

上海大学

SHANGHAI UNIVERSITY

毕业论文（设计）

UNDERGRADUATE THESIS (PROJECT)

题目：基于机器学习的子宫肉瘤患者的临床病理  
及预后分析

学院 计算机工程与科学学院

专业 计算机科学与技术

学号 19120171

学生姓名 黄奕恺

指导教师 沈文枫

起讫日期 2023.02.21– 2023.06.03

# 目 录

摘 要 .....	III
ABSTRACT .....	IV
第1章 绪论 .....	1
§1.1 研究背景与意义 .....	1
§1.2 子宫肉瘤研究发展现状 .....	1
§1.2.1 在子宫肉瘤方面使用的图像模型 .....	1
§1.2.2 在子宫肉瘤方面使用的数据模型 .....	2
§1.3 本文的研究内容及目标 .....	2
§1.3.1 研究内容 .....	2
§1.3.2 研究目标 .....	2
§1.4 本文组织结构 .....	2
第2章 子宫肉瘤预后模型相关技术综述 .....	4
§2.1 环境 .....	4
§2.2 技术栈 .....	4
§2.2.1 Docker .....	4
§2.2.2 Vue.js .....	5
§2.2.3 Flask .....	6
§2.3 生存分析方法 .....	6
§2.3.1 KM曲线 .....	7
§2.3.2 Cox单/多因素分析 .....	7
§2.4 其他预后模型 .....	7
§2.4.1 决策树 .....	7
§2.4.2 随机森林 .....	7
§2.4.3 聚类+分类 .....	8
第3章 子宫肉瘤预后模型的数据提取 .....	9
§3.1 数据源介绍 .....	9
§3.1.1 SEER癌症数据库 .....	9
§3.1.2 医院验证数据库 .....	9
§3.2 数据字段对应编码与含义 .....	9

第4章 参考文献 .....	10
致 谢 .....	11
参考文献 .....	12
附录 A 经典不等式 .....	13

# 基于机器学习的子宫肉瘤患者的临床病理及预后分析

## 摘 要

子宫肉瘤是源于子宫体部位的一组独立的高度恶性肿瘤，根据原发部位又分为子宫平滑肌肉瘤、子宫内膜间质肉瘤等。该肿瘤属于较为罕见的肿瘤，通常发病于45岁以上的绝经后妇女，虽然已经有一些相关文献对其进行了科普与介绍，关于其临床病理及预后分析的研究依然较少。本文基于SEER癌症数据集使用KM曲线与COX多因素比例风险回归模型对影响其患者生存率的各个因素进行了分析与总结；并根据其预后相关数据进行了多种聚类算法的使用，分析聚类后得到簇所代表的表型具有的特征，从而分析得到子宫肉瘤患者的不同表型特点。由于聚类算法属于无监督学习，难以运用在新的数据上。所以本文建立了相关预后模型，使用决策树与随机森林对SEER上的聚类得到的表型进行训练，从而得到能区分患者所对应表型的预后模型，从而更好地评估患者的生存情况。最后结合医院提供的数据对上述模型进行了验证，总结全文工作与创新点，并展望后续工作。

**关键词：**机器学习, 比例风险回归模型, 决策树, 随机森林, 层次聚类

# Machine learning-based clinicopathological and prognostic analysis of patients with uterine sarcoma

## ABSTRACT

Uterine sarcoma is an independent group of highly malignant tumors originating from the body of the uterus, which are classified into smooth muscle sarcoma and endometrial mesenchymal sarcoma according to the site of origin. It is a rare tumor that usually develops in postmenopausal women over 45 years of age. Although it has been popularized and introduced in the literature, there are still few studies on its clinicopathological and prognostic analysis. In this paper, we analyzed and summarized the factors affecting the survival rate of patients based on the SEER cancer dataset using the KM curve and COX multifactor proportional risk regression model; and used various clustering algorithms to analyze the characteristics of the phenotypes represented by clusters after clustering to analyze the different phenotypic characteristics of patients with uterine sarcoma based on their prognosis-related data. Since clustering algorithms are unsupervised learning, they are difficult to apply on new data. Therefore, in this paper, a relevant prognostic model is developed and the phenotypes obtained from clustering on SEER are trained using decision trees and random forests to obtain a prognostic model that can distinguish the phenotypes corresponding to the patients and thus better assess the survival of the patients. Finally, the above models were validated with the data provided by the hospital, summarizing the full work and innovations and looking forward to the follow-up work.

**Keywords:** Machine learning, proportional risk regression models, decision trees, random forests, hierarchical clustering

## 第1章 绪论

本章主要介绍了子宫肉瘤研究的相关研究的研究背景及意义，分析了相关课题的研究方法与现状，最后列举了本文的研究目标内容与结构。

### §1.1 研究背景与意义

子宫肉瘤是一种罕见的子宫恶性肿瘤，其相关死亡率在子宫恶性肿瘤中的比例在16%以上<sup>[1]</sup>，目前的主要治疗方式是以手术为主，手术后一般会通过化疗（少数情况使用放疗）辅助，以确保最后效果。根据SEER数据分析显示，子宫肉瘤患者的中位年龄在65岁左右，发病原因有很多，根据对文献中的数据显示，子宫肉瘤可能与某些遗传基因的突变有关。而子宫肉瘤的复发情况较为常见。子宫肉瘤的复发是指在手术和治疗完成后，术后阶段肿瘤重新生长并再次出现的一系列过程。子宫肉瘤的复发并不局限于原发病灶，也可能出现在周边组织内，包括子宫、输卵管、卵巢等等。按照文献显示，I II期患者5年生存率为59%，而III期则为22%，IV期为9%。<sup>[1]</sup>从数据中可以看出，子宫肉瘤的复发率较高，预后效果较差。在预防子宫肉瘤的复发中，一些研究表明了术后辅助治疗和放疗能够达到降低复发率的作用，并提高患者的生存率。然而作为较为少见的恶性肿瘤，目前子宫肉瘤方面尚未有比较公认的预后处理模型出现。为了处理近年各个医院积累的子宫肉瘤的随访与预后数据，利用数据对患者术后复发状况或生存率进行预测与评估，能够更好地帮助患者了解自身身体状况，也能帮助医生对于危险程度较高的患者进行重点关注，从而能够让医疗资源能够更加有效地被利用。因此，本文旨在建立一套使用患者预后数据训练而成的预后模型，并建立相应的UI界面，让医学人士能够方便地管理并使用该系统对于患者的将来的生存状况进行评估。

### §1.2 子宫肉瘤研究发展现状

目前，子宫肉瘤的研究尚处于开展阶段，使用的模型主要分为以下两类：图像模型、数据模型。

#### §1.2.1 在子宫肉瘤方面使用的图像模型

图像模型主要用于子宫肉瘤的诊断与治疗方案的提出。其中比较常用的图像源是计算机断层扫描（CT）与核磁共振成像（MRI）。图像模型主要通过人工智能技术对于图像进行分隔，提取出ROI，并用于分类模型。分类模型是用于分类和预测的模

型，目前常用于子宫肉瘤的治疗决策中。其中，Transformer与生成对抗网络（GAN）是目前比较常用的模型。Transformer可以实现多模态的医学图像分类。以往使用的深度神经网络由于是基于卷积架构形成，它在图像像素较为清晰，或内容较为复杂时，难以对于图像中结构的远端依赖性有较为明确的认知，对于复杂情况分类效果并不好。然而在使用了自注意力机制的Transformer后，它对远端结构的编码让它拥有了更强的学习表达能力，从而有了更好的分类效果。而生成对抗网络则可以生成与真实数据相似的模拟图像，通过模拟图像与真实图像的对比，生成模型与判别模型的不断迭代提升。GAN生成的图像可以为有限的图像数据添加标注后的新数据，同时其附带的判别模型也可以用于对于医生训练数据的扩充。

### §1.2.2 在子宫肉瘤方面使用的数据模型

最常用的分类模型是支持向量机（SVM）和随机森林（Random Forest）。这些模型可以用于预测子宫肉瘤的侵袭能力和转移风险，从而为临床医生提供重要的治疗决策参考。

## §1.3 本文的研究内容及目标

### §1.3.1 研究内容

本文旨在设计并使用Vue与Flask作为前后端技术栈构建一个简易的患者病情预测平台，调用Python实现的预后模型。从而可以帮助医生预判患者病情，使得医生倾注医疗资源来为高风险患者进行进一步的病情随访，推动医学诊断的数字化，让医生能在这个更便携的平台上开展一系列工作，同时也为医疗服务的集成提供了一条较为有效的工程实践经验。

### §1.3.2 研究目标

对于本文的研究内容，我制定了以下几条目标：

- 1) 从SEER数据库获得数据，验证各类模型的前置条件是否满足，测试各类模型方法，评估各类方法的作用、优缺点与效果。
- 2) 实现一个具有登录、授权功能的前后端系统，添加工单功能，同时引入数据导出与导出功能，在后端中集成Python实现的预后模型方法，并提供在线的训练与预测功能，从而能够满足医生与管理员的共同使用。

## §1.4 本文组织结构

整篇论文一共分为七章。

第一张介绍了子宫肉瘤相关研究的背景与意义，阐述了当前子宫肉瘤相关研究的内容与方向，并说明了本文的研究目标与具体内容。

第二章主要介绍了本文使用的代码环境与技术，描述了相关的曲线或参数的具体含义，并阐释了为何使用这些技术与指标。

第三章是本文筛选并处理数据集的过程，描述了本文中如何从SEER癌症数据集与医院数据集中分别下载并处理数据，从而能在下面的章节中分析并使用。

第四章分析了数据集中的内容，利用数据进行了生存分析，使用KM曲线与COX比例风险模型研究各影响因素对生存的影响，并进而验证数据的有效性。

第五章首先使用无监督聚类模型对患者的表型进行分类，并分析了各表型所具有的不同特征与预后效果的不同。用分类得到的患者表型数据作为监督学习的数据源，使用决策树与随机森林模型以判断患者所处的表型，并与使用三年生存率的传统方法进行了对比，评估了两者的效果。

第六章介绍了根据模型实现的项目主要功能及实现过程中的技术细节。

第七章对全文进行了总结，归纳了本文的创新点与具体内容，并指出了本文使用的模型的局限性与改进方向。



## 第2章 子宫肉瘤预后模型相关技术综述

本文的子宫肉瘤预后模型使用Python搭建模型，力求模型能够具有更好的可移植性，并依据这一点构建了相关系统。系统主要使用Vue.js作为核心框架，并配合以TypeScript和Sass作为技术补充；服务端考虑到模型算法基于Python，使用轻量级并高度可定制的Flask实现，并使用Docker容器技术让后端的部署更加稳定、便携。

同时，本文亦使用了一些医学方面的常用模型与分析方法，本章主要对使用到的相关技术、生存分析方法与预后模型进行介绍。

### §2.1 环境

本文使用WSL2(Windows Subsystem Linux)Arch发行版进行开发，Python版本为3.10.9，Vue.js版本为5.0.8，主要依赖为Element-plus。WSL2是基于Windows开发的Linux子系统，使用子系统可以在性能不受限的情况下使用Windows的大部分计算与存储资源，且传统的虚拟机与双启动系统的开销也不存在。相较于直接使用Windows进行开发，使用WSL进行开发能够使用apt、pacman等包管理器，且Arch还具有详细的官方WIKI与社区支持；而比较使用Linux开发，WSL的Remote连接更加稳定，能够更好地使用Windows独占的部分软件，例如本文使用的SEERStat数据库官方软件，让开发更加便利。

### §2.2 技术栈

#### §2.2.1 Docker

Docker是一个现今广泛使用的用于开发，运输和运行容器化应用程序的开放平台。使用Docker能够让我们专注于应用程序，而不是花费大量时间调试基础架构。Docker相比较于之前的虚拟机服务，Docker不需要占用多余的磁盘IO，能够有效减少计算资源的消耗。同时，Docker自包含程序依赖，这意味着Python项目的使用中也可以像js前端框架中一样，使用类似package.json的Dockerfile记录所需要的依赖，不同点是Dockerfile不直接写明依赖，而是用requirements.txt等文件来进行存储。这些优势意味着Docker无论在开发还是后期维护中都提供着很强的便携性与健壮性。Docker的核心概念是镜像、容器与仓库。其中镜像不难理解，类似于Linux中的镜像。我们一般使用一个基础镜像，譬如Win10、Linux等，这些基础镜像中包含着能够运行容器的最低限度的底层环境，而我们则基于这个基础镜像编写相关的配置，

比如导入依赖等，然后将这些配置逐层地添加到镜像中，使用Union FS技术对其进行分层与合层记录。镜像的层化技术能让我们具体地分出环境的各层结构，并依据其共享来减少重复镜像的拉取，从而最大化资源的利用率。而容器则包含程序运行需要的一切环境，轻量化地提供可共享可复制地一致服务，容器层的一切修改都不会作用于底层环境，而容器销毁时随着其生命周期的结束所有更改也会消失，从而提供了copy on write的安全特性。仓库类似Github中的仓库，可以使用Docker命令拉取。本文使用Docker搭配代码部署平台后，程序可以在上传到Github仓库之后自动部署程序于服务器上，从而实现方便便携的开发。

### §2.2.2 Vue.js

Vue.js（简称Vue）是一个用于构建用户界面的JavaScript框架。它建立于HTML、CSS、JavaScript之上，提供声明式和基于组件的编程模型。声明式表明Vue可以使用模板语法动态渲染HTML，让我们可以基于JavaScript中的各个状态动态地描述输出的HTML文本；而组件化意味着Vue的各个模块是可以高度可重用的，Vue SFC将HTML、CSS、JavaScript组合在一起并封装在一个文件中。Vue使用渐进式框架，可以根据需求使用其支持的特性，包括：去构建化的HTML增强、可以在任何页面上嵌入的组件、单页应用程序（SPA）、全栈服务器端渲染（SSR）、静态站点生成（SSG）、可以面向多种应用（包括桌面、移动、WebGL与终端）。

#### §2.2.2.1 Vite

Vite与Vue-Cli类似，是一个提供项目脚手架与开发服务器的构建工具。不过区别在于Vite并不是构建在Webpack上的，而是使用浏览器中的ES模块，这让Vite项目提供了很低的延迟和很高的速度，在大型项目中，Vite拥有着远超Vue-Cli的构建与启动速度。在日常使用中，随着项目的不断增大，基于Webpack的Vue-Cli构建速度一般在20秒左右，而相同体量的Vite项目往往恒定在1秒以下，这在需要经常修改的前端项目中会提供很大的便利。这是由于Vite不绑定服务端而是使用浏览器的原生支持。

但是这样的设计也会带来一些问题，Vite的开发环境需要基于现代浏览器，也就是说至少要支持ES2015，在版本较老甚至只使用CommonJS的浏览器中兼容不全，可能会带来一定问题，不过在如今的生产环境中这个问题很少见；同时它暂时也不支持Vue2；脚手架功能中相较于Vue-Cli有一定删节；最后就是开发与构建工具不同可能会导致一些程序页面构建后与开发服务器上的内容不一致。

### §2.2.2.2 Element-Plus

Element-Plus是一套基于Vue实现的常用组件库，它提供了较为丰富的PC组件。使用这些封装好的组件能够一定程度上减少开发者自己对于常用组件的再实现与重封装，从而提升开发的速度与稳定性。

Element-Plus的设计原则一共有四条：

- 一致 Consistency: 这表示不但组件的流程与逻辑与生活中使用的一致，而且所有元素和结构亦保证有一致的风格与逻辑
- 反馈 Feedback: 表示用户操作与页面状态都可以让用户感知到不同，从而让用户对网页状态有清晰的认知
- 效率 Efficiency: 说明组件的操作流程直观清晰，用户可以快速而直接地认知到各个结构的用途而不是花额外的时间回忆
- 可控 Controllability: 所有操作都交由用户自身来决策，而且用户可以对已完成或正在进行的操作撤销或终止

Element-Plus受到ES2018以及以上的浏览器支持，如果需要支持旧版本的浏览器，则需要用到Babel或者其他工具进行版本控制。

### §2.2.3 Flask

Flask是由Python实现的轻量化Web框架，它的实现是先基于底层的HTTP与Web服务器功能的封装，再使用WSGI(Python Web Server Gateway Interface)来建立Web应用。这样的结构让Flask在传递请求前，需要先将HTTP报文转换为WSGI所需的字典、响应头部的结构体，前者包含请求的全部信息，而后者则是将要调用的函数。而请求信息则由一个显示应用对象处理，在中小型的Flask后端中，大部分的接口都可以定义在这一个对象中。使用显示对象的原因是在Python中，隐式对象只能包含一个实例，所以使用显示对象能够让应用程序集中在一个文件中。这样的设计让小型的Flask服务不会过于臃肿，而且使用者可以使用寥寥几十行代码构造服务，整体结构也相当清晰。而且相当一部分人工智能模型都在Python中有较为简单与便携的实现方式，使用Python来实现后端可以更为简单地嵌入这些模型，让前后端使用的语言数量减少，从而让系统更加健壮。

## §2.3 生存分析方法

生存分析是一系列统计分析方法，用于探讨人在特定情况下的，及生存时间的分布由于时间或其他因素的变化趋势。但是生存分析并不仅仅可以用在医学领域，它还可以在商业等多种环境中使用。比如使用生存分析可以探讨会员、订阅等机制

的用户使用情况，并让厂商对于如何留住自身的客户有一定作用。其中生存时间并非单纯表示对象的存活时间，在医学数据中，由于数据需要得到用户的允许才能使用，往往数据的获取都是通过随访得到的。这意味着对象可能生存了更长的时间，而我们的数据只能确认对象在一段时间的存活，这种现象被称为数据的右删失。而这需要通过模型的修改或者后期调整来去除。

### §2.3.1 KM曲线

Kaplan-Meier曲线是用来进行时向统计的良好方法，它可以用来评估患者群体的健康状况和治疗效果。在生存分析中，我们可以将用户群体按照特征的不同，比如年龄段、肿瘤大小等，分为多个群体，通过比较群体间生存率的区别，我们可以对特征对于患者的影响有较好的初步结论。同时，KM曲线可以用来判断数据是否能使用COX比例风险模型。

### §2.3.2 Cox单/多因素分析

Cox单因素分析一般用来研究单变量中的各个值的关系，它通过类似多项式模型的方法来分析该自变量对因变量的影响，而不纳入其他变量的影响。它主要用来研究该因素对于死亡风险的影响程度。Cox模型定义了风险函数，用来表示一个实体具有相应自变量值时，因变量发生的概率。而Cox多因素得到的p值用来检验该变量对于因变量的影响是否显著，一般p值在0.05以下则表示影响显著，可以纳入。Cox多因素分析则与单因素相仿，研究多个变量对于结果变量的影响程度。

## §2.4 其他预后模型

### §2.4.1 决策树

决策树是一种在医学中常用的分类回归模型，它使用树形结构表示决策过程。树上的每一个节点都表示特定范围的特征值，而树的每一个叶子节点都对应一个特定的类。决策树的建立过程中，模型首先把全体数据分为大小相近的若干个子集，然后不断在这些子集中选择最佳的子集，让这些子集具有尽量多的相似特征，直到回归达到最大深度停止。决策树可以同时分类与回归问题使用。

### §2.4.2 随机森林

随机森林结合了决策树与随机梯度下降算法，随机地从数据集中抽样，组合成多颗决策树。具体地说，模型为每棵决策树划分了它使用的数据样本空间，并使用梯度下降算法让它在样本空间上训练，随后得到一系列决策树组成的森林。由于

随机森林分为多颗树训练，在样本数较大的大规模数据集中训练速度快于单棵决策树，同时也可以减小样本的方差。随机森林在NLP、医学诊断学、金融分析学中都有较为广泛的应用。

### §2.4.3 聚类+分类

先使用聚类方法得到带标签的新数据，分析聚类得到的患者表型对应集群的生存表现，并根据这个无监督学习的标签分类来预测患者生存率是本文根据文献<sup>[2]</sup>使用的新方法。在这个模型中，聚类方法的引入让模型的分类更有可解释性，同时不同患者的表型数据虽然在生存率方面会有不同，但是相近的特征依然存在，从而让模型对于相近的数据会有更相似的结果，连续性更好。相比单纯使用二分类模型的概率，该模型也可以使用聚类统计得到的概率作为参考，整体上得到的分类结果更加细致。

## 第3章 子宫肉瘤预后模型的数据提取

### §3.1 数据源介绍

数据提取过程如图3.1所示：

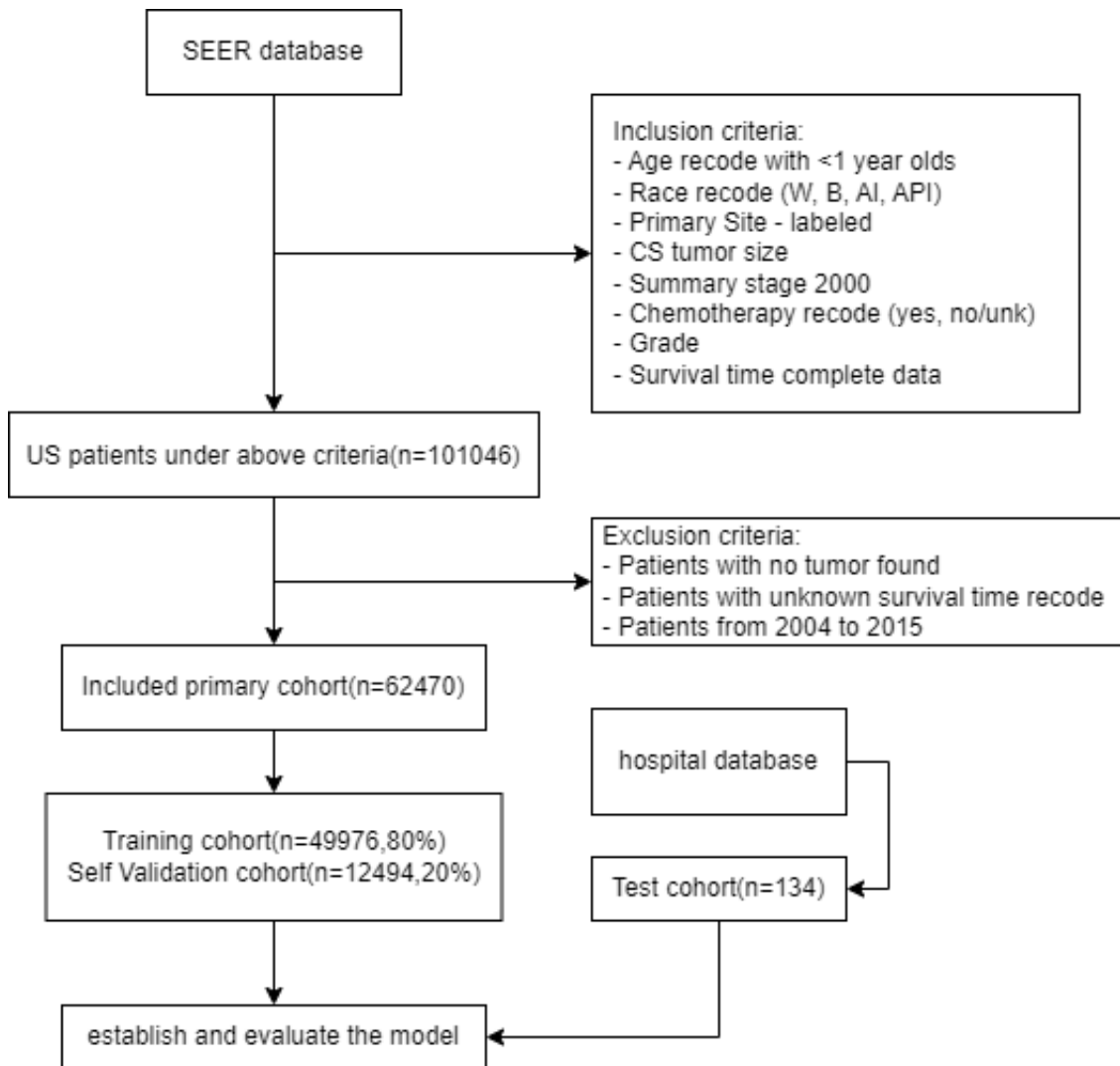


图 3.1 数据提取流程图

由图可见，在SEER数据库中，首先提取

#### §3.1.1 SEER癌症数据库

#### §3.1.2 医院验证数据库

### §3.2 数据字段对应编码与含义

## 第4章 参考文献

参考文献可以直接写在 thebibliography 环境里, 利用 `\bibitem` 罗列文献条目. 虽然费点功夫, 但是好控制, 各种格式可以自己随意改写.

本模板推荐使用 **BIB<sub>T</sub>E<sub>X</sub>**, 样式文件为 `shuthesis.bst`, 基本符合学校的参考文献格式. 看看这个例子: 关于书的<sup>[3,4]</sup>, 还有这些<sup>[5-12]</sup>.

有时候一些参考文献没有纸质出处, 需要标注 URL. 缺省情况下, URL 不会在连字符处断行, 这可能使得用连字符代替空格的网址分行很难看. 如果需要, 可以将模板类文件中

```
\RequirePackage{hyperref}
```

一行改为:

```
\PassOptionsToPackage{hyphens}{url}  
\RequirePackage{hyperref}
```

使得连字符处可以断行. 更多设置可以参考 `url` 宏包文档.

## 致 谢

衷心感谢导师沈文枫教授对本人的精心指导。感谢上海大学开源社区提供的Latex模板。



## 参考文献

- [1] 中国医师协会微无创医学专业委员会妇科肿瘤专业委员会(学组)中国优生科学协会生殖道疾病诊治分会, 中国优生科学协会肿瘤生殖学分会. 子宫癌肉瘤诊治中国专家共识(2020年版) [J]. 中国癌症防治杂志, 2020, 12 (6): 599–605.
- [2] Zhou X, Nakamura K, Sahara N, et al. Exploring and Identifying Prognostic Phenotypes of Patients with Heart Failure Guided by Explainable Machine Learning [J/OL]. Life, 2022, 12 (6). <https://www.mdpi.com/2075-1729/12/6/776>.
- [3] Knuth D E. The  $\text{\TeX}$  Book [M]. 15th ed. Reading, MA: Addison-Wesley Publishing Company, 1989.
- [4] 聂灵沼, 丁石孙. 代数学引论 [M]. 北京: 高等教育出版社, 2000.
- [5] Nikiforov V. Analytic methods for uniform hypergraphs [J]. Linear Algebra and its Applications, 2014, 457: 455–535.
- [6] Bu C, Fan Y, Zhou J. Laplacian and signless Laplacian Z-eigenvalues of uniform hypergraphs [J]. Frontiers of Mathematics in China, 2016, 11 (3): 511–520.
- [7] Hu S, Qi L, Shao J. Cored hypergraphs, power hypergraphs and their Laplacian H-eigenvalues [J]. Linear Algebra and its Applications, 2013, 439: 2980–2998.
- [8] Kang L, Nikiforov V. Extremal problems for the  $p$ -spectral radius of graphs [J]. Electronic Journal of Combinatorics, 2014, 21 (3): P3.21.
- [9] Lin H, Zhou B. Distance spectral radius of uniform hypergraphs [J]. Linear Algebra and its Applications, 2016, 506: 564–578.
- [10] Lu L, Man S. Connected hypergraphs with small spectral radius [J]. Linear Algebra and its Applications, 2016, 509: 206–227.
- [11] Nikiforov V. Hypergraphs and hypermatrices with symmetric spectrum [J]. Linear Algebra and its Applications, 2017, 519: 1–18.
- [12] Qi L.  $H^+$ -eigenvalues of Laplacian and signless Laplacian tensors [J]. Communications in Mathematical Sciences, 2014, 12 (6): 1045–1064.

## 附录 A 经典不等式

论文中用到的经典不等式.

**(Hölder Inequality)** 设  $a_i \geq 0, b_i \geq 0, i = 1, 2, \dots, n$ , 且  $p > 1, q > 1$  满足  $1/p + 1/q = 1$ . 则有

$$\sum_{i=1}^n a_i b_i \leq \left( \sum_{i=1}^n a_i^p \right)^{\frac{1}{p}} \cdot \left( \sum_{i=1}^n b_i^q \right)^{\frac{1}{q}},$$

等号成立当且仅当存在一个常数  $c$  满足  $a_i^p = c b_i^q$ .

**(PM Inequality)** 设  $x_1, x_2, \dots, x_n$  是  $n$  个非负实数. 如果  $0 < p < q$ , 那么

$$\left( \frac{x_1^p + x_2^p + \dots + x_n^p}{n} \right)^{\frac{1}{p}} \leq \left( \frac{x_1^q + x_2^q + \dots + x_n^q}{n} \right)^{\frac{1}{q}},$$

等号成立当且仅当  $x_1 = x_2 = \dots = x_n$ .

**(AM-GM Inequality)** 设  $x_1, x_2, \dots, x_n$  是  $n$  个非负实数. 则有

$$\frac{x_1 + x_2 + \dots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \dots x_n},$$

等号成立当且仅当  $x_1 = x_2 = \dots = x_n$ .