

“疲劳可堪”——疲劳分析预警系统

设计文档

目录

- 一、 所在系列及赛项1
- 二、 目标问题与意义价值1
 - 2.1 项目设计背景1
 - 2.2 项目设计目的与基本功能2
 - 2.3 应用领域与应用价值2
- 三、 设计思路与方案2
 - 3.1 项目设计思路2
 - 3.1.1 设计灵感与理念2
 - 3.1.2 作品层次划分和流程设计2
 - 3.2 技术路线3
 - 3.2.1 作品技术方向3
 - 3.2.2 总体技术设计4
 - 3.2.3 核心情感分析算法原理4
 - 3.3 详细设计方案4
 - 3.3.1 数据层4
 - 3.3.2 业务层7
 - 3.3.3 网络层10
- 四、 实现10
 - 4.1 环境配置10
 - 4.2 数据集的收集与处理11
 - 4.2.2 数据集的处理11
 - 4.3 模型训练与测试12
 - 4.4 Web 界面优化12
- 五、 运行结果/效果13
 - 5.1 网页界面效果13
 - 5.2 检测模型预测效果13
- 六、 创新与特色15

6.1 综合多种传感器数据，疲劳检测和判断更准确.....	15
6.2 使用模糊决策树，保证准确率的前提下，具有更快的检测速度	16
6.3 使用 python matplotlib 等多种手段，实现结果的可视化	16
6.4 基于云构建和部署平台，提高系统的可迁移性、可扩展性和可维护性.....	16

一、 所在系列及赛项

A 系列

二、 目标问题与意义价值

2.1 项目设计背景

1) 现代社会职业竞争激烈，生活压力大。部分人处于高强度的学习或工作状态。长此以往，身体会产生持久的难以缓解的疲劳感。疲劳工作和学习不仅影响工作效率而且会危及生命安全。近年来大学生猝死的案例频发，疲劳手术的案例层出不穷，交通方面疲劳更是造成了大量的人员伤亡。根据统计，有 30%左右的交通事故是由疲劳驾驶引起的。尤其在高速公路上，疲劳驾驶引发的事故占比达到 30%以上。由此可见，疲劳已成为威胁生命安全的一个重要因素。

2) 目前，我国在疲劳作业方面的控制仍处在起步阶段。仅交通方面在车辆上装有疲劳测试仪，且覆盖范围不广。其他领域的疲劳检测则更少。

3) 根据我国关于印发《“互联网+”人工智能三年行动实施方案》的通知（发改高技[2016]1078 号）文件，“推动基于感知数据、多媒体、自然语言等大数据的深度学习技术研发，开展类脑神经计算系统、类脑信息处理等类脑智能领域的前沿理论和技术研究。支持人工智能领域的芯片、传感器、操作系统、存储系统、高端服务器、关键网络设备、网络安全技术设备、中间件等基础软硬件技术开发，支持开源软硬 3 平台及生态建设。加快基于人工智能的计算机视听觉、生物特征识别、复杂环境识别、新型人机交互、自然语言理解、机器翻译、智能决策控制、网络安全等应用技术研发和产业化”。由此可见，国家目前鼓励各行各业的智能化发展。而本系统便是采用信息化的手段实现疲劳探测的智能化。

目前传感器在各个领域应用广泛，未来我国各个行业将朝着“可感”的信息化方向不断发展。详见图 1、图 2。



图 1：国内传感器应用领域分布（单位：%）



图 2：2015-2020 中国传感器应用领域市场占比变化图

2.2 项目设计目的与基本功能

本项目是以疲劳状态检测为核心功能并能够及时警示人们处于何种疲劳程度的智能信息化系统。项目旨在通过对人们疲劳状态的检测，警示使用者是否处于疲劳状态，从而减少因疲劳作业而造成的事故的发生。

本系统可以通过对由使用者佩戴的摄像头采集的图片视频进行眼动分析，对由脑电波传感器采集的脑神经活动信息进行脑神经分析，对由脉搏传感器采集脉搏信息进行心跳分析，对由声音采集器采集的音频信息进行气息分析，再综合分析上述多种传感器数据来判断使用者的身体处于是否处于疲劳状态，进而评判出使用者是否适合继续学习工作，及可继续学习工作的时长等。本系统结构完整、功能全面、识别准确、实时效果好。

2.3 应用领域与应用价值

本系统市场定位为体感疲劳测试仪，可应用于包含交通、医疗、教育、工业生产等多个领域在内的社会生产生活各方面，应用范围较广泛，并且系统使用较为方便，对硬件要求仅有摄像头和传感器，成本低下且效果良好，可通过对使用者的疲劳状态进行检测，从而判断其是否适合继续进行工作，提醒使用者注意自己的身体状态，能够有效减少由于疲劳作业而导致的安全事故的发生，保障人们的生命安全。

三、 设计思路与方案

3.1 项目设计思路

3.1.1 设计灵感与理念

疲劳是一种主观不适的感觉，但客观上会在同等条件下，失去其完成原来所从事的正常活动或工作能力。每个人对于疲劳都不陌生，当你从事过于繁重的工作时，可能你就会产生疲劳的感受。疲劳因工作而产生，也影响于工作，因此避免因疲劳产生的不利影响至关重要，严重时甚至会产生极其严重的后果。基于疲劳对于人的潜移默化的影响，本项目旨在实现一种结合多传感器的疲劳检测系统，能够准确分析出即将给病人做手术的医生的疲劳程度，从而尽可能的减小手术风险，确保病人的人身安全，具有很大的实际应用意义。

3.1.2 作品层次划分和流程设计

系统按照层次结构大体分为三层，数据层、业务层、网络层，如图 3 所示。

数据层利用不同传感器模块来采集数据并向上传递。业务层接收并处理数据，并网络层建立相应会话。网络层使用 HTML、flask 等框架与技术完成 Web 端的数据展示和结果分析。

项目总体设计流程图如图 4 所示。

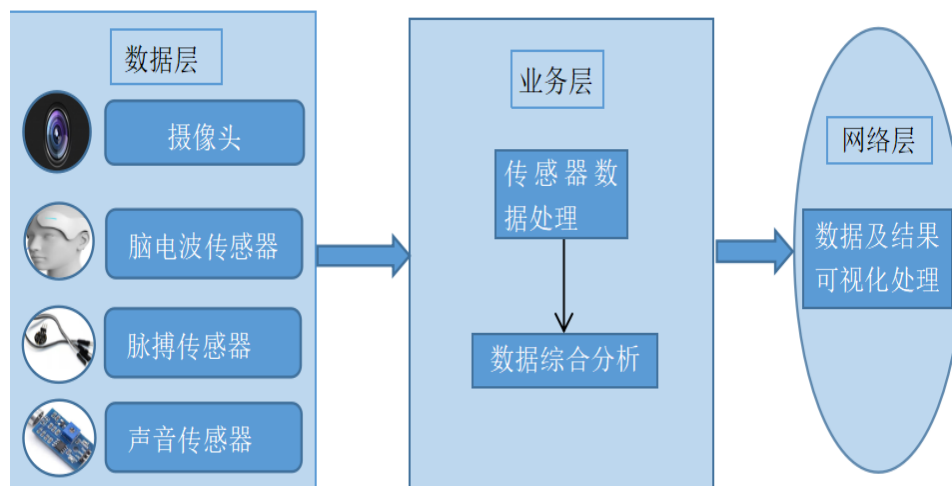


图 3 系统层次图

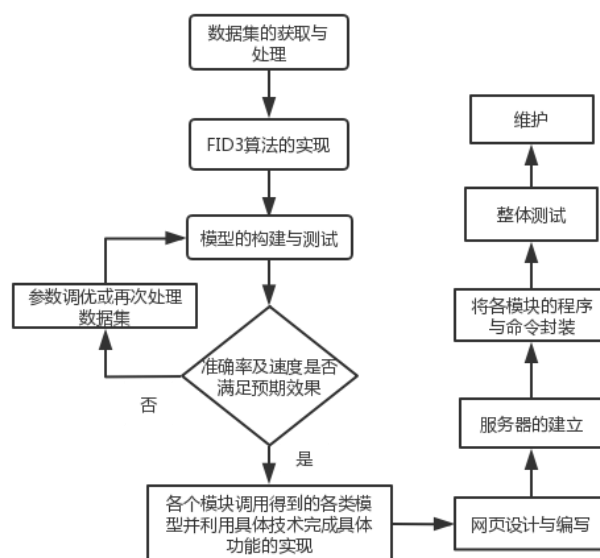


图 4 项目总体设计流程图

3.2 技术路线

3.2.1 作品技术方向

通过摄像头、头戴式脑电波传感器、脉搏传感器以及声音传感器采集人体状

态信息，并将数据通过 SM3 算法加密并存入数据库。然后利用基于模糊决策树的算法实现数据分析。使用 python matplotlib 模块实现数据及决策树的可视化。Web 方面使用 Flask 框架完成搭建。

3.2.2 总体技术设计

数据层：以数据库为结点建立传感器与业务层的联系，提供沟通的桥梁。

业务层：功能的代码实现，利用情感分析、数据加密等相关算法实现对应的业务功能。

网络层：利用 Web 页面展示友好的界面并简化用户的操作。

3.2.3 核心情感分析算法原理

本系统采用模糊决策树 FuzzyID3 算法，即使用模糊信息熵作为分割依据，每次选择最大的信息增益值的属性进行分割，以此构建出模糊决策树。

具体步骤为：（1）数据预处理（2）归纳建立决策树（3）把得到的决策树转换成一组模糊规则（4）把得到的模糊规则用于分类。

模糊决策树是通过把决策树以及模糊理论相结合而产生的算法，从总体来讲，模糊决策树的优点一般表现在如下几个方面：

（1）它具有和传统的决策树一样的构建的自上而下的过程，只是用到了模糊集。对于模糊决策树而言，在构造某个节点上某个属性也许不存在满足的分支，那么该属性会被模糊化进入多个分支，而不像传统的决策树一样，只允许一条数据进入一个分支；

（2）模糊集合和近似推理可以处理噪声、非一致和非完全的数据，它比清晰决策树更加的精确；

（3）模糊决策树使用了模糊隶属度函数，可以处理连续型和离散性数据。

3.3 详细设计方案

3.3.1 数据层

传感器数据采集：通过实体的传感器，实时监测用户的各项数据，汇总到主控芯片进行预处理，然后通过 WLAN 实时上传 JSON 文件到云服务。

数据库及数据处理：将传感器采集到的数据利用 SM3 算法加密并保存到数据库中。

图 6 为总体流程，图 7 为 SM3 加密的具体实现，图 8 为 SM3 中 CF 的具体流程。

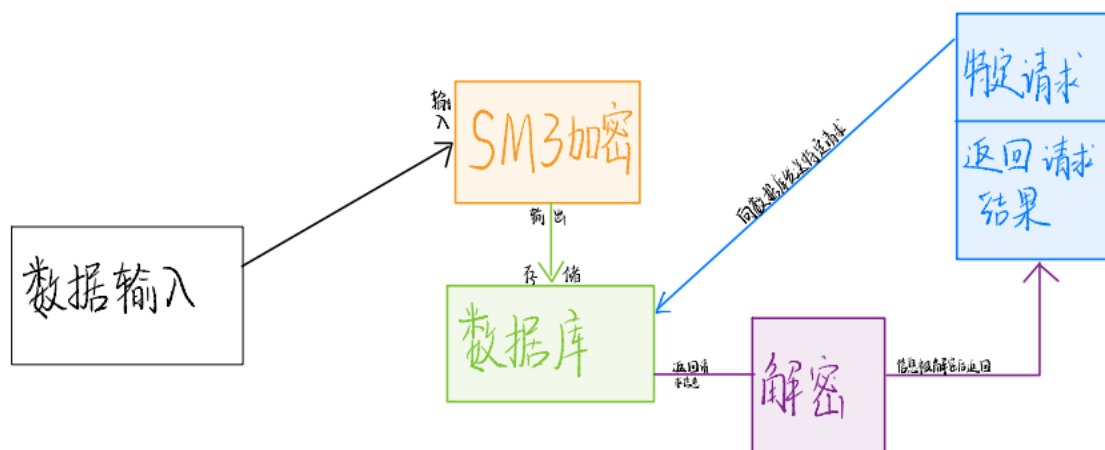
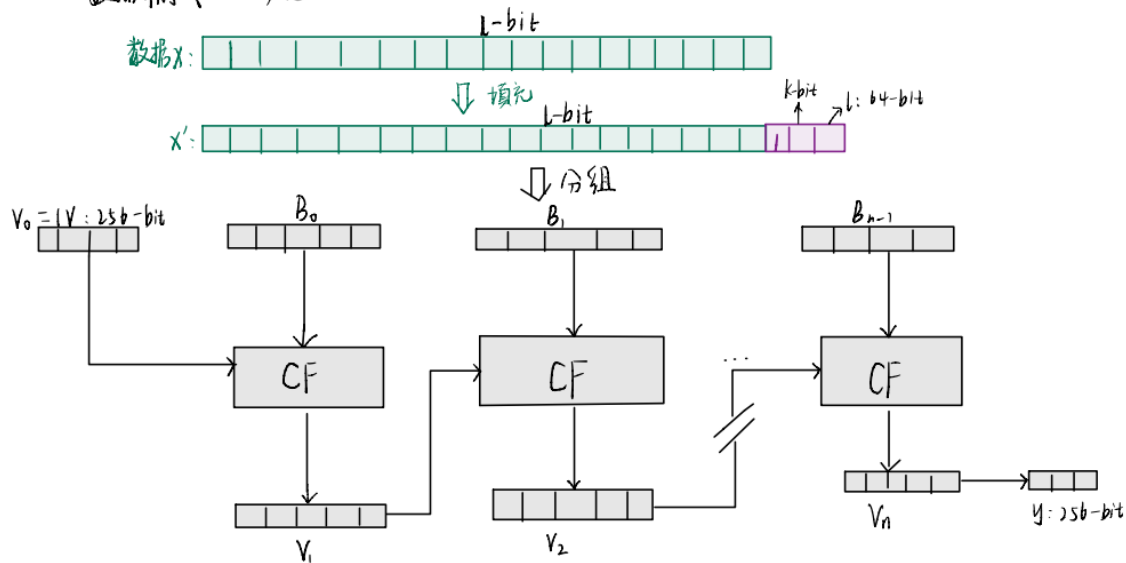


图 5 数据库总体流程

SM3算法流程:

数据输入 \rightarrow 填充 \rightarrow 分组 \rightarrow 消息扩展 \rightarrow 迭代压缩 \rightarrow 输出。如下图:



后附 CF 中的执行流程 \rightarrow

图 7 SM3 算法加密数据流程

图 8 SM3 算法中 CF 的具体实现

3.3.2 业务层

数据处理模型：本项目的数据处理模型算法采用基于模糊数学的模糊决策树算法。模糊决策树是传统清晰决策树的推广。它将模糊理论应用于训练与匹配过程中，结合了决策树的可理解性和模糊集合的表示能力用来处理模糊性和不确定性信息，使决策树拥有更好的健壮性，提高了决策树的可理解性，并使决策树归纳算法的扩展能力增强。

本项目使用的是 Fuzzy ID3 算法流程图如下。

Algorithm 1 FID3 learning

```

1: Generate the root node with a fuzzy set comprising all instances with the membership degree 1;
2: if a node  $t$  with a fuzzy set of instances  $S$  satisfies one of the following conditions:
    (a) the proportion of instances of a class  $c_k$  is greater than or equal to a threshold  $\tau_1$ , that is,  $\frac{S^{c_k}}{|S|} \geq \tau_1$ ;
    (b) the number of instances is less than a threshold  $\tau_2$ , that is,  $|S| < \tau_2$ ;
    (c) there are no attributes for further splitting;
    then
3:   The node  $t$  is a leaf and is assigned class labels with membership degrees;
4: else
5:   for each  $x_j (j = 1, \dots, p)$  do
6:     Calculate the information gain  $G(x_j, S)$  following Eq. 3;
7:     Select the test attribute  $x_{max}$  that maximizes the information gain;
8:   end for
9:   Divide  $S$  into fuzzy subsets  $S_1, \dots, S_m$  according to  $x_{max}$ , where the membership degree of every
    instance  $\mathbf{x}^i$  in  $S_v$  is the product of  $\mathbf{x}^i$ 's membership degree in  $S$  and the value of  $\mu_{max, F_v}(\mathbf{x}^i)$ ;
10:  for each fuzzy subset  $S_v (v = 1, \dots, m)$  do
11:    Generate a new node  $t_v$  corresponding to  $S_v$ ;
12:    Label the membership functions  $\mu_{max, v}$  on the edge that connects  $t_v$  to  $t$ ;
13:    Replace  $S$  by  $S_v$ ;
14:    Repeat from Line 2 recursively;
15:  end for
16: end if
  
```

图 8 SM3 算法中 CF 的具体实现

下文将以实例介绍 FID3 的实现流程。

示例结构图如下。

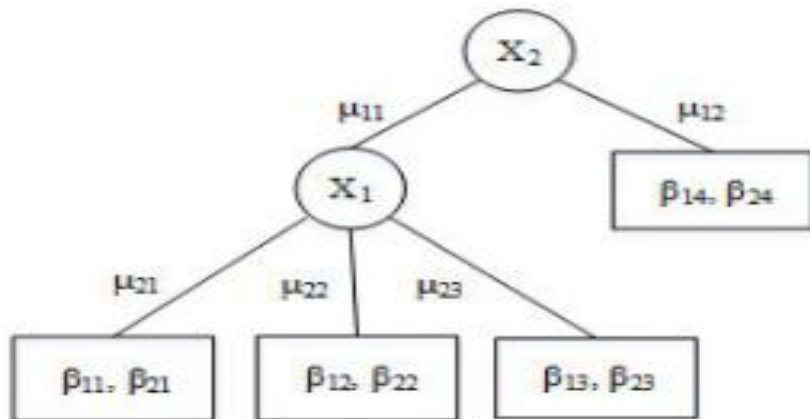


图 9 示例结构图

1. 数据集【以两组数据为例】格式：脉搏、眨眼频率、用户状态

示例：

【70、0.35、安全】

【80、0.3、疲劳】

【65、0.1、疲劳】

2. 确定隶属函数→【用来将原始连续数据模糊化（概率化）】

示例：Y1 = F（脉搏数、快慢）、Y2 = F（眨眼频率，快慢）

示例：Y3 = F（脉搏数，P(安全)）、Y4 = F（眨眼频率，P(安全)）

数据处理：→将原始连续数据模糊化（概率化）【使用隶属函数】

【P(脉搏快)，P(脉搏慢)，P(眨眼快)，P(眨眼慢)，安全/疲劳】

以处理第一个数据为例：【0.3，0.7，0.4，0.6，安全/疲劳】

3. 计算特征属性值的模糊熵【决定分类使用何种特征】—— G（脉搏快）

$$G(\mathbf{x}_j, S) = I(S) - E(\mathbf{x}_j, S)$$

S：总体样本集合、X_j：特征 X 的属性值为 j【脉搏、快】

$$I(S) = - \sum_{k=1}^q (P(c_k, S) \cdot \log_2 P(c_k, S))$$

P(c_k, S) : S 中某类样本的隶属度的和 / 集合总隶属度的和

$$E(\mathbf{x}_j, S) = \sum_{v=1}^m (P(S_{jv}, S) \cdot I(S_{jF_v}))$$

$$P(S_{jv}, S) = \frac{|S_{jF_v}|}{\sum_{v=1}^m |S_{jF_v}|}$$

S_{jv}：S 中 第 v 个 特征 X 的属性值为 j 的 样本

S_{jF_v}：S 中 特征 X 的属性值 j 的隶属度为 F_v 的样本的集合

| S_{jF_v} |： S 中 特征 X 的属性值 j 的隶属度为 F_v 的样本的隶属度的和

I(S_{jF_v})：S_{jF_v} 集合中某一类样本的隶属度的和 / 集合总隶属度的和

4. 计算特征属性值的权重：【决定特征的属性值下有无子叶节点】

$$W(\text{脉搏快}) = G(\text{脉搏快}) / (G(\text{脉搏快}) + G(\text{脉搏慢}))$$

根据根节点修改数据集

$$\text{例: } G(\text{脉搏慢}) > G(\text{脉搏快})$$

5. 将原始数据中所有脉搏慢项取出并删去脉搏特征，以此构成新的数据子集。

用新子集递归 5.6 点 直至所有特征被删掉，构建出决策树。

6. 根据 决策树计算模糊规则即某条路径的真实度 S

S(脉搏快, 安全)、S(脉搏快, 疲劳) S'(眨眼快, 安全)、S'(眨眼慢, 安全)、S'(眨眼快, 疲劳)、S'(眨眼慢、疲劳)

S' : 由脉搏慢构成的新的子集

$$S(A, B) = \frac{M(A \cap B)}{M(A)} = \frac{\sum_{x \in X} \min\{\mu_A(x), \mu_B(x)\}}{\sum_{x \in X} \mu_A(x)}$$

A: 脉搏快、B 安全

X: 数据集的所有样本 x: 数据集中某个样本

$\mu_A(x)$: 样本脉搏数对应 脉搏快 的概率。【隶属函数 Y1】

$\mu_B(x)$: 样本脉搏数对应 用户安全 的概率。【隶属函数 Y3】

-----模糊决策树构建完成-----

数据可视化流程图如下。

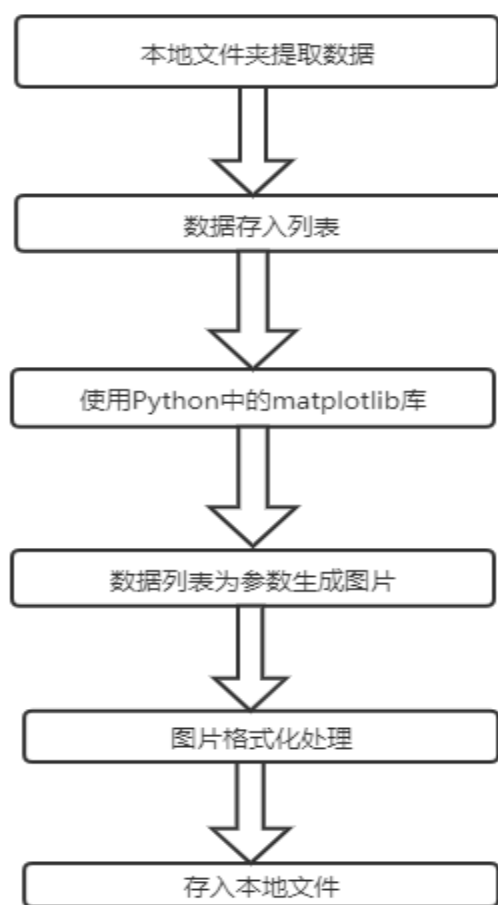


图 10 数据库数据及结果的可视化处理

3.3.3 网络层

前端与云服务设计：为了使用户更加方便的使用，本系统制作了网页界面将各个模块功能的实现进行封装，其中云服务利用 Nginx 实行了请求的负载均衡，之后利用 Flask 框架接受并响应来自客户端的请求。客户端的具体界面效果详见本文 5.1（运行效果-网页界面）

四、 实现

4.1 环境配置

本系统在 windows 上运行，首先配置环境并下载相关的依赖包、安装框架以及各模块。

4.2 数据集的收集与处理

（一）测试数据集的获取：

（1）数据收集场景：5 分钟打字工作，5 分钟驾驶游戏。

（2）两个样本的选取：

“安全”样本：分别收集上午起床后两个小时和下午 4 点左右的用户数据，作为“安全”的样本；

“疲劳”样本：分别收集熬夜睡眠不足后和半夜入睡前的用户数据，作为“疲劳”的样本

（3）测试数据的获取：每个时间段对 4 个不同的同学进行 10 分钟连续的数据记录

（4）整理测试数据。

（二）验证数据集的获取：

同测试数据集的获取。

4.2.2 数据集的处理

对数据集的处理：

（1）对收集得到的用户测试数据进行整合分类；

（2）对得到的测试数据集进行数据分析；

（3）对收集得到的用户验证数据进行整合分类；

（4）对得到的验证数据集进行数据分析；

（5）比对测试数据集和验证数据集的数据分析情况，分析本项目的准确度。

测试样例：脉搏数，眨眼数，声音大小，专注度，分析结果

81, 11, 67, 748, “Safe”

83, 12, 60, 709, “Safe”

89, 18, 45, 398, “Fatigue”

92, 16, 49, 415, “Fatigue”

例：

[89, 18, 45, 398, "Fatigue"] [81, 11, 67, 748, "Safe"]

验证样例：脉搏数，眨眼数，声音大小，专注度，分析结果

87, 15, 84, 793, “Safe”

85, 16, 59, 766, “Safe”

92, 15, 51, 481, “Fatigue”

62, 24, 47, 668, “Fatigue”

例：

[98, 15, 64, 435, "Fatigue"] [87, 15, 84, 793, "Safe"]

4.3 模型训练与测试

系统采用前期采集并处理完成的训练集和测试集进行训练和测试。最终模型结构图如下。

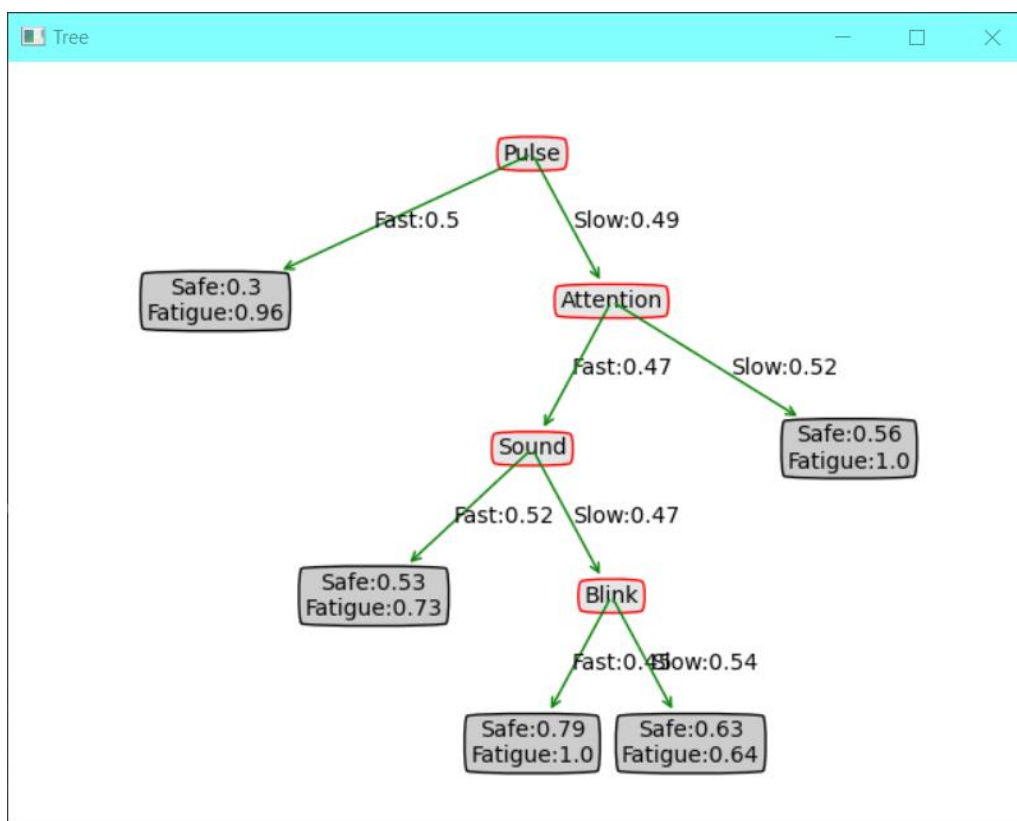


图 11 疲劳分析决策树可视化结果

4.4 Web 界面优化

在完成各项功能之后，为了使复杂的功能系统和执行方式简单化，利用 python、html、flask 等技术 with 框架制作了简洁的前端界面，直观的显示模型分析的结果和传感器实时测量的各项数据，完善用户体验。

五、 运行结果/效果

5.1 网页界面效果

疲劳感知系统通过用户名和密码进行登录验证，运行页面如下图：

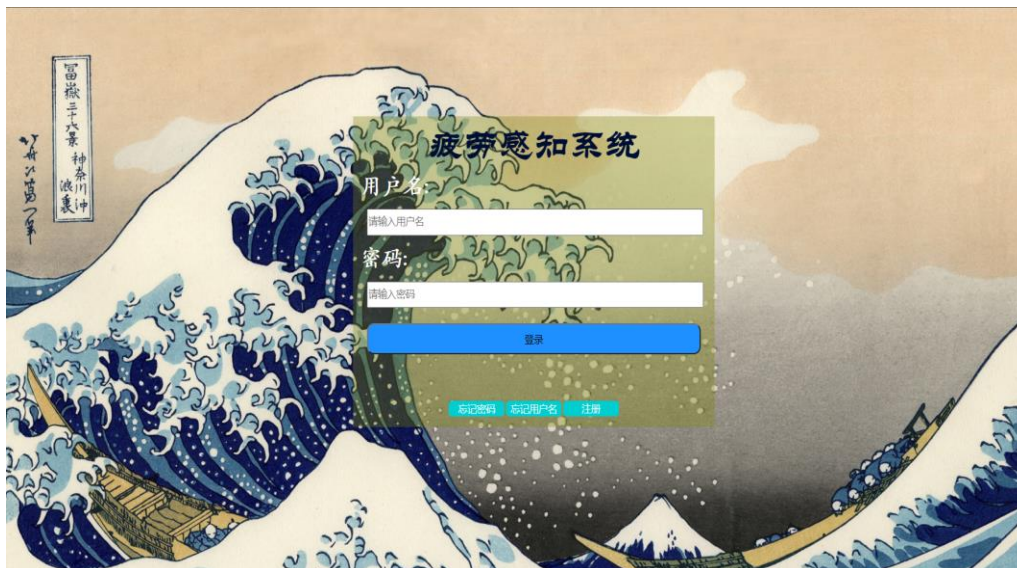


图 12：网页登录页面



图 13：等待检测页面

5.2 检测模型预测效果

本系统检测模型主要通过四个方面对用户进行检测，分别是脉搏检测、声音检测、眼动检测、脑电波检测，并得出最后的疲劳认证结果。检测模型如下：



图 14: 实时脉搏检测

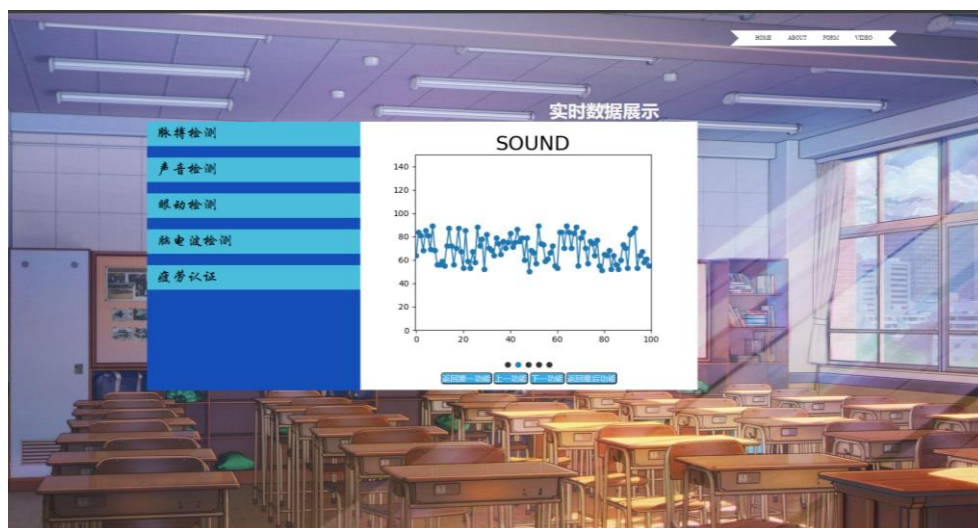


图 15: 实时声音检测



图 16: 实时眼动检测

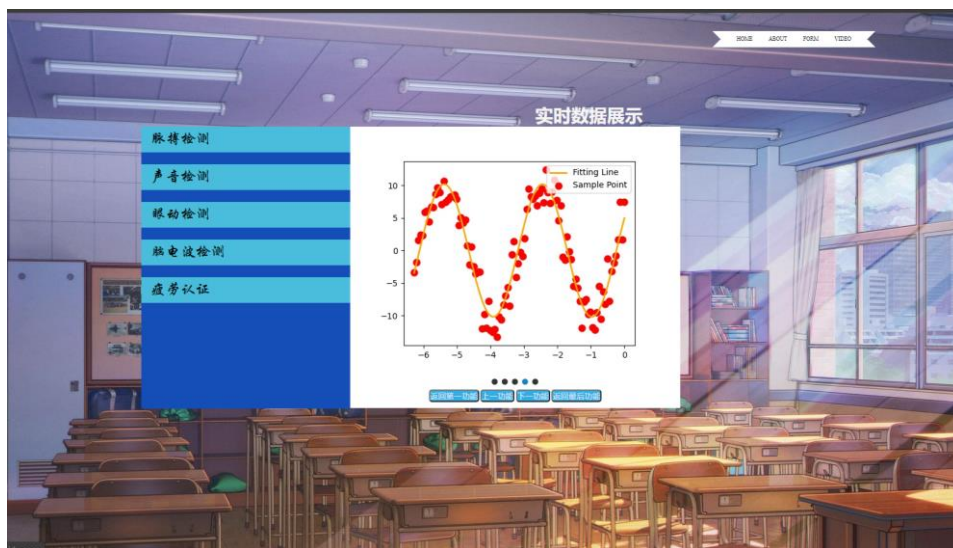


图 17：实时脑电波检测



图 18：疲劳认证结果展示

六、 创新与特色

6.1 综合多种传感器数据，疲劳检测和判断更准确

本系统使用了摄像头观察使用者的行为特征，利用脑电波传感器、脉搏传感器、声音采集器等多种传感器获取包括脉搏数据、眼动情况数据、声音数据、脑电波数据在内的多种数据集，综合多种传感器数据，多方面判断，使疲劳检测和判断更准确。

6.2 使用模糊决策树，保证准确率的前提下，具有更快的检测速度

由于应用领域，环境的不同，传感器采集的信息可能会收到预期之外的信息，影响数据的分析。

在技术方面，本系统使用模糊决策树分析算法对数据进行综合评判分析，合理地处理了由各种环境造成的不精确信息，有极强的分类能力。经模糊决策树进行信息分类，云端对信息的分析处理更加精确，同时也提高了系统检测的速度。

6.3 使用 python matplotlib 等多种手段，实现结果的可视化

基于本系统能够应用于多领域的优点，不同的领域有不同的要求，直观的数据展示显得非常重要。

本系统使用 python graphviz 等多种手段，使抽象数据具体化，结果可视化，满足不同领域的不同需求。

6.4 基于云构建和部署平台，提高系统的可迁移性、可扩展性和可维护性

本系统基于云端构建和部署平台，可以使用 Docker 技术进行容器化和虚拟化，易于部署和迁移。同时由于云端的特性，系统在可扩展性和可维护性上也具有显著优势。