**哈 尔 滨 理 工 大 学**

**毕业设计开题报告**

**题 目：**

**院、 系：**

**姓 名：**

**指导教师：**

**系 主 任： 姚登举**

**2023年10月**

**一、选题的依据、意义，理论或实际方面的价值**

**1.选题的依据和意义**

lncRNA-disease关联预测在生物信息学领域具有重要意义。长链非编码RNA（lncRNA）在细胞过程中起着关键作用，与多种疾病的发生和发展密切相关。通过深入研究lncRNA与疾病之间的关联，可以揭示其在疾病发生机制中的作用，并为疾病的早期诊断和治疗提供新的方向。深度学习作为一种强大的机器学习方法，已经在多个领域取得了显著的成果。其能够从大规模数据中学习和挖掘复杂的模式和规律，因此被广泛应用于生物信息学和医学领域。

通过基于深度学习的预测平台，可以揭示lncRNA与疾病之间的潜在关联，帮助理解lncRNA在疾病发生机制中的具体作用。这有助于揭示疾病的发病机制，为疾病的早期诊断和治疗提供新的方向和靶点。通过建立lncRNA-disease关联预测模型，可以为疾病的预测和诊断提供有力的支持。通过分析患者的lncRNA表达谱和临床数据，可以快速准确地预测患者的疾病风险和诊断结果，有助于提前干预和治疗。基于深度学习的lncRNA-disease关联预测平台可以根据个体的遗传背景、环境因素和临床特征，为每个患者量身定制个性化的诊断和治疗方案，提高治疗效果和生存质量。该研究将生物信息学和深度学习相结合，探索双方的交叉应用。这将为生物信息学和机器学习领域的研究人员提供新的思路和方法，促进两个领域的交流和发展。

因此，构建一个基于深度学习的lncRNA-disease关联预测平台具有重要的科学和应用价值。

**2.理论或实际应用方面的价值**

基于深度学习的lncRNA-disease关联预测平台具有重要的理论和实际应用价值。它不仅有助于揭示lncRNA-disease关联的机制和提高预测模型性能，还可以在疾病预测与诊断、个性化医学和新药研发等方面，为生物医学研究和临床应用提供重要的支持和指导。

1.理论价值

（1）探索lncRNA-disease关联的机制：通过深度学习算法挖掘lncRNA与疾病之间的模式和规律，可以揭示其关联的潜在机制，有助于深入理解lncRNA在疾病发生和发展中的作用。

（2）探索深度学习在生物医学领域的应用：本研究将深度学习算法应用于lncRNA-disease关联的预测，探索深度学习在生物医学领域的应用潜力，为其他相关研究提供理论参考和方法借鉴。

（3）优化预测模型性能：通过深度学习算法的使用，可以提高lncRNA-disease关联预测模型的准确性和可解释性，推动预测模型的发展和优化。

2.实际应用方面价值

（1）疾病预测与诊断：通过该平台，可以根据患者的lncRNA表达谱及临床特征，快速准确地预测患者的疾病风险和诊断结果，为早期预防、诊断和治疗提供参考依据。

（2）候选疾病标志物的筛选：通过该平台，可以识别出与特定疾病相关的lncRNA，为疾病标志物的筛选提供线索，有助于开发新的临床诊断和治疗方法。

（3）个性化医学的推进：基于深度学习的lncRNA-disease关联预测平台可以根据个体的遗传背景、环境因素和临床特征，为每个患者量身定制个性化的诊断和治疗策略，提高治疗效果和生存质量。

（4）新药研发的辅助工具：通过该平台的预测结果，可以为新药研发提供指导，揭示潜在的药物靶点和治疗策略，加速新药的研究和开发过程。

**二、国内外研究现状**

国内外的研究表明，利用深度学习方法进行lncRNA-disease关联预测已经取得了一定的成果。例如，一些研究利用深度学习模型对lncRNA和疾病的表达谱进行建模，预测它们之间的关联。此外，还有研究者结合生物信息学和网络分析等方法，构建了一些lncRNA-disease关联预测的数据库和工具。这些研究通过不同的深度学习算法和模型设计，提高了预测模型的性能和可解释性。然而，仍然存在一些挑战，如数据稀疏性和特征选择等，需要进一步的研究和改进。因此，本课题可以在已有研究的基础上，进一步探索和完善基于深度学习的lncRNA-disease关联预测平台，为生物医学研究和临床应用提供更有效的工具和方法。

**三、课题研究的内容及拟采取的方法**

**3.1编写目的**

开发系统的过程中，去调查用户的功能诉求，对需要存在的功能进行需求分析是特别重要的，且对于系统的开发有着实际的意义，通过对用户的需求进行分析，结合实际情况进行开发研究，对用户的所有需求做出一个完整的基本的框架，然后一步一步的完成、实现。需求分析可以为系统的开发提供一个目标，只有按照这个目标进行开发设计，才能进行完整的开发，这样设计出的系统才有使用的意义，才能在竞争激烈的软件市场中生存，才能真正的帮助人们解决问题，提高实际的效率。

**3.2系统的目标**

基于深度学习的lncRNA-disease关联预测平台以用户获取预测结果和管理员对数据的管理为基础展开设计。

1.开发深度学习的lncRNA-disease关联预测平台，使所有使用者可以浏览自己个人信息，也可以修改自己的个人信息。

2.专家可以依据需求自行修改通告内容、数据集更新，并于普通用户在线交流。

3.普通用户可以通过输入自己的关联数据让平台进行自主预测疾病，平台可以对预测结果低于正常值的用户进行预警，普通用户可以与专家进行线上交流。

4.管理员权限最多，并可以修改其他用户权限并管理全站信息。

**3.3运行环境**

操作系统：Windows

支持环境：JDK8

数 据 库：MySQL 8

**3.4系统功能需求**

通过与家人、同学、医护人员等用户交流、问卷调查等方式调研，分析结果后明确了基于深度学习的lncRNA-disease关联预测平台的功能需求，将用户分为普通用户、专家管理员3个角色，管理员账户由系统自行配置，专家由管理员进行添加和维护，普通用户自行注册登录。这样操作使得系统里的信息更加安全。下面就系统主要功能分别阐述它们的需求。

**3.4.1用户登陆注册**

普通用户若第一次访问平台，用户需要进行注册获得账号。若用户已有账号，则通过登录页面提前输入账号和密码进行登录，如果登录成功，用户进入系统。如果输入的信息不正确，则提示给用户错误信息。本系统支持三种角色登录。

**3.4.2通告管理**

系统应具有以下功能：

1.通告列表，主要显示通告标题名称、操作按键；

2.按通告名称查询通告信息；

3.添加通告，主要内容包括通告标题名称、通告内容。

4.单个删除、批量删除，根据选中的通告标题名称，进行单个或批量删除操作；

5.对通告进行按时间发布排序。

**3.4.3疾病预测**

系统应具有以下功能：

1.支持用户上传数据；

2.根据数据进行关联预测，生成结果；

3.对结果数据进行可视化处理。

**3.4.4数据监测**

系统应具有以下功能：

1.对于异常数据的报警；

2.支持数据在不同角色间的传递。

**3.4.5数据集管理**

系统应具有以下功能：

1.添加数据集名称；

2.编辑已有数据；

3.单个删除、批量删除数据。

**3.4.6专家问诊**

系统应具有以下功能：

1.普通用户可以与已有专家进行交流；

2.用户可以上传自己的预测结果给专家。

**3.4.7信息管理**

系统应具有以下功能：

1.显示个人信息列表，主要包含账号ID、用户名名称、密码、描述、操作；

2.编辑个人信息；

3.修改登录密码；

**3.4.8用户及权限管理**

1.显示用户名称列表，主要内容包含账号ID、用户名名称；

2.单个删除、批量删除，根据选中的用户名称，进行单个或批量删除操作；

3.对用户进行按名称首字母排序；

4.对用户所拥有权限进行管理，开启或关闭权限。

**3.5数据描述**

**3.5.1静态数据**

用户：用户ID，用户名，密码，是否被删除，用户状态，注册时间，上次登录时间，注册时IP地址，上次登录IP。

专家信息：专家ID、专家姓名、专家账号、专家工号、专家职称、登记日期。

通告信息：通告编号ID、通告标题名称、通告描述。

上报信息：用户ID、用户名称、预测结果。

预测标准指标信息：标准指标信息ID、标准指标观测点ID、标准指标内容、备注信息。

图表：主要观测点ID、主要观测点名称、属性、备注信息。

专家咨询：用户ID、专家ID、专家名称、用户名称、信息发布时间。

登录日志，登录日志ID，登录URL，登录IP，登录是否成功。

系统管理：用户名、密码、权限。

**3.5.2动态数据**

输入数据：

鼠标对按钮的点击；

   查询方式、查询关键字；

新建通告、数据集；

数据集、用户信息更改；

备份数据恢复所有的数据备份文件；

受限操作所需的密码；

输出数据：

查询关键字确定的数据库记录；

   统计结果及其格式化文件；

   信息录入，删改结果（成功或失败）；

展示关联预测结果；

   数据备份输出的数据备份文件；

内部数据：查询操作建立的索引。

**3.6系统功能模块**

根据上述需求得出功能模块划分如表3-1所示。

表3-1 功能划分表

|  |  |  |
| --- | --- | --- |
| 模块划分 | 说明 | 优先级 |
| 通告管理模块 | 显示当前系统公告，可对公告内容进行上传、修改。 | 高 |
| 疾病预测模块 | 上传身体数据，系统进行计算预测，显示结果并生成可视化图。 | 高 |
| 疾病监测模块 | 系统根据预测结果进行异常提示或疾病预警，用户可选择将结果上报给专家。 | 高 |
| 数据集管理模块 | 可对当前数据集进行上报、分类、删除。 | 高 |
| 专家咨询模块 | 普通用户和专家可在此模块实现在线线上文字交流。 | 高 |
| 信息管理模块 | 所有用户在此模块可以实现个人信息管理。 | 高 |
| 系统配置模块 | 管理员在此模块管理用户信息及用户权限 | 高 |
| 图表分析子模块 | 将预测结果的数据整理分析形成折线图、饼状图等可视化图。 | 低 |
| 个人信息管理子模块 | 包括注册、登录、修改个人信息和修改密码等。 | 低 |

**3.6.1系统的功能模块划分**

本系统功能模块图如图3-1：

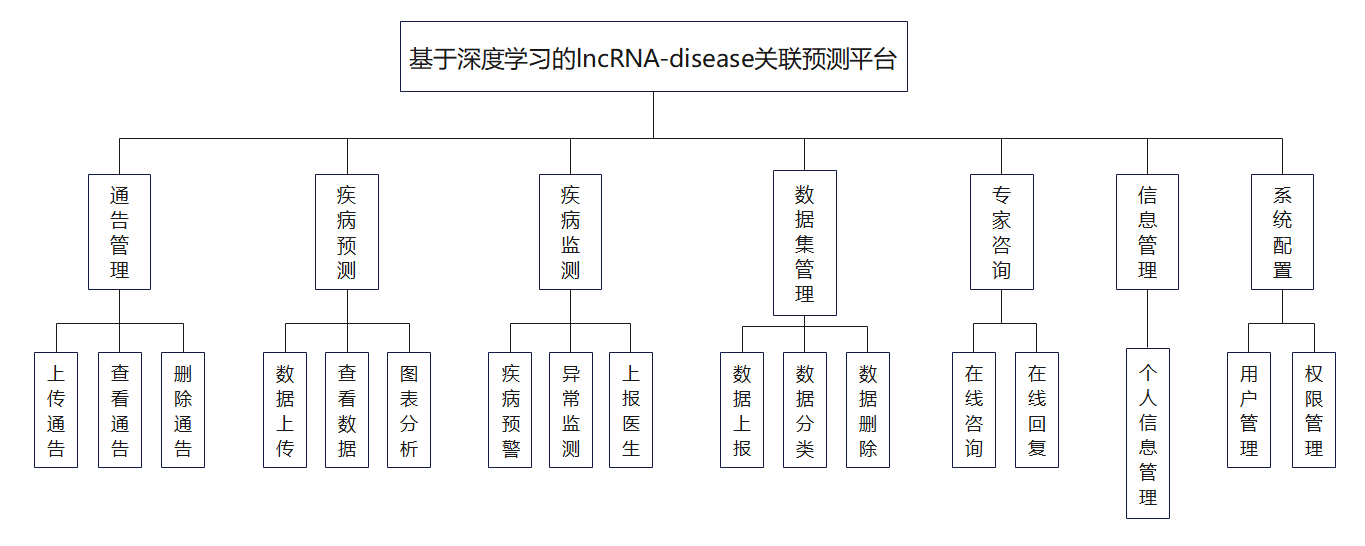


图3-1 功能模块图

本系统分为七个模块，以下是对各功能模块的概述：

1.通告管理模块

在本模块中，管理员和专家可以依据需求来增加、删除通告内容，普通用户只可以查看已发布的通告。通告默认以发布时间排序。本模块用例图3-2如下：

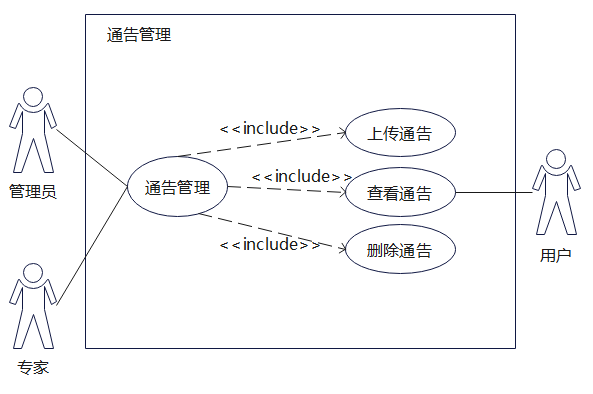


图3-2 通告管理模块用例图

2.疾病预测模块

本模块中，普通用户可以上传自己的身体各项数据，系统自动预测并显示预测结果，并生成数据相关图表实现数据可视化，供普通用户方便查看理解。本模块用例图3-3如下：

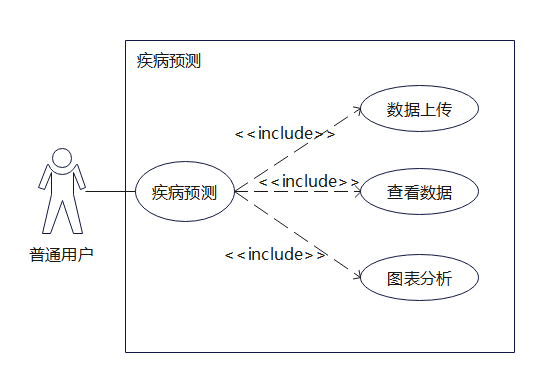


图3-3 疾病预测用例图

3.疾病监测模块

本模块中，系统将根据普通用户预测的结果，自主对比数据库中的标准健康数据标准，若该普通用户某一项指标低于标准下限，平台将对该用户发出提示异常，若用户检查数据上传无误，则系统将发出疾病预警，该用户可以将结果发送给专家。本模块用例图3-4如下：

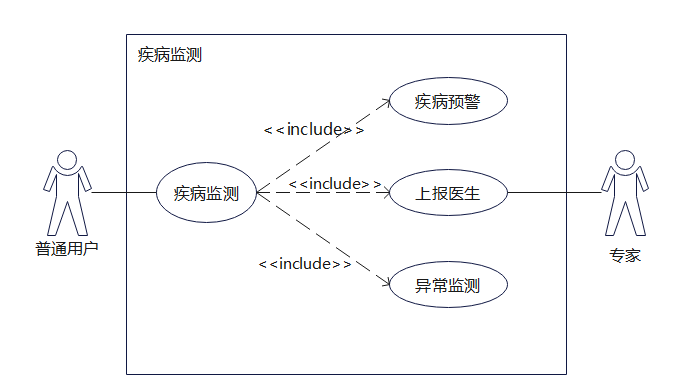


图3-4 疾病监测用例图

本模块与疾病预测模块相关联，具体顶层数据流图3-5，第0层数据流图3-6，第1层数据流图3-7如下：

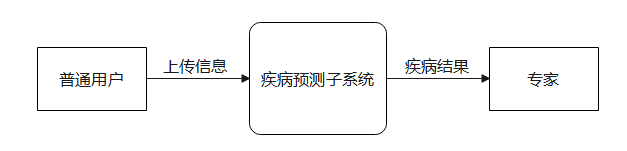


图3-5 顶层数据流图

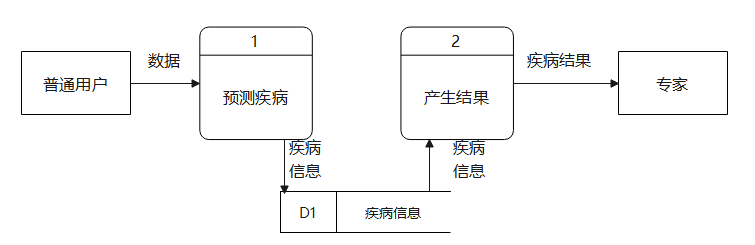


图3-6 第0层数据流图

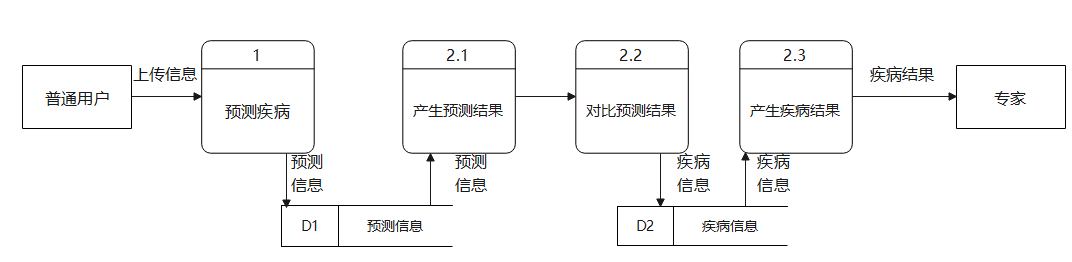


图3-7 第1层数据流图

4.数据集管理模块

本模块中，只有管理员和专家可以对预测模型数据集进行修改、更新，并且可以对数据集进行分类。本模块用例图3-8如下：

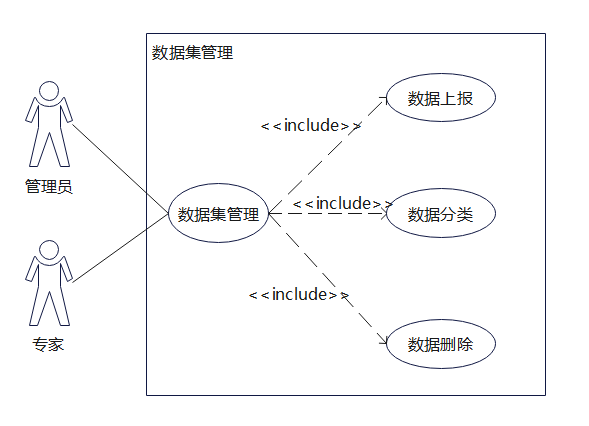


图3-8 数据集管理用例图

5.专家咨询模块

本模块中，普通用户可以与专家进行线上文字交流，专家也可以线上回复用户信息。本模块用例图3-9如下：

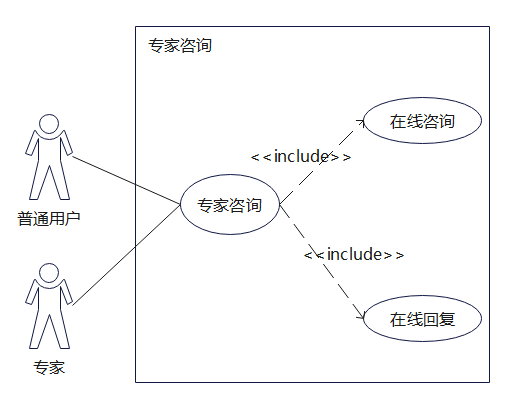


图3-9 专家咨询用例图

6.信息管理模块

本模块中，用户实现登陆注册信息，所有用户可以对自己的个人信息进行管理，如用户名、密码等，可以修改密码。本模块用例图3-10如下：

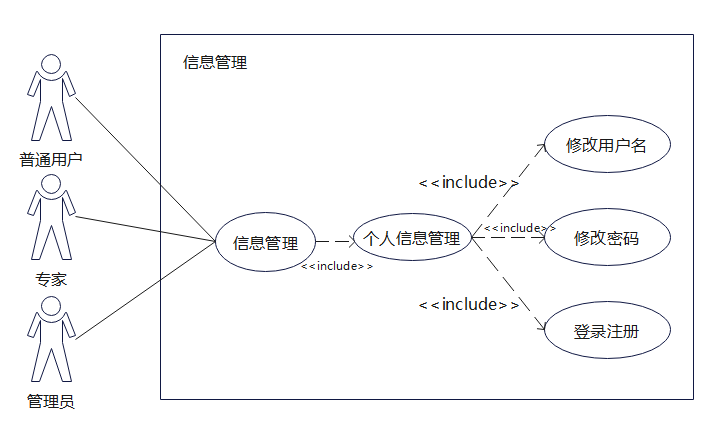


图3-10 信息管理用例图

7.系统配置模块

本模块中，只有管理员可以对所有用户进行管理，例如删除用户等，也可以对用户权限进行管理，打开或关闭用户权限。本模块用例图3-11如下：

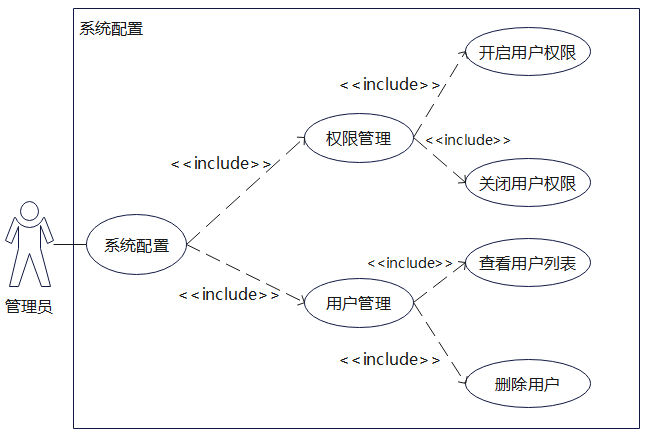


图3-11 系统配置用例图

**3.7系统的主要参与者分析**

出于操作方面和管理方面等考虑，系统分为三类角色，分别为普通用户、专家、管理员。

普通用户可以利用系统查看预测疾病、查看通告，收到来自系统的疾病预警并可以与专家线上交流。管理员和专家可以对数据集进行修改、对通告进行更新等管理。专家可以收到来自普通用户的预测结果信息并可以与普通用户进行线上交流。管理员权限最大，可以管理其他角色权限并可以修改系统相关数据。

系统分为前台运营管理和后台数据处理。后台数据管理有通告管理子系统、数据集管理子系统、信息管理子系统、系统配置子系统等。前台营运管理有个疾病监测子系统、疾病监测子系统。

本系统用户用例图3-12如下：

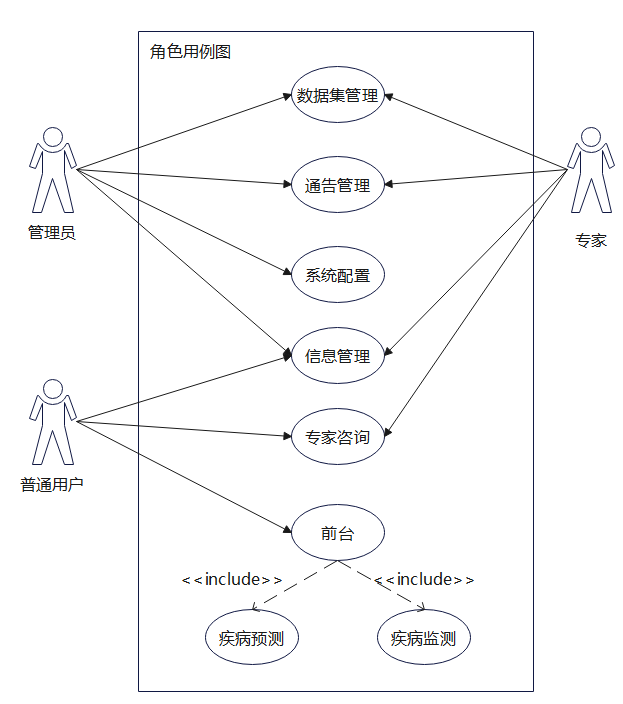


图3-12系统用户用例图

本系统设置三种登录身份：管理员、专家和普通用户。

1.普通用户：普通用户对自己的基本信息进行管理，并且可以使用平台上传自己的相关医疗数据，系统会据此进行疾病预测生成结果，并生成图表更清晰展示数据。对于某些数据异常系统会提示是否重新输入数据，若用户上传数据无误且结果处于低于健康标准最差值则系统发出预警提醒用户注意，用户可以将结果上报给医生，并可以与医生发消息线上交流。普通用户用例图，如图3-13所示。

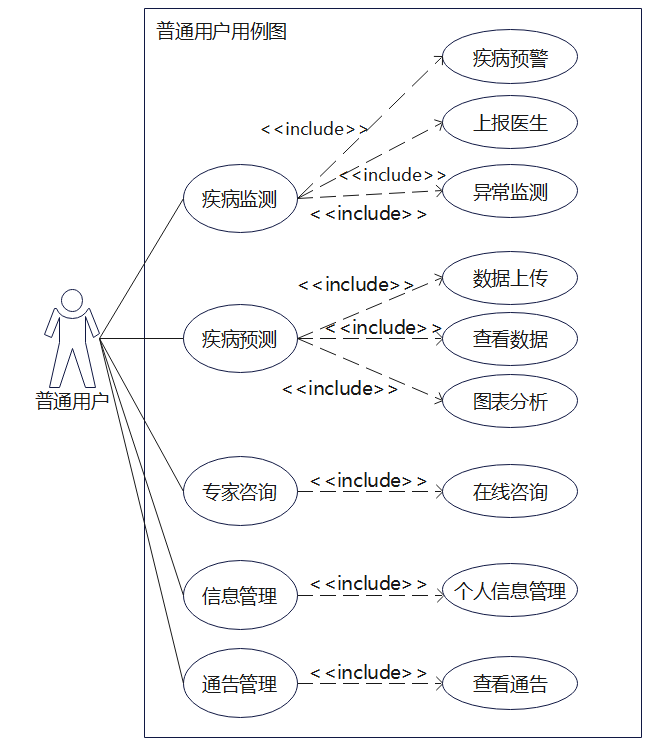


图3-13 普通用户用例图

2.管理员：管理员拥有独立账号，拥有通告和数据集管理权限，可对其进行操作，可对个人信息进行管理，拥有独立权限系统设置，可以管理用户账号以及分配权限。管理员用例图，如图3-14所示。

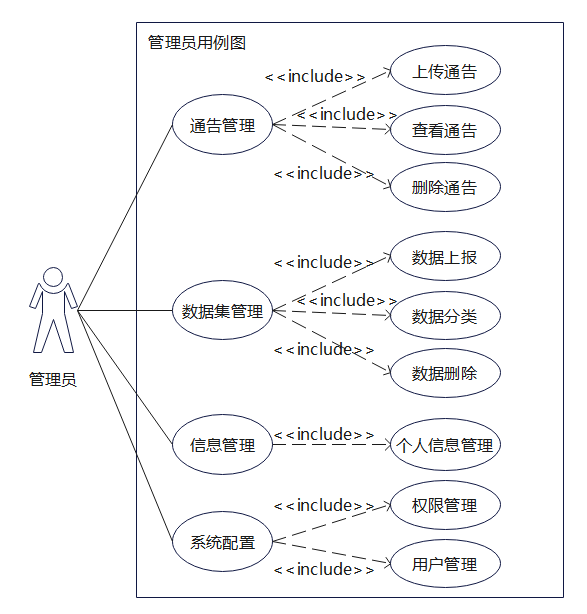


图3-14 管理员用例图

3.专家：专家与管理员一样拥有通告和数据集的管理权限，专家依据需要更改上传新的数据，专家也可以进行个人信息管理，并可以在线回复普通用户，专家用例图，如图3-15所示。

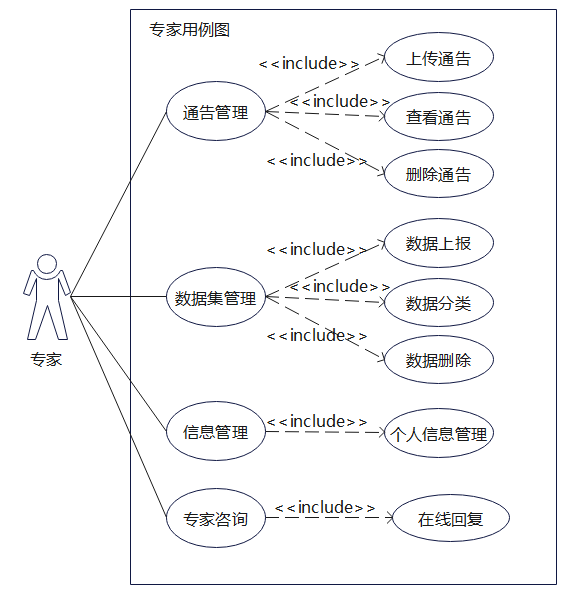


图3-15 专家用例图

**3.8可行性分析**

基于深度学习的lncRNA-disease关联预测平台主要目标是实现智能预测疾病。在系统的开发之前先确定了目标，之后从技术、操作、经济、法律四个方面进行可行性分析。

1.技术可行性

基于深度学习的lncRNA-disease关联预测平台主要采用了Python、Java的技术，后台开发数据库运用MySQL数据库，构建这样一个可靠的预测平台，关键是拥有高质量且充足的数据集。目前，公共数据库如LncRNAdb、LncRNADisease等提供了大量的lncRNA-disease关联数据，可以用于模型训练和验证。此外，还可以通过生物实验获得更多的数据，进一步提高平台的精确性。

深度学习在许多领域都取得了巨大的成功，对于lncRNA-disease关联预测也显示出潜力。可以采用卷积神经网络（CNN）、递归神经网络（RNN）或Transformer等模型，通过学习lncRNA和疾病之间的复杂关系，实现精准的预测。在lncRNA-disease关联预测中，使用基于深度学习的方法，如自编码器或生成对抗网络，从lncRNA和疾病的表达数据中提取有用的特征。此外，还可以考虑整合其他生物信息学数据，如基因表达、网络拓扑和蛋白质相互作用等，进一步提高预测准确性。

lncRNA-disease关联预测平台的目标是为生物学研究人员提供一个方便、高效的工具。通过将模型整合到用户友好的界面中，研究人员可以输入lncRNA和疾病信息，获取预测结果。此外，平台还可以提供其他功能，如数据可视化、关联分析和交互式探索等，进一步提升用户体验和应用价值。

2.操作可行性

操作可行性也就是系统的可用性，一个系统的操作是否容易决定着这个系统的使用度，在系统的操作方面的设计我都是采取简洁易懂的方式，操作的整个菜单界面整齐有序，所有的功能都有序的排列，不会出现重叠或者需要转换的现象，用户想要哪方面的操作都可以直接进行操作，所以该系统任何人都可以进行操作，不需要有相关专业的技术这样用户在操作起来就容易很多。

3.经济可行性

本项目开发的初衷就是为了节约，因为系统开发的所有过程都是自主开发，而且在开发过程使用到的技术也是市面上常见的容易操作的，不需要请专业的人士花资金来进行系统的开发。开发软件可以在网络中进行免费下载，对计算机的软硬件没有很高要求，因此这个项目是非常实惠的，在经济方面是完全可行的。

4.法律可行性

本系统本设计所用开发软件全为正版，数据信息均可保证合法来源，并且本系统全部为自主研发设计，不存在侵权行为，技术资料为官方开发文档，不构成社会危害性，所以法律是可行的。

**3.9系统性能要求**

**3.9.1数据精确度**

保证查询的查全率和查准率为100%，所有在相应域中包含查询关键字的记录都能查到，所有在相应域中不包含查询关键字的记录都不能查到。

**3.9.2系统响应时间**

系统对大部分操作的相应时间应在1—2秒内。

**3.9.3其他需求**

1.适应性需求

满足运行环境在允许操作系统之间的安全转换和与其他应用软件的独立运行要求。

2.输入输出需求

输入输出数据类型，对软件的数据输出及必须标明的控制输出量进行解释并举例，包括对硬拷贝报告（正常结果输出、状态输出及异常输出）以及图形或显示报告的描述。

3.安全性需求

安全性是一个系统的基本需求。用户进入系统都必须通过用户名及密码验证，登录成功后才能浏览存于此网站的个人信息，这样可以保证系统的安全性。在具体实现中对不同的权限进行设定，不同权限的用户在系统中登陆后，不可以进行自身权限以外的操作。在实际使用的过程中系统都会涉及到很对需要进行保密的数据，系统出现一些漏洞将这些信息泄露出去将会对用户产生很大的损失，所以在开发系统时我们充分确保了本系统的安全性、可靠性。

4.界面需求

系统对界面的要求窗口布局清晰，颜色搭配合理，色调柔和，各窗体主题风格一致同时采用windows 的通用图形界面，同时对鼠标和键盘提供支持，界面的设计应遵循如下规则：  
 （1）界面具有一致性，接你按规范应遵循MS Windows软件界面概念规范。  
 （2）提供简单的错误处理。  
 （3）提供信息反馈，用多种信息提示用户当前的系统运行状态。  
 （4）操作可逆，其动作可以是单个操作，或者是一个相对对立的操作序列。

5.时间特殊性需求

当管理员向系统添加数据资源时需要在一定时间内处理数据并将数据录入数据库中。

6.稳定性需求

系统部署后，在硬件条件和支持软件条件没有变化的情况下，能够一直保持运行状态，直到系统被升级或代替。

7.扩展性需求

系统应该支持功能扩展与支持环境的扩展。功能扩展就是在现有的功能模块的基础上可以添加信息的功能模块。

8.故障处理能力需求

系统可能遇到的软件故障是数据库与应用程序服务器。为了满足信息处理的需求，可以采取数据恢复数据来解决。

**四、课题研究中的主要难点及解决办法**

难点1：

数据稀疏性和噪声干扰：lncRNA-disease数据的获取和标注存在一定的困难，关联数据通常是稀疏的，且存在噪声和缺失值，这可能导致模型的准确性下降。

解决办法：

使用数据清洗和预处理技术，例如填充缺失值、去除噪声、降低维度等，以提高数据的质量和可靠性。同时采用合适的数据预处理和增强方法，可以考虑使用数据增强技术来扩充数据集，增加训练样本的多样性。

难点2：

特征选择和融合：lncRNA-disease关联的特征种类繁多，如何从大量的特征中选择最相关的特征，以提高模型的预测性能。

解决办法：

使用深度学习技术，如卷积神经网络（CNN）、循环神经网络（RNN）等，来提取lncRNA和疾病的相关特征。同时，可以使用特征选择算法，如基于稀疏模型的特征选择、相关系数分析等，来筛选最具信息量和相关性的特征。

难点3：

数据集不平衡问题：lncRNA-disease关联数据集通常存在类别不平衡的问题，即某些lncRNA-disease关联样本数量较少，而其他样本数量较多，可能导致模型对少数类别的预测性能较差。

解决办法：

可以使用采样技术，如欠采样、过采样等方法来平衡数据集中的类别分布。另外，可以使用类别加权的损失函数，如加权交叉熵损失函数，来给予少数类别更大的权重，以提高模型对少数类别的预测能力。

难点4：

模型优化和调参：深度学习模型通常有大量的参数需要进行优化和调参，如网络结构、学习率、批次大小等，而不同的参数组合可能对模型的性能产生不同的影响。

解决办法：

使用交叉验证等技术来评估不同参数组合下模型的性能，选择最优的参数组合。可以使用自动调参算法，如网格搜索、贝叶斯优化等，来加速参数搜索的过程。

难点5：

平台开发和性能优化：构建一个用户友好的lncRNA-disease关联预测平台，并优化平台的性能和响应速度，以提高用户体验。

解决办法：

采用合适的开发框架和技术，如Web开发框架，来开发一个用户友好的平台。同时，使用性能优化技术，如模型压缩、并行计算等，来提高平台的运行效率和响应速度。

**五、进度安排**

1、选题与需求分析阶段（2023.09.11-2023.10.22）

2023.09.11-2023.09.24：完成技术调查与选题。

2023.09.23-2023.10.22：完成《开题报告》、《项目需求说明书》。

2、系统设计阶段（2023.10.23-2023.12.17）

2023.10.23-2023.11.19：进行概要设计并撰写《概要设计说明书》。

2023.11.20-2023.12.17：进行详细设计并撰写《详细设计说明书》、《中期检查报告》。

3、系统实现阶段（2023.12.18-2024.03.31）

2023.12.18-2024.03.17：完成系统编码和单元测试。

2024.03.18-2024.03.31：完成系统的集成测试并修改bug。

1. 毕业设计答辩阶段（2024.04.01-2024.06.20）

2024.04.01-2024.05.31：撰写毕业设计论文。

2024.06.01-2024.06.14：修改毕业论文，制作答辩PPT。

2024.06.13-2024.06.20：毕业答辩。