|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | 训练集 | 测试集 | *P*值 |
| 性别 | 男 | 65 | 20 | 0.232 |
|  | 女 | 19 | 2 |  |
| 年龄 |  | 59.37±11.55 | 59.55±7.95 | 0.946 |
| **乙肝或丙肝病毒感染** | 是 | 60 | 18 | 0.325 |
|  | 否 | 24 | 4 |  |
| **TBIL****（nmol/L）** |  | 15.13±8.43 | 19.551±11.96 | 0.135 |
| **ALB（g/l)** |  | 40.12±7.67 | 41.00±4.77 | 0.471 |
| **AFP(ng/ml)** | 正常 | 31 | 6 | 0.122 |
|  |  | 53 | 15 |  |
| **CA199(ng/ml)（U/ml）** |  | 25.85±30.30 | 35.69±33.81 | 0.169 |
| **CEA(ng/ml)0-5** |  | 2.89±1.61 | 2.56±1.14 | 0.488 |
| **最大径** |  | 8.02±4.51 | 7.56±4.25 | 0.700 |
| **肿瘤边界光滑** | 是 | 32 | 10 | 0.530 |
|  | 否 | 52 | 12 |  |
| **包膜完整或不完整** | **是** | 67 | 20 | 0.351 |
| **是** | **否** | 17 | 2 |  |
| **包膜** | **有** | 35 | 6 | 0.469 |
|  | **无** | 49 | 15 |  |
| **动脉期瘤周强化** | 有 | 7 | 2 | 1.000 |
|  | 无 | 77 | 20 |  |
| **瘤内坏死** | 有 | 68 | 14 | 0.094 |
|  | 无 | 16 | 8 |  |
| **肿瘤破裂** | 是 | 5 | 0 | 0.364 |
|  | 否 | 79 | 22 |  |
| **静脉癌栓** | 有 | 13 | 4 | 0.750 |
|  | 无 | 71 | 18 |  |
| **肝硬化** | 有 | 38 | 13 | 0.247 |
|  | 无 | 46 | 9 |  |
| **瘤内动脉** | 有 | 75 | 20 | 1.000 |
|  | 无 | 9 | 2 |  |
| **瘤周低密度环** | 有 | 42 | 9 | 0.447 |
|  | 无 | 42 | 13 |  |
| **瘤与肝脏边界的低密度影** | 有 | 58 | 16 | 0.738 |
|  | 无 | 26 | 6 |  |
| **RIV** |  | 39 | 10 | 1.000 |
|  |  | 44 | 12 |  |

205例是自己收的，246例是加上雅慧的；排序就按照这个顺序拍；108和109重复，删掉了108

普通填充缺失值

Arg1：整体：肿瘤破裂 小于0.001

实验组同上

CK19：整体：单因素：包膜是否完整 0.006、肿瘤边界光滑 0.009，多因素：无，同前所述两者相互影响

Ki67：整体：单因素：**1**.静脉癌栓 0.035；**2.**有无包膜：0.16；**3.**包膜是否完整：0.015；**4.**肿瘤边缘是否光滑：0.005；多因素：2、3、4三者相互影响，Logistic去掉其中两个剩下的一个就有意义。

GPC3：整体：单因素：TBIL：0.30；CA199：0.03；RIV：0.001；瘤周低密度环：0.004；瘤内动脉：0.012；肿瘤分化程度；肝硬化：＜0.001；；静脉癌栓；包膜是否完整：0.005；AFP：0.003

多因素：CA199（连续）；肝硬化；包膜完整或不完整。

多重插补后：

Arg1：整体：单因素：**①**TBIL *P*=0.003；**②** 肿瘤破裂 0.014；**③** 包膜是否完整：0.022；

多因素：TBIL 0.008；肿瘤破裂 0.023或包膜是否完整：0.013；**②③**相互影响

**①+②****AUC值0.735 ①+③** **AUC值：0.764**

CK19：整体：单因素：**①**包膜是否完整 0.006、**AUC值：0.648；②**肿瘤边界光滑 0.004，**AUC值：0.622；**

多因素：无，同前所述两者相互影响

KI67：整体：单因素：**①**瘤内动脉0.024；**②**静脉癌栓 0.010；**③.**有无包膜：0.030；**④.**包膜是否完整：0.021；**⑤.**肿瘤边缘是否光滑 0.003。

多因素：**①+④** **AUC值0.692**

GPC3：整体：单因素：TBIL：0.30；CA199：0.03；RIV：0.001；瘤周低密度环：0.004；瘤内动脉：0.012；肿瘤分化程度；肝硬化：＜0.001；；静脉癌栓；包膜是否完整：0.005；AFP：0.003

多因素：CA199；肝硬化；包膜完整或不完整。**AUC值0.872**

分化程度：整体：单因素：ALB 0.44；静脉癌栓 0.36；多因素：无

然后分析了一下临床影像特征与个病理指标之间的相关性，发现相关性都不太高，最高r=0.35

多重插补一般不要用于分类变量

**20220414**

205例ki67删除缺失值：

CD34：单：reader1：1.RIV；2.瘤周低密度环；3.包膜是否完整；4.边界是否光滑；5.最大径

多：RIV+肿瘤边界 0.720

Ki6714：单：reader1：瘤内动脉；瘤内坏死；最大径 多：瘤内动脉瘤；内坏死；分化程度；多：瘤内动脉；瘤内坏死；分化程度，ROC 0.739

CK19：瘤内动脉

246例插补：Ki6714：单：reader1：瘤内动脉；静脉癌栓；最大径 多：静脉癌栓；瘤内动脉

Ki6714：单：reader2 瘤内动脉；肿瘤边界；静脉癌栓；最大径；多：无

CD34：单：reader1：RVI；肝脏边界低密度；低密度环；最大径；多：RVI

单：reader2 ：RVI；肝脏边界低密度；低密度环；瘤内动脉；最大径；包膜是否完整；多：RVI

CK19：单：reader1：瘤内动脉；reader2：瘤内动脉

GPC3：单：reader1：肝硬化；静脉癌栓；肿瘤破裂；多：无；单：reader2：肝硬化；静脉癌栓；肿瘤破裂；多：无

246删除缺失值：

Ki6714：单：reader1：瘤内动脉；静脉癌栓；肿瘤边界光滑；多：瘤内动脉静脉癌栓肿瘤边界光滑。AUC 0.754

CD34：单：reader1：单：reader1：RVI；低密度环；肿瘤边界；包膜是否完整；瘤内动脉；最大径；多：肿瘤边界光滑最大径.AUC 0.689

GPC3：单：reader1：肝硬化，静脉癌栓；瘤周低密度影；多：肝硬化，静脉癌栓；AUC：0.679

CK19：单：reader1：肝硬化，瘤内动脉；瘤内坏死；多：肝硬化；瘤内坏死；AUC0.631

**20220416**

缺失值全部删除189例

KI67：reader1：单：瘤内动脉，静脉癌栓；多：瘤内动脉，静脉癌栓 AUC0.725

CD34：单：reader1：RVI；低密度环；瘤内坏死；最大径；（包膜是否完整；瘤内动脉；）多：RVI+（包膜是否完整）AUC：0.712

CK19：单：瘤内动脉；

GPC3：肝硬化，静脉癌栓。瘤周低密度；（瘤内动脉）多：肝硬化；静脉癌栓（瘤内动脉），AUC：0.751

primary拆分训练集和测试集

CD34为标签：

训练集：CD34：单：RVI；低密度影；低密度环；肝硬化；瘤内坏死；边界光滑；(瘤内动脉)；多：

KI67:单:瘤内动脉（静脉癌栓）

CK19：单：瘤内动脉（肝硬化）

GPC3：单：肝硬化（RVI）

验证集：CD34：单：动脉期瘤周强化

CK19为标签：

训练集：CD34：单：最大径；边界光滑；瘤内坏死；瘤内动脉；低密度环（包膜是否完整；肝硬化，低密度影）；多

KI67:单:瘤内动脉；静脉癌栓；

CK19：单：（动脉期瘤周强化）

GPC3：单：肝硬化，静脉癌栓；肿瘤破裂；（低密度影）；多；肝硬化（静脉癌栓）

验证集：CD34：单RVI

KI67：单：RVI；低密度环；（低密度影；瘤内坏死）

GPC3无；CK19：瘤内动脉；（RVI；包膜完整或不完整）

KI67为标签：

训练集：CD34：RVI；低密度环；卫星灶；

ki67：瘤内动脉；静脉癌栓

CK19：瘤内动脉；

GPC3：肝硬化；静脉癌栓

验证集：CD34：边界光滑

ki67：

CK19：肝硬化瘤周强化最大径

GPC3：肝硬化

GPC3为标签：

训练集：CD34：RVI；瘤周低密度环，瘤内坏死；（静脉癌栓）多：瘤内坏死静脉癌栓

ki67：瘤内动脉（瘤内坏死；肿瘤破裂）

CK19：无

GPC3：低密度影；肝硬化；（静脉癌栓）

验证集：CD34：包膜是否完整

ki67：静脉癌栓（瘤周低密度环）

CK19：瘤内坏死；包膜；（瘤内动脉）；

GPC3：瘤周强化

**20220417**

secondary拆分训练集和测试机

CD34为标签：

训练集：CD34单：最大径；肿瘤边界，瘤内坏死；肝硬化；瘤内动脉；低密度环；低密度影；RVI；肿瘤边界光滑；RVI；AUC0.724

KI67：单：瘤内动脉；静脉癌栓；AUC0.722

CK19：肝硬化

GPC3：静脉癌栓；肝硬化；瘤内动脉；多：瘤内动脉；肝硬化AUC：0.705

验证集：CD34：瘤周动脉强化

KI67：静脉癌栓；肿瘤边界强化；（瘤内动脉）；CK19：无

GPC3：肝硬化（静脉癌栓）

KI67为标签：

训练集：CD34：RIV；低密度环；包膜是否完整；多：RVI；包膜是否完整 AUC：0.713

KI67：瘤内动脉。静脉癌栓 0.753

CK19：瘤内动脉

GPC3：静脉癌栓；肝硬化；多同前。0.726

验证集：CD34：肿瘤边界是否光滑

KI67：无

CK19：肝硬化；瘤周动脉期强化

GPC3：肝硬化；（瘤内动脉）

CK19为标签：

训练集：CD34：肿瘤边界；瘤内坏死；瘤内动脉；低密度环；RVI；多（肿瘤边界+RVI）

KI67：瘤内动脉；静脉癌栓；多：同前0.732

CK19：（动脉瘤周强化；肝硬化；瘤内动脉）

GPC3：肝硬化；静脉癌栓；肿瘤破裂；多：静脉癌栓；肝硬化；0.734

验证集：CD34：（RVI）

Ki67：低密度环；RVI（瘤内坏死；静脉癌栓）

CK19：无

GPC3：

GPC3为标签：

训练集：CD34：瘤内坏死；瘤周低密度环；RVI（瘤内动脉；）多:无

KI67：瘤内动脉；静脉癌栓；

CK19：瘤内动脉

GPC3：静脉癌栓；低密度影；肝硬化

多：静脉癌栓；肝硬化

验证集：CD34：

KI67：

CK19：

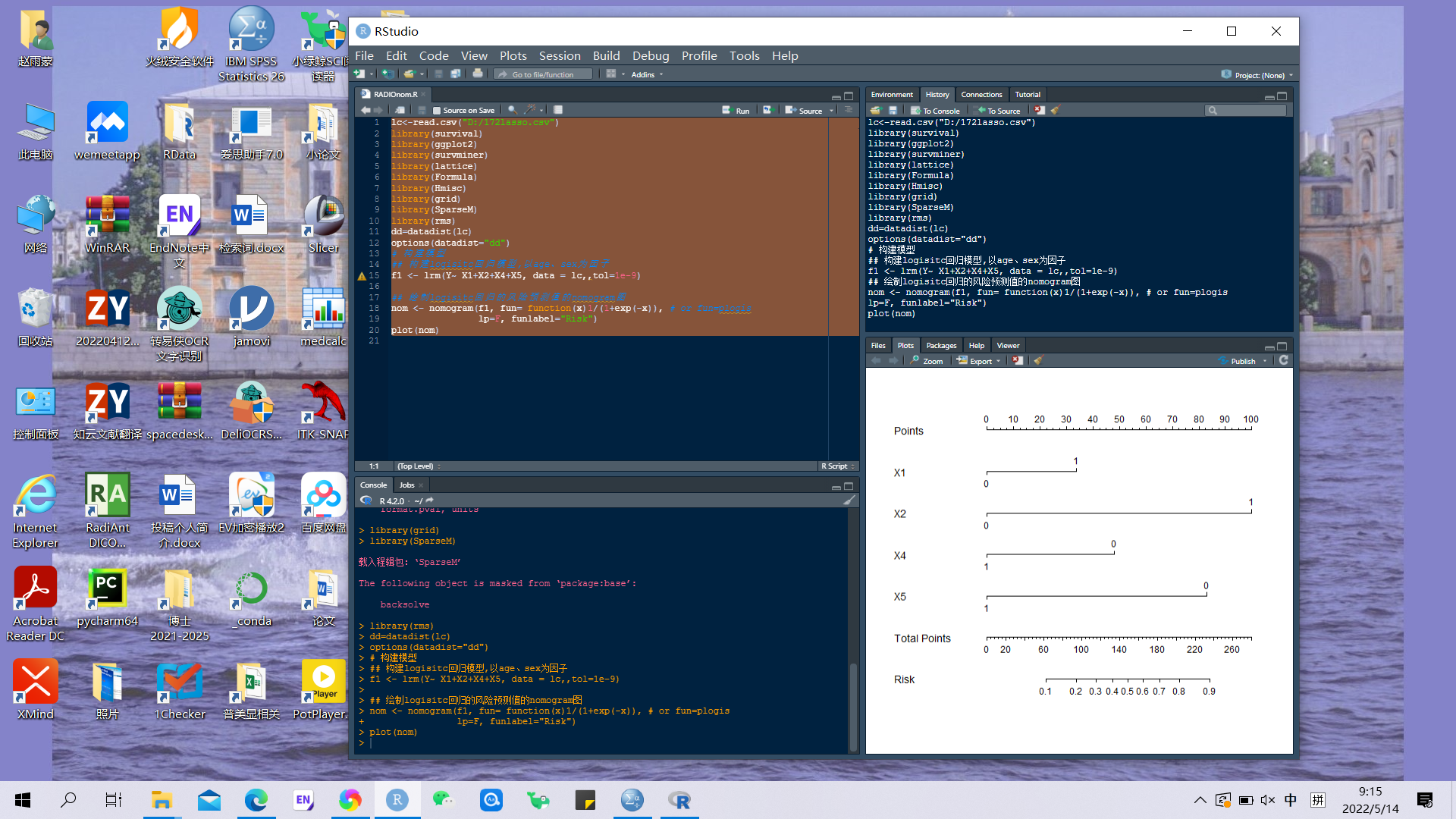
GPC3：

以secondarydataki67为标签；进行分组特征提取

**思路1.R语言-logistics回归分析-诺莫图**

2022-05-02：D:\DOCZ 2021-2025\Gd-EOB-DTPA ralated\Gd-EOB-DTPA DATA\RadiomicDATA20220404\ML-DATA\188LI-RADS\172SPSS

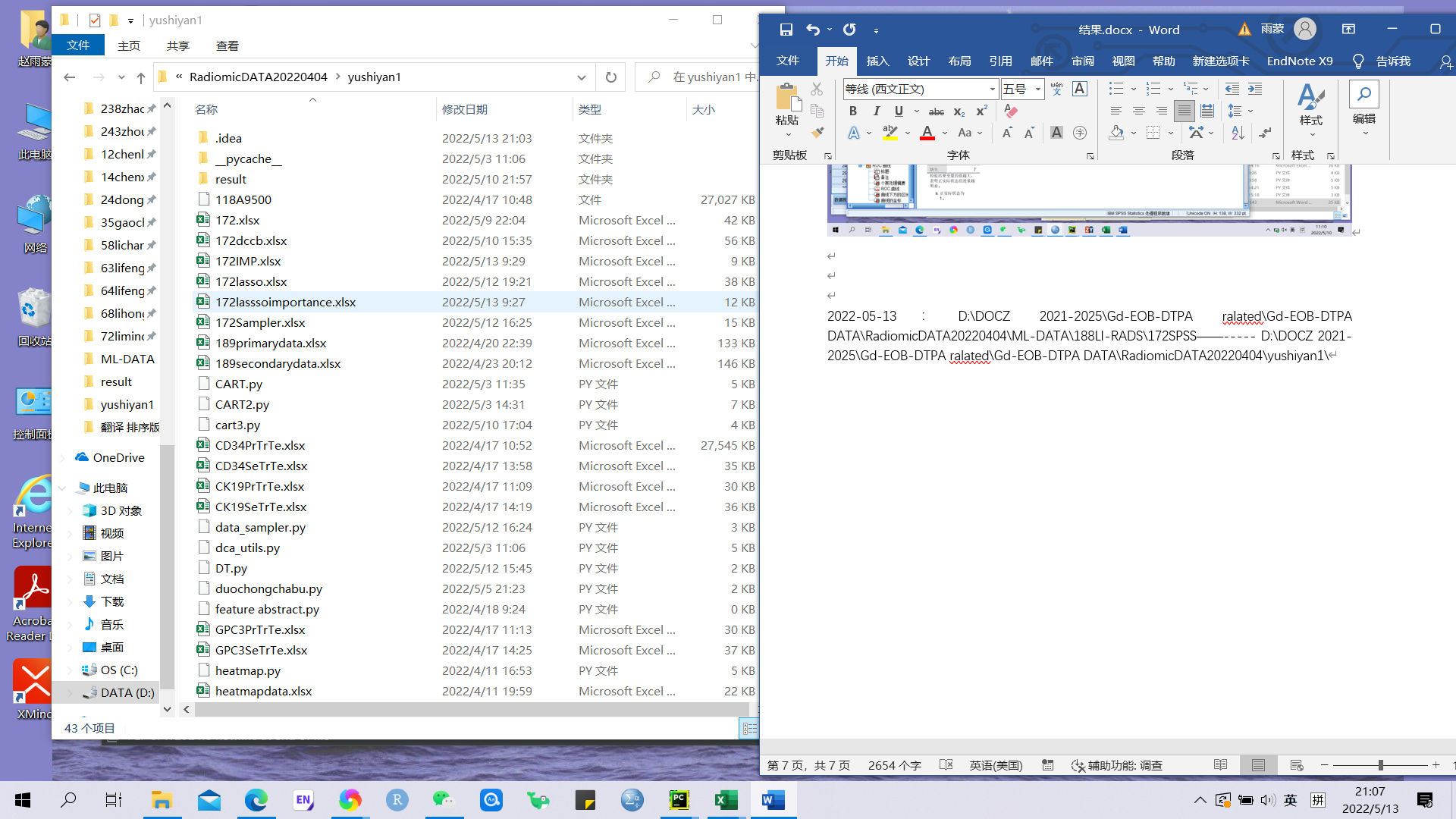
Lirads影像学特征：KI-67：最大径、5cm，静脉癌栓、瘤周卫星灶、瘤内动脉（0.05）；是否廓清、周边廓清、环状强化、肿瘤分化程度（0.2）；多：瘤周卫星灶；是否廓清；动脉期高强化，最大径5cm；AUC：0.705

临床特征：单AFP

分化程度：瘤内坏死；最大5cm；（0.05）；环状强化、瘤内动脉、周边廓清、结中结、瘢痕征、静脉癌栓，多：瘤内坏死、结中结



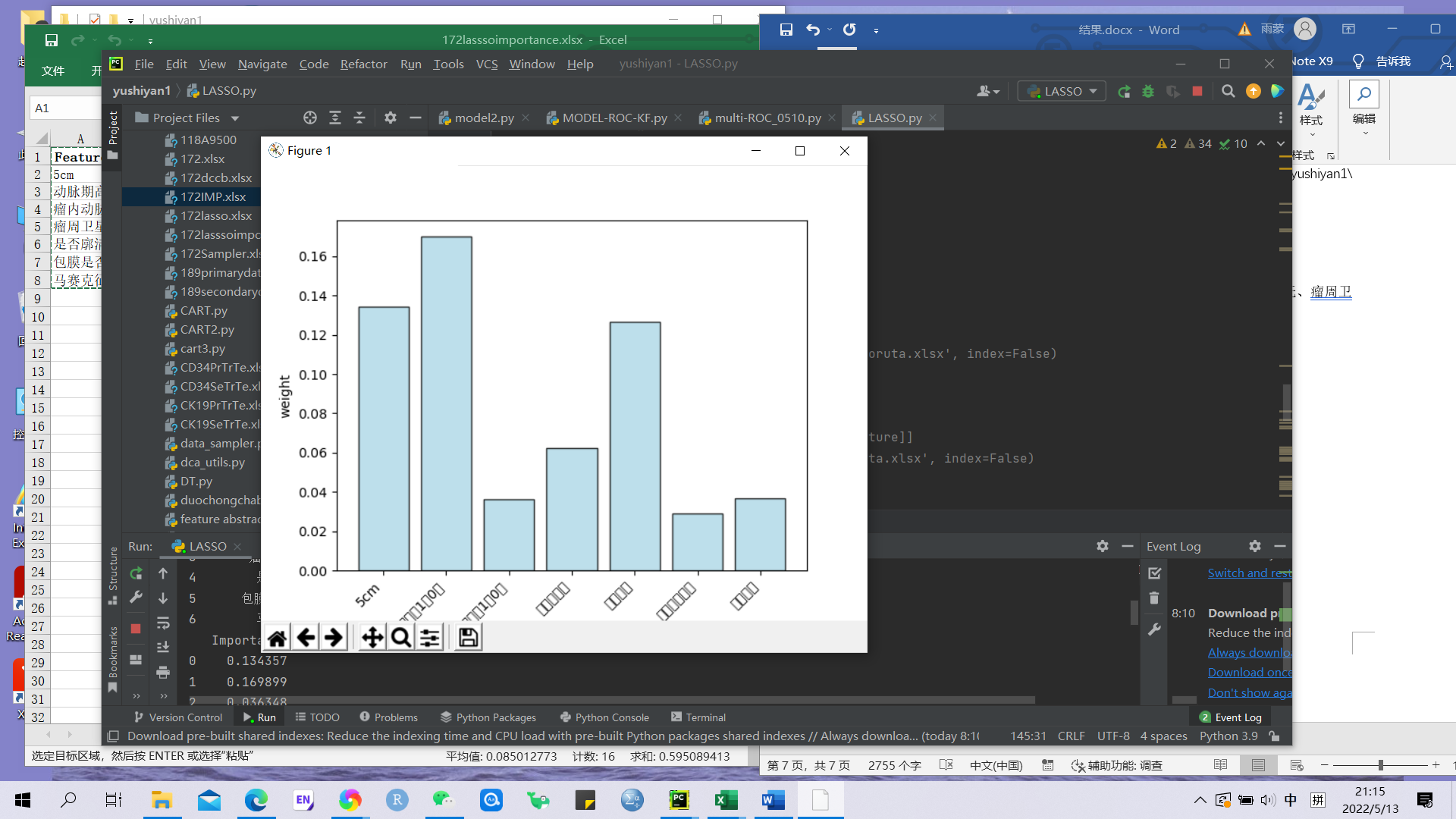


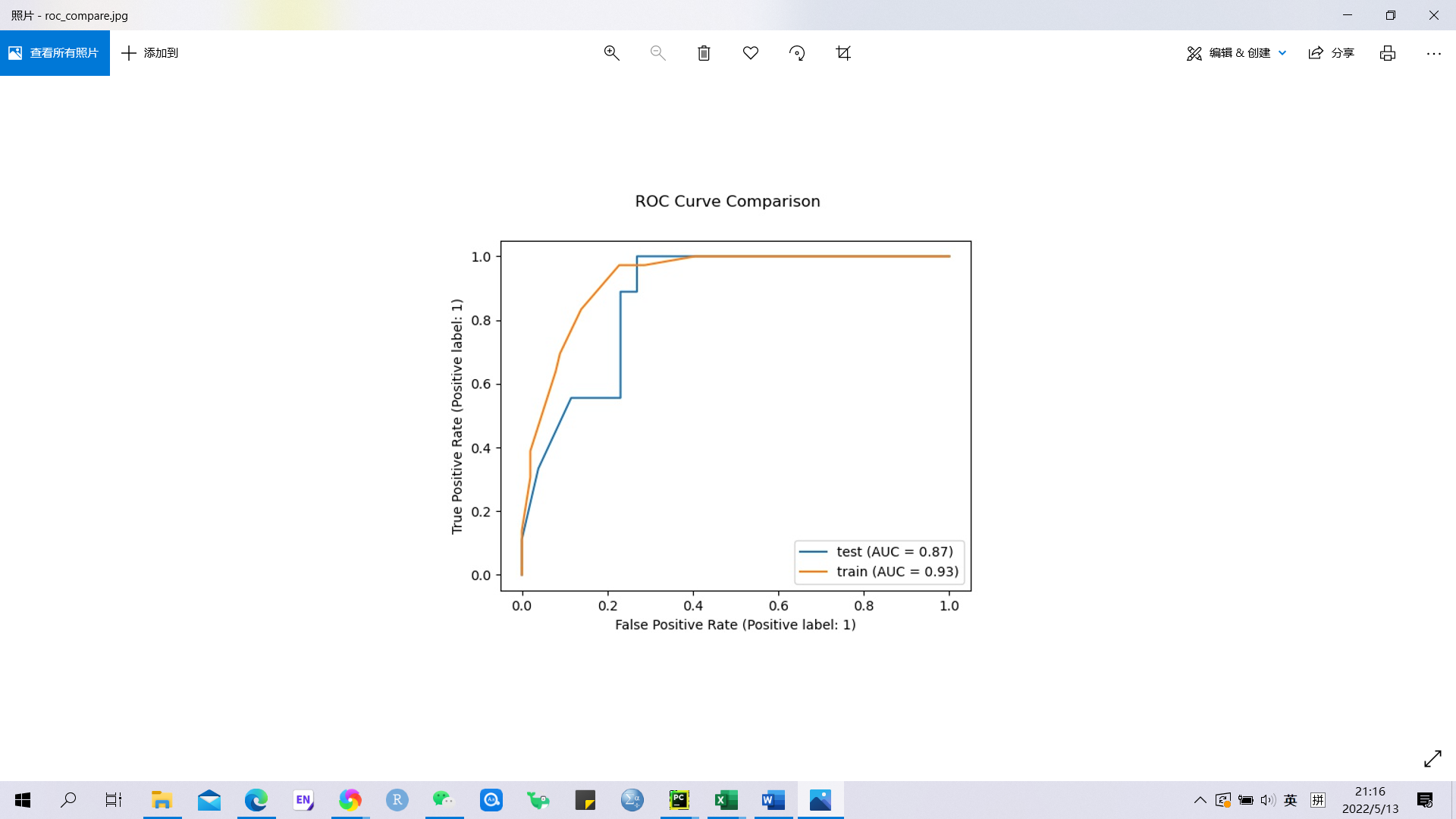
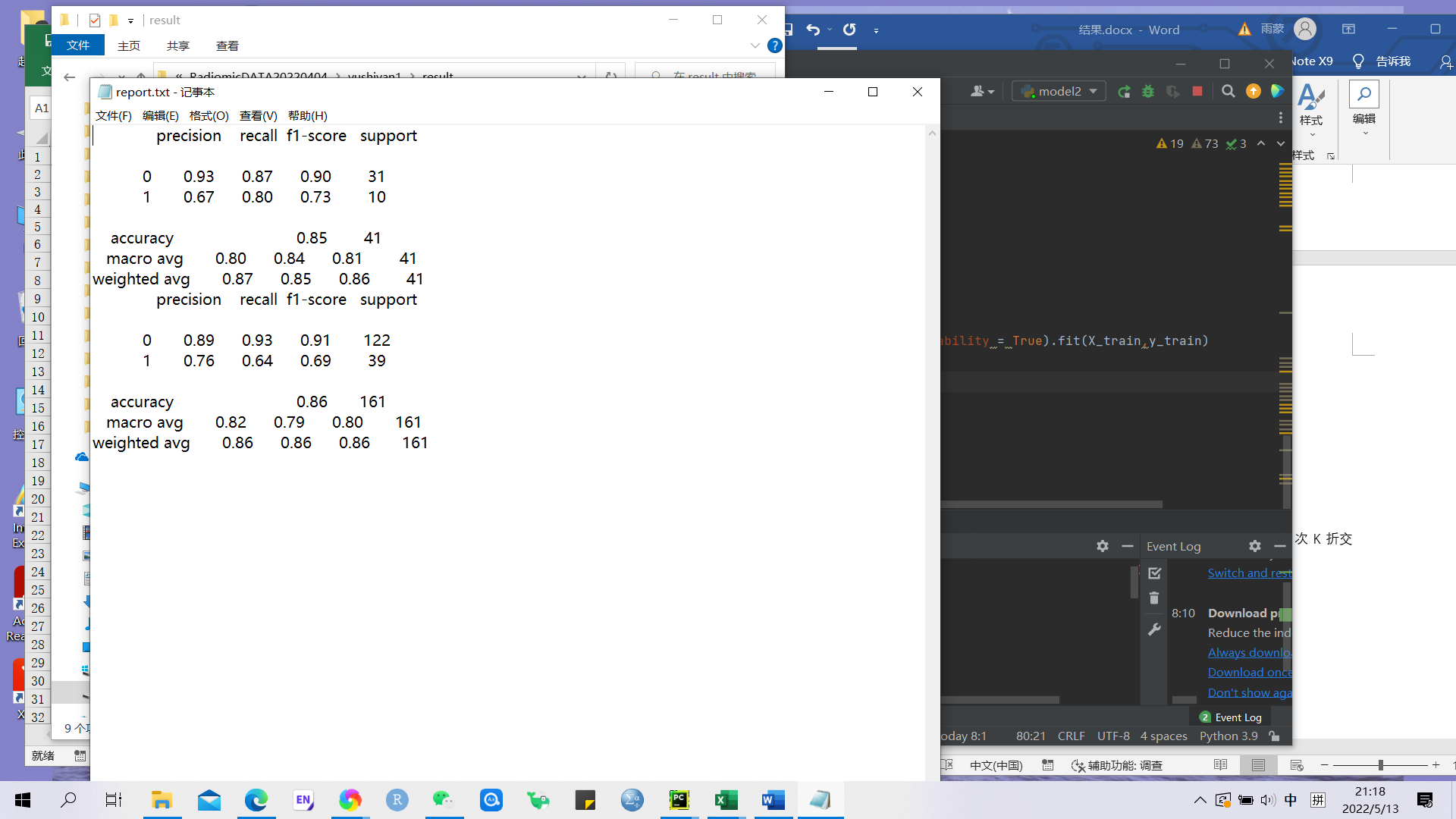
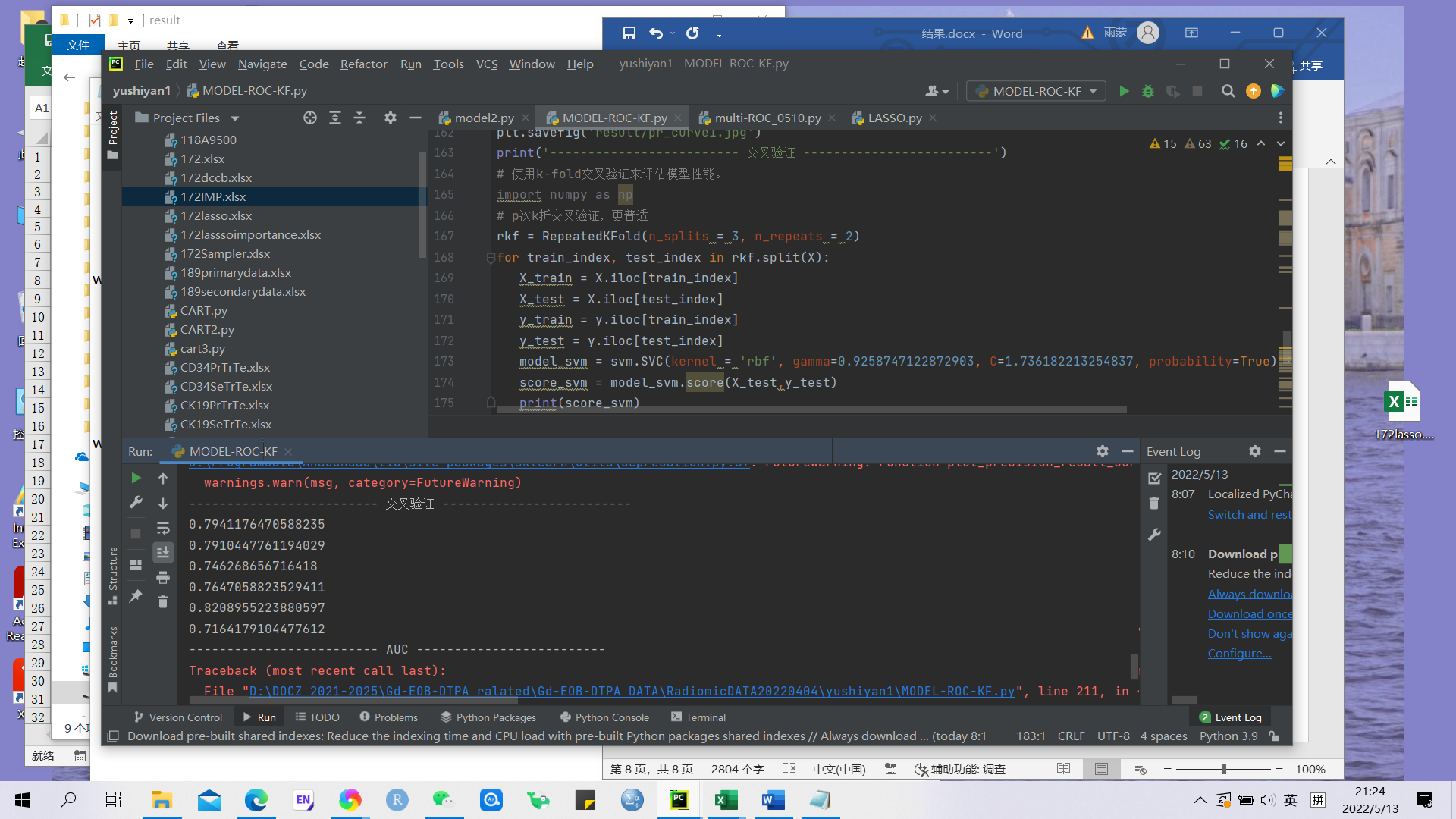
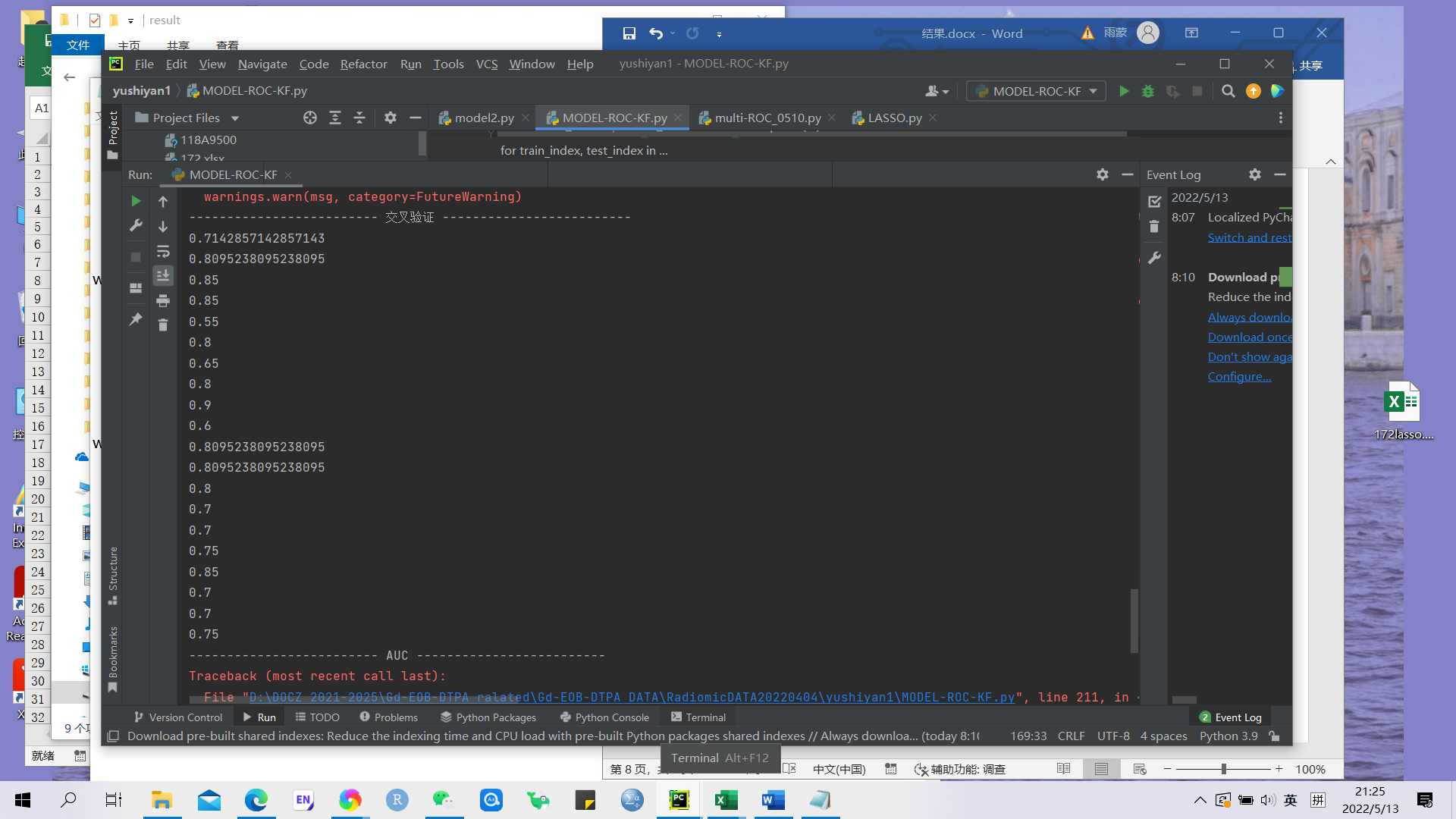
2022-05-13：D:\DOCZ 2021-2025\Gd-EOB-DTPA ralated\Gd-EOB-DTPA DATA\RadiomicDATA20220404\ML-DATA\188LI-RADS\172SPSS——----- D:\DOCZ 2021-2025\Gd-EOB-DTPA ralated\Gd-EOB-DTPA DATA\RadiomicDATA20220404\yushiyan1\ 

172lasso-sheetname（SPSS）

|  |  |
| --- | --- |
| **Feature** | **Importance** |
| 5cm | 0.134357 |
| 动脉期高强化1是0否 | 0.169899 |
| 瘤内动脉1有0无 | 0.036348 |
| 瘤周卫星灶 | 0.062321 |
| 是否廓清 | 0.126474 |
| 包膜是否强化 | 0.028897 |
| 马赛克征 | 0.036793 |

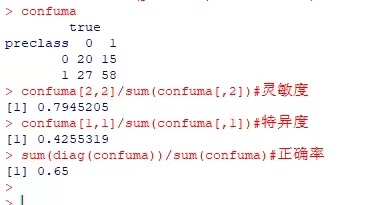
**思路2：LASSO筛选-特征-机器学习模型-交叉验证**

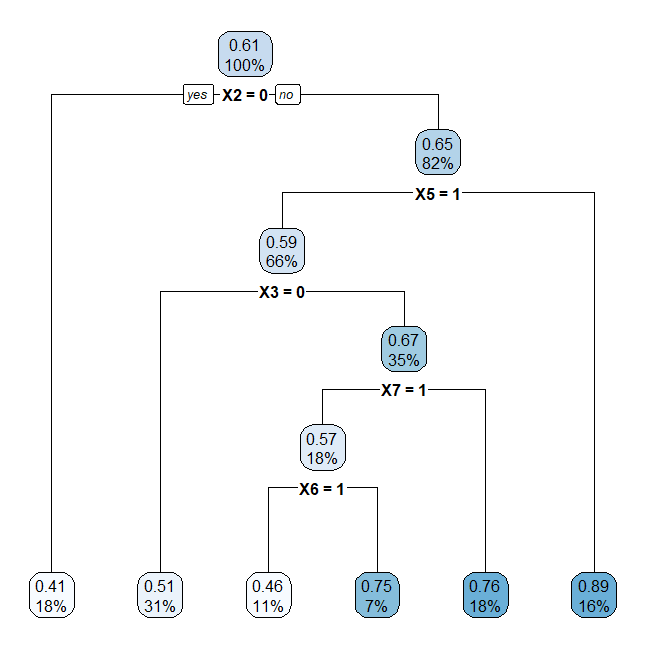
，5cm，动脉期高强化1是0、瘤内动脉1有0无、瘤周卫星灶、是否廓清、包膜是否强化、马赛克征； 

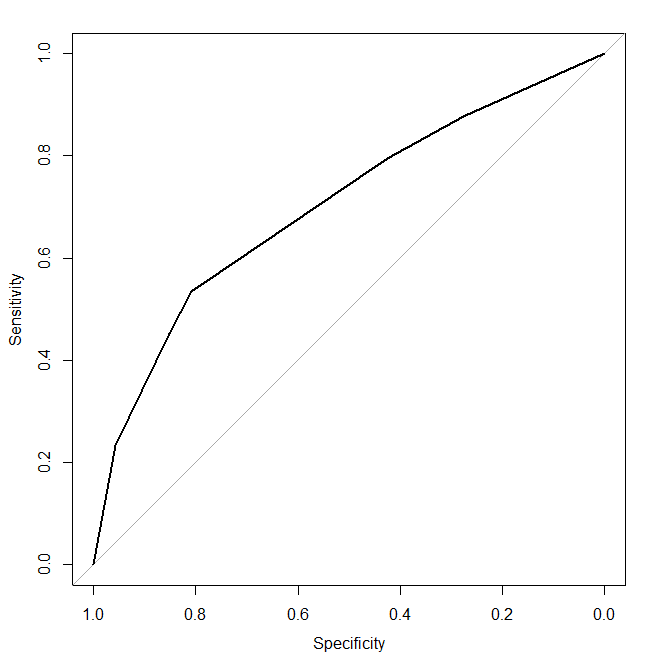
随机森林拆分训练集测试集并建模，，特异度灵敏度以及准确性 ，P次K折交叉验证模型性能：（3次2折）：2次10折 

**思路3：logistic筛选-特征-cart决策树**

**结果不好**

****

****



0529：发现python中标签放置错误，应该在最后一列，导致结果出问题

修改后：**7个指标**

1.SVM0.78\0.72;LR ：0.75/0.79;0.76\0.72;RF:0.83/0.73;0.83\0.75(5050150200);XGB:0.81/0.72;DT:0.84/0.72;KNN:0.83/0.66

2.SVM0.82\0.76;LR 0.74/0.72;RF:0.78/0.70;XGB:0.85/0.71;DT:0.84/0.75;KNN:0.71/0.79

6个指标：1.SVM0.82\0.76;LR 0.74/0.73;RF:0.83/0.75（50100150200）;XGB:0.81/0.73（100150150200）;DT:0.82/0.73;KNN:0.73/0.81

2.SVM0.82\0.76;LR 0.74/0.73;RF:0.83/0.75（50100150200）;XGB:0.81/0.73（100150150200）;DT:0.82/0.73;KNN:0.73/0.81

Logistic回归：1.SVM:0.35/0.18；2.LR: 0.68/0.81；3.RF：0.70/0.83；4.XGB：0.70/0.83；5.DT0.79/0.83；6.KNN：0.65/0.66

LASSO回归（7个指标）：1.SVM:0.82/0.76；2. RF：0.86/0.76；3. XGB：0.84/0.72；4. LR: 0.74/0.71；5.DT0.78/0.71；6.KNN：0.71/0.79

Logistic回归(4个指标)：1.SVM:0.73/0.70；2.LR: 0.71/0.79；3.RF：0.76/0.73；4.XGB：0.74/0.73；5.DT0.76/0.75；6.KNN：0.72/0.78

20220602：

与师姐数据一致性检验后：

Logistic回归分析（3个指标：动脉期高强化、瘤周低密度灶、静脉癌栓）

1.SVM:0.54/0.46；2. RF：0.81/0.79；3. XGB：0.81/0.79；4. LR: 0.77/0.79；5.DT0.81/0.79；6.KNN：0.75/0.72

Lasso回归（4个指标：动脉期高强化、瘤内动脉、瘤周低密度灶、静脉癌栓）

）

1.SVM:0.76/0.70；2. RF：0.86/0.0；3. XGB：0.83/0.77；4. LR: 0.82/0.76；5.DT0.81/0.79；6.KNN：0.79/0.78

Lasso回归（5个指标：动脉期高强化、瘤内动脉、有无包膜、瘤周低密度灶、静脉癌栓）

）

1.SVM:0.76/0.71；2. RF：0.84/0.82；3. XGB：0.83/0.77；4. LR: 0.82/0.76；5.DT0.81/0.79；6.KNN：0.79/0.78

20220606结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **检验统计a** | | | | | | |
| Train | | 年龄 | ALB（g/l) | TBIL（nmol/L高于20.6异常） | CEA(ng/ml)0-5 | CA199(ng/ml)（U/ml） |
| 0 | 曼-惠特尼 U | 281.500 | 302.000 | 295.000 | 277.000 | 281.500 |
| 威尔科克森 W | 512.500 | 798.000 | 791.000 | 773.000 | 777.500 |
| Z | -.642 | -.438 | -.569 | -.905 | -.821 |
| 渐近显著性（双尾） | .521 | .661 | .569 | .366 | .412 |
| 精确显著性（双尾） | .527 | .667 | .579 | .371 | .418 |
| 精确显著性（单尾） | .264 | .334 | .289 | .186 | .209 |
| 点概率 | .003 | .003 | .006 | .002 | .003 |
| 1 | 曼-惠特尼 U | 1594.500 | 1539.500 | 1685.000 | 1551.500 | 1704.000 |
| 威尔科克森 W | 2819.500 | 2764.500 | 2910.000 | 2776.500 | 4189.000 |
| Z | -.525 | -.948 | -.162 | -.883 | -.059 |
| 渐近显著性（双尾） | .600 | .343 | .871 | .377 | .953 |
| 精确显著性（双尾） | .602 | .345 | .873 | .379 | .954 |
| 精确显著性（单尾） | .301 | .173 | .436 | .190 | .477 |
| 点概率 | .001 | .001 | .001 | .001 | .001 |
| a. 分组变量：Ki67 | | | | | | |

20220614：和双双姐讨论开题计划

20220617：小组会文献汇报，存在问题：文献level不够高，下次要拔高，前言少说，重点关注研究方法

20220618：1.筛查MRI部分病例，找出了132例，后续进行剔除(术前TACE，手术、没有图等等)，同时拷图。2.筛查肿瘤医院病例

20220703

Ck19 单：靶征；包膜完整；周边廓清；廓清；环状强化；5cm；肝硬化；

多

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **方程中的变量** | | | | | | | |
|  | | B | 标准误差 | 瓦尔德 | 自由度 | 显著性 | Exp(B) |
| 步骤 1a |  | .885 | .444 | 3.982 | 1 | .046 | 2.424 |
| 环状强化1是0否 | 1.462 | .656 | 4.973 | 1 | .026 | 4.316 |
| 廓清 | -1.054 | .528 | 3.989 | 1 | .046 | .349 |
| 肝硬化 | -.948 | .449 | 4.466 | 1 | .035 | .387 |
| 常量 | -1.010 | .514 | 3.862 | 1 | .049 | .364 |
| a. 在步骤 1 输入的变量：, 环状强化1是0否, 廓清, 肝硬化。 | | | | | | | |

ROC:0.703

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **方程中的变量** | | | | | | | | | |
|  | | B | 标准误差 | 瓦尔德 | 自由度 | 显著性 | Exp(B) | EXP(B) 的 95% 置信区间 | |
| 下限 | 上限 |
| 步骤 1a | 5cm | .886 | .449 | 3.899 | 1 | .048 | 2.426 | 1.007 | 5.849 |
| 环状强化1是0否 | 1.355 | .684 | 3.922 | 1 | .048 | 3.875 | 1.014 | 14.805 |
| AFP(ng/ml)0-20正常 | 1.107 | .514 | 4.635 | 1 | .031 | 3.026 | 1.104 | 8.290 |
| 廓清 | -1.115 | .537 | 4.308 | 1 | .038 | .328 | .114 | .940 |
| 肝硬化 | -1.053 | .460 | 5.232 | 1 | .022 | .349 | .142 | .860 |
| 常量 | -1.720 | .625 | 7.565 | 1 | .006 | .179 |  |  |
| a. 在步骤 1 输入的变量：5cm, 环状强化1是0否, AFP(ng/ml)0-20正常, 廓清, 肝硬化。 | | | | | | | | | |

AUC:0.747

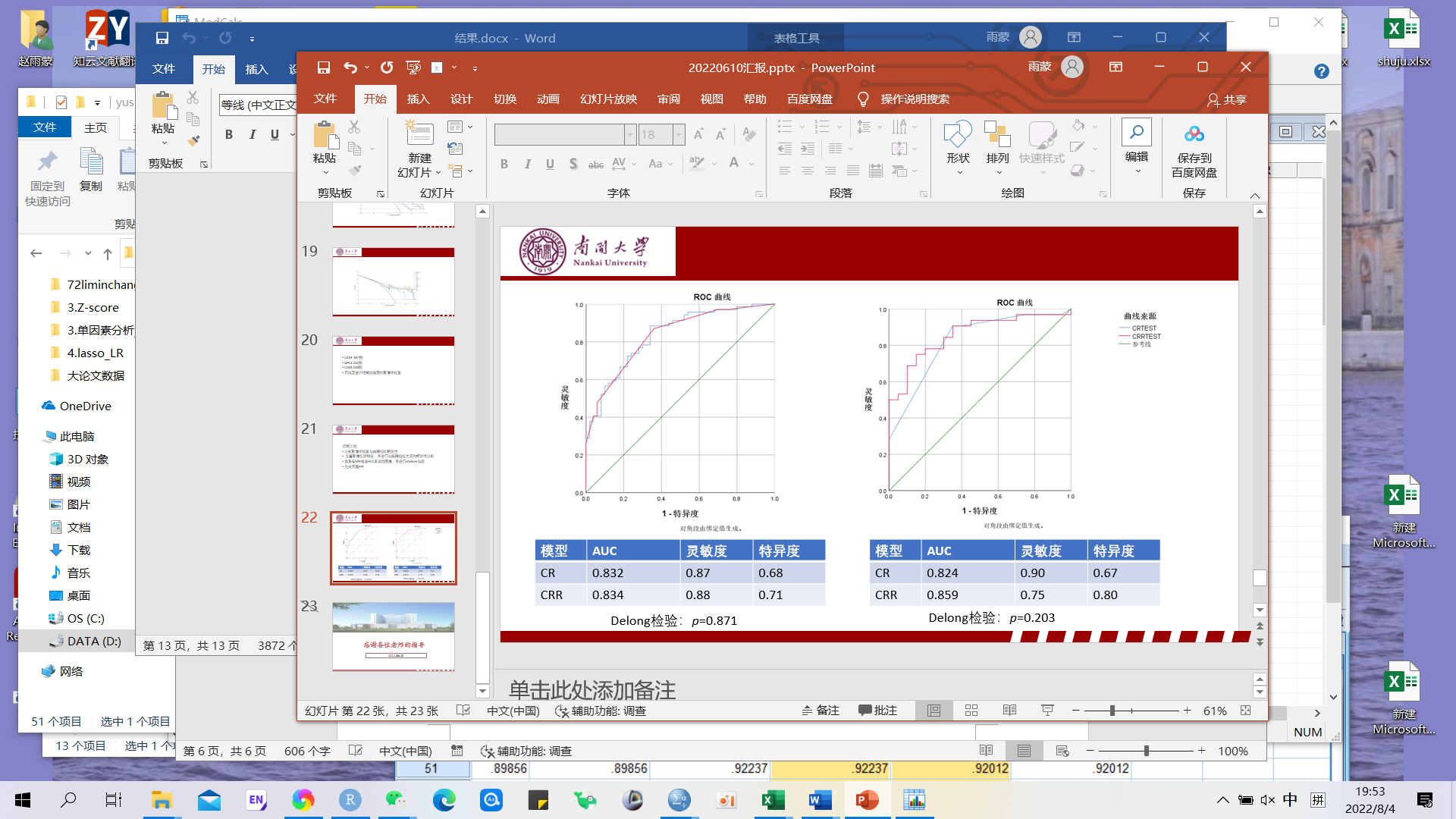
GPC3

多

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **方程中的变量** | | | | | | | |
|  | | B | 标准误差 | 瓦尔德 | 自由度 | 显著性 | Exp(B) |
| 步骤 1a | AFP(ng/ml)0-20正常 | 1.992 | .420 | 22.494 | 1 | .000 | 7.327 |
| 脂肪肝背景肿瘤不含脂肪（是1） | -1.268 | .517 | 6.018 | 1 | .014 | .281 |
| 病灶破裂（是1 | -2.611 | .806 | 10.488 | 1 | .001 | .073 |
| 常量 | .787 | .283 | 7.719 | 1 | .005 | 2.196 |
| a. 在步骤 1 输入的变量：AFP(ng/ml)0-20正常, 脂肪肝背景肿瘤不含脂肪（是1）, 病灶破裂（是1。 | | | | | | | |

0.792

20220804



训练集：CR：afp 高强化、静脉癌栓、PPTVI

CRR：radscore：1se

测试集：CR：afp 高强化、PPTVI;如果加上静脉癌栓auc就太大了，超过训练集

CRR：radscore：1se

结果保存spss：D:\DOCZ 2021-2025\Gd-EOB-DTPA ralated\Gd-EOB-DTPA DATA\RadiomicDATA20220404\SPss、20220804ki67CR-CRR.mc1；

GPC3

廓清病灶破裂afp