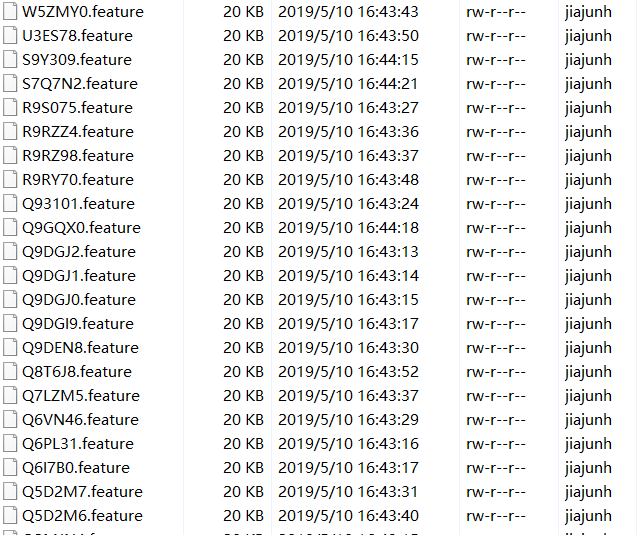
## 1.pssm的批量处理

把9大类的氧结合蛋白和非氧结合蛋白的pssm文件进行批量处理，生成.feature文件。、



## 2.弄懂模型的层次和参数。

## 处理完成后，参考学习学姐给的模型，（2层CNN和双向GRU）

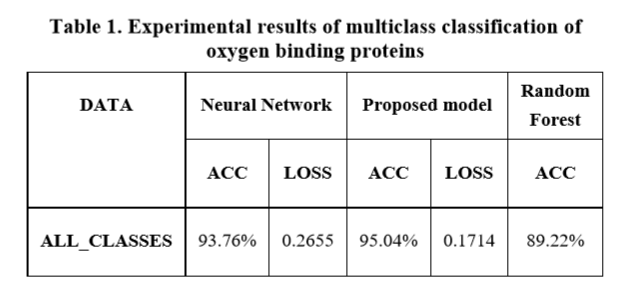
输入->嵌入层->一层CNN->dropout层（防止过拟合）->一层CNN->dropout层->池化层->双向GRU->全连接层->输出层（使用softmax函数）

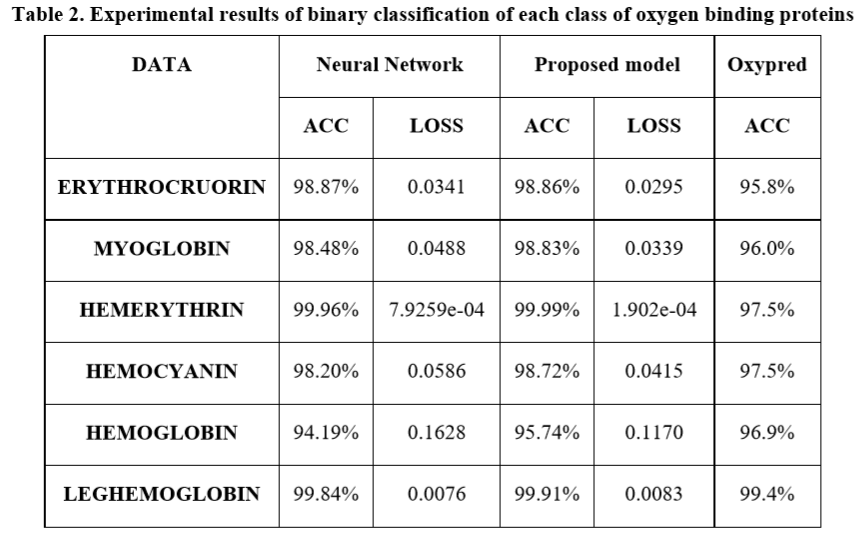
## 3.pssm的重新处理

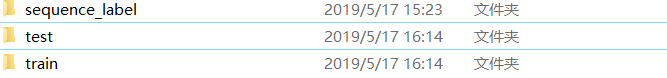
由于先前处理的pssm方法是把矩阵转化成了400维的向量，不能使用CNN，因而只能重新处理pssm文件，生成700\*20（序列长度小于700，不足补满）的矩阵，（遇到了一点小问题，还未测试完成）。

## 4.对训练集测试集的划分和打标记：

论文中的评估



对每一个类，取出序列，并对每一个序列打标签，标签号为文件名字（然后替换为数字1-10）。把生成的列表打印到一个txt文件中。手动划分每个类的训练集和测试集（另外把所有的放入一个文件，作为all classes的精确度测试）。



Sequence\_label为all classes。

效果图：前为序列，后为标签。

