 支持向量机（ <code>rbf</code> 核）    | 判断为发病  | 0.96 | 0.95 | 0.97 |
|                               | 判断为未发病 | 0.95 | 0.96 |      |
| <code>AdaBoost</code> 算法      | 判断为发病  | 0.93 | 0.90 | 0.92 |
|                               | 判断为未发病 | 0.91 | 0.93 |      |
| <code>XGBoost</code> 算法       | 判断为发病  | 0.97 | 0.98 | 0.98 |
|                               | 判断为未发病 | 0.98 | 0.97 |      |

通过表 3-1 可以得出，无论是准确率、精确率还是召回率，逻辑回归算法与支持向量机算法（线性核）的训练结果几乎没有差异，AdaBoost 算法的训练结果要比前两种算法更优，支持向量机算法（rbf 核）的训练结果要比前三种算法更优，XGBoost 算法的训练效果是最优的，使用 XGBoost 监测癫痫发作的准确率和召回率均是最高的。

需要注意的是，在这个实验中训练得到的算法模型均保存到一个文件夹中，后续进行测试实验过程中直接调用这个文件夹中对应的模型文件，这样保证了前后使用的模型是一致的。

## 3.6 测试实验

### 3.6.1 边界函数测试实验

在工程运用中，需要尽可能减少运算成本，不能每次都对大量数据进行模型训练，不仅时间成本高，而且内存消耗也大。虽然使用之前保存的模型进行癫痫监测不会造成运算成本，但不方便移植到癫痫手环上的嵌入式系统中。如果训练得到的模型能够用函数表达式来表示，这样就可以将函数表达式移植到癫痫手环中，十分方便。无论运行逻辑回归算法还是支持向量机算法，都能得到可以用表达式表示的模型结果，因此可以由输出的表达式中的系数和截距得到边界函数。由于 AdaBoost 算法与 XGBoost 算法都属于集成算法，虽然它们的最终模型是由若干个基学习器加权组合而成，但由于基学习器使用的是默认的 CART 决策树模型，不能用确切的函数表达式来表示，因此这训练两个算法并不能得到边界函数。逻辑回归算法模型的边界函数是一个线性函数，当支持向量机算法使用的核函数是线性核函数时，其边界函数也是一个线性函数，当支持向量机使用的核函数是径向基函数时，其边界函数则是一个非线性函数，但是也能确切表示出来。

误报率，表示假正类样本（即将负类样本划分到正类这种情况，本文中发病为正类，未发病为负类）在所有样本中的占比；漏报率，表示假负类样本（即将正类样本划分到负类这种情况）在所有样本中的占比。准确率越高越好，误报率和漏报率越低越好，在此处以提高准确率和降低误报率与漏报率为目标。

通过训练相关模型得到的边界函数来判断癫痫发作情况，结果如表 3-2 所示。

通过这个表可以得出，支持向量机（线性核）的边界函数判断癫痫发作的准确率要比逻辑回归算法高，漏报率要比逻辑回归算法低，误报率比逻辑回归

算法略高。总体来说，支持向量机（线性核）的边界函数进行癫痫监测的效果最优。

表 3-2 边界函数判断癫痫发作结果对比表

| 边界函数判断结果        | 准确率     | 误报率     | 漏报率     |
|-----------------|---------|---------|---------|
| 逻辑回归算法          | 0.87901 | 0.05033 | 0.07065 |
| 支持向量机（linear 核） | 0.88498 | 0.05141 | 0.06361 |

### 3.6.2 临床数据验证

由于 10 名患者的数据已经用来进行算法的学习训练了，因此在这个测试实验中重新选取了其他 2 名患者的生理信号数据集作为样本，这里将发病数据集和未发病的数据集分开来测试，在采集这 2 名患者数据的过程中一共监测到了 4 次癫痫发作。由于癫痫发作状况检测出来越迅速就越能保护患者的生命健康安全，所以在短时间内根据生理信号数据判断出发病情况非常有必要。本小节研究了根据 5 秒钟的生理信号数据集检测癫痫发作情况，也就是以 5 秒钟的数据为一个窗口，这个窗口中包含 100 条数据，根据这 100 条数据来判断癫痫发作情况。四种算法的运行结果如表 3-3 所示。

表 3-3 四种算法的运行结果对比表

| 算法运行结果          | 准确率     | 判断为发病的精确率 | 判断为未发病的精确率 |
|-----------------|---------|-----------|------------|
| 逻辑回归算法          | 0.64142 | 0.45567   | 0.97978    |
| 支持向量机（linear 核） | 0.70124 | 0.49145   | 0.97512    |
| 支持向量机（rbf 核）    | 0.62741 | 0.33437   | 0.99844    |
| AdaBoost 算法     | 0.68964 | 0.44479   | 0.99533    |
| XGBoost 算法      | 0.72153 | 0.50544   | 0.99844    |

通过上表可以看出，从判断为发病的精确率这一指标来看，四种算法检测的结果都不是很理想，准确率仅能保持在 50%左右；从判断为未发病的精确率这一指标来看，检测的结果都比较理想，都能超过 95%。从准确率来看，四种算法的训练效果从好到差依次是：XGBoost 算法、支持向量机（线性核）、AdaBoost 算法、逻辑回归算法、支持向量机（rbf 核）；从判断为发病的精确率

来看，四种算法的训练效果从好到差依次是：XGBoost 算法、支持向量机（线性核）、逻辑回归算法、AdaBoost 算法、支持向量机（rbf 核）；从判断为未发病的精确率来看，四种算法的训练效果从好到差依次是：XGBoost 算法、支持向量机（rbf 核）、AdaBoost 算法、逻辑回归算法、支持向量机（线性核）。综上所述，在临床数据验证中使用 XGBoost 算法的效果最好。

由于判断发病的精确率都不高，因此这个问题有待进一步研究，所以接下来将进行根据连续 20 分钟数据集来判断癫痫发作的测试。

选取其中一名患者的数据集，这名患者在监测过程中共有 2 次癫痫发作，并且这两次发作时间中间间隔数小时，因此选取 2 段时长 20 分钟的连续数据集，这两段数据集中分别包括两段癫痫发作时间在内，分别记为时间段 1 和时间段 2，如表 3-4 所示：

表 3-4 两个时间段的示意表

| 序号 | 测试时间段               | 其中的发病时间段            |
|----|---------------------|---------------------|
| 1  | 20:30:00 ~ 20:50:00 | 20:38:44 ~ 20:40:45 |
| 2  | 22:50:00 ~ 23:10:00 | 22:56:51 ~ 23:04:03 |

两段数据分析的过程和最终结论都类似，下面以分析时间段 1 中的数据为例，图 3-5 为这些数据三种信号的波形图，其中最上面的波形为皮肤电信号的波形图，中间的波形为心率信号的波形图，最下面的波形为加速度信号的波形图，图中的方框表示方框两侧边界以内的数据均为发病数据。

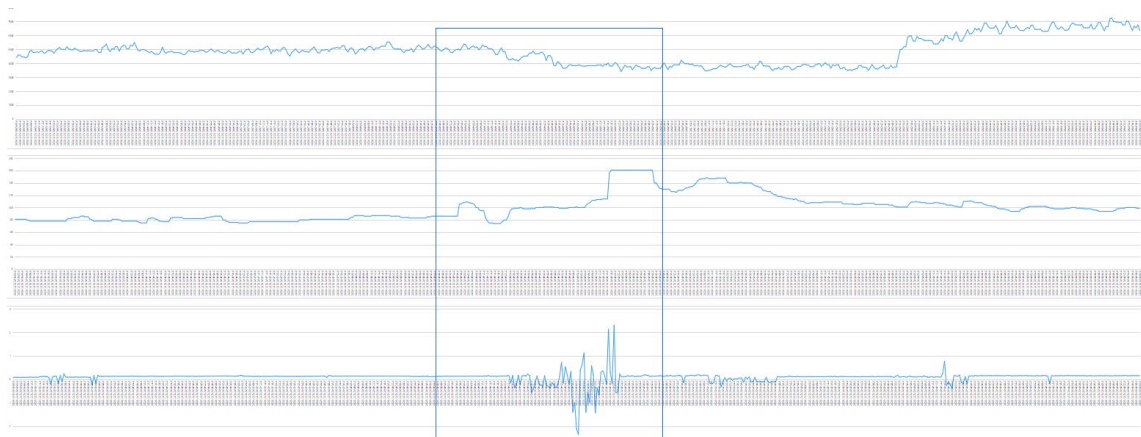


图 3-5 时间段 1 中数据三种信号的波形图

Figure 3-5 Waveforms of the three signals in time period 1

因为采用的窗口重合率为 80%，因此一共会有 1195 个窗口，由此产生对应

的预测结果，其中发病标记为 1，未发病则标记为 0。根据表 3-4 中的发病时间段可知，第 524 - 624 个窗口属于发病数据，若用逻辑回归算法预测，第 533 - 699 个窗口中七成都会被标记为 1；若用支持向量机算法预测，第 533 - 699 个窗口中七成都会被标记为 1；若用 AdaBoost 算法预测，第 543 - 701 个窗口中七成都会被标记为 1；若用 XGBoost 算法预测，第 526 - 689 个窗口中七成都会被标记为 1。由上述结果可知，使用 XGBoost 算法预测的效果是最好的。可以发现，这些预测为 1（即预测发病）的数据段几乎都是在图 3-5 加速度波形中波动幅度大的部分，因此有理由说明，在 47 个特征中加速度的时域特征所占的权重比较大；并且预测为 1 的数据量几乎在对应的大窗口中都是七成，这也可以解释表 3-3 中判断发病的精确率都在 50%左右。

### 3.7 本章小结

本章将围绕癫痫监测的动态数据算法展开，共研究了四种机器学习算法，其中包括逻辑回归算法、支持向量机算法、AdaBoost 算法与 XGBoost 算法。逻辑回归算法的理论基础是线性回归算法，但解决的不是回归问题，而是分类问题；支持向量机主要是得到一个超平面分割正负样本点，支持向量是两个边界上的点；后两者属于集成学习算法，这两种学习器的机理都是由若干个基学习器以训练好的权重组合而成。

每个算法的研究过程以及在本论文中的介绍方法都是类似的，也就是先介绍相关算法的基础理论，然后详细阐述该算法在本课题中的训练过程。

在算法模型对比仿真实验中，对比了逻辑回归算法、支持向量机（线性核）、支持向量机（rbf 核）、AdaBoost 算法与 XGBoost 算法训练数据对最终结果的影响，最终得出 XGBoost 训练数据的效果最好。

最后进行了测试实验，测试了边界函数预测癫痫发作，并进行了临床数据验证。在边界函数的测试实验中，得出支持向量机（rbf 核）的边界函数预测癫痫发作的效果最优；在临床数据验证中，得出使用 XGBoost 算法进行癫痫预测的效果最优。

## 第4章 癫痫监测系统软件设计与开发

除了癫痫监测的动态数据算法研究之外，开发癫痫监测系统配套的软件也是一项十分重要的工作。算法分析所需要的原始数据，需要从数据库中获取到，那么将数据传输到数据库中以及从数据库中下载下来分析的过程都需要配套软件的支持。患者佩戴手环采集生理数据到数据上传至服务器，再到下载分析数据，这一系列过程体现在数据流上，主体包括癫痫手环、云服务器以及手机。除此之外，配套软件还可以支持其他一系列功能，比如管理员（比如算法开发人员、医生）的登录验证，普通用户的登录验证，数据分析（在后台对数据进行简单处理得到前台可视化所需要的数据），合作医生信息介绍、健康推送与新闻推送、产品介绍、APP 操作介绍等功能所需要的增删改查操作。本课题开发了云平台软件与手机安卓端软件，本章将对这部分内容展开介绍。

### 4.1 癫痫监测软件功能模块设计

本项目中的软件开发将分解为多个功能模块分别进行开发，也即采用层次化的设计思想<sup>[42]</sup>。

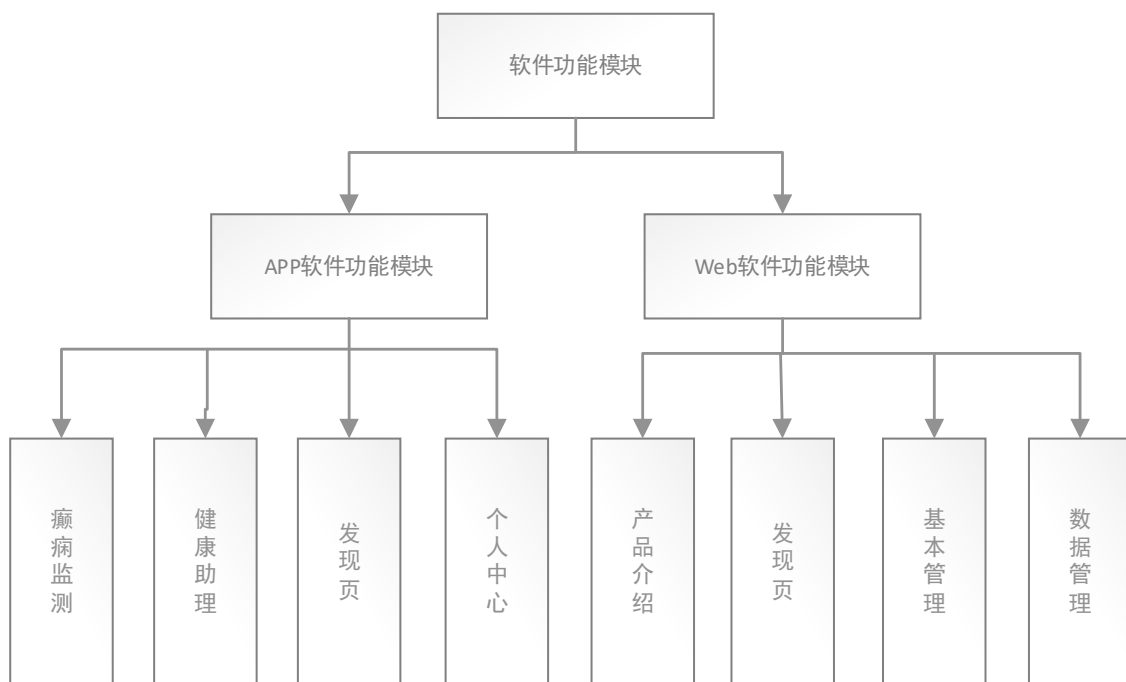


图 4-1 功能模块图

Figure 4-1 Function module diagram

癫痫监测软件具体包括 APP 软件和云平台，APP 软件和云平台有分别包含多个功能模块。功能模块图如图 4-1 所示。下面具体介绍各模块<sup>[43]</sup>。

#### （1）癫痫监测模块

癫痫监测模块主要包括发病监控、心率分析、运动统计、睡眠分析这四个模块。发病监控主要用来显示指定时间范围内的发病情况，包括发病记录数据以及统计图；运动统计与睡眠分析模块也是如此，显示指定时间范围内的数据并给出统计图。至于心率分析模块，用来显示半小时内心率有关的统计值，以及统计图实时更新心率值，并且提供了不同年龄段和静息心率所对应的运动心率区间给用户作参考。

#### （2）健康助理模块

健康助理模块主要包括医生推荐、产品推荐这两个模块。其中，医生推荐模块用来给患者推荐合适的医生以及医院，并提供了医生与医院的相关信息；产品推荐模块用来向患者推荐癫痫监测手环以及合适的药品，以及提供相关产品的信息。

#### （3）发现页模块

发现页主要包括健康新闻、健康推送这两个模块。其中，健康新闻模块可以提供与健康相关的新闻给用户阅读，让用户第一时间了解；健康推送模块提供与健康相关特别是与癫痫相关的推送或者说是健康常识给用户学习了解。

#### （4）个人中心模块

个人中心主要包括设置和软件升级这两个模块。其中，设置模块主要包括修改账户密码、变更手机号以及其他基本信息等功能，软件升级模块包括升级 APP 安装包以及升级癫痫监测手环的固件两个功能。

#### （5）基本管理模块

基本管理模块主要包括网页与 APP 信息编辑、用户管理和通信管理这三个模块。其中，网页与 APP 信息编辑模块包括增删改查网页与 APP 软件上所显示文字图片信息功能；用户管理模块包括用户数量统计、用户信息检索以及用户意见反馈等功能；通信管理包括上传 APP 安装包、编辑 APP 安装包信息、上传癫痫监测手环固件包以及编辑固件包信息等功能。

#### （6）数据管理模块

数据管理模块主要包括统计数据检索、癫痫记录数据检索、秒级生理数据检索（秒级生理数据指的是每秒钟的所有生理信号数据集合）、算法模型训练这四个模块。其中，统计数据检索、癫痫记录数据检索、秒级生理数据检索这三个模块可以方便管理员根据相关患者账号以及时间范围检索得到对应的生理

数据，相关生理数据显示在表格与统计图当中，以便直观的观察其中规律，并且提供导出 Excel 按钮，管理员可以将检索得到的生理数据导出成一个文件，使得相关算法工程师或者医生可以方便使用数据；算法模型训练模块可以根据检索用户账号、发病时间范围与未发病时间范围让服务器运行相关算法，训练出模型，并将运行的结果显示在网页当中。

## 4.2 数据库设计

云服务器中的数据存储模块采用 MySQL 关系型数据库管理系统。数据库中主要存储用户的基本信息以及用户的生理数据。本小节将介绍数据库的 E-R 模型以及数据库表的结构。

### 4.2.1 数据库的 E-R 模型

E-R 模型，又称为实体联系模型，在数据库设计的前期阶段进行数据建模，通过 E-R 模型可以清晰的看出数据间的联系，为之后的数据库设计提供便利<sup>[44]</sup>。

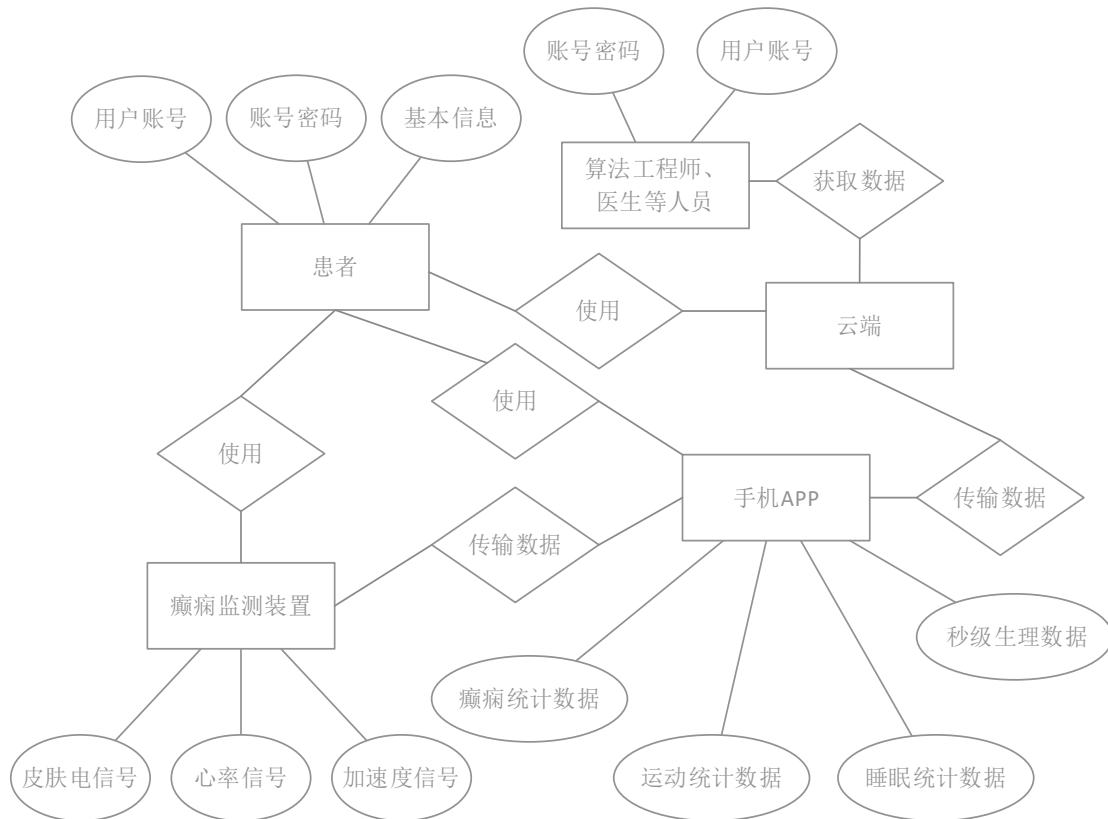


图 4-2 数据库 E-R 图

Figure 4-2 E-R Diagram of Database



本项目中共包含以下实体：患者、算法工程师或医生等人员、癫痫监测装置、手机端软件、云端软件。数据库 E-R 图如图 4-2 所示。从图中可以看出这五个主要实体之间的关系，患者包括用户账号、账号密码以及一些基本信息，使用癫痫监测装置、手机端软件与云端软件；癫痫监测装置采集皮肤电信号、心率信号与加速度信号，给手机端传输数据；手机端软件包含处理分析统计数据、运动统计数据、睡眠统计数据以及秒级生理数据的功能，可以向云端传输数据；云端服务器保存了生理信号数据，算法工程师或医生等人员可以从云端获取所需要的原始数据，从而算法工程师可以依次研究算法，医生可以依次分析数据，给患者进行诊断。

#### 4.2.2 数据结构设计

在 E-R 模型的基础上进一步设计数据库的数据结构，确定出表中字段的名称、数据类型以及约束一系列信息。以下列出常用的 8 个表的基本信息，如表 4-1 所示。

表 4-1 常用表的基本信息

| 序号 | 表名                    | 描述信息         |
|----|-----------------------|--------------|
| 1  | table_consult         | 存储在线问诊模块医生信息 |
| 2  | table_epilepsy_record | 存储癫痫记录数据     |
| 3  | table_epilepsy_stat   | 存储癫痫统计数据     |
| 4  | table_feedback        | 存储用户反馈信息     |
| 5  | table_sleep_stat      | 存储睡眠统计数据     |
| 6  | table_sport_stat      | 存储运动统计数据     |
| 7  | table_user            | 存储用户信息数据     |
| 8  | table_normal_second   | 存储秒级生理数据     |

表 table\_consult 是与在线问诊模块医生信息相关的表，在线问诊模块在 APP 中，在网页中没有此模块，包含问诊医生 id 号、医院名称、医院简介、医生姓名、医生描述、医生联系方式等关键字段。其中，id 表示问诊医生 id 号，此值不能重复，设置 UNIQUE 键。

表 table\_epilepsy\_record 是与癫痫记录相关的表，包含用户账号、皮电、心率、加速度、癫痫发作标记、发病日期等关键字段。

表 table\_epilepsy\_stat 示与癫痫统计数据相关的表，包含用户账号、全天发

病次数、发病最长时间、发病最长时间、首次发病时间、末次发病时间、全天发病时间点明细、发病日期等关键字段。

表 `table_feedback` 表示与用户反馈信息相关的表，包含序号、用户账号、反馈正文、联系人邮箱、反馈时间、处理意见等关键字段。其中 `fdId`（序号）没有实际意义，仅用来在数据库中占用位置，为主键。

表 `table_sleep_stat` 是与睡眠统计数据相关的表，包含用户账号、总睡眠时长、深睡时长、浅睡时长、浅睡时长、记录日期等关键字段。

表 `table_sport_stat` 是与运动统计数据相关的表，包含用户账号、步数、久坐次数、记录日期等关键字段。

表 `table_user` 是与用户数据相关的表，包含用户的 ID、用户账号、用户密码、用户姓名、性别、身高、体重、出生日期、监护人姓名、监护人手机号、注册时间等关键字段。其中，`uId` 表示用户的 ID，也有定位在数据库中位置的作用，为主键。

表 `test_normal_second` 仅在测试中用，记录秒级生理数据。包含用户账号、皮电、心率、加速度、步数、久坐次数、记录日期等关键字段。

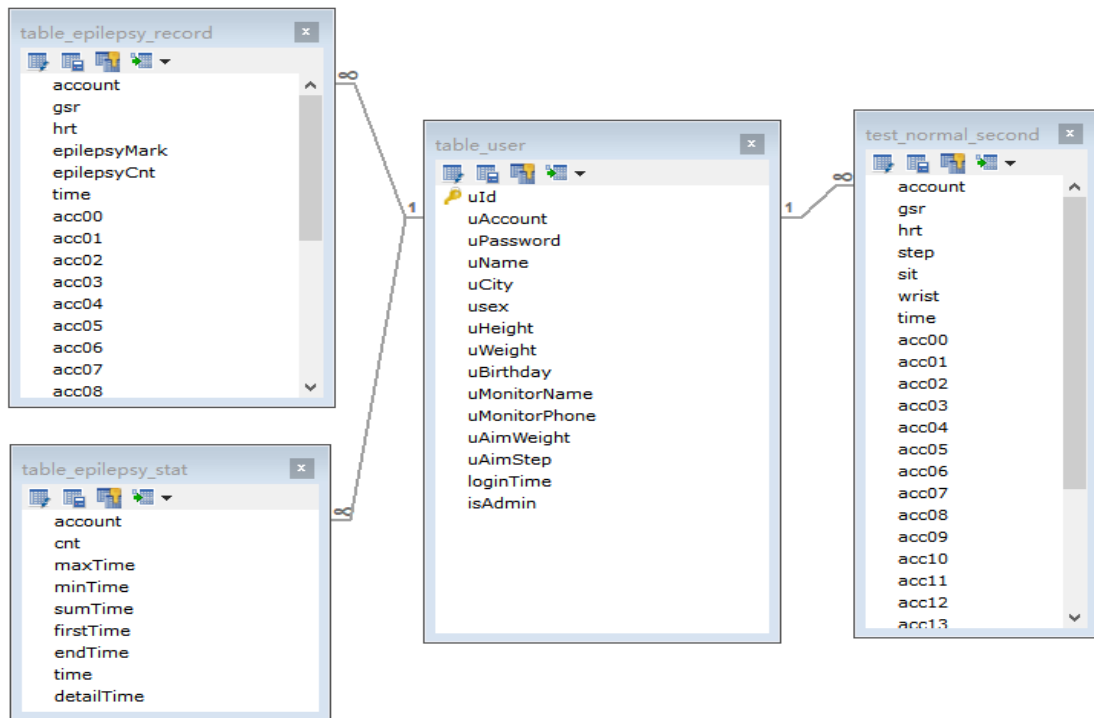


图 4-3 数据库中表的关系图 1

Figure 4-3 Table relationships in the database Figure 1

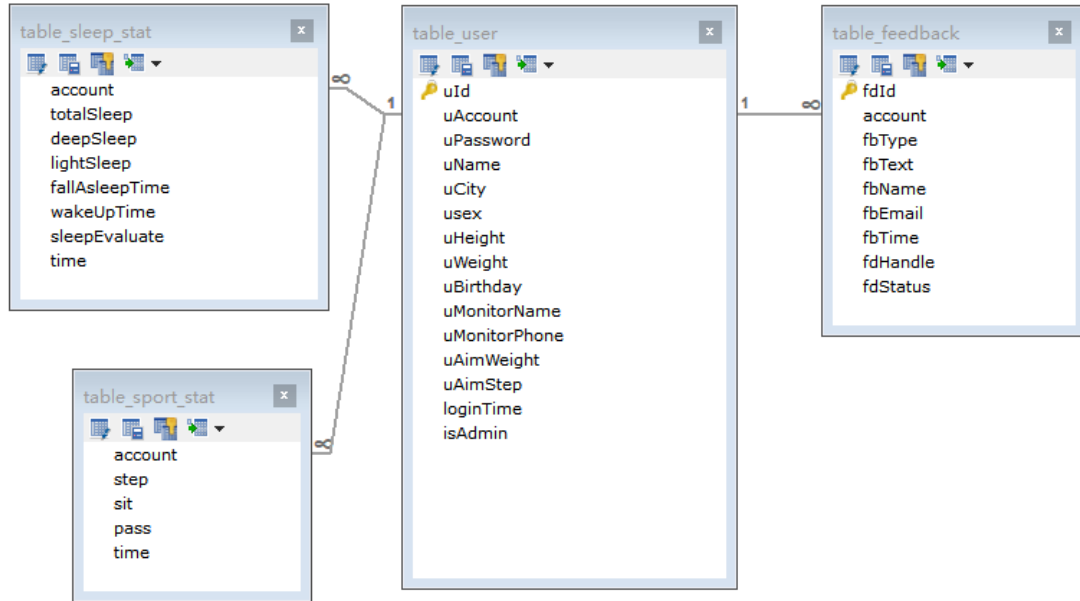


图 4-4 数据库中表的关系图 2

Figure 4-4 Table relationships in the database Figure 2

各表之间的关系图如图 4-3 和图 4-4 所示。由图 4-3 可知，表 table\_user 与表 table\_epilepsy\_record、表 table\_epilepsy\_stat、表 test\_normal\_second 为一对多的关系，即每个用户都对应着对应表中的多条生理信号数据。由图 4-4 可知，表 table\_user 与表 table\_sleep\_stat、表 table\_sport\_stat、表 table\_feedback 为一对多的关系。

### 4.3 数据流设计

基于 STM32F103 芯片设计的腕带式癫痫监测手环通过外接蓝牙传感器及其外围电路，配置了蓝牙模块，通过此蓝牙模块与手机进行配对绑定，使其与手机相连。

此外，该腕带式癫痫监测手环外接了心率传感器、皮肤电传感器与三轴加速度传感器，采集患者的生理信号，通过蓝牙将这些生理信号的数据传输到手机上。

手机通过 HTTP 协议将生理信号数据以 json 格式发送到云服务器端，服务器得到生理数据后，将其存储到 MySQL 数据库中。

当需要分析数据时，可以登录云平台管理系统，服务器中先将所需要的数据简单处理，以表格和折线图的形式呈现给管理员，并且算法工程师或者医生等相关人员如果需要拿到这些数据，可以在云平台导出包含生理数据的 Excel 文件，十分方便简捷。

当然，该数据流还有另外一个流向。该腕带式癫痫监测手环嵌入的算法程序会直接分析采集得到的生理信号数据，再通过蓝牙将分析的结果传输到手机上，手机 APP 中的程序则会根据分析结果来判断是否需要向患者的监护人发出报警，也即判断患者的发病状态。

数据流的流程图如图 4-5 所示。

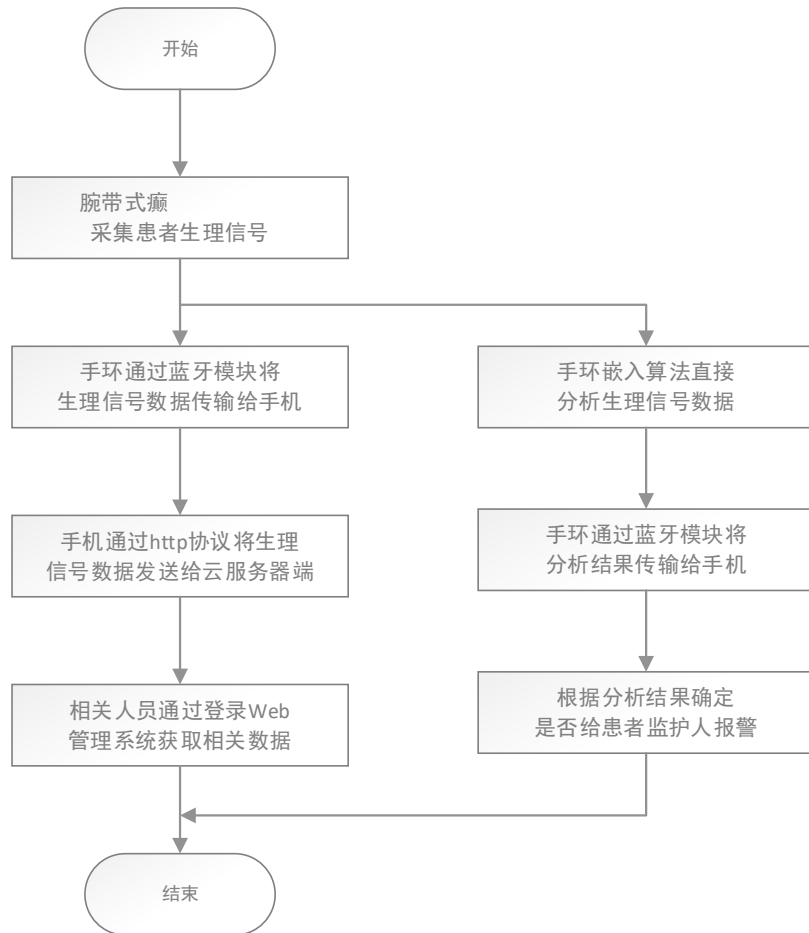


图 4-5 数据流的流程图

Figure 4-5 Flow chart of data flow

#### 4.4 软件数据接口设计

在本课题中，关于云平台与 APP 端的数据接口需要实现如下功能：管理员（比如算法开发人员、医生）的登录验证，普通用户的登录验证，数据上传，数据下载（导出 Excel 文件的形式），数据分析（在后台对数据进行简单处理得到前台可视化所需要的数据），合作医生信息介绍、健康推送与新闻推送、产品介绍、APP 操作介绍等功能所需要的增删改查操作。

在实际开发中选择的开发环境为 JDK9.0.4，开发的语言为 JAVA，开发框架

选择 SSM（Spring + Spring MVC + MyBatis）框架，选择的云服务器为华为云，云服务器的系统为 Linux 系统，数据库软件选择 Mysql 5.5.40。

#### 4.4.1 云平台数据接口设计

云平台中主要的数据接口可以分为生理信号数据处理接口、算法模型训练接口、普通增删改查接口、管理员登录相关接口这四大类，下面将具体介绍<sup>[45]</sup>。

##### （1）生理信号数据处理接口

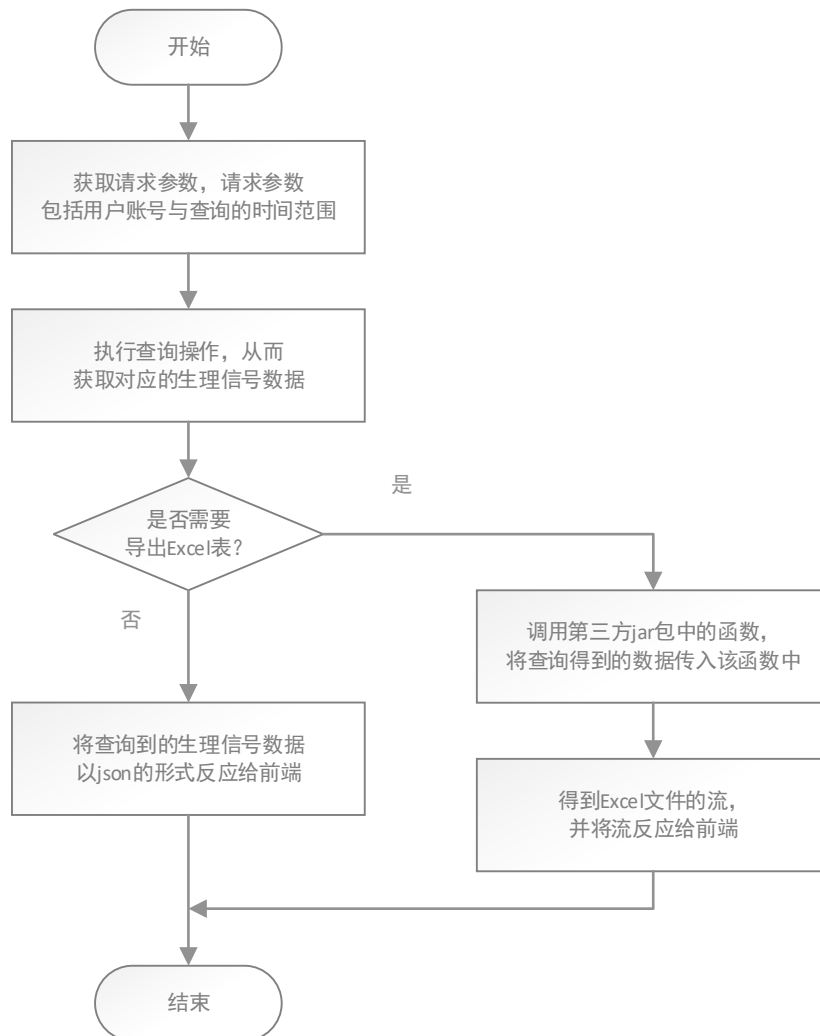


图 4-6 生理信号数据处理接口的程序流程图

Figure 4-6 Program flow chart of physiological signal data processing interface

此接口主要是用来检索生理信号数据，检索到的生理信号数据以表格或者统计图显示，并可以导出为 Excel 表。导出 Excel 表需要用到第三方 jar 包 poi 和 poi-ooxml。程序流程图如图 4-6 所示。

##### （2）算法模型训练接口

主要逻辑是使用系统中的 python 环境运行指定路径下包含算法模型的 python 文件，从而将运行得到的结果响应给前端界面。核心语句为：

```
Process pr = Runtime.getRuntime().exec(strs)
```

就相当于使用 command 命令行执行 strs 语句。程序流程图如图 4-7 所示。

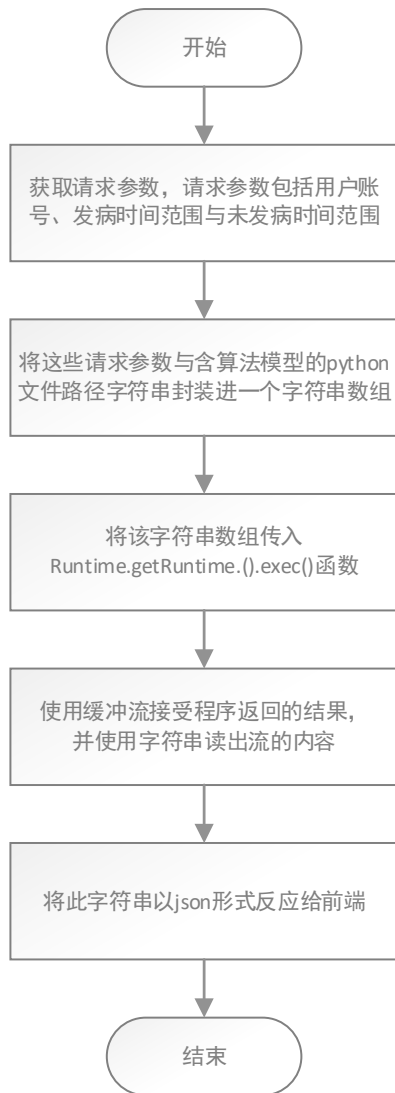


图 4-7 算法模型训练接口的程序流程图

Fig. 4-7 Program flow chart of training interface of algorithm model

### (3) 普通增删改查接口

普通的增删改查接口是最常见的接口，无论是编辑信息、删除信息还是查询信息都与其相关。因为此类接口最为常见，所以绘出程序流程图，如图 4-8 所示。

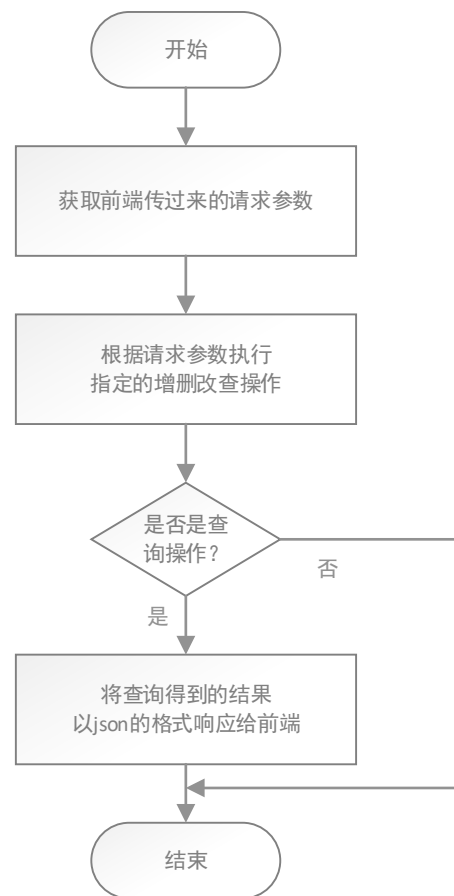


图 4-8 普通增删改查接口的流程图

Fig. 4-8 Flow chart of common add, delete, change and search interface

#### （4）管理员登录相关接口

分为管理员登录接口、管理员退出登录接口、登录状态验证接口。

其中管理员登录接口的逻辑是，首先得到请求传来的参数（账号、密码），调用方法判断密码是否正确、账号是否属于管理员账号，两条都满足则返回 `true`，判断为 `true` 则创建两个键值对，一个的键为“`isLogin`”，初始值为“`true`”，另一个键值对的键为“`username`”，值为登录的账号，将这两个键值对存入 `session` 中，并响应“`OK`”；如果判断为“`false`”，则响应“`ERROR`”。

管理员退出登录接口的逻辑是，将 `session` 中键为“`isLogin`”的键值对的值设为“`false`”，用于标识未登录状态。

登录状态验证接口的逻辑是，判断 `session` 中是否有键为“`isLogin`”的键值对或是否键为“`isLogin`”的键值对的值为“`true`”，若都满足则意味着登录状态，响应头“`status`”设为“`OK`”，并将键为“`username`”的键值对中存储的账号响应给前端；若为“`false`”则响应头“`status`”设为“`ERROR`”。

### 4.4.2 APP 端数据接口设计

在 APP 接口中，一个 URL 地址可能对应多个接口，需要实现多个功能，因此需要通过得到的 `intention` 请求头信息来辨别实现的是哪一个功能。并且，访问接口前需要判断 `User-Agent` 请求头信息，主要是确保请求是从安卓端发出来的，如果 `User-Agent` 的值不是规定的值，则不能执行任何操作，比如不能在浏览器中直接访问 APP 端的接口。APP 端主要的接口可以分为生理信号数据处理接口、用户操作相关接口、安装包相关接口这三大类，下面将具体介绍。

#### （1）生理信号数据处理接口

生理信号数据处理接口分为上传生理信号数据接口和获取生理信号数据接口。

其中，上传生理信号数据接口的逻辑是，首先获取用户账号参数以及需要上传的生理信号数据数组。然后对这些对象数组循环操作，循环语句里分别更新每一条对象数据进数据库，这个方法如果更新成功则返回的值为 1，否则为 0。返回值为 1 表示在已经存在用于表示这一时间的生理数据了，所以只需要更新操作即可；返回值为 0 则表示目前还没有表示这一时间的生理数据，程序流程图如图 4-9 所示。

获取生理信号数据接口的逻辑是，首先获取请求参数，包括用户账号与所需数据的时间范围，调用业务层的方法将这些参数传进去获得相应的对象数组，这个方法的逻辑就是普通的查询数据库的逻辑。如果能正常得到数据，则返回

响应头“status”的值为“OK”，否则为“ERROR”，并将查到的数据序列化成 json 数据并返回。程序流程图如图 4-10 所示。

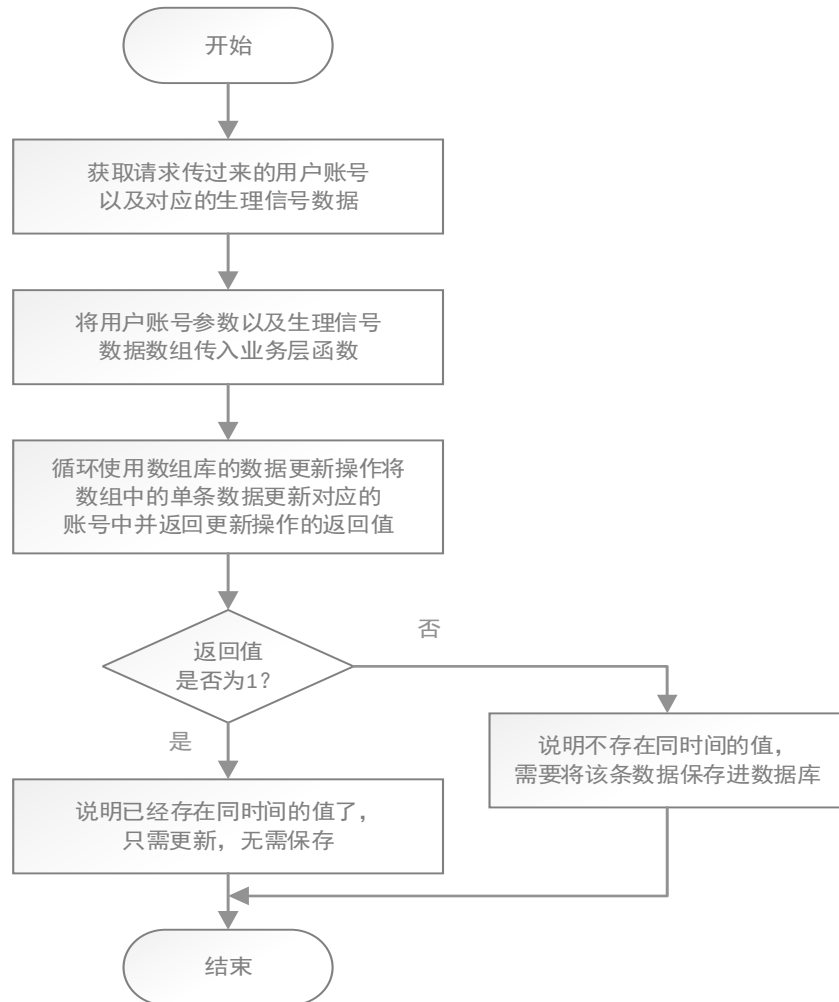


图 4-9 上传生理信号数据接口的程序流程图

Figure 4-9 Program flow chart of uploading physiological signal data interface

## (2) 用户操作相关接口

与用户操作相关的接口包括提交注册接口、登录验证接口、获取验证码接口、修改密码接口、变更手机号接口、上传用户个人信息接口、获取用户个人信息接口等。这类接口除了调用第三方的接口外，均属于基本的增删改查操作，不赘述。

## (3) 安装包相关接口

此类接口包括检查是否有新版本的接口、获取 APP/固件新版本信息的接口、获取最新 APP/固件安装包的接口。这类接口均属于基本的增删改查操作，不赘述。



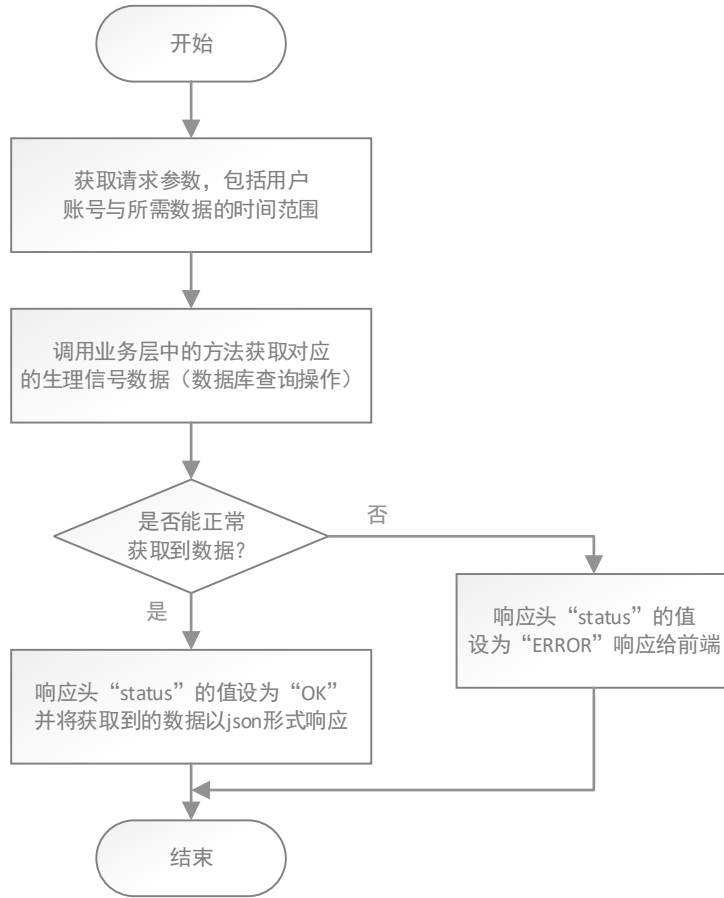


图 4-10 获取生理信号数据接口的程序流程图

Figure 4-10 program flow chart of the interface for obtaining physiological signal data

## 4.5 云平台主要模块界面

在本论文中，始终是围绕生理信号数据与算法来展开的，因此下面重点展示这两个模块的界面。

### （1）查询获取生理信号数据界面

在云平台的后台管理系统中，可以输入用户的账号、需要获取生理信号数据的时间范围来查询对应的生理信号数据，并且可以显示出生理信号数据的统计图。此模块属于数据管理模块，该模块的界面如图 4-11 所示。

### （2）训练算法模型界面

同在云平台的后台管理系统中，还可以通过输入用户账号、发病时间范围和未发病时间范围来训练算法模型，网页中则可以对应显示出训练的结果，并输出边界函数的系数以及截距。此模块同属于数据管理模块，该模块的界面如图 4-12 所示。

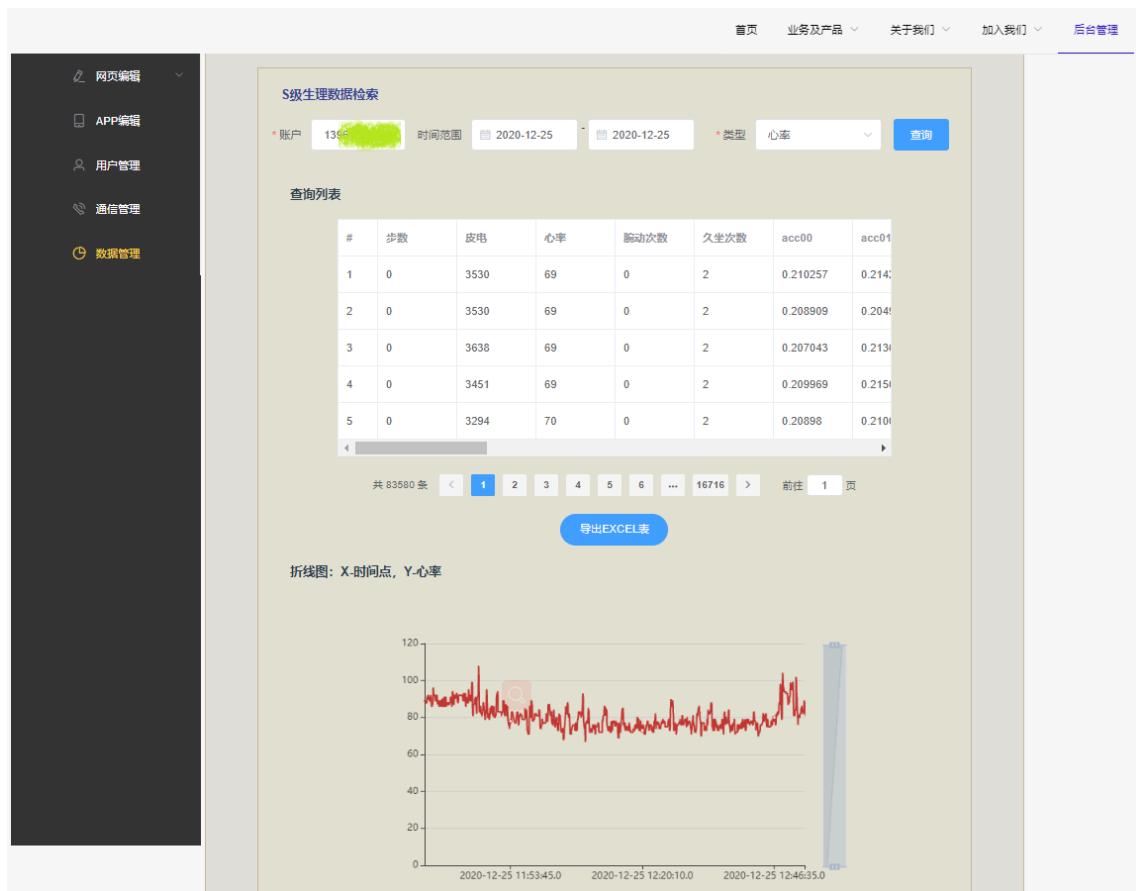


图 4-11 查询获取生理信号数据界面

Figure 4-11 Interface for Query and Obtain Physiological Signal Data

## 4.6 软件平台测试

软件开发很重要的一环是软件测试，这决定着软件开发的成功与否。在这里主要通过实际使用来验证软件是否能正常运行。

### (1) 云平台测试

云平台可以正常查询获取网页数据，包括用户的生理数据，也可以正常下载数据并导出为 Excel 表格，管理系统中对信息的增删改查功能也能正常使用，各项功能均能正常使用。

### (2) APP 端测试

癫痫监测装置发送给 APP 端的生理信号数据能够正常上传到云服务器的数据库中，而且 APP 也能够正常获取到云端的生理信号数据、普通文本信息及图片。

APP 的最新安装包和 DFU 固件包也能够正常更新下载，文本信息也能够正

常上传到云服务器中，各项功能均能正常使用。

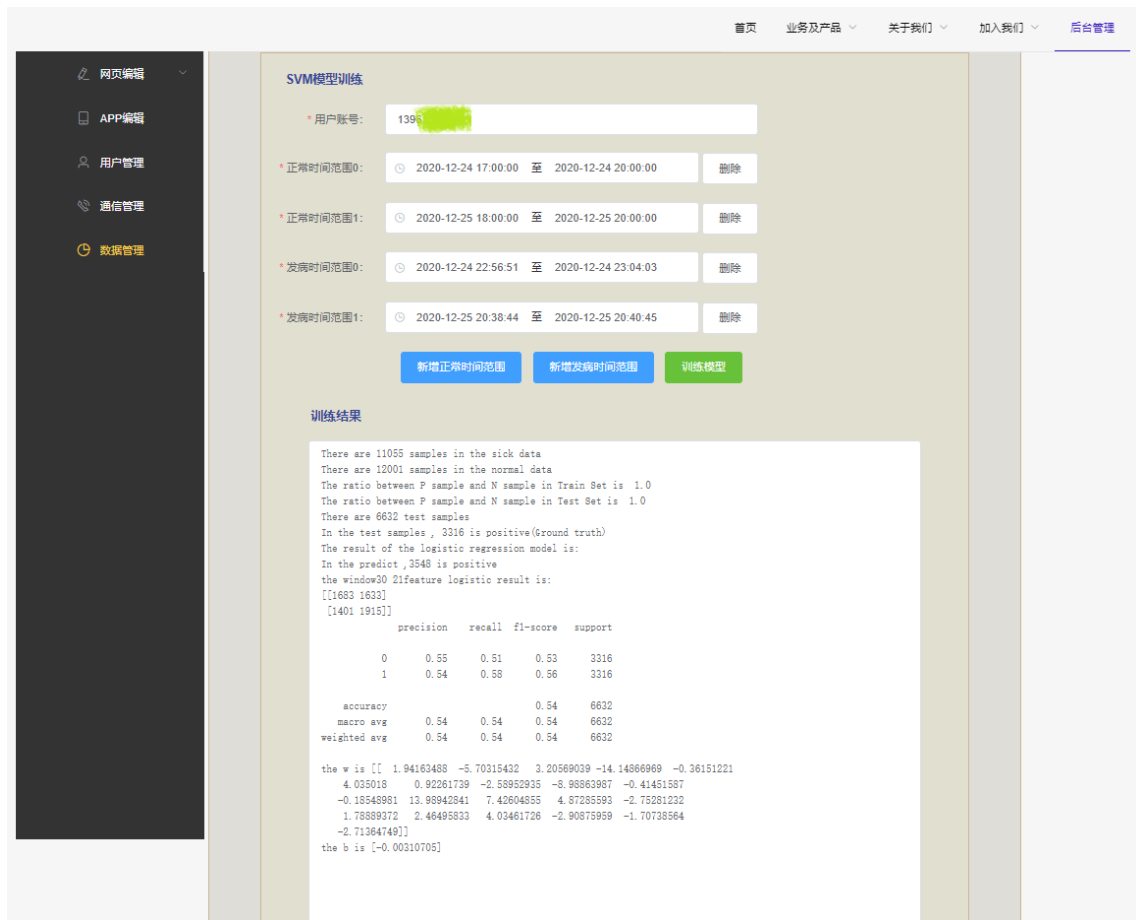


图 4-12 训练算法模型界面

Figure 4-12 Interface of training algorithm model

## 4.7 本章小结

本章围绕癫痫监测系统软件设计与开发展开，其中囊括了数据接口的设计与研究。该系统软件开发的过程属于模块化开发，首先介绍的便是软件的主要功能模块。其中，云平台软件包含了癫痫监测模块、健康助理模块、发现页模块和个人中心模块，安卓软件包含了基本管理模块、数据管理模块、产品介绍模块和发现页模块，并分别对这些模块进行了详细的介绍。

接着阐述了数据库的设计过程。在数据库设计前期阶段为了更清楚的看出数据间的关系，设计了 E-R 模型，在本文中进行了图形展示。接着介绍了数据库中的主要表，这些表储存的数据包括生理信号数据和用户数据等等，并以图形的形式展示各表之间的关系。

随后介绍了整个癫痫监测系统的数据流。首先通过癫痫手环采集患者的生

理信号数据，其次将手环中的数据通过蓝牙通信模块传输到手机上，手机 APP 再将这些数据通过 HTTP 协议发送给云服务器中的数据库，用户则可以通过手机和网页形式查阅或处理数据。

本章着重介绍了数据接口的设计。云平台的数据接口包含生理信号数据处理接口、算法模型训练接口、普通增删改查接口和管理员登录相关接口，安卓软件的数据接口则包含生理信号数据处理接口、用户操作相关接口和安装包相关接口，本文对这些数据接口进行了详细的介绍并且绘制了相应的程序流程图。

最后展示了云平台主要模块的界面，其中展示了查询获取生理信号数据界面和训练算法模型界面。

通过软件使用的测试证明云平台软件和安卓软件均可正常运行使用，实现了预期的效果。

## 结 论

我国癫痫患者基数大，并且每年新增的数量也是非常大，癫痫病不仅影响患者的日常生活，更是给他们的生命健康安全带来了严重的威胁。对于癫痫患者来说癫痫症状随时随地都有可能发生，并且癫痫发作带来的危害具有不可逆性，需要尽可能早的得到介入帮助，因此癫痫的即时监测显得尤为重要。目前癫痫监测主要是借助脑电信号，而依据脑电信号进行监测的设备体积都比较大，只能在专业医院进行。本课题采集癫痫患者生理数据所使用的监测设备是可穿戴腕带式手环，研究动态数据算法分析采集到的三轴加速度信号、皮肤电信号与心率信号，由此进行监测癫痫的发作情况。在此基础上，本文提出了四种基于机器学习算法判断癫痫发作情况的方法，对数据集预处理中的相关方法与四种机器学习算法进行了对比实验，并进行了仿真测试实验和基于临床试验数据的验证实验。除此之外，还研究开发了癫痫监测系统配套软件，方便用户在云平台或安卓客户端上进行查阅癫痫相关的统计数据等操作。本文的研究成果可分为以下四点：

(1) 得到了一套用于癫痫监测的数据预处理与特征提取方法，这套方法的原始数据集由本课题自研的癫痫手环采集而得，其中包括统一频率、数字滤波去噪、数据分割、特征提取、数据集标记与分割、数据集标准化。统一频率用的是插值方法，使得皮肤电信号和心率信号的频率与加速度的频率保持一致，本文对比了二阶 B 样条插值与三阶 B 样条插值的效果，结果是三阶 B 样条插值的效果最好；数字滤波去噪过程中选用了五点中值滤波算法，本文对比了巴特沃斯滤波、切比雪夫滤波以及五点中值滤波的效果，结果是五点中值滤波的效果最好；特征提取过程中共使用了包含时域特征与频域特征在内的 47 个特征，其中包括三种熵值非线性特征；数据集标记与分割过程中使用的是滑窗分割的方法，本文对比了窗口重复率为 80% 的情况与窗口重复率为 50% 的情况，结果是窗口重复率为 80% 时的效果最好。经过这一系列数据预处理与特征工程步骤后，处理后的数据集可输入到癫痫监测的动态数据算法中。

(2) 研究了四种基于机器学习的动态数据算法，其中有逻辑回归、基于线性核函数或径向基核函数的支持向量机算法、AdaBoost 算法与 XGBoost 算法，支持向量机学习模型的参数通过网格搜索法确定，而另外三种学习模型均使用的是默认的参数设定，经过四种算法的对比实验可以看出，通过 XGBoost 算法检测癫痫发病的准确率为 98%，与其他算法相比效果是最好的。

(3)对逻辑回归算法和支持向量机(线性核函数)的边界函数进行了测试,并且进行了临床数据验证。在边界函数的测试实验中比较了两种算法边界函数判断癫痫发作的准确率、误报率和漏报率,支持向量机(线性核函数)的准确率更高、漏报率更低,因此监测效果最好,具体的准确率为 88.50%,误报率为 5.14%,漏报率为 6.36%;在临床数据验证中,也分别使用了四种算法进行了测试,得到:判断为发病的精确率均在 50%左右,判断为未发病的精确率均在 90%以上,其中 XGBoost 算法的训练效果最好,准确率最高,达 72.153%。

(4)开发了一套癫痫监测系统软件,包括云平台软件与手机安卓端软件,主要侧重研究开发其中的数据接口。其中云平台软件包含生理信号数据处理接口、算法模型训练接口、普通增删改查接口、管理员登录相关接口这四大类接口,安卓端软件包含生理信号数据处理接口、用户操作相关接口、安装包相关接口这三大类接口,这两部分软件均实现了预期功能。

基于以上的研究成果,对今后的工作有以下三点展望:

(1)限于实验机器的性能条件,本文中研究 AdaBoost 算法与 XGBoost 算法监测癫痫发作的过程中设定的参数都选择了默认值,这一点在今后可以使用网格搜索法搜索得到局部最优的参数,观察参数对准确率、误报率与漏报率的影响。

(2)在特征提取方面,提取的时域特征数目远多于频域特征数目,导致数据时域波动对结果的影响权重比较大,因此在今后可适度增加频域特征的数目,进一步研究特征。

(3)在测试实验中,可以监测患者连续多天的数据进行分析,宏观上确定算法的监测准确率、误报率与漏报率,这里宏观指的是癫痫发作的次数,一次癫痫发作包含若干秒的数据,因此在今后需要采集患者连续多天的生理数据。

## 参考文献

- [1] H.Adeli, Z.Zhou,N. Dadmehr, Analysis of EEG records in an epileptic patient using wavelet transform[J]. Journal of Neuroscience Methods, 2003, 123(1): 69-87.
- [2] Garry H, Mcginley B, Jones E, et al. An evaluation of the wavelet coefficient quantisation in transform based EEG compression[J]. Computers in biology and medicine, 2013, 43(6): 661-669.
- [3] M. Gheryani, O. Salem, A. Mehaoua. Epileptic Seizures Detection based on Inertial and Physiological Data from Wireless Body Sensors.[C]. ICC 2020 - 2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 2020:1-6.
- [4] A. V. de Vel, K. Cuppens, B. Bonroy, et al. Non-EEG seizure-detection systems and potential SUDEP prevention: State of the art[J], Seizure - European Journal of Epilepsy, 2013(22): 345-355.
- [5] 耿东云. 基于改进小波神经网络的自动癫痫检测[D]. 济南: 山东大学, 2016:38-39.
- [6] 李明阳, 陈万忠, 张涛. 基于 DD-DWT 和 Log-Logistic 参数回归的癫痫脑电自动识别方法[J]. 仪器仪表学报, 2017(38): 1368-1377.
- [7] 吴浩博. 基于小波变换和样本熵的癫痫信号分类研究[D]. 兰州: 兰州大学, 2018:54-55.
- [8] 尹倡隆. 癫痫脑电信号特征提取及分类研究[D]. 兰州: 兰州理工大学, 2020:42-43.
- [9] Gotman J. Automatic Recognition of Epileptic Seizures in the EEG[J]. Electroencephalography and Clinical Neurophysiology, 1982, 52(5): 530-540.
- [10] ErdemYavuz, Mustafa Cem, CanEyüpoğlu, et al. An epileptic seizure detection system based on cepstral analysis and generalized regression neural network[J], Biocybernetics and Biomedical Engineering, 2018(38): 201-216.
- [11] Lal Hussain. Detecting epileptic seizure with different feature extracting strategies using robust machine learning classification techniques by applying advance parameter optimization approach[J]. Cognitive Neurodynamics, 2018(12): 271-294.
- [12] S. Raghu, Natarajan Sriraam, YasinTemel, et al. Performance evaluation of DWT based sigmoid entropy in time and frequency domains for automated detection of epileptic seizures using SVM classifier[J]. Computers in Biology and

- Medicine, 2019(110): 127-143.
- [13]Nabil Dib, Benali Radhwane, Reguig Fethi Bereksi. Epileptic seizure recognition using EEG wavelet decomposition based on nonlinear and statistical features with support vector machine classification[J]. Biomedical Engineering-Biomedizinische Technik, 2020(65): 133-148.
- [14]A. Sharmila, P. Geethanjali. Evaluation of time domain features on detection of epileptic seizure from EEG signals[J]. Health and Technology, 2020(10): 711-722.
- [15]S. Raghu, Natarajan Sriraam, Shyam Vasudeva Rao, et al. Automated detection of epileptic seizures using successive decomposition index and support vector machine classifier in long-term EEG[J]. Neural Computing and Applications, 2020(32): 8965-8984.
- [16]H. Anila Glory, C. Vigneswaran, Sujeet S. Jagtap, et al. AHW-BGOA-DNN: a novel deep learning model for epileptic seizure detection[J]. Neural Computing and Applications, 2021(33): 6065-6093.
- [17]Hanan Al-Hadeethi, Shahab Abdulla, Mohammed Diykh, et al. Adaptive boost LS-SVM classification approach for time-series signal classification in epileptic seizure diagnosis applications[J]. Expert Systems with Applications, 2020(161): 113676.
- [18]Rout, Susanta Kumar, Sahani, et al. Multifuse multilayer multikernel RVFLN+ of process modes decomposition and approximate entropy data from iEEG/sEEG signals for epileptic seizure recognition[J]. Computers in Biology and Medicine, 2021(132): 104299.
- [19]张根选, 张莉, 等. 一种用于癫痫发作预测的可穿戴无线传感器[J]. 中国医疗器械杂志, 2016(4): 257-259.
- [20]吴书峰. 便携式癫痫发作监测系统设计与研究[D]. 秦皇岛: 燕山大学, 2016:62-63.
- [21]Bideaux A, Anastasopoulou P, Hey S, et al. Mobile monitoring of epileptic patients using a reconfigurable cyberphysical system that handles multi-parametric data acquisition and analysis[C]. Proceedings of the Wireless Mobile Communication and Healthcare, 2014 EAI 4th International Conference, 2014: 377-380.
- [22]A. Ahmed, W. Ahmad, M. J. Khan, et al. A wearable sensor based multi-criteria-decision-system for real-time seizure detection[C]. Seogwipo: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017: 2377-2380.



- [23]A. Zsom, et al. Ictal autonomic activity recorded via wearable-sensors plus machine learning can discriminate epileptic and psychogenic nonepileptic seizures[C]. Berlin: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019: 3502-3506.
- [24]A. A. Kabanov, A. I. Shchelkanov. Development of a Wearable Inertial System for Motor Epileptic Seizure Detection[C]. Novosibirsk: 2018 XIV International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE), 2018: 339-342.
- [25]K. Vandecasteele, T. De Cooman, Y. Gu, et al. Automated Epileptic Seizure Detection Based on Wearable ECG and PPG in a Hospital Environment[J]. Sensors, 2017 (17): 2338.
- [26]S. Boeckx, W. van Paesschen, B. Bonte, et al. Live Demonstration: SeizeIT - A wearable multimodal epileptic seizure detection device[C]. Cleveland: 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2018: 1-1.
- [27]D. Sopic, A. Aminifar, D. Atienza. e-Glass: A Wearable System for Real-Time Detection of Epileptic Seizures[C]. Florence: 2018 IEEE International Symposium on Circuits and Systems (ISCAS), 2018: 1-5.
- [28]Romigi A, Rizzo G, Izzi F, et al. Heart Rate Variability Parameters During Psychogenic Non-epileptic Seizures: Comparison Between Patients With Pure PNES and Comorbid Epilepsy[J]. Frontiers in Neurology. 2020(11):713.
- [29]周立鹏. 基于机器学习的摔倒检测算法的研究与应用[J]. 杭州: 浙江工业大学, 2020:13-15.
- [30]杨舒涵. 基于脑电信号的分析算法研究及其在癫痫检测方面的应用[D]. 长春: 吉林大学, 2020:13-17.
- [31]陈泽龙, 谢康宁. 脑电 EEG 信号的分析分类方法[C]//中国医学装备大会暨 2019 医学装备展览会论文汇编. 中国医学装备, 2019: 300-307.
- [32]Francesco Onorati, Giulia Regalia, et al. Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors[J]. Epilepsia, 2017(11):1870-1879.
- [33]蔡菁. 皮肤电反应信号在情感状态识别中的研究[D]. 重庆: 西南大学, 2010:17-19.
- [34]李明阳. 基于复合域分析的癫痫脑电信号检测算法研究[D]. 长春: 吉林大学, 2018:58-59.
- [35]王宇桥. 基于广义 S 变换和随机森林算法的癫痫脑电信号特征提取与分类研究[D]. 长春: 吉林大学, 2018:15-20.
- [36]李航. 统计学习方法方法[M]. 北京: 清华大学出版社, 2012:91-94.

- [37]朱塞佩·博纳科尔索. 机器学习算法[M]. 北京: 机械工业出版社, 2020:143-154.
- [38]Paola Vargas, Miguel Angel Moreles, Joaquin Peña. Estimation and SVM classification of glucose-insulin model parameters from OGTT data: a comparison with the ADA criteria Estimation and SM classification of glucose-insulin model parameters[J]. International Journal of Diabetes in Developing Countries, 2021(41): 54-62.
- [39]Pooya Tavallali, Mehran Yazdi, Mohammad Reza Khosravi. Robust cascaded skin detector based on AdaBoost[J]. Multimedia Tools and Applications, 2019, 78(2): 2599-2620.
- [40]Deandra Aulia Rusdah, Hendri Murfi. XGBoost in handling missing values for life insurance risk prediction[J]. SN Applied Sciences, 2020(8): 1336.
- [41]Zhifen Zhang, Yiming Huang, Rui Qin, et al. XGBoost-based on-line prediction of seam tensile strength for Al-Li alloy in laser welding: Experiment study and modelling[J]. Journal of Manufacturing Processes, 2021(64): 30-44.
- [42]钟琳倩. 基于云平台的睡眠监护软件的设计与实现[D]. 重庆: 重庆邮电大学, 2019:31-58.
- [43]徐诗怡. 老年人情绪监测智能手环创新设计研究[D]. 西安: 西安理工大学, 2019:37-38.
- [44]张仁昊. 建筑设备监控系统的云端设计与实现[D]. 济南: 山东建筑大学, 2020:15-27.
- [45]杨思豪. 基于可穿戴手环多感知特征融合的身心健康评估方法研究[D]. 成都: 电子科技大学, 2020:20-24.

## 哈尔滨工业大学学位论文原创性声明和使用权限

### 学位论文原创性声明

本人郑重声明：此处所提交的学位论文《基于癫痫监测手环的数据处理算法与软件开发》，是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。

作者签名：余锡彪

日期：2021 年 6 月 17 日

### 学位论文使用权限

学位论文是研究生在哈尔滨工业大学攻读学位期间完成的成果，知识产权归属哈尔滨工业大学。学位论文的使用权限如下：

(1)学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；(2)学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；(3)研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名单位为哈尔滨工业大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。

本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名：余锡彪

日期：2021 年 6 月 17 日

导师签名：黎连伟

日期：2021 年 6 月 17 日

## 致 谢

转眼之间两年的研究生时光就快要过去了，即将毕业走向社会，在本篇论文即将付梓之际，内心感触颇多。此时心里最想感谢的是我的导师梁廷伟教授，本篇论文的完成离不开梁老师的悉心指导。梁老师治学严谨，工作认真负责，对待学生平易近人，诲人不倦，是我学习科研上的引路人，在两年时间里梁老师教会了我许许多多，不仅有学习科研上的知识，还有做人的道理。

同时也要感谢课题组中的金显吉老师、吴宇宇师兄和已经毕业的孔得慧师兄和李业鸿师兄，他们在课题方面给了我很多宝贵的指导意见和建议，还要感谢王凯同学，在生活和学习科研中都给了我许多帮助。

此外，非常感谢飞行器动力学与控制研究所中给与过我帮助的老师们的和师兄师姐们、同学们，我在研究所里的成长离不开大家的帮助。

还要感谢可爱的室友们，有爱看书的王瑞星同学，坚持运动的王凯同学，还有刻苦学习准备攻读博士的聂成洋同学，两年来大家互帮互助，一同成长了很多，让我收获到了真挚的友情。

另外特别感谢我的父母，他们给与了我无微不至的照顾。特别是去年由于新冠疫情不能返校，整个研一下学期都在家中，是他们给我提供了良好的学习科研环境，让我能够安心学习工作，没有后顾之忧。

最后，感谢母校哈尔滨工业大学的培养，非常荣幸在两年求学期间见证了百年校庆，祝愿母校在新的一百年里蒸蒸日上，继续创造辉煌。