

# 智能软件缺陷预测与排序系统 设计说明书

软件名称：智能软件缺陷预测与排序系统

硬件环境（填写 CPU、内存、硬盘等最低要求）

开发硬件环境：个人计算机，CPU：Intel PentiumIII以上等级，内存：1G 以上

运行硬件环境：个人计算机，CPU：Intel PentiumIII以上等级，内存：512M 以上

硬盘：40G 以上；

软件环境

开发软件环境：PyCharm 2024

运行软件环境：Windows 11

编程语言 Python

主要功能和技术特点

智能软件缺陷预测与排序系统是一款基于深度学习与机器学习技术的软件工具，专为解决当前软件工程中代码缺陷预测与排序的挑战而设计。系统结合了先进的代码大模型 CodeBERT 与线性回归算法，能够对代码中的潜在缺陷进行智能分析与排序，帮助开发人员优先处理高风险缺陷，从而提高软件质量和开发效率。该系统不仅能高效处理大规模代码数据，还能够应对数据稀缺与不平衡问题，使预测结果更加准确和具有针对性。

技术上，本系统使用 Python 作为开发语言，确保了其灵活性和跨平台兼容性。用户界面设计采用 PySide2 技术，提供简洁、直观的图形化交互方式，便于用户对缺陷检测结果进行管理与操作。系统整体架构注重模块化设计，结合深度学习模型与线性回归，形成了一套完善的缺陷预测与排序解决方案，适合各种规模的软件项目开发需求。

## 1 系统设计说明书

本系统旨在提升软件开发过程中代码质量，降低潜在缺陷带来的风险，保障软件的稳定性与安全性。无论是为提升开发团队的生产效率，优化项目管理中的缺陷处理流程，还是为了探索智能化软件质量保障的新领域，本系统都提供了精准且高效的解决方案。

本系统将帮助开发者迅速识别并优先处理高风险缺陷，提升软件产品的整体可靠性，并为对软件质量保障感兴趣的研究者和开发人员提供有力的技术支持。

### 1.1 软件架构

本系统采用 Python 语言开发，并使用 PySide2 进行界面设计。系统依托深度学习模型 CodeBERT 与线性回归技术，完成对软件缺陷的智能预测与排序。系统架构如图 1 所示。

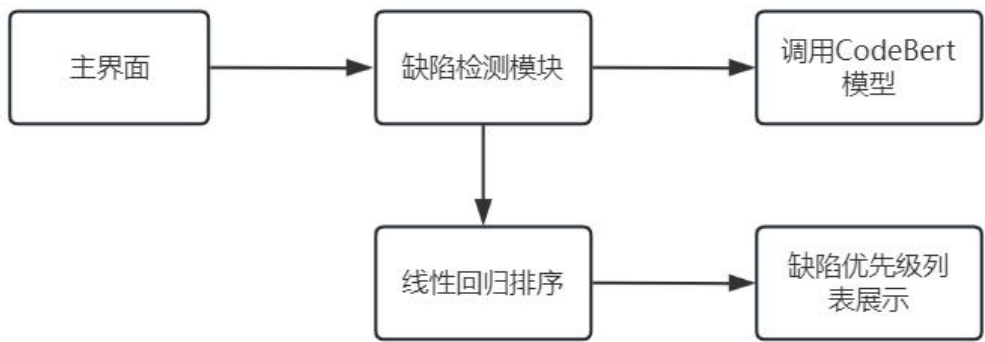


图 1 系统架构图

从图中可以看出，系统的操作从主界面开始，用户可以通过该界面进入缺陷检测模块。在主界面中，用户能够输入待分析的代码，点击相关按钮后，系统将自动调用 CodeBERT 模型对代码进行分析，并通过线性回归模型对潜在缺陷进行优先级排序。用户可以根据系统提供的缺陷排序列表，逐步进行优化和修复。此外，系统支持批量处理多个代码文件，以提升用户工作效率。

系统架构设计强调模块化与可扩展性，确保各个模块独立运行，同时通过统一接口实现数据传输与调用。无论是对小规模项目，还是大规模代码库的缺陷检测，本系统都能够灵活应对，满足不同规模的需求。

### 1.2 设计理念

设计思想流程如图 2 所示。

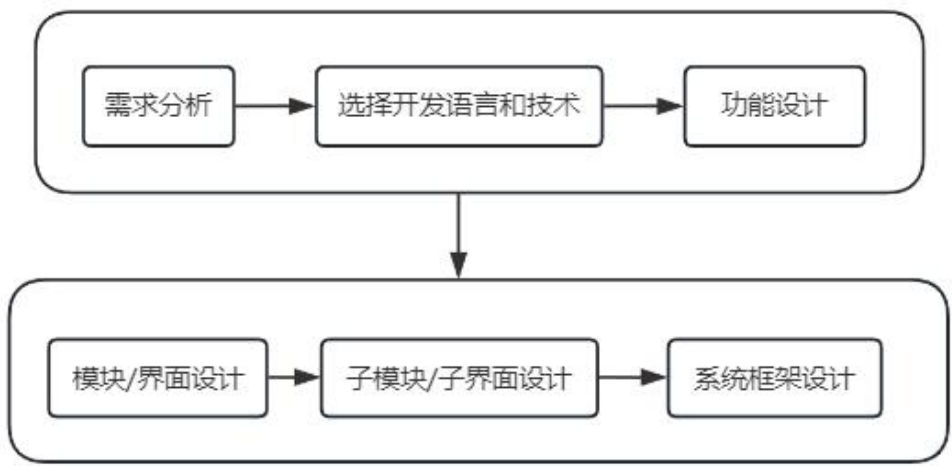


图 2 设计思想流程图

本系统的设计理念主要体现在两大核心阶段：首先是需求分析阶段，深入了解用户对代码缺陷预测与排序的实际需求，进而确定采用 Python 作为开发语言，并选择 PySide2 作为图形用户界面开发框架，确保系统具备良好的用户交互体验。同时，在技术选型上，系统结合了 CodeBERT 深度学习模型与线性回归算法，确保预测的准确性与效率。

其次是系统架构与界面设计，注重界面的简洁与模块的清晰划分。在主界面设计中，布局各类功能按钮和输入区域，用户可以方便地操作各模块并实时查看缺陷分析结果。在此基础上，系统对各功能模块进行详细的划分，确保每个子模块能够独立运作，并通过主界面将各子模块紧密连接，提供流畅的用户体验。

最终，将功能模块与界面布局紧密结合，形成完整的软件架构，为用户提供高效、准确、可扩展的智能软件缺陷预测与排序解决方案。

## 1.3 技术特点

### 1.3.1 Python-PySide2 技术

本系统采用了 Python 编程语言作为核心开发语言，结合 PySide2 库构建用户图形界面。Python 以其简洁、易读、丰富的库生态以及强大的处理能力，成为现代软件开发中的主流语言之一，特别是在人工智能和数据分析领域。PySide2 作为 Qt 框架的官方 Python 绑定，提供了丰富的界面控件和工具，帮助开发者快速创建高效、美观的图形用户界面（GUI）。

Python 语言以其动态特性、广泛的第三方库支持和高度的可扩展性，为本系统提供了丰富的工具链及算法实现。Python 的面向对象编程（OOP）能力与其灵活的模块化设计，使得整个系统在设计、开发和维护时都具备良好的扩展性和可维护性。

其主要技术特点包括：

#### （1）面向对象编程（OOP）

Python 具有强大的 OOP 特性，支持类、继承、多态等高级特性，这使得开发者可以灵活地对系统进行模块化设计。通过使用类来封装系统的功能模块，各个模块之间保持低耦合，使得代码的复用性和可维护性大大提高。

#### （2）丰富的第三方库支持

Python 具有大量的开源第三方库，尤其是在人工智能、数据分析领域，如用于机器学习的 scikit-learn 库、深度学习的 transformers 库和自然语言处理的 NLTK 库。本系统借助这些丰富的库资源，大大简化了模型开发、数据处理等复杂流程。

### **(3) 事件驱动编程与 GUI 设计**

使用 PySide2 构建的图形界面能够通过拖拽式布局和信号槽机制，实现用户界面的直观操作与事件响应。通过将控件直接拖放到窗口中，并通过信号与槽连接，开发者能够轻松实现交互功能，提高用户体验的友好性。

### **(4) 跨平台支持**

Python 及其 PySide2 库具有良好的跨平台特性，能够支持 Windows、Linux 和 macOS 等多种操作系统，这使得本系统能够在多种环境下灵活部署和运行。

## **1.3.2 CodeBERT 技术**

本系统核心缺陷预测功能由 CodeBERT 大模型提供支持。CodeBERT 是一款基于 Transformer 架构的深度学习模型，经过大规模代码数据集预训练，专门用于理解和生成代码。该模型能够解析各种编程语言，并通过上下文信息进行准确的缺陷预测。

CodeBERT 的主要技术优势在于其强大的代码语义理解能力。它能够在不同编程语言间进行迁移学习，并根据上下文信息预测代码中的潜在缺陷，具有极高的准确性。同时，CodeBERT 模型采用了自监督学习的预训练策略，能够在缺乏大量标注数据的情况下，依然保证较高的预测性能。

其技术特点包括：

### **(1) 代码语义理解**

CodeBERT 模型通过对大规模代码语料库进行预训练，能够捕捉到代码的深层次语义信息，不仅理解代码的语法结构，还能通过上下文分析识别出代码中的潜在问题。

### **(2) 跨语言支持**

CodeBERT 能够处理多种编程语言的数据，包括 Python、Java、C++ 等。这使得系统不仅限于处理某一种语言的代码缺陷，能够对多语言的代码库进行预测和排序。

### **(3) 自监督学习**

CodeBERT 采用自监督学习策略，能够在无监督的数据下进行训练。这一特性特别适用于缺陷预测场景，尤其是在标注数据稀缺的情况下，CodeBERT 仍然能保证高效的预测能力。

### 1.3.3 线性回归排序技术

本系统通过线性回归模型对检测出的代码缺陷进行排序，以帮助开发者优先处理严重的缺陷。线性回归是一种经典的统计学习方法，通过拟合一条最优直线来预测目标变量（缺陷的严重性）。本系统根据不同缺陷的影响程度、代码上下文信息以及缺陷的历史修复数据等多个维度，训练线性回归模型，对缺陷的优先级进行精准排序。

其技术特点包括：

#### （1）数据驱动的优先级排序

线性回归模型能够根据训练数据中的历史缺陷和修复记录，动态调整每个缺陷的权重，从而对新检测出的缺陷进行科学排序，帮助开发者快速定位优先修复的代码问题。

#### （2）高效且易于解释

线性回归模型具有较低的计算复杂度，能够快速处理大规模代码库中的缺陷排序问题。同时，模型的输出结果易于解释，开发者能够直观理解缺陷排序背后的逻辑。

#### （3）可扩展性

线性回归模型具有良好的可扩展性，可以根据不同项目的需求灵活调整特征维度。例如，可以引入更多关于代码结构、开发者修改历史等维度的数据，以进一步提升排序的准确性和合理性。

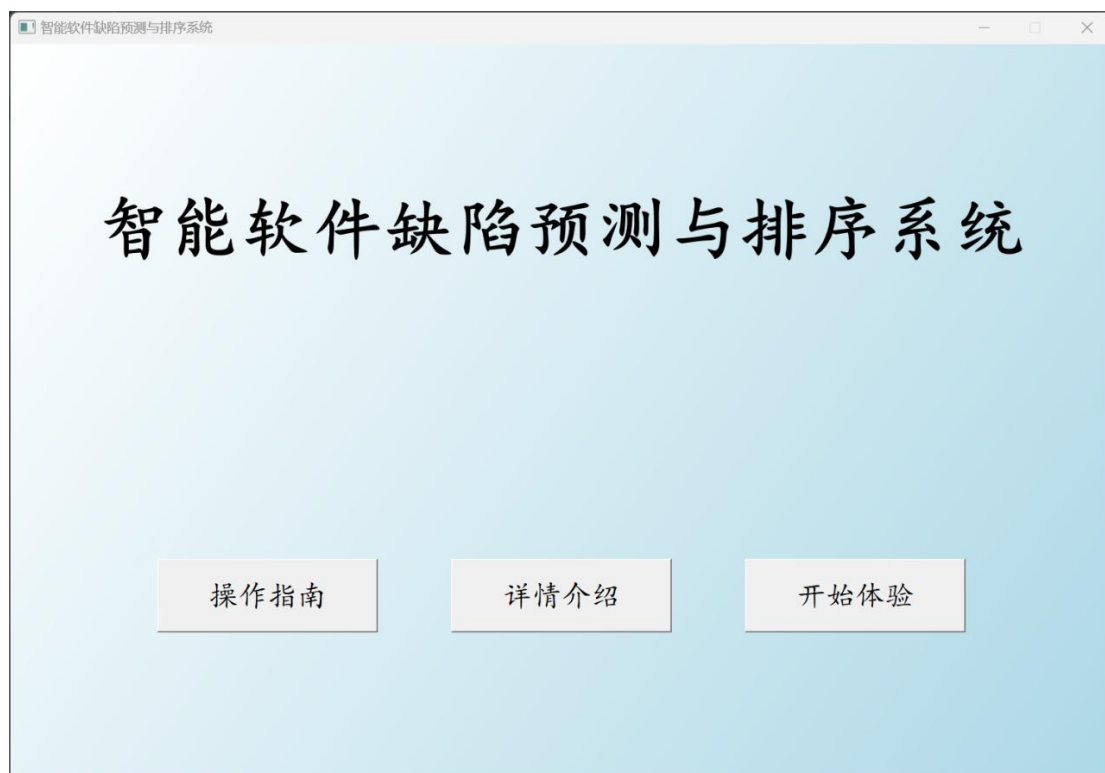
## 2 使用说明书

### 2.1 使用前相关准备

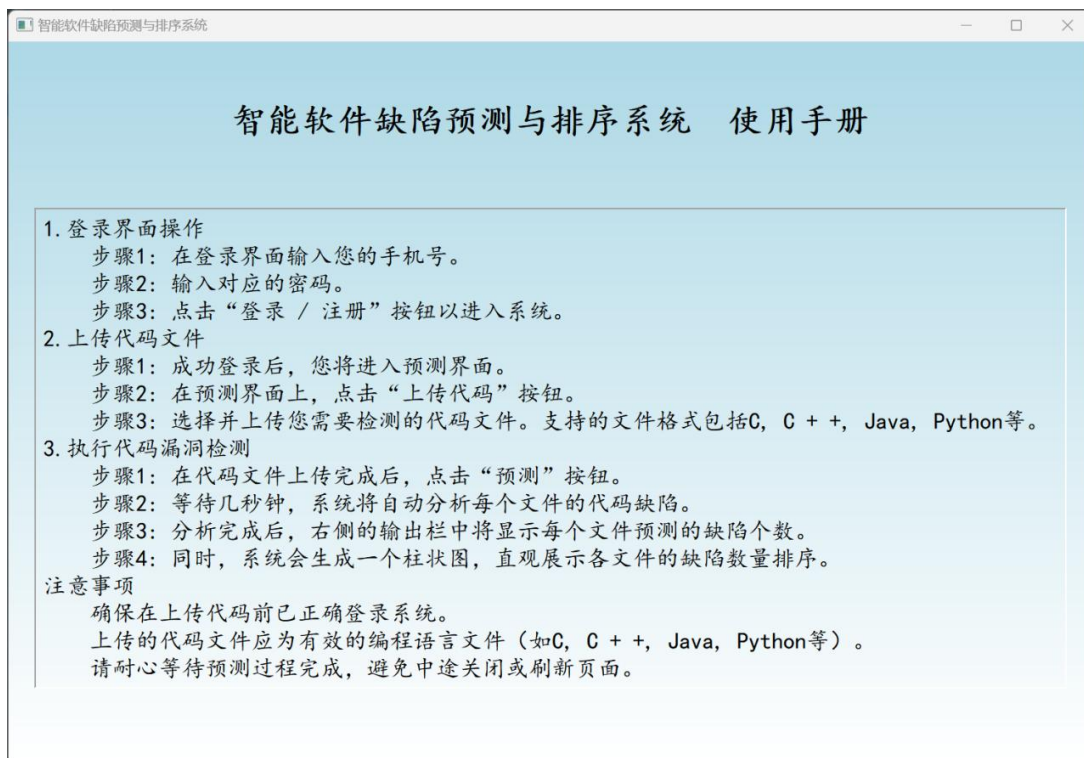
在 Window 11 或者 Window XP 等系统上运行 `run.bat` 文件，软件即可打开。

### 2.2 使用介绍

1. 软件的初始界面具有操作指南、详情介绍和开始体验三个按钮选项。界面如下所示：



3. 点击“操作指南”按钮，跳转界面，可查看软件系统的使用手册，界面如下所示：



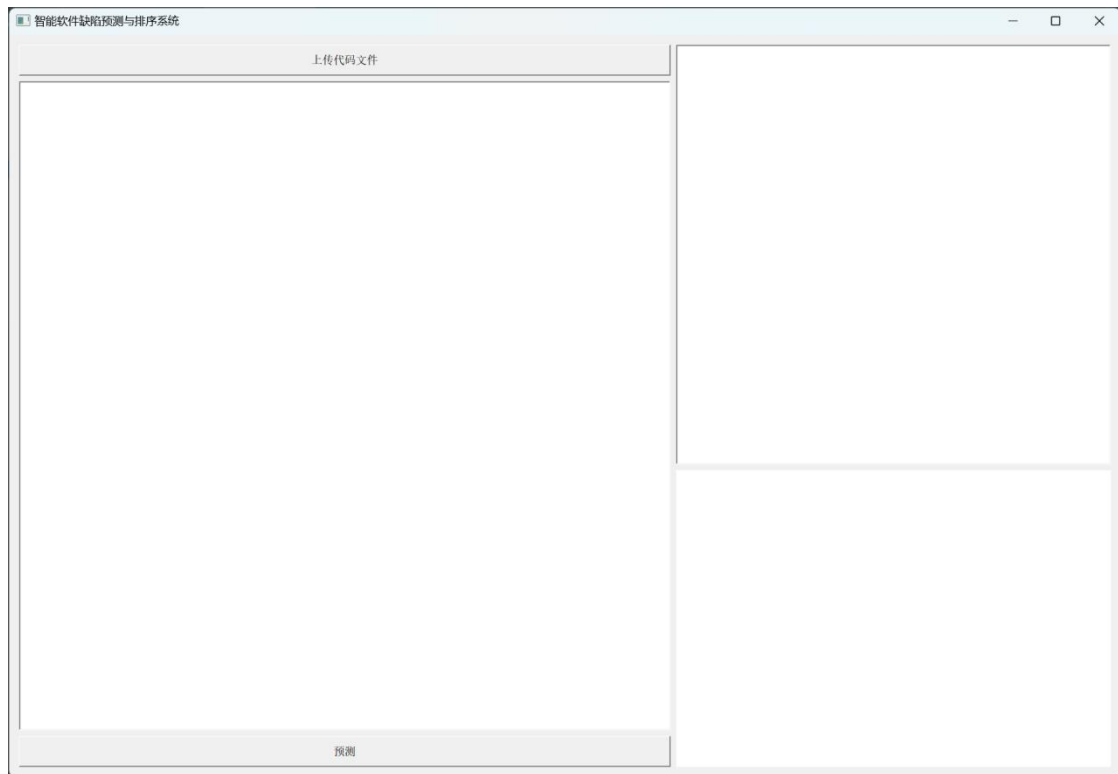
4. 点击“详情介绍”按钮，跳转界面，可查看关于智能软件缺陷预测与排序系统的详情介绍，界面如下所示：



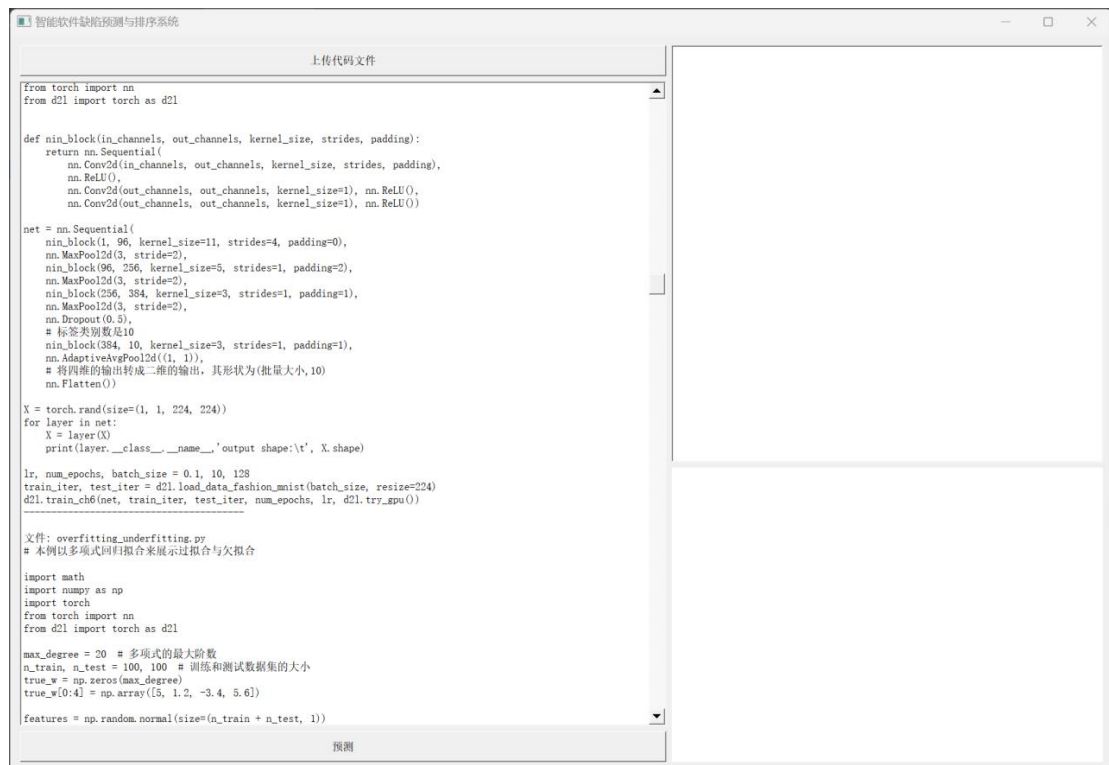
5. 点击“开始体验”按钮，跳转界面，用户可进行登录和注册。用户输入手机号和密码，新用户自动注册账号并登录，已注册过的用户输入手机号和密码进行检测，手机号和密码正确则登录成功。界面如下所示：



6. 用户登录成功后进入软件缺陷预测与排序系统界面，界面如下所示：

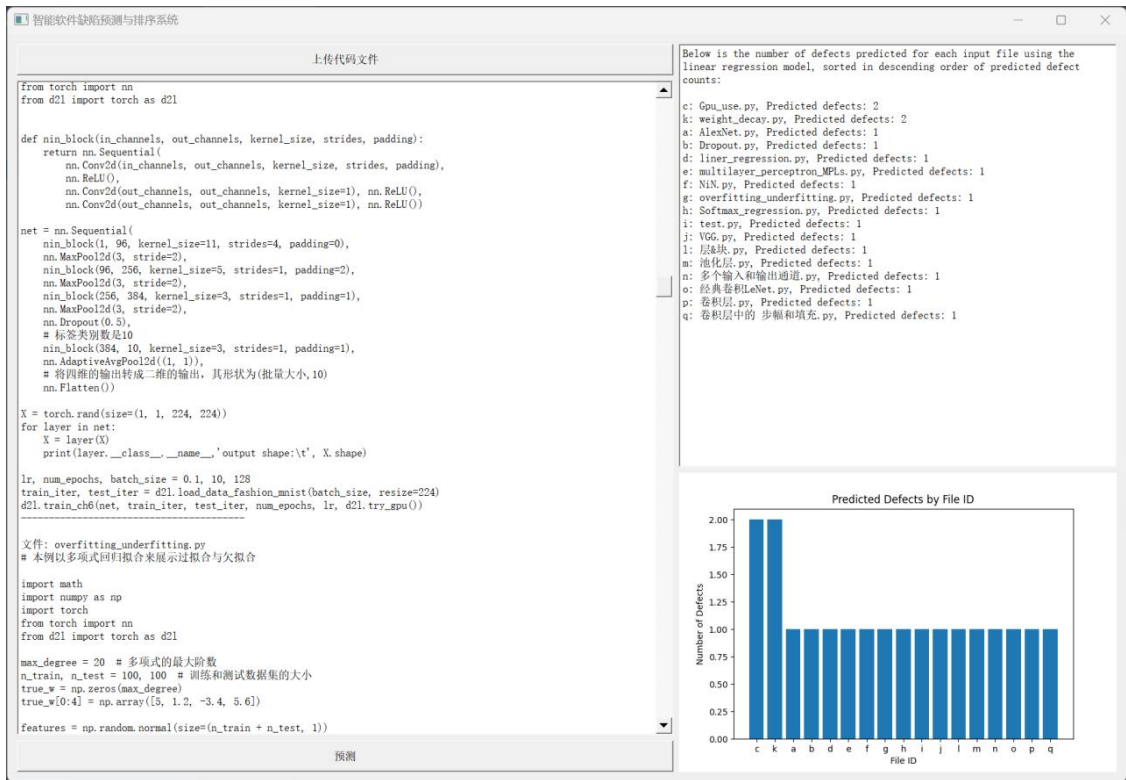


7. 点击上传代码软件，可打开文件管理器选择多个需要检测的代码文件，上传成功后，文件内容会在左边的文本框中显示，运行效果如下：





8. 点击预测按钮，后台会运行训练好的模型对上传代码文件进行预测排序，预测与排序结果会在右上部分的文本框中显示，同时也会绘制成柱状图在右下部分区域显示，运行效果如下：



## 2.3 注意事项

如有问题请联系作者进行交流。