

多病种关联关系挖掘工具软件 操作手册

1. 软件介绍

多病种关联关系挖掘工具软件是一款通过已知的疾病-症状关联关系来挖掘潜在的、未知的疾病之间、症状之间以及疾病-症状之间的关联关系的软件。用户可以根据需求上传已知的疾病-症状关联关系数据集，并选择合适的算法进行关联挖掘。最终挖掘得到的关联关系将以图表的形式进行展示，用户也可以在结果展示页面进行自定义查询。

由于多病种关联关系模块所需的数据不能直接从医院获取，该软件额外设置了一个疾病危险因素挖掘模块。用户可以通过此模块去挖掘单种疾病与危险因素之间的关联关系，并将挖掘得到的关联关系存入到多病种知识库中，从而用于进一步的多病种关联关系挖掘。

2. 数据集介绍

该软件目前使用的数据集来源于公开知识图谱 Disease-KB，该知识图谱的数据是从权威医药网站上爬取的。我们从中提取了疾病-症状关联关系，并进行了精简处理。其中，数据样式如表 1 所示。

表 1 关联表

关联 id	疾病 id	危险因素 id	疾病名称	危险因素名称
1	0	0	糖尿病	头痛
2	0	1	糖尿病	消化不良
3	1	1	心绞痛	消化不良
4	2	0	胃癌	头痛

上传多病种数据集时疾病字段和危险因素字段为必须字段，且命

名为 disease_name 和 factor_name，另外可添加权重字段。

除了多病种关联数据库，还设有一个多病种知识库，由疾病危险因素挖掘模块得到的结果构建。

3. 软件操作说明

本章介绍系统功能的操作，主要分为三类，任务管理、数据管理和系统管理，任务管理包含多病种关联挖掘、疾病危险因素挖掘及对应的任务管理；数据管理分别管理多病种关联挖掘和疾病危险因素的数据；系统管理包含用户管理、信息发布、数据管理、算法管理、日志查看、病种设置，具体操作流程如下。

3.1 登录与注册

系统分为管理员和普通用户，普通用户注册后需要管理员激活后才能使用，登录后进入各自的页面，管理员相比于普通用户多了系统管理功能用户登录和注册页面见图 3-1、图 3-2。

用户忘记密码可以通过密保问题找回。

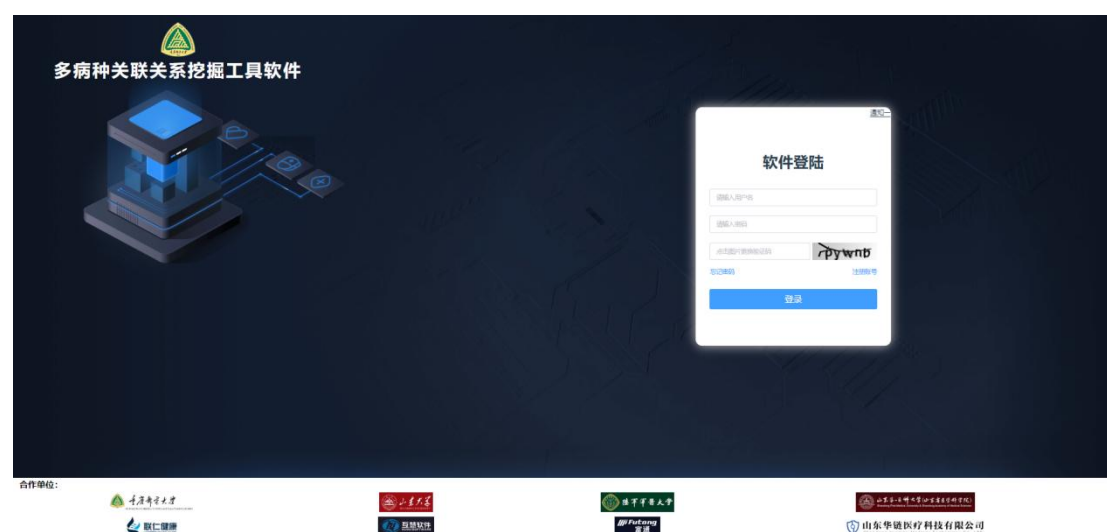


图 3-1

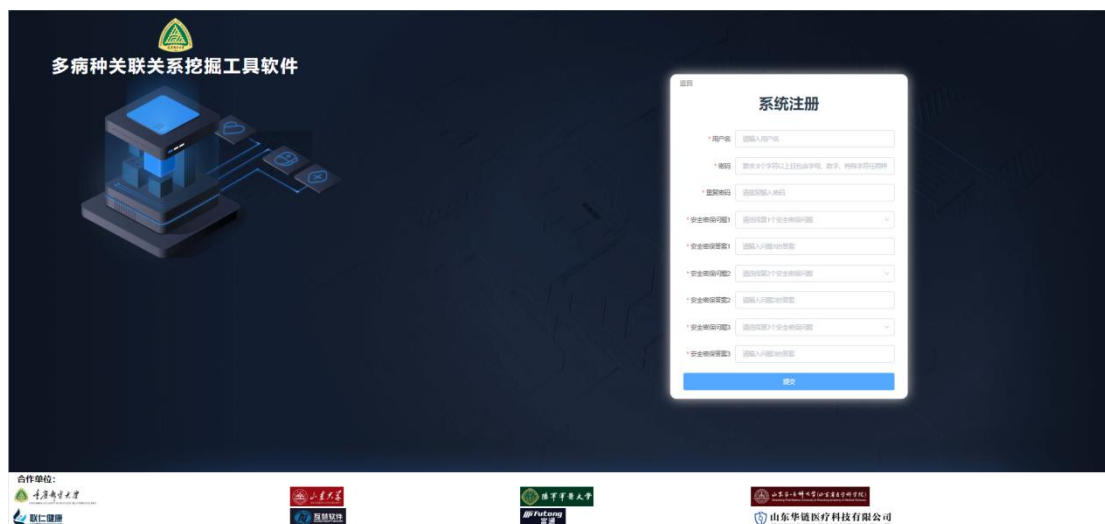


图 3-2

3.2 任务管理

3.2.1 多病种关联挖掘

多病种关联挖掘首先需要选择数据，上方一个步骤条显示所处了步骤并简要介绍每个步骤的任务，共分为四个步骤，基本信息填写、数据选择、病种选择和症状选择；下方显示每个步骤需要填写的参数。基本信息填写需要选择要使用的算法和任务类型，并填写任务名称、任务描述等信息，见图 3-3；数据选择可以预览选择的数据，此模块的数据由公开数据集和多病种知识库构成，见图 3-4；病种选择会自动将所选数据表中含有的疾病展示出来以供选择，由于病种数量较多，使用折叠面板按疾病种类进行收纳，支持疾病的搜索，见图 3-5；危险因素选择会显示选择的数据表中与选择的危险因素相关联的症状，危险因素同样按照症状分类进行收纳，支持危险因素的搜索，见图 3-6。

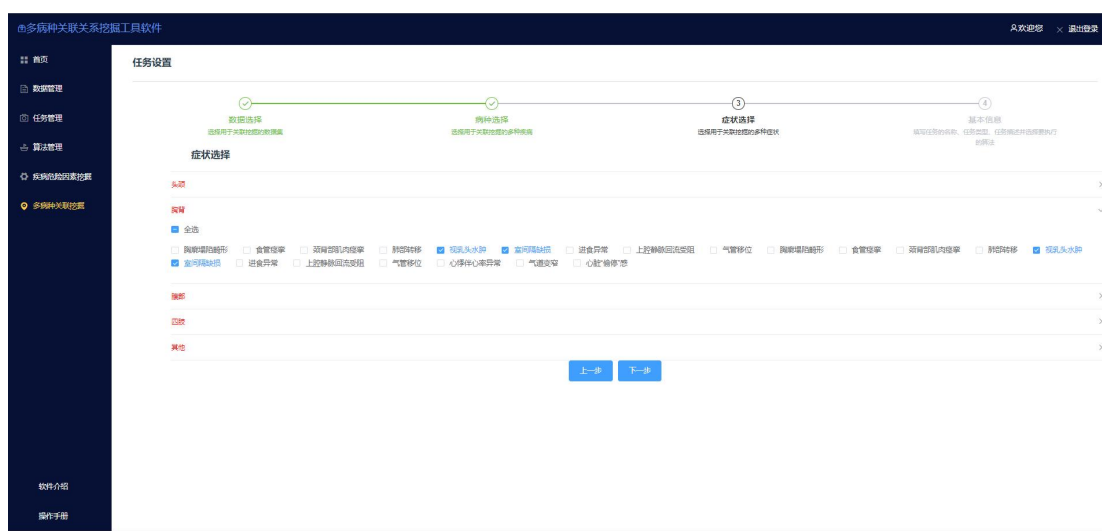


图 3-6 任务设置：症状选择

任务设置完成后提交任务即可根据任务的类型跳转到不同的展示界面，分别为疾病关联关系挖掘、症状关联关系挖掘和疾病-症状关联关系挖掘。

疾病关联关系挖掘展示界面上方为查询展示模块，左边多选框内选择要查询的疾病，多选框下面的表格显示了选择的每一种疾病和其他的疾病的关联情况，可以对每一行（即某一种疾病和其他疾病的关联关系）用关联图在边卡片内展示，中间的关联图展示了所选的集中疾病之间的关联关系。多选框中的疾病由下方的总体关联图中的疾病节点决定。下方是一个总体的疾病关联图谱，剔除了孤立节点，在图谱上方的疾病种类块可以对图谱按节点进行精简，也可以按权重大小对边进行精简，同时展示了图谱的统计信息，即病种数量和关联边数量。疾病关联挖掘的按搜索局部展示见图 3-7，总体展示见图 3-8。



图 3-7 疾病关联挖掘按搜索局部展示

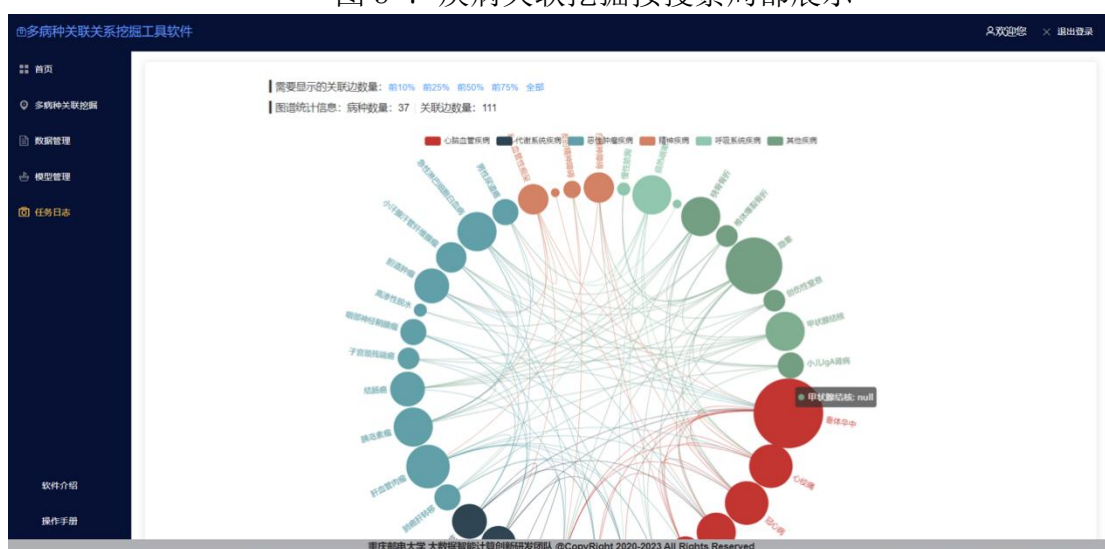


图 3-8 疾病关联挖掘总体展示

症状关联关系挖掘的布局与疾病关联关系挖掘基本一致，上方为按搜索局部展示，下方为总体展示，这里不过多介绍。

疾病-症状关联关系挖掘上方为多病种关联查询和多症状关联查询，前面的疾病关联挖掘和症状关联挖掘只能展示疾病-疾病之间，症状-症状之间的关联，此模块展示疾病-症状之间的关联。多病种关联查询仍是先选择要查询的疾病，下面用文本框展示所选疾病的共有症状，用表格展示所选择的每种疾病与症状的关联关系，中间的图为共有症状的图谱展示，右边的图为所选择的每种疾病与症状的关联关

系的图谱展示。多症状关联查询的展示同多疾病关联查询，显示症状的共有疾病和每种症状与疾病的关联情况。下方同样为疾病-症状关联关系的总体展示，按节点颜色对疾病和症状进行了区分，深色为症状节点，鲜艳色为疾病节点，可以按节点和关联边对图谱进行精简，精简后上方可以选择的疾病和症状也会相应减少。疾病-症状关联挖掘的按搜索局部展示见图 3-9，总体展示见图 3-10。

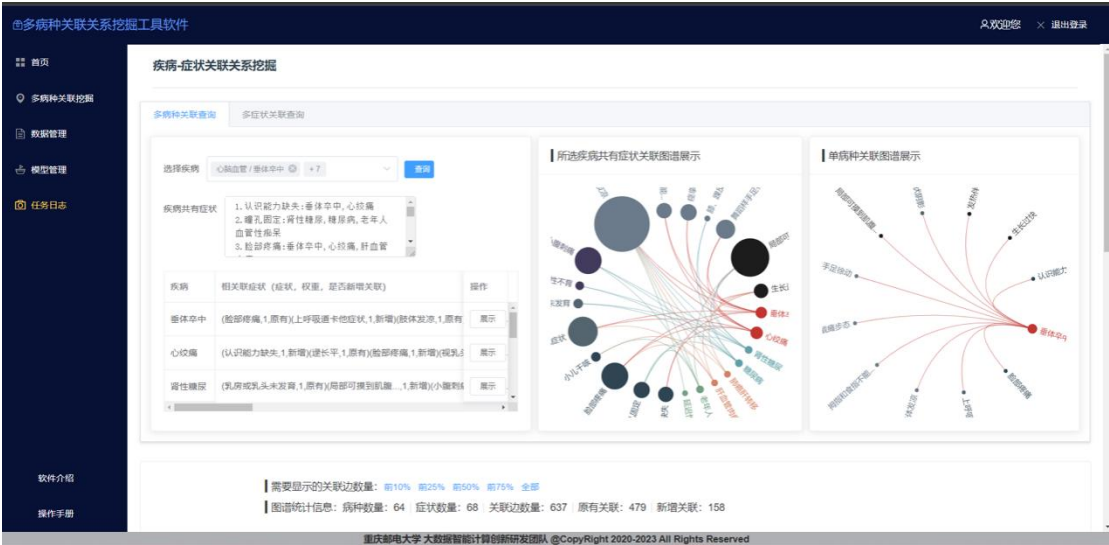


图 3-9 疾病-症状关联挖掘按搜索局部展示

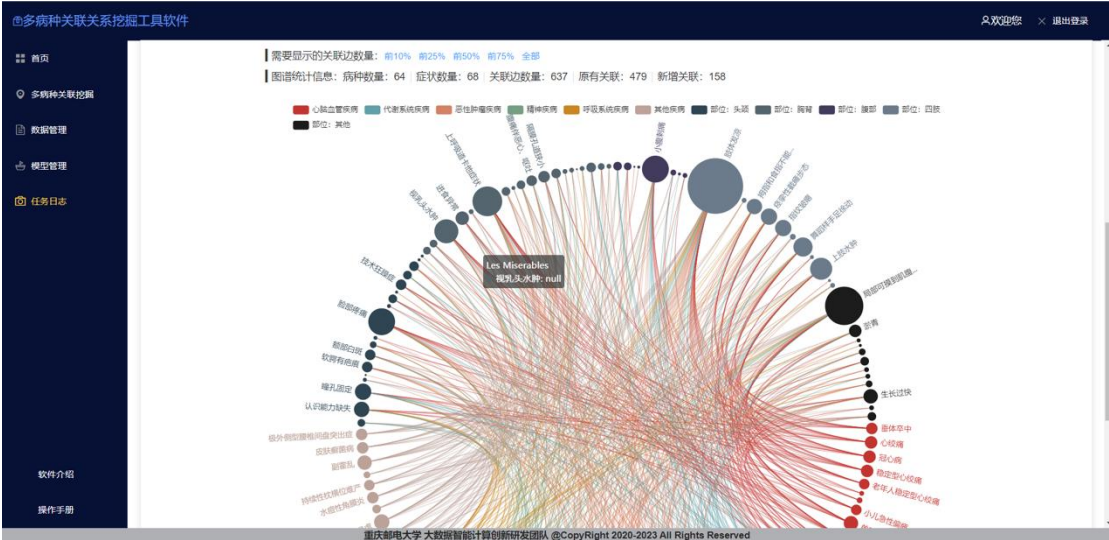


图 3-10 疾病-症状关联挖掘总体展示

3.2.2 疾病危险因素挖掘

多病种关联挖掘模块完成单种疾病的危险因素挖掘，主要由五个

步骤，任务信息填写、数据选择、特征选择、算法选择和运算结果。

首先填写任务信息，见图 3-11；选择挖掘需要使用的数据表，见图 3-12；特征选择阶段会罗列出所选择的数据表中所有的字段，并选择要挖掘的疾病和危险因素，见图 3-13；之后选择使用的算法并设置参数，见图 3-14；最后将算法挖掘出的结果进行展示，并可选择将结果加入到多病种知识库中，见图 3-15。

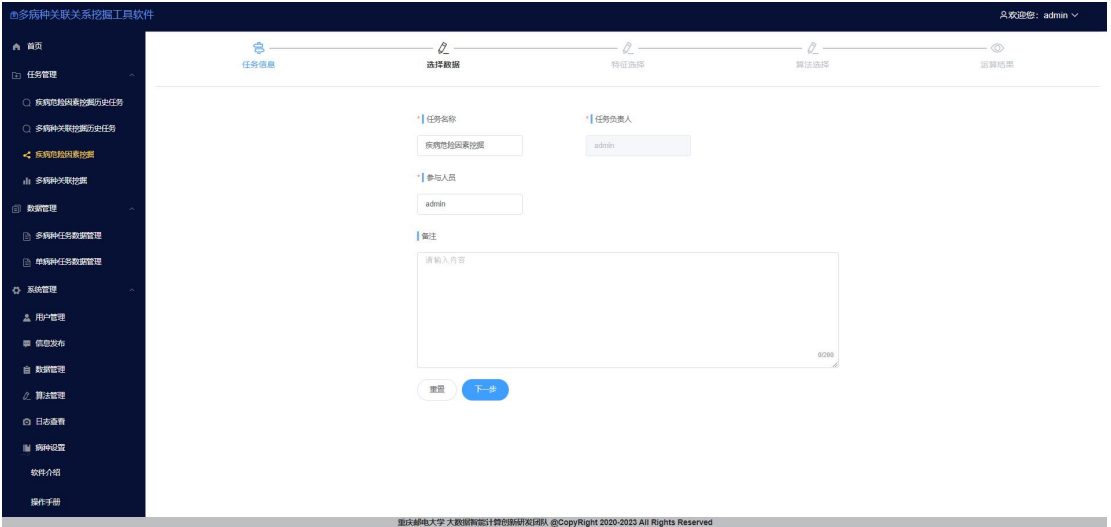


图 3-11



图 3-12 数据选择

3.2.3 多病种关联挖掘历史任务

多病种关联挖掘历史任务完成任务的记录、运行和删除操作，页面展示效果如图 3-16 所示。任务日志存储用户之前任务设置中的各个参数，用户的提交的每次任务自动添加到任务日志中。在上方用文本框完成查询操作，可以通过右侧的任务表、任务负责人、任务类型和任务名称进行查找，下方用卡片展示每次任务的参数信息，下方为运行和删除操作，删除操作会删除对应的任务日志，运行操作会按照之前选定的参数重新运行算法并跳转到结果展示界面，编辑操作会进入任务编辑页面，修改选中的任务的基本信息。

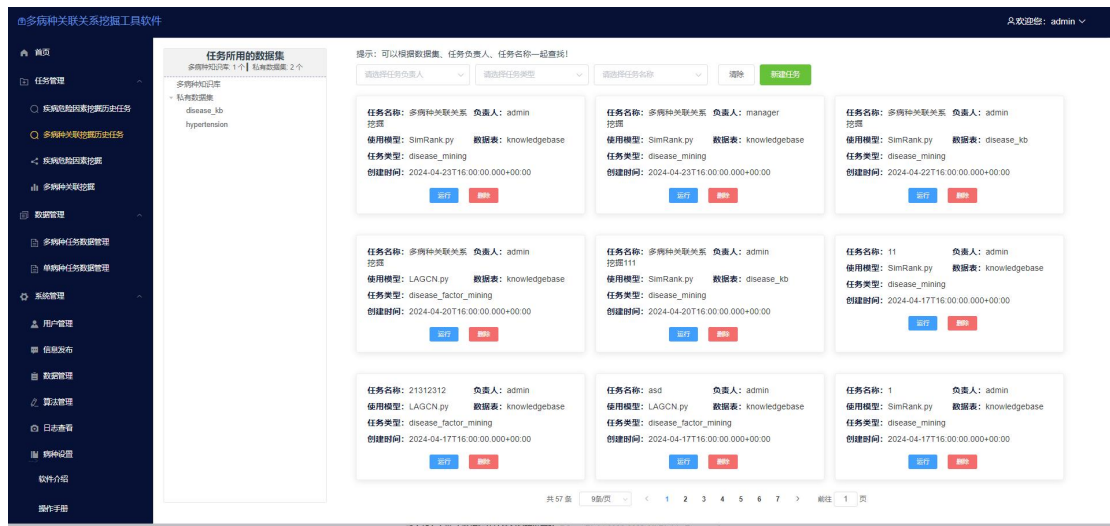


图 3-16 多病种关联挖掘历史任务

3.2.4 疾病危险因素挖掘历史任务

疾病危险因素挖掘历史任务完成任务的记录、查看和删除操作，页面展示效果如图 3-17 所示。任务存储了挖掘的结果，因此可以直接查看任务结果。支持按右侧任务表、任务负责人和任务名称查找。

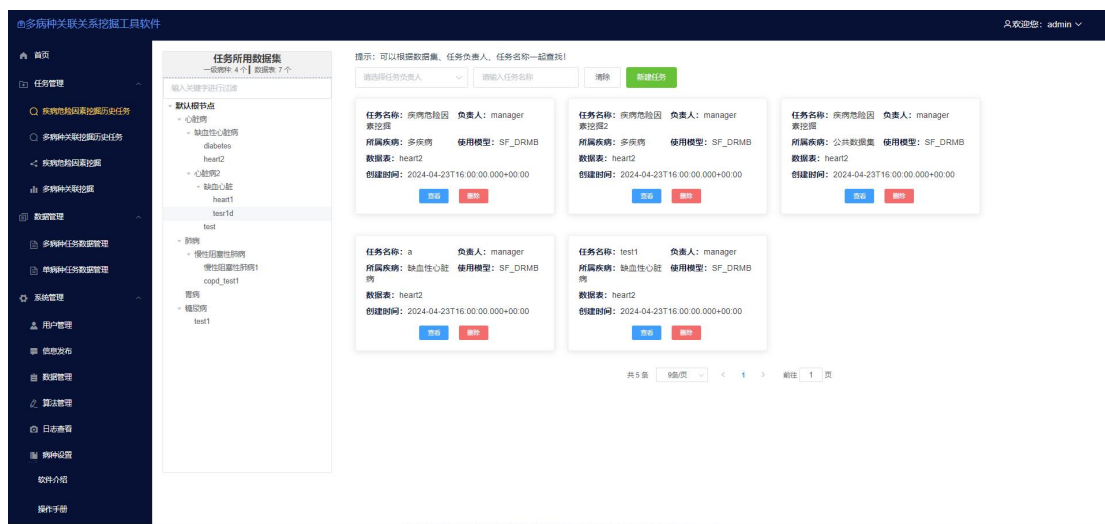


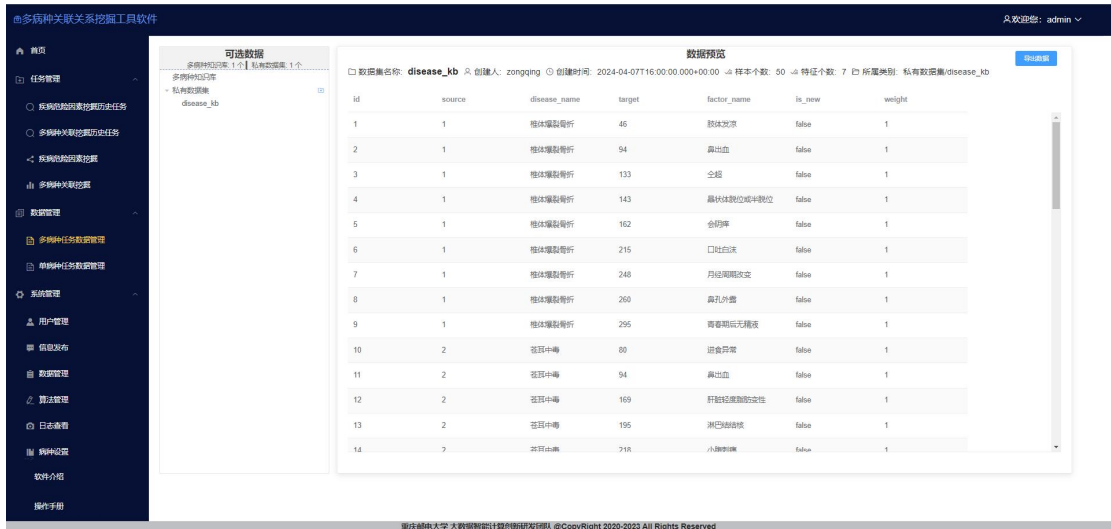
图 3-17 疾病危险因素挖掘历史任务

3.3 数据管理

数据分为两部分，多病种关联挖掘所需的数据表和疾病危险因素挖掘所需的数据表。

3.3.1 多病种任务数据管理

多病种任务数据管理完成多病种关联挖掘的数据查看、上传和导出，页面如图 3-18 所示。数据分为多病种知识库和公开数据集，多病种知识库由疾病危险因素挖掘中结果的插入得到，公开数据集为用户上传的数据集。导出数据时选择要导出的字段后以表格的形式导出，如图 3-19。



3-18 多病种任务数据管理



3-19 导出数据

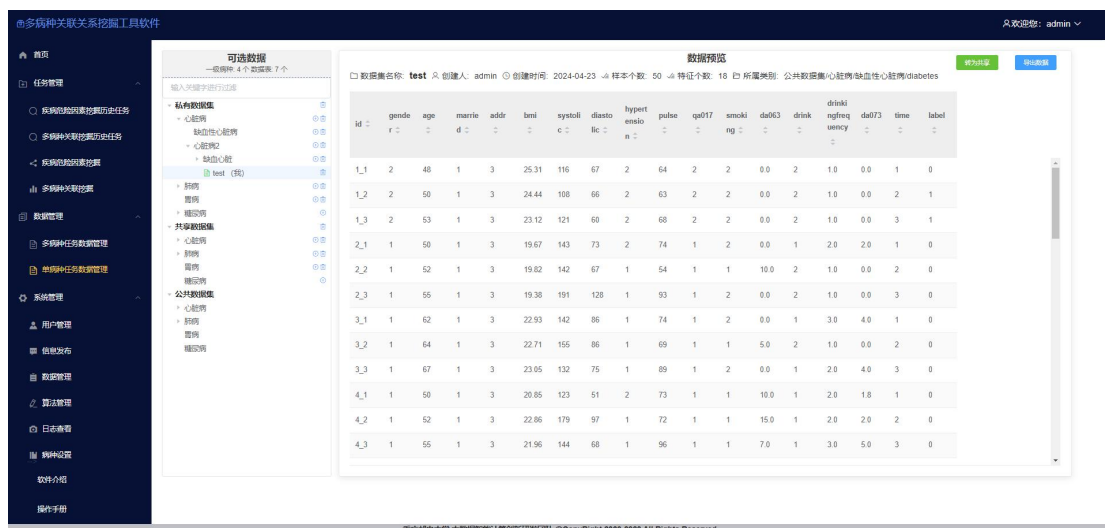
3.3.2 多病种任务数据管理

数据管理页面，页面效果如图 3-20、图 3-21、图 3-22 示。该页面主要有上传公开数据集与筛选专病数据集的功能。

3.3.2.1 上传公开数据集

页面效果如图 3-21 所示。用户可以在公开数据集处选择新建病种或上传数据集。新建病种代表数据集的分类，用户可以直接上传数据集或在对应的病种下进行上传。上传界面如图所示，用户输入相关数据集信息，点击确定，实现上传功能。

上传成功后可在左侧对应位置处查看所上传的文件。



页面效果如图 3-22 所示。当用户想从数据库中筛选相应的疾病进行研究时，可以在相应的疾病下面进行筛选（若无相应的病种，可先新建病种，新建病种如图所示），筛选页面如图所示，输入新增数据集的相关信息后，在下方选择相应的特征特征作为数据集的筛选条件，点击筛选病例，会展示根据筛选条件得出的数据（展示前 50 条），点击新建表，筛选出的数据将会被保存到数据库中。

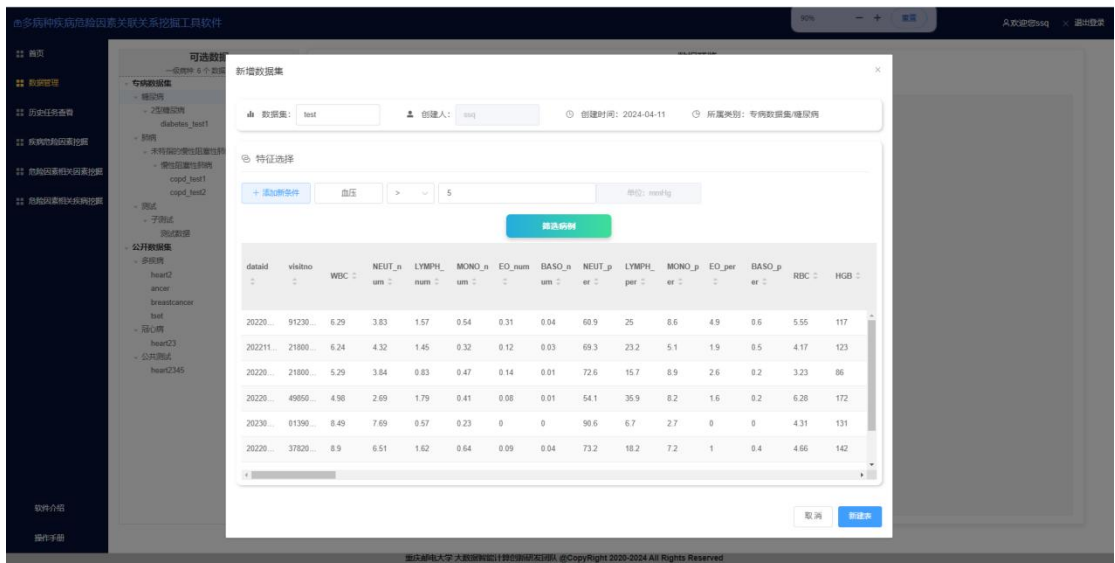


图 3-22 筛选页面

3.4 系统管理

3.4.1 用户管理

用户管理对所有的软件用户的信息进行管理，其中包含用户名称，用户权限以及用户可以用容量等信息。管理员可以对用户信息进行编辑修改删除，如图 3-23 所示。

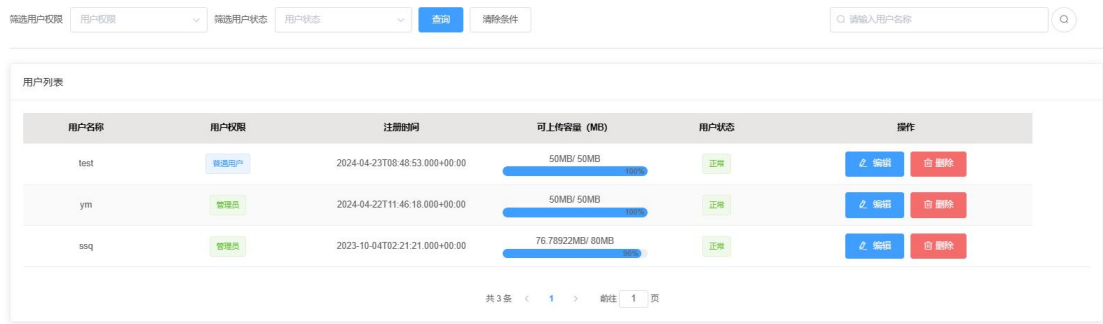


图 3-23 用户管理

3.4.2 信息发布

信息发布主要对登录首页中通知栏部分的信息进行管理。管理员可以对通知栏信息进行编辑修改删除，如图 3-23 所示。

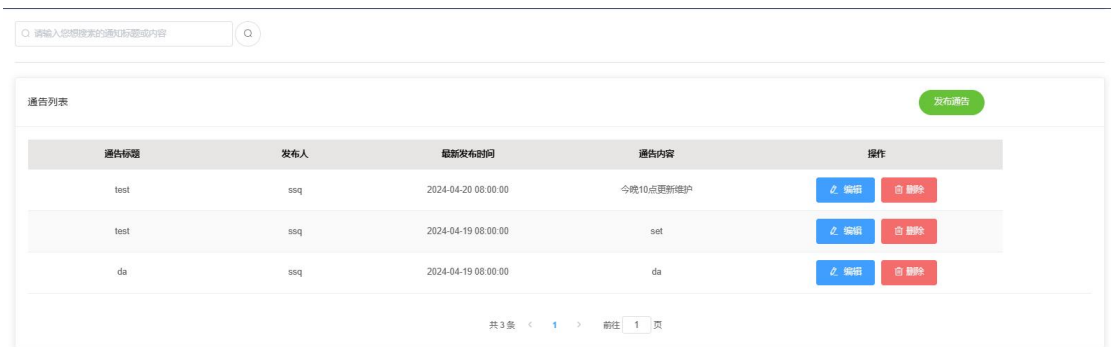


图 3-23 信息发布

3.4.3 系统数据管理

系统数据管理针对系统中所有数据集进行管理，管理员可以对数据集的状态进行修改删除，并管理员在系统数据管理处进行公共数据集上传，如图 3-24 所示。

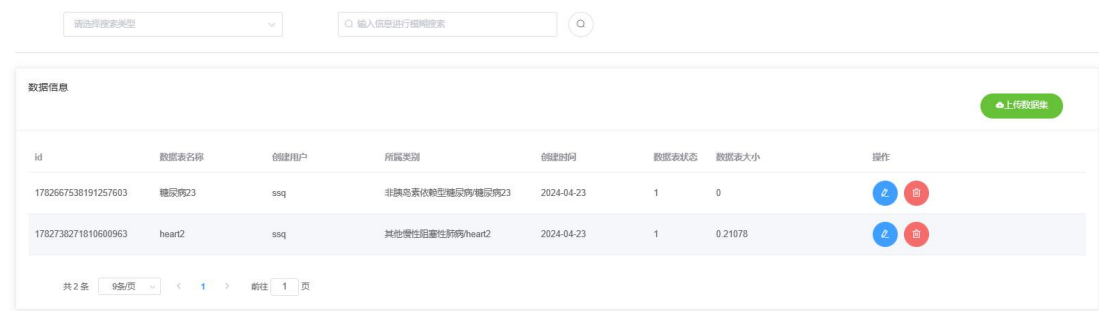


图 3-24 系统数据管理

3.4.4 日志查看

日志查看可供管理员用户查看所有用户的操作日志，如图 3-25 所示。

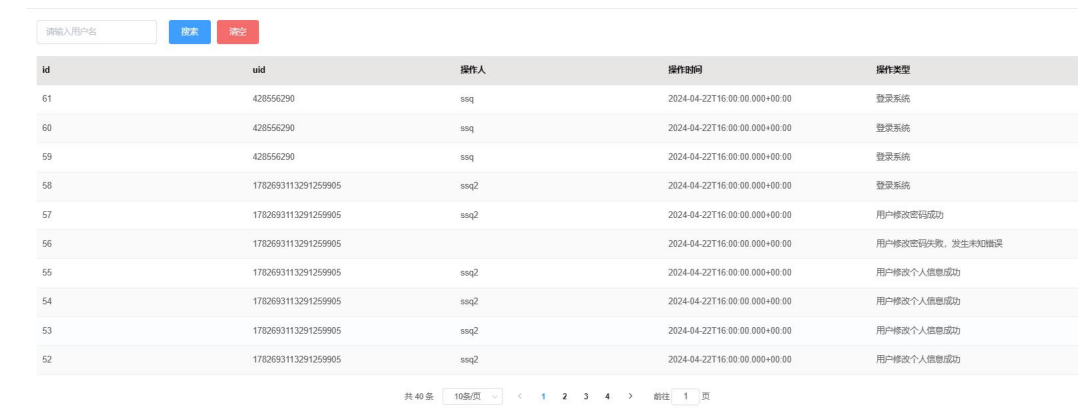


图 3-25 日志查看

3.4.5 算法管理前端界面

算法管理完成对算法的编辑、查询操作，页面展示效果如图 3-26 所示。进入此页面会显示后台所有的算法信息，上方用文本框完成查询操作，可以通过算法的 id 或者算法名进行查找，下方用表格完成查询结果的展示。每行最右一列为编辑操作，编辑操作会进入编辑页面，修改选中的算法的基本信息。

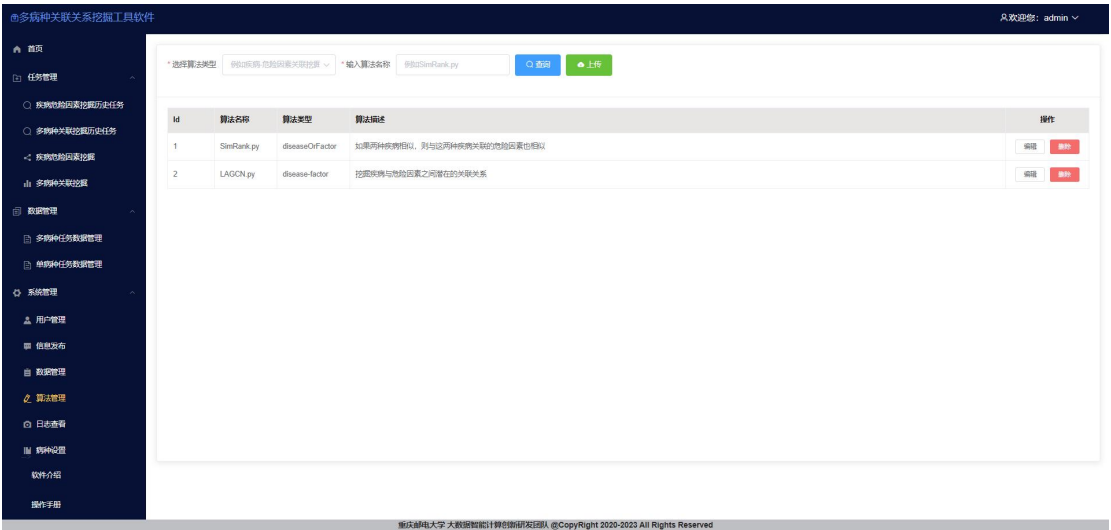


图 3-26 算法管理

3.4.6 病种设置

病种设置可供管理员对于软件所有涉及的病种进行设置，病种设置严格依照 ICD10 编码进行设置，如图 3-27 所示。

疾病名称:

疾病列表 添加 - 新增病种

注意: 删除上级病种时会同时删除其下所有病种!

病种名称	icd编码	集数(公共数据集)	集数(私有数据集)	本数据集(数据集)	创建者	操作
糖尿病	1	0	0	0	admin	<input type="button" value="修"/> <input type="button" value="删"/> <input type="button" value="加"/>
↳ 心脏病	1	0	2	2	admin	<input type="button" value="修"/> <input type="button" value="删"/> <input type="button" value="加"/>
心脏病I	11	0	1	0	admin	<input type="button" value="修"/> <input type="button" value="删"/> <input type="button" value="加"/>
↳ 缺血性心脏病	1	0	1	2	admin	<input type="button" value="修"/> <input type="button" value="删"/> <input type="button" value="加"/>
缺血性心脏病	1	0	0	0	manager	<input type="button" value="修"/> <input type="button" value="删"/> <input type="button" value="加"/>
↳ 肝病	1	1	0	0	admin	<input type="button" value="修"/> <input type="button" value="删"/> <input type="button" value="加"/>
慢性阻塞性肝病	1	1	0	0	admin	<input type="button" value="修"/> <input type="button" value="删"/> <input type="button" value="加"/>
肝病	1	0	0	0	admin	<input type="button" value="修"/> <input type="button" value="删"/> <input type="button" value="加"/>

图 3-27 病种设置