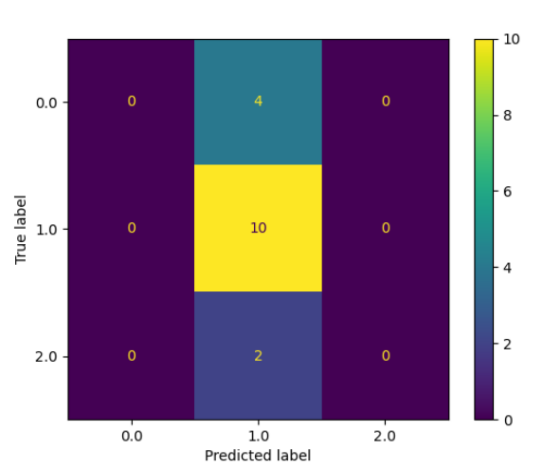
# 无OCTA相关数据的糖肾分类

**全部特征分类**

以AUC为评价标准，采用网格搜索法得到最佳参数。多次测试发现结果并不稳定，以下分析取最优结果。以151个病人为研究对象，取21个特征进行分类时，在糖肾训练集上准确率为0.64，在糖肾测试集准确率为 0.625。无糖肾、糖肾、二者合并，这三类的支持向量的数量分别是30、87、18。分类结果如下表：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 精确度 | 召回率 | F1分数 | 测试样本数量 |
| 无糖肾 | 0.00 | 0.00 | 0.00 | 4 |
| 糖肾 | 0.62 | 1.00 | 0.77 | 10 |
| 二者合并 | 0.00 | 0.00 | 0.00 | 2 |

混淆矩阵结果如下图：



采用不同的模型参数，进行10次特征选择排序，分数越低重要性越高，结果如下表：

|  |  |  |
| --- | --- | --- |
| 特征得分 | 特征名称 | 重要性排名 |
| 44 | HBA1C（糖化血红蛋白） | 6 |
| 204 | 空腹血糖GLU（静脉抽血葡萄糖） | 16 |
| 248 | AST谷草转氨酶 | 19 |
| 196 | TG甘油三酯 | 15 |
| 51 | CDEC低密度脂蛋白胆固醇 | 7 |
| 242 | ALT谷丙转氨酶 | 18 |
| 144 | 白蛋白ALB（溴甲酚绿法） | 12 |
| 70 | CREA肌酐(酶法） | 9 |
| 19 | TC总胆固醇 | 1 |
| 180 | HDEC高密度脂蛋白胆固醇 | 14 |
| 41 | 24h尿蛋白定量 | 5 |
| 24 | 转铁蛋白 | 3 |
| 99 | al-微球蛋白（晨尿) | 10 |
| 19 | 免疫球蛋白IgG（晨尿) | 1 |
| 21 | 微量白蛋白（晨尿) | 2 |
| 54 | 游离T3 | 8 |
| 25 | 游离T4 | 4 |
| 224 | 促甲状腺素 | 17 |
| 145 | 血红蛋白HGB | 13 |
| 105 | 收缩压 | 11 |
| 286 | 舒张压 | 20 |

只以最优结果的模型参数为例进行特征排序：

|  |  |
| --- | --- |
| 特征重要性 | 特征名称 |
| 1 | HBA1C（糖化血红蛋白） |
| 11 | 空腹血糖GLU（静脉抽血葡萄糖） |
| 17 | AST谷草转氨酶 |
| 13 | TG甘油三酯 |
| 15 | CDEC低密度脂蛋白胆固醇 |
| 16 | ALT谷丙转氨酶 |
| 14 | 白蛋白ALB（溴甲酚绿法） |
| 5 | CREA肌酐(酶法） |
| 1 | TC总胆固醇 |
| 10 | HDEC高密度脂蛋白胆固醇 |
| 4 | 24h尿蛋白定量 |
| 2 | 转铁蛋白 |
| 12 | al-微球蛋白（晨尿) |
| 1 | 免疫球蛋白IgG（晨尿) |
| 3 | 微量白蛋白（晨尿) |
| 1 | 游离T3 |
| 7 | 游离T4 |
| 9 | 促甲状腺素 |
| 8 | 血红蛋白HGB |
| 6 | 收缩压 |
| 18 | 舒张压 |

运用多因素方差分析的结果如下表：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| df | sum\_sq | mean\_sq | F | PR(>F) |  |  |  |
| A1 | 1.0 | 1.717774 | 1.717774 | 5.215757 | 0.024018 | 有显著差异 | HBA1C（糖化血红蛋白） |
| A2 | 1.0 | 0.011938 | 0.011938 | 0.036249 | 0.849302 | 无显著差异 |  |
| A3 | 1.0 | 0.689519 | 0.689519 | 2.093618 | 0.150341 | 无显著差异 |  |
| A4 | 1.0 | 0.452096 | 0.452096 | 1.372720 | 0.243504 | 无显著差异 |  |
| A5 | 1.0 | 0.615113 | 0.615113 | 1.867694 | 0.174117 | 无显著差异 |  |
| A6 | 1.0 | 0.026561 | 0.026561 | 0.080649 | 0.776875 | 无显著差异 |  |
| A7 | 1.0 | 0.609134 | 0.609134 | 1.849540 | 0.176209 | 无显著差异 |  |
| A8 | 1.0 | 0.252212 | 0.252212 | 0.765802 | 0.383145 | 无显著差异 |  |
| A9 | 1.0 | 0.286011 | 0.286011 | 0.868430 | 0.353131 | 无显著差异 |  |
| A10 | 1.0 | 0.016126 | 0.016126 | 0.048963 | 0.825228 | 无显著差异 |  |
| A11 | 1.0 | 0.631176 | 0.631176 | 1.916470 | 0.168636 | 无显著差异 |  |
| A12 | 1.0 | 2.027612 | 2.027612 | 6.156531 | 0.014379 | 有显著差异 | 转铁蛋白 |
| A13 | 1.0 | 0.003326 | 0.003326 | 0.010100 | 0.920103 | 无显著差异 |  |
| A14 | 1.0 | 0.790834 | 0.790834 | 2.401247 | 0.123688 | 无显著差异 |  |
| A15 | 1.0 | 0.053422 | 0.053422 | 0.162207 | 0.687800 | 无显著差异 |  |
| A16 | 1.0 | 1.378670 | 1.378670 | 4.186119 | 0.042788 | 有显著差异 | 游离T3 |
| A17 | 1.0 | 0.076930 | 0.076930 | 0.233586 | 0.629696 | 无显著差异 |  |
| A18 | 1.0 | 0.052791 | 0.052791 | 0.160292 | 0.689550 | 无显著差异 |  |
| A19 | 1.0 | 0.187638 | 0.187638 | 0.569733 | 0.451741 | 无显著差异 |  |
| A20 | 1.0 | 0.257531 | 0.257531 | 0.781953 | 0.378190 | 无显著差异 |  |
| A21 | 1.0 | 0.080293 | 0.080293 | 0.243799 | 0.622316 | 无显著差异 |  |

通过多因素方差分析结果看出HBA1C（糖化血红蛋白）、转铁蛋白、游离T3三个特征重要，对分类结果有显著影响。

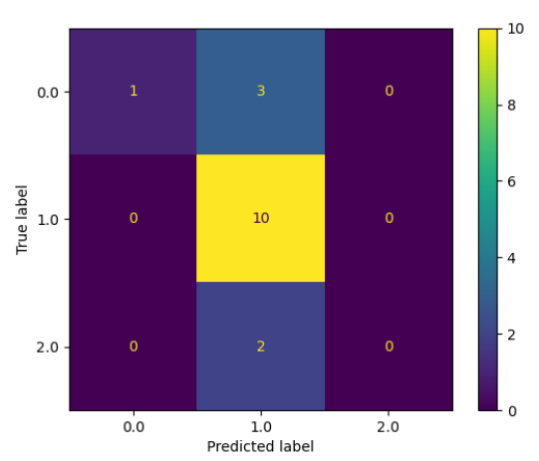
结合以上三个表大致可以看出HBA1C（糖化血红蛋白）、TC总胆固醇、免疫球蛋白IgG（晨尿)、游离T3、转铁蛋白、微量白蛋白（晨尿)等特征相对更重要。

**部分特征分类**

下面以上述6个重要特征为研究对象进行糖肾分类。同样以网格搜索法取最优参数，并以最好结果为例。在糖肾训练集上准确率为0.74，在糖肾测试集准确率为 0.69。得到的分类结果如下表：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 精确度 | 召回率 | F1分数 | 测试样本数量 |
| 无糖肾 | 1.00 | 0.25 | 0.40 | 4 |
| 糖肾 | 0.67 | 1.00 | 0.80 | 10 |
| 二者合并 | 0.00 | 0.00 | 0.00 | 2 |

混淆矩阵结果如下图：



**总结**

结合采用21个特征或者6个特征的分类结果发现，准确率在0.6到0.7之间。采用其他特征分类时准确率也一般在0.6到0.7之间，但是有的结果会更差。