**氧结合蛋白的分类预测**

1.3CNN测试结果：

测试集的各项指标如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | loss | precision | recall | f1score | acc |
| 第一折 | 1.00847 | 0.692 | 0.556 | 0.582 | 0.978 |
| 第二折 | 0.71071 | 0.466 | 0.457 | 0.265 | 0.869 |
| 第三折 | 0.69121 | 0.550 | 0.532 | 0.510 | 0.953 |
| 第四折 | 0.44332 | 0.689 | 0.768 | 0.723 | 0.972 |
| 第五折 | 0.38534 | 0.787 | 0.678 | 0.703 | 0.969 |
| 平均值 | 0.64781 | 0.637 | 0.598 | 0.557 | 0.948 |

每一类的分类性能的评价：

Cytoglobin（107）

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1score |
| 第一折 | 0.632 | 0.857 | 0.727 |
| 第二折 | 0.185 | 0.952 | 0.310 |
| 第三折 | 0.769 | 0.455 | 0.571 |
| 第四折 | 0.448 | 0.619 | 0.52 |
| 第五折 | 0.556 | 0.455 | 0.5 |
| 平均值 | 0.481 | 0.477 | 0.464 |

Erythrocruorin（13）

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1score |
| 第一折 | 0 | 0 | 0 |
| 第二折 | 0.008 | 0.333 | 0.016 |
| 第三折 | 0 | 0 | 0 |
| 第四折 | 0 | 0 | 0 |
| 第五折 | 1.0 | 0.333 | 0.5 |
| 平均值 | 0.2 | 0.067 | 0.1 |

Flavohemoprotein（943）

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1score |
| 第一折 | 0.935 | 0.994 | 0.964 |
| 第二折 | 0.947 | 0.937 | 0.941 |
| 第三折 | 0.932 | 0.947 | 0.939 |
| 第四折 | 0.873 | 0.989 | 0.928 |
| 第五折 | 0.902 | 0.979 | 0.939 |
| 平均值 | 0.928 | 0.939 | 0.93 |

Hemerythrin（35）

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1score |
| 第一折 | 1.0 | 0.429 | 0.6 |
| 第二折 | 0 | 0 | 0 |
| 第三折 | 0 | 0 | 0 |
| 第四折 | 0.875 | 1.0 | 0.933 |
| 第五折 | 0.778 | 1.0 | 0.875 |
| 平均值 | 0.536 | 0.686 | 0.492 |

Hemocyanin（31）

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1score |
| 第一折 | 0 | 0 | 0 |
| 第二折 | 0 | 0 | 0 |
| 第三折 | 1.0 | 0.667 | 0.8 |
| 第四折 | 0.75 | 1.0 | 0.857 |
| 第五折 | 1.0 | 0.857 | 0.923 |
| 平均值 | 0.6 | 0.705 | 0.596 |

Hemoglobin（1374）

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1score |
| 第一折 | 0.947 | 0.916 | 0.931 |
| 第二折 | 1 | 0.007 | 0.014 |
| 第三折 | 0.945 | 0.756 | 0.840 |
| 第四折 | 0.975 | 0.836 | 0.900 |
| 第五折 | 0.941 | 0.869 | 0.904 |
| 平均值 | 0.762 | 0.675 | 0.715 |

Leghemoglobin（28）

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1score |
| 第一折 | 1.0 | 0.4 | 0.571 |
| 第二折 | 0.053 | 1 | 0.102 |
| 第三折 | 0.75 | 0.6 | 0.667 |
| 第四折 | 1.0 | 1.0 | 1.0 |
| 第五折 | 0.75 | 0.5 | 0.6 |
| 平均值 | 0.7 | 0.5 | 0.568 |

Myoglobin（120）

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1score |
| 第一折 | 0.719 | 0.958 | 0.821 |
| 第二折 | 1.0 | 0.833 | 0.909 |
| 第三折 | 0.275 | 0.917 | 0.423 |
| 第四折 | 0.733 | 0.917 | 0.815 |
| 第五折 | 0.944 | 0.708 | 0.810 |
| 平均值 | 0.575 | 0.883 | 0.64 |

Neuroglobin（100）

|  |  |  |  |
| --- | --- | --- | --- |
|  | precision | recall | f1score |
| 第一折 | 1.0 | 0.45 | 0.621 |
| 第二折 | 1 | 0.05 | 0.095 |
| 第三折 | 0.281 | 0.45 | 0.346 |
| 第四折 | 0.55 | 0.55 | 0.55 |
| 第五折 | 0.211 | 0.4 | 0.276 |
| 平均值 | 0.408 | 0.37 | 0.359 |

修改：总表中的precision,recall,fscore都是用的9类的均值，但是显然样本的数量差别很大，所以选择使用（数量\*各类的指标）/总数来求平均。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | loss | precision | recall | f1score | acc |
| 第一折 | 1.00847 | 0.909 | 0.900 | 0.896 | 0.978 |
| 第二折 | 0.71071 | 0.912 | 0.412 | 0.386 | 0.869 |
| 第三折 | 0.69121 | 0.862 | 0.790 | 0.810 | 0.953 |
| 第四折 | 0.44332 | 0.885 | 0.875 | 0.875 | 0.972 |
| 第五折 | 0.38534 | 0.883 | 0.862 | 0.868 | 0.969 |
| 平均值 | 0.64781 | 0.89 | 0.768 | 0.767 | 0.948 |

2CNN+2Bigru:

修改后的指标：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | loss | precision | recall | f1score | acc |
| 第一折 | 0.27383 | 0.950 | 0.914 | 0.928 | 0.981 |
| 第二折 | 0.27424 | 0.925 | 0.931 | 0.927 | 0.985 |
| 第三折 | 0.32414 | 0.918 | 0.790 | 0.917 | 0.982 |
| 第四折 | 0.29305 | 0.931 | 0.925 | 0.920 | 0.983 |
| 第五折 | 0.23949 | 0.871 | 0.882 | 0.870 | 0.973 |
| 平均值 | 0.28095 | 0.919 | 0.914 | 0.912 | 0.981 |

2.Autokeras

autokeras：开源，自动调参，寻找最优模型的工具。

支持的模型类别：用于分类的卷积神经网路（CNN）、循环神经网络（RNN）、长短期记忆网络（LSTM）

使用技术：高效神经架构搜索，贝叶斯优化