基于多模态信息融合的药物分子活性预测系统v1.0

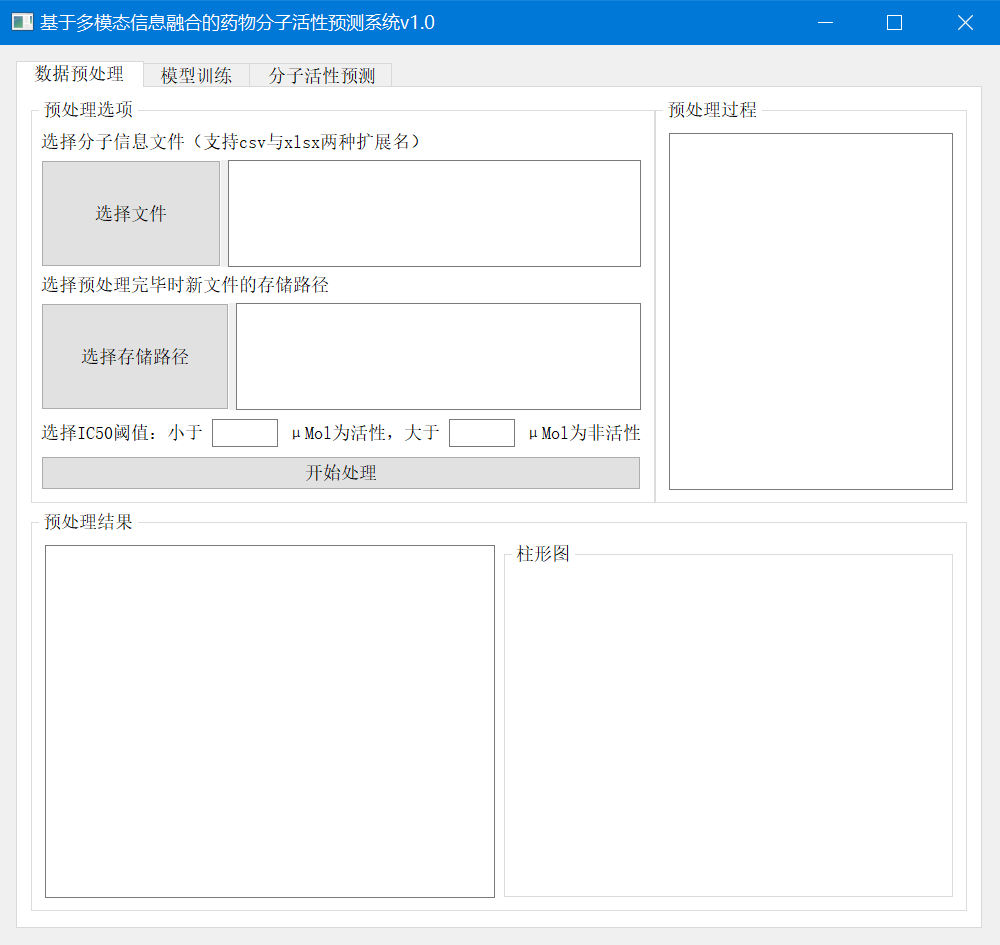
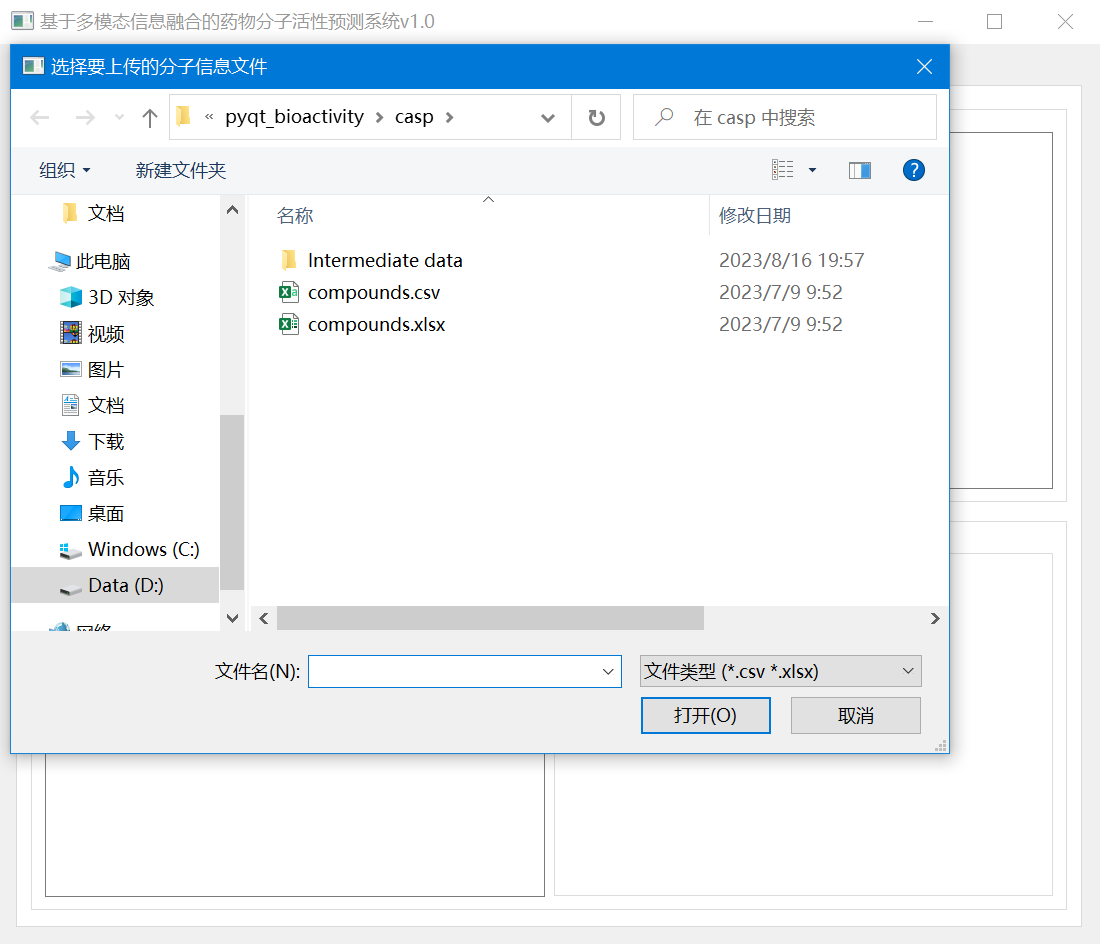
该系统由三部分组成，分别是数据预处理、模型训练、分子活性预测。系统中数据预处理的主界面如图1所示。

图1 系统数据预处理功能的主界面

在数据预处理部分，首先需要选择一个分子信息文件，该文件中至少要包含：分子SMILES式、IC50值两项。文件可以是csv或xlsx后缀。点击“选择文件”按钮，将弹出对话框，如图2所示。

图2 导入分子文件的选择地址对话框

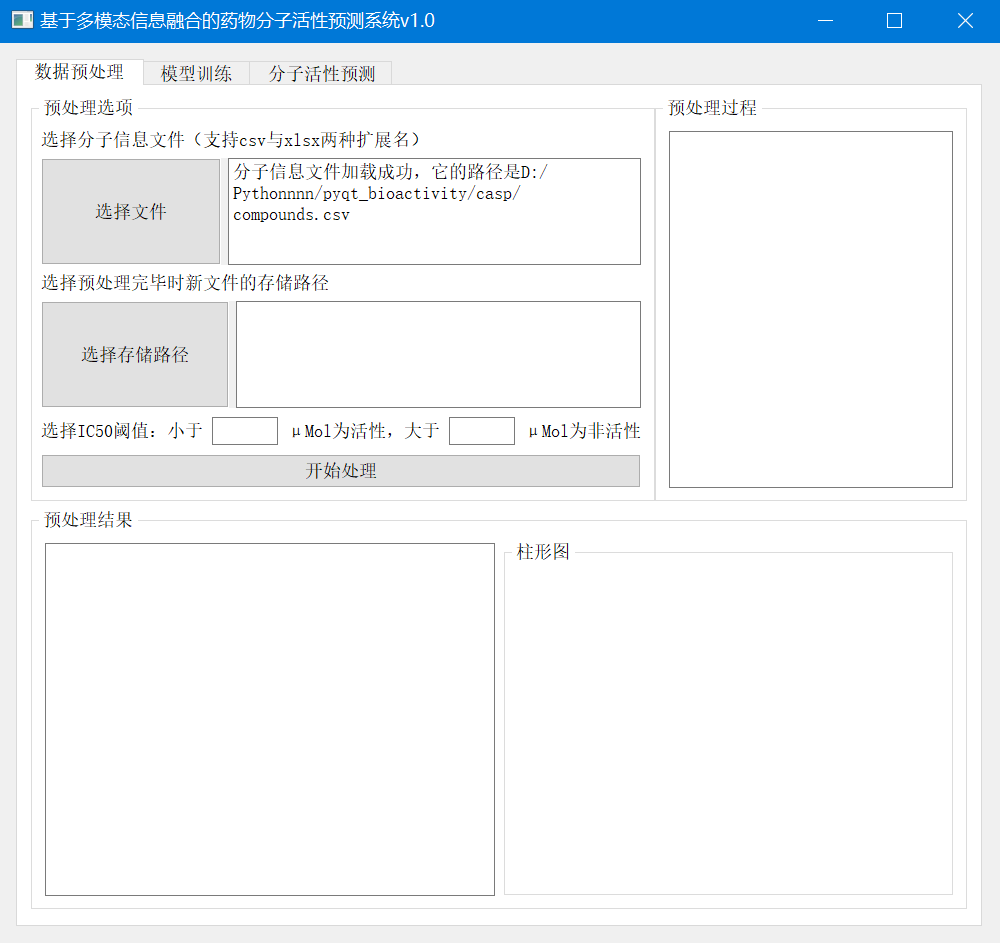
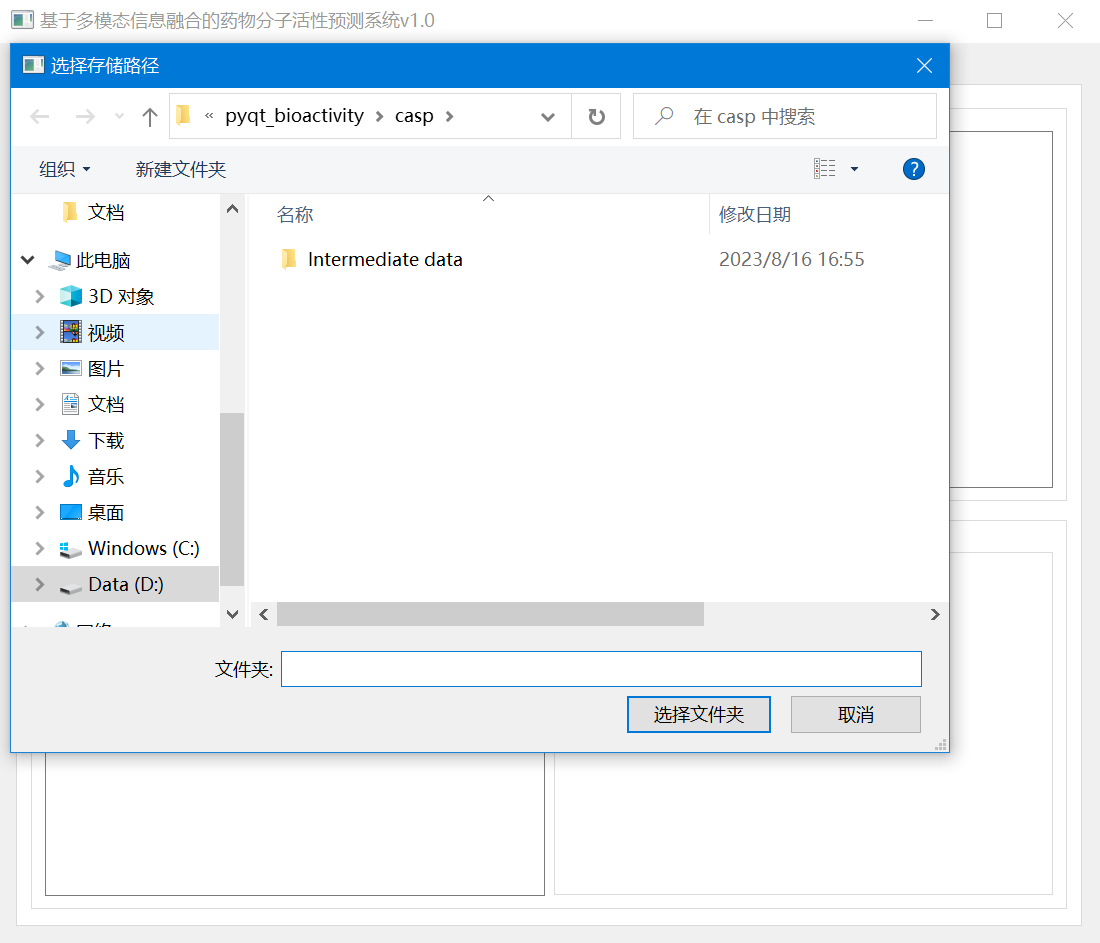
选择完分子文件之后，系统“选择文件”按钮的右侧将提示操作成功，并显示出所选择分子文件的路径，如图3所示。

图3 分子文件选择完毕后，按钮右侧文本框中提示了此操作成功，并显示出分子文件的路径

在选择完文件之后，需要选择一个路径来存储系统预处理之后的分子文件。点击“选择存储路径”将弹出路径选择的对话框，如图4所示。

图4 选择预处理后的分子文件存储路径的对话框

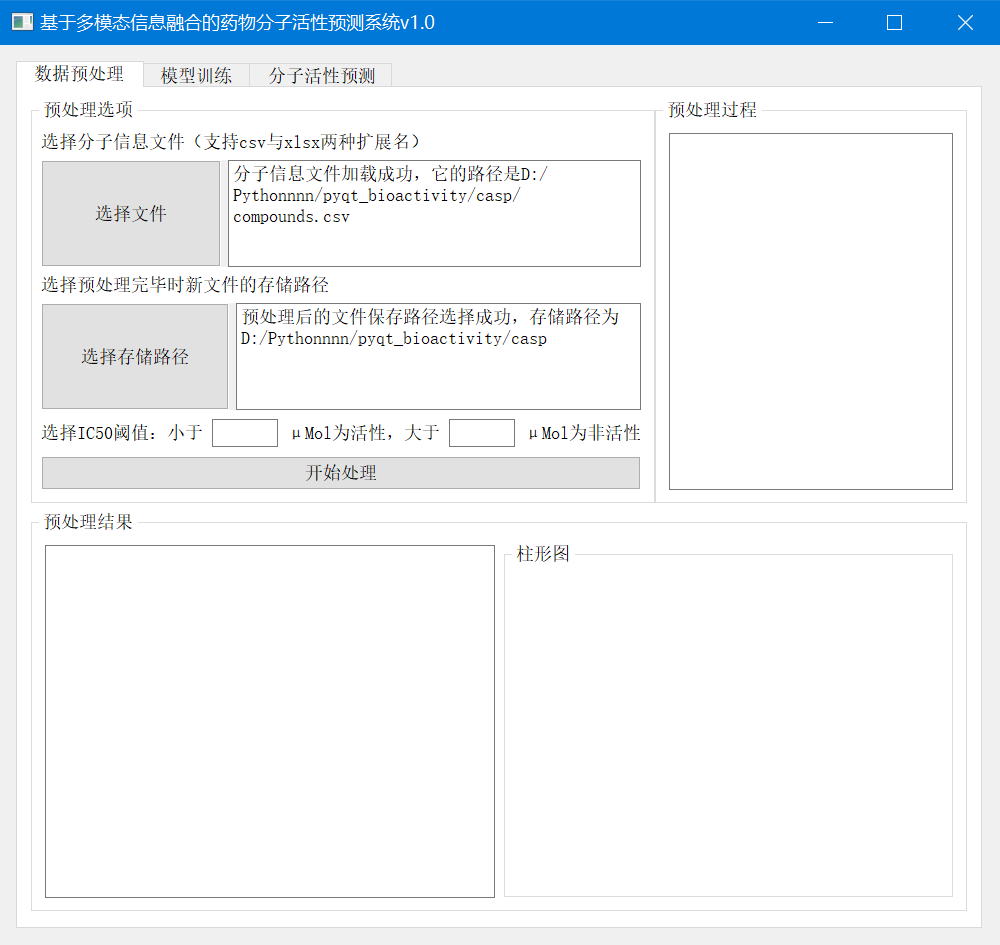
选择完路径之后，该按钮的右侧文本框中将提示此操作成功，并显示出所选择的路径地址，如图5所示。

图5 存储路径选择完毕后，按钮右侧文本框中提示了此操作成功，并显示出分子文件的路径

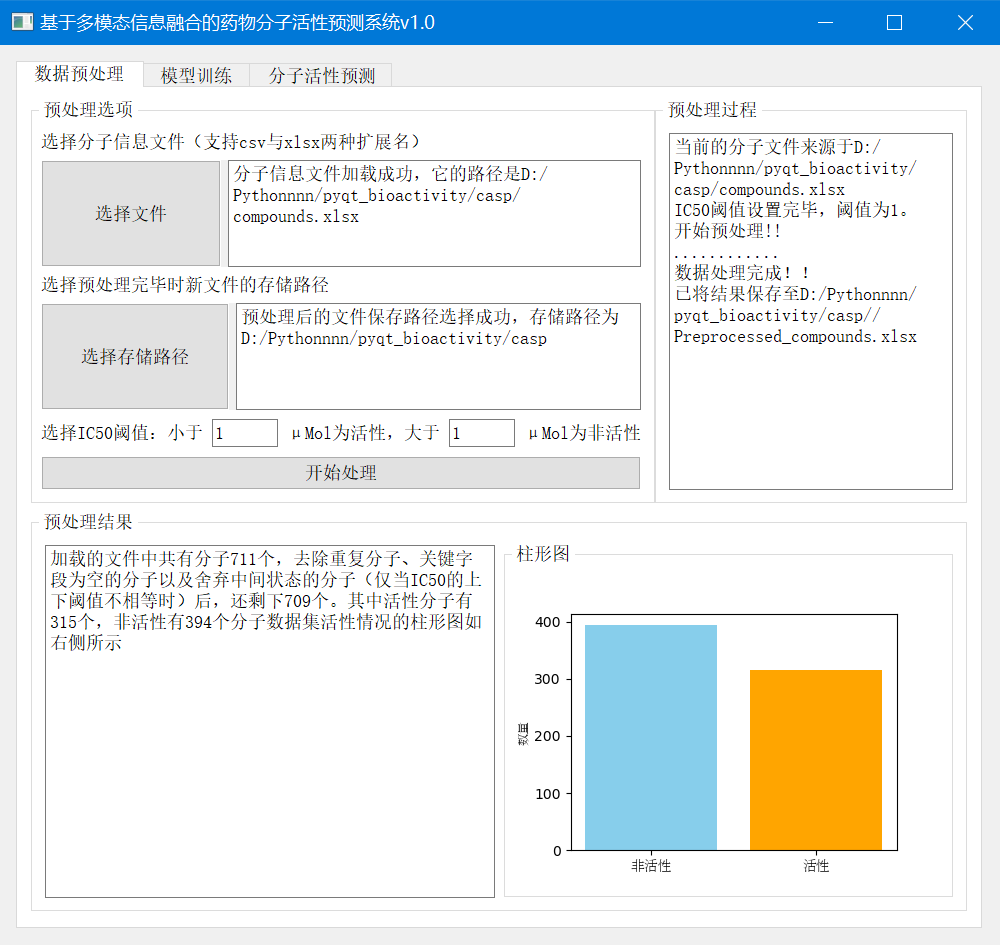
最后，需要填写IC50的两个阈值。两个阈值可以相等，也能不相等。当填写的2个阈值均为1μMol时，系统将会把IC50值低于1μMol的分子视为活性，高于则视为非活性。输入完毕后点击“开始处理”按钮，预处理过程文本框中将显示数据处理过程中的一些细节。预处理结果中将显示分子文件的最终划分结果，它右侧将会显示文件中活性与非活性的分子的柱形图。把阈值设为1μMol时的处理过程与结果，如图6所示。

图6 IC50阈值填写1μMol时的预处理结果

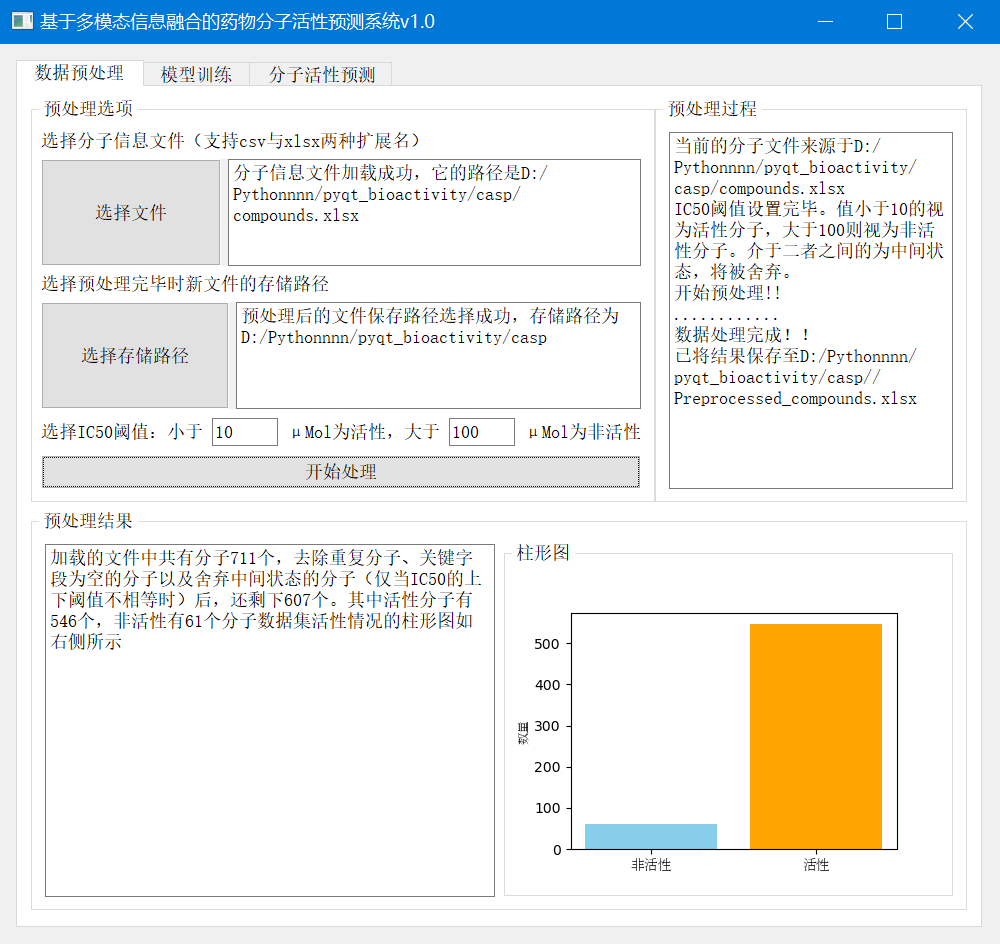
系统数据预处理部分，IC50的两个阈值若不相等，如下阈值选择为10μMol，上阈值选择100μMol时，系统会把IC50值小于10μMol的视为活性，大于100μMol的视为非活性，而介于二者之间的（10μMol<IC50<100μMol）分子视为中间状态，中间状态的分子将被舍弃，数据处理结果如图7所示。这种做法能够提升分子数据集的质量，从而提高模型的训练效果。但是对于分子较少的文件，这样划分会使小分子数据集变得更小，因此不推荐小数据集使用这种分类方式。

图7 IC50阈值分别填写10μMol、100μMol时的预处理结果

系统第二部分的功能是模型训练，该部分包括训练设置、训练过程和训练结果。它的主界面如图8所示。

图8 系统模型训练部分的主界面

在模型训练部分，首先需要选择训练设置。点击“选择分子数据集”按钮，会弹出对话框，在对话框中选择一个使用系统“数据预处理”部分处理过的分子文件。选择完毕后，右侧文本框会提示选择成功，并显示选择的分子文件路径。点击“模型存储地址”按钮之后，会弹出一个路径选择对话框。选择完存储路径后，按钮右侧文本框会显示出选择的路径，如图9所示。

图9 分别点击“选择分子数据集”按钮、“模型存储地址”按钮之后，文本框中显示出了所选择的文件地址和存储路径

训练设置中需要设置训练轮数、批数大小、学习率三个参数的值，设置完毕后，点击“开始训练”按钮，模型则会在上一步选择的分子数据集中进行训练和测试。“训练过程”中的文本框将显示训练时的详细信息，如图10所示。

图10 训练设置选择完毕后，点击“开始训练”按钮，模型在所选择的数据集上进行训练，训练的详细过程将显示在“训练过程”文本框内

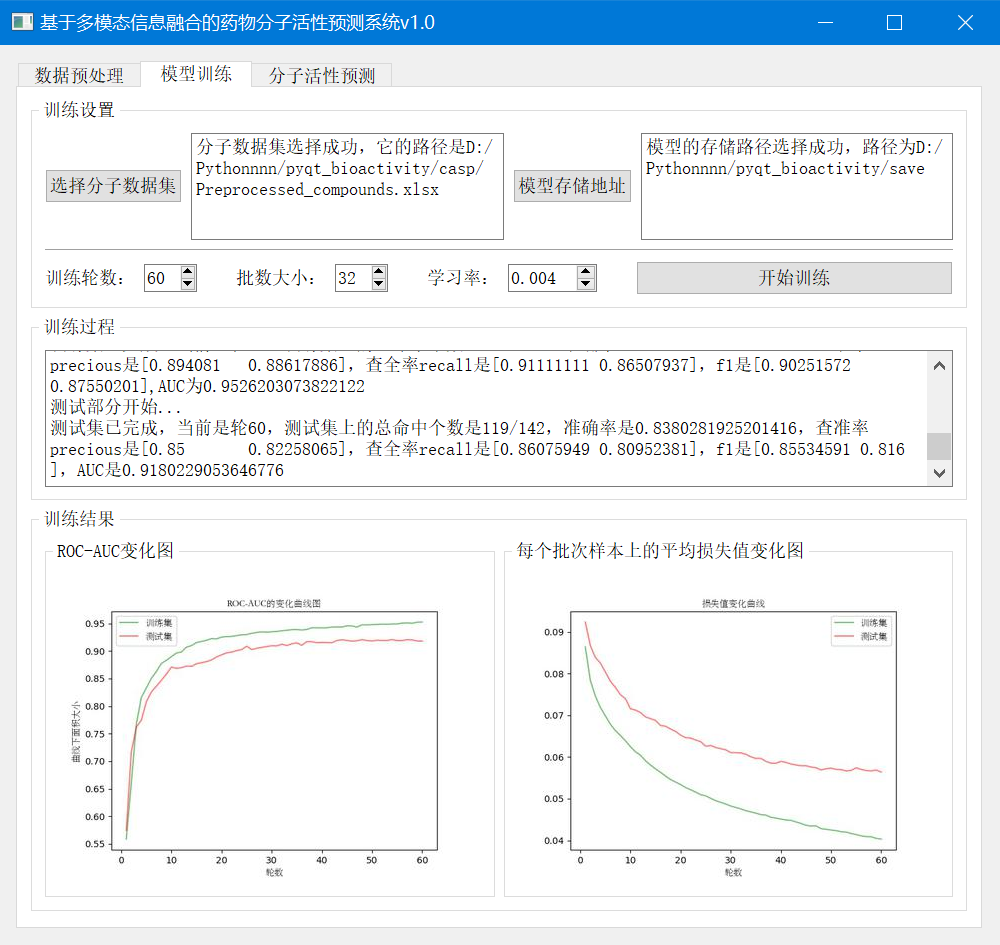
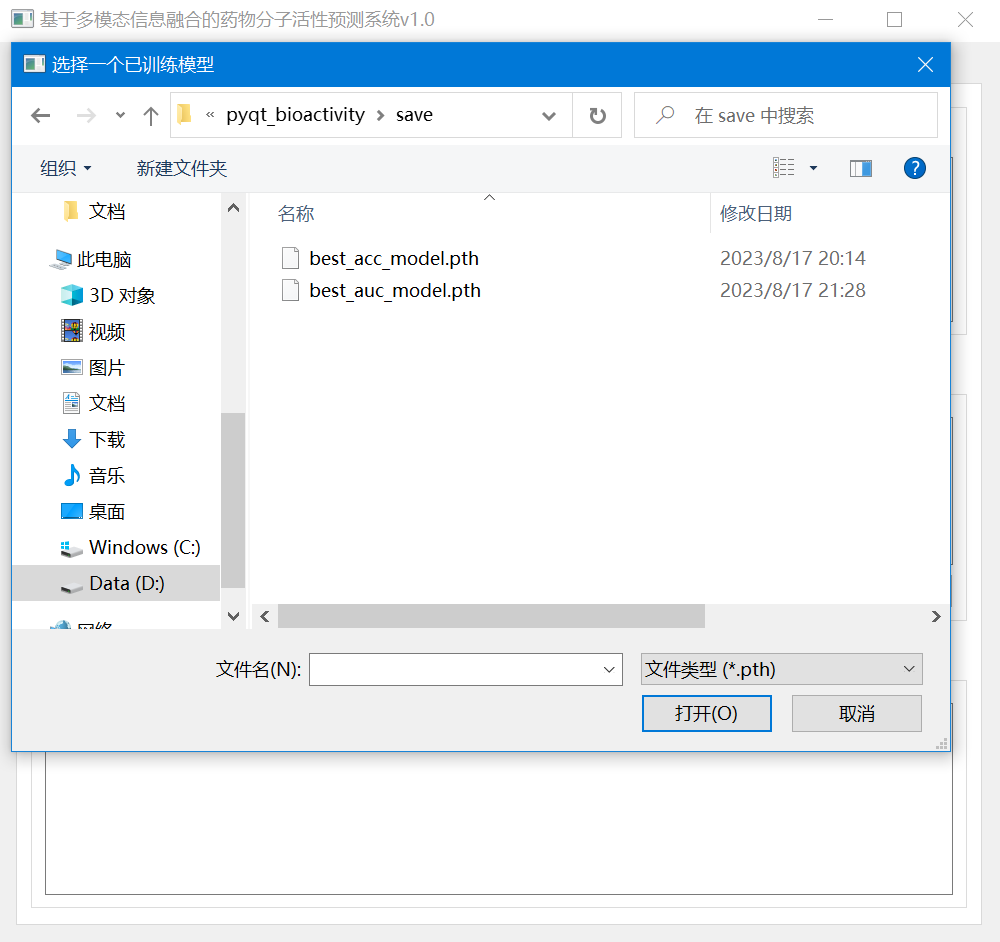
训练完毕之后，训练时ROC-AUC与损失值的变化曲线图将显示在训练结果之中，如图11所示。为了方便观察，在损失值变化图中，纵坐标是批次上的平均损失而非总损失。

图11 以0.004为学习率，批数32，训练60轮时得到的结果。损失值变化曲线图纵坐标为平均损失

系统第三部分的功能是分子活性预测，该部分包括选择已训练的模型、输入SMILES式、展示预测结果。它的主界面如图12所示。

图12 系统分子活性预测部分的主界面

在分子活性预测部分，首先需要选择一个已训练完毕的模型。点击“选择模型”按钮，会弹出对话框，在对话框中选择一个后缀为pth的模型，如图13所示。

图13 点击“模型选择”按钮后出现的对话框

选择完毕后，右侧文本框会提示选择成功，并显示选择的模型路径，如图14所示。

图14 选择完模型之后，“选择模型”按钮右侧的文本框中会提示选择成功，并显示所选择模型的路径

接下来，输入需要预测的分子的SMILES式至第二个文本框内，点击“预测”按钮，系统将会在第三个文本框中显示预测结果，如图15所示。

图15 输入一个SMILES式并点击预测之后，系统将显示该分子活性状态的预测结果

V2.0 更新内容



