# 基于片段的关于房颤的二分类的模型

## 不加任何特征进行训练的模型

### 代码和模型存放的位置

hupeng/2019-10-23/DenseNet\_Classification\_2\_Whether\_AF，模型都是以“.h5”结尾的文件，模型的文件命名以它的深度和验证集上的准确率在命名，dimension\_of\_2和dimension\_of\_1 分别代表的是输入的数据是采用一维的还是二维的，关于代码的具体解释说明已经写成了具体的README文档放在了此文件夹内部。

### 模型的测试结果(采用的输入数据是一维的)



## 加入特征(心博数、RR间期、RR间期差值)之后进行训练的结果

### 代码的和模型存放的位置

代码和模型都存放在hupeng/2019-10 -23/DenseNet\_Classification\_2\_Whether\_AF\_RR\_Interphase这个文件夹中，模型都是以“.h5”结尾的文件，文件的具体功能参照该文件夹内对应的README文档，Standardization表示该输入数据是采用的是均值方差归一化进行数据的预处理的过程的，而Normalization表示该输入数据是采用最值归一化的方式进行数据的预处理的

### 模型的测试结果

#### 使用最值归一化进行数据预处理的模型的结果



在加入心博数、RR间期、RR间期差值，再接着对每一条数据进行了最值归一化之后，效果并没有一个较好的提升，在此不建议使用均值方差归一化对数据进行数据的预处理工作。

#### 使用均值方差归一化进行数据预处理的模型的结果



在加入心博数、RR间期、RR间期的差值，再接着对输入数据进行了均值方差归一化，对比不加任何特征，在测试集上的准确率提升了约 1%。

# 基于波形分类的模型

## 2,1 利用4个二分类的模型来模拟四分类的模型

### 2.1.1 代码和模型存放的位置

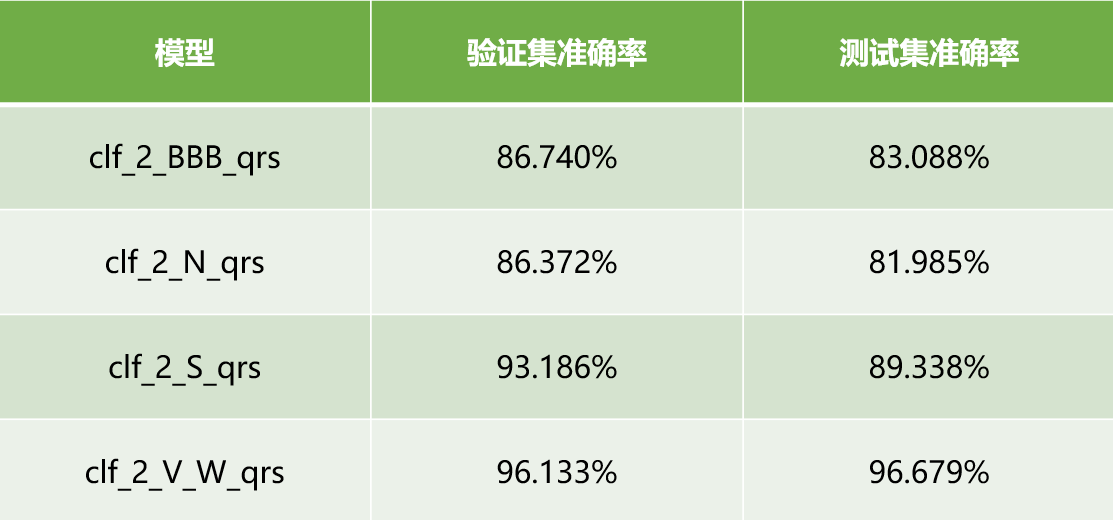
4个波形的类别分别为 BBB\_qrs(束支阻滞)、N\_qrs(窦性)、S\_qrs(房性)、V\_W(室性)这4个类别，4的代码的所有内容都存放在DenseNet\_Classification\_2\_Whether\_QRS这个文件夹中，在此文件夹中，每一个文件都以各自的波形名称作为自己的代码的模型的。每一个文件夹内都有自己的模型以及模型对应的训练代码。

### 2.1.2 思路流程

采用的是 OvR（即 One VS Rest，把其中的一类单独作为一类，剩下的几种类别作为一种类别）的方法进行的用4个二分类的模型去模拟四分类的模型，这样得出的结果，OvR的方法是速度快，但是准确率相比较 OvO方式(即One vs One)来说的话略低。如果是有N个类别的数据的话，采用OvR方式就应该训练出 N 个模型出来。

经过跟公司的交流讨论，现拟定了如下2种测试模型最终准确率的方案：

1. 串行测试：把一条数据放进4个二分类的模型中，把4个模型的结果全部输出，模型顺序为：室性、房性、束支阻滞、窦性（这个排列顺序是根据模型的正确率来进行排序的），如果是室性，就不放进下一个模型中了；如果不是，就放进下一个模型中；以此类推；
2. 并行测试：5个模型跑了之后取概率最大的为真
3. 只看单个模型的概率
4. 模型的准确度乘以预测出的概率得出最终的结果取最大值

2.1.3 模型的测试结果  


上图是 4 个二分类的模型的训练结果，从上至下依次是BBB\_qrs（束支阻滞）、N\_qrs(窦性)、S\_qrs(房性)、V\_W\_qrs（室性）在验证集和测试集上面的准确率

测出按照上述所示的 3 种方案得出的准确率如下图所示



### 2.1.3 结果分析

从上图中的结果中分析可知，在使用OvR（即One vs Rest）这种方式中，采用并行测试的第一种方案是得到的测试集上面的准确率是最高的，并行测试的第一种方案的思路就是通过把每一条数据放入这4 个模型之中，把4个模型的预测结果中判断为是的概率最大的那一种类别作为最终模型输出的结果的类别。

## 2.2 只用一个四分类的模型来做关于波形的四分类的模型

### 2.2.1 代码和模型存放的位置

关于单独使用一个四分类的模型来判断是否为BBB\_qrs(束支阻滞)、S\_qrs(房性)、N\_qrs（窦性）、V\_W\_qrs(室性)这4种类别，模型的训练代码和训练好的模型都放在了DenseNet\_Classification\_4\_Whether\_qrs这个文件夹中，文件夹内有各自对应的README.md文档，详细的说明了模型的训练使用方式。在这个模型中采用的是一维的数据输入。

### 2.2.2 模型的测试结果



上图为进行了训练之后的几个四分类的判断波形的训练结果，从上图中训练结果中可以看出，均衡测试集上面的准确度和批量预测时间来说，推荐使用上图中的clf\_4\_depth\_16这一个模型

### 结果分析

从上图中的结果对比采用4个二分类的模型来模拟四分类的模型来说，准确率还是直接采用一个四分类的模型来做更高一点，这一点的原因有可能如下：在用4个二分类的模型来模拟四分类的模型中，我们采用的方式是OvR(这种方式的优点是效率高，但是就准确率来说，这种方式并不如 OvO方式高)

## 除去BBB\_qrs之后的三分类的模型

### 2.3.1 代码和模型存放的位置

由于之前无论是使用4个二分类的模型来做模拟四分类的模型还是做的只是一个单独的四分类的模型做的关于波形的分类的模型的准确率都上不去，所以打算去掉其中数据量最少的BBB\_qrs(束支阻滞波形)，只做剩下的关于N\_qrs(窦性)、S\_qrs(房性)、V\_W\_qrs(室性)这3种类的波形的模型的训练结果，模型和代码存放在DenseNet\_Classification\_3\_Whether\_QRS这个文件夹中

### 2.3.2 模型测试的结果



### 2.3.4 测试结果分析

从上图中的测试结果中可以看出，在我们将 BBB\_qrs(束支阻滞波形)去掉了之后，在测试集上面的准确率可以看到有一个显著的提高，产生这种现象的一个原因是BBB\_qrs(束支阻滞)这种波形的数据量较少，深度学习模型无法学习到充分的参数。

# 测试代码

## 3.1 测试准确率、精准率、召回率、混淆矩阵、F1分数

### 3.1.1 代码存放的位置

关于准确率、精准率、召回率、混淆矩阵、F1分数的代码都放在hupeng/2019-10-23/Test\_Code/Test\_Accuracy\_Code这个文件夹中，关于测试时间的代码存放在hupeng/2019-10-23/Test\_Running\_Time这个文件夹中，代码的使用的说明文档都写了对应的README.md文档放在了此文件夹中。

### 3.1.2 方法上面的改进

在之前我们一直使用的是逐条数据逐条预测的方式，这种方式导致的一个结果时就是数据的预测速度较低，无法适应一个比较庞大的数据量，比如一个人7天的心电数据量，当数据量较大时，这时的预测速度就会变得较慢，逐条数据预测的方式是读取一条，预测一条，再接着读取下一条，再预测下一条。我们后来是采用的是批量预测的方式进行的数据的预测，这样做的好处在于我们现在是一次性读取所有的数据，再一次性对所有的数据进行预测，在时间上面就会有较大的提升。

### 3.1.3 测试结果及其分析



上图为四分类的模型的测试预测时间的结果，从上图中可以看出，我们在采用了批量预测时候，在测试时间上面有了一个质的飞跃。

# 波形识别

## 4.1 代码存放的位置

所有的代码都存放在GeneFeng这个文件夹中，注释和README.md也放在了对应的文件夹内。

## 4.2 代码的测试结果



上面是第一版的测试结果



上面是第三版的测试结果

# 展望

## 5.1 关于二分类的基于片段的模型

此模型的泛化能力仍然有待提高，由于公司更新了一批新的数据集，AF(房颤)的数据由原来的286条到了现在的786条，因此可以在原来的基础上进行新模型的训练，之前由于数据量较少，导致我们的模型有一些参数还没有学习到，深度学习发挥效用的一个前提条件就是数据量要具有一定的规模

## 5.2 关于基于波形的判断是否是四分类的模型

### 5.2.1 关于数据预处理方面

经过我们之前的实验的结果可知，使用均值方差归一化比使用最值归一化的效果要来得的好，另外我们是可以在原始数据的基础上加入新的特征组成新的训练数据再输入到模型之中进行训练，这样得出的结果应该会更好一点

### 5.2.2 关于用4个二分类的模型来模拟四分类的模型

在我们之前尝试的方式中，是采用的OvR（即One vs Rest）方式，即把其中的一类作为一类，把所有剩下的类别作为一类进行的二分类的模型的训练，这种方式的好处就是耗时较短，但存在的一些小问题就是准确率会偏低一点。后期其实可以采用 OvO(即 One vs One)方式，即选其中的两类作为两类，有多少个这样的组合就做多少个这样的组合，这样做的好处是准确率会上升，但是速度上可能会有所下降，后期其实可以尝试这样的

其次就是加入一些特征与原数据组成新的训练数据，进行模型的训练。