1.给你一个癌症检测的数据集。你已经建好了分类模型，取得了96%的精度。为什么你还是不满意你的模型性能？你可以做些什么呢

答：如果你分析足够多的数据集，你应该可以判断出来癌症检测结果是不平衡数据。在不平衡数据集中，精度不应该被用来作为衡量模型的标准，因为96%可能只有正确预测多数分类，但我们感兴趣的是那些少数分类（4%），是那些被诊断出癌症的人。

因此，为了评价模型的性能，应该用灵敏度（真阳性率），特异性（真阴性率），F值用来确定这个分类器的“聪明”程度。如果在那4%的数据上表现不好，我们可以采取以下步骤：

1.我们可以使用欠采样、过采样或SMOTE让数据平衡。

2.我们可以通过概率验证和利用AUC-ROC曲线找到最佳阀值来调整预测阀值。

3.我们可以给分类分配权重，那样较少的分类获得较大的权重。

4.我们还可以使用异常检测。