

# 利用光电容积脉搏波描记法预测血压

## 一. 任务背景

血压是高血压的直接指标，是多种心血管疾病（CVD）的一个重要危险因素，根据世界卫生组织（WHO）的数据，心血管疾病是最常见的死亡原因之一。因此，定期血压监测对一般人群很重要，但对已经患有高血压或相关疾病的人尤其重要，因为这些人特别容易受到血压升高的影响。家用血压计是袖带的，使用的是听诊法。就是将压力计（一般称血压计）的臂带绑扎于上臂肱动脉搏动位置，充气加压将肱动脉压瘪，然后再放气减压。听诊法只能间断测量，需要时间和精力。

我们需要一种可以连续又方便的血压监控手段。这里我们考虑光电容积脉搏波描记法（英语：Photoplethysmography, PPG）。光电容积脉搏波描记法是一种相当简单和廉价的技术，在测量心率和血氧饱和度方面得到了广泛的医学应用。它通常使用发光二极管（LED）照亮皮肤，然后使用光电探测器（光电二极管）测量穿过皮肤或从皮肤反射的光的量。如果 PPG 用于测量血流速度，即通常所说的脉搏波速度，那么 PPG 和血压之间的关系就容易建立了。当血管变得更硬或更收缩时，血液流动更快，产生更大的压力。同样地，当血管更放松或更具弹性时，血液的流动会更慢，产生的压力也会更小。大量的临床实测结果证实，脉搏波的波形特征与心血管疾病有着密切的关系。脉搏波所表现出的形态（波的形状）、强度（波的幅值）、速率（波的速度）与节律（波的周期）等方面的综合信息的确在相当程度上反映出人体心血管系统的许多生理和病理特征。光电容积脉搏波描记法并不需要复杂而昂贵的仪器设备，且操作简单、性能稳定，具有无创伤和适应性强等诸多优点。

## 二. 问题定义

我们可以基于光电容积脉搏波提取血压特征参量，然后结合性别年龄身高体重等特征对血压进行预测。这里使用的数据是相关专业人员已经提取好的特征参量。当然我们也可以使用卷积神经网络来处理 PPG 信号，目标是对血压值进行预测。

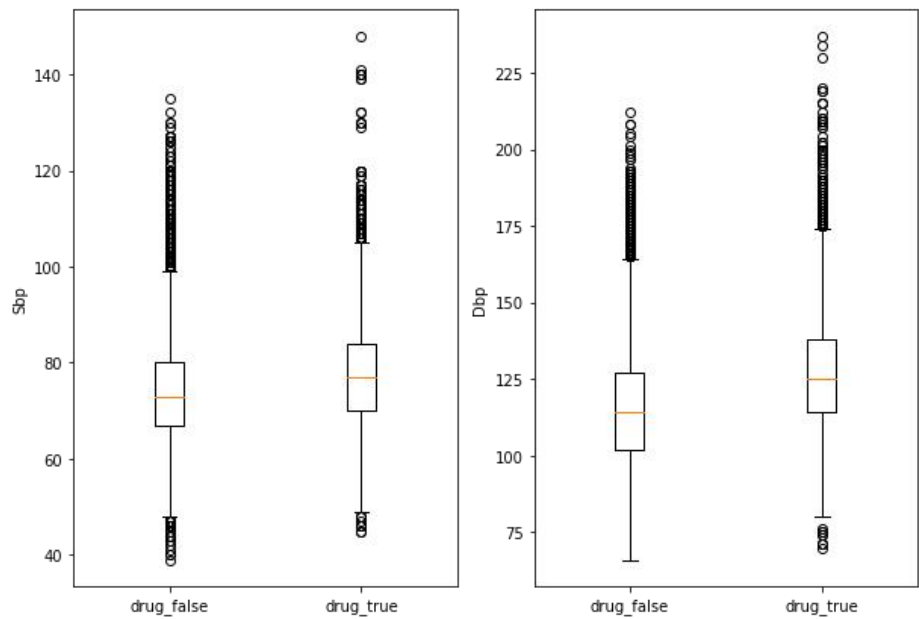
## 三. 数据描述

使用的数据有 57993 条，分别来自 4972 人。需要预测的是收缩压和舒张压，用来预测的特征是性别、年龄、身高、体重、是否使用降压药、人的 id 以及专

业人士提取好的关于来自 PPG 信号的 45 个特征。

|   | Gender | Age  | Height | Weight | Drug |
|---|--------|------|--------|--------|------|
| 0 | 1      | 58.0 | 158    | 55.0   | 0    |
| 1 | 1      | 58.0 | 158    | 55.0   | 0    |
| 2 | 1      | 58.0 | 158    | 55.0   | 1    |
| 3 | 1      | 62.0 | 158    | 60.0   | 0    |
| 4 | 1      | 62.0 | 158    | 60.0   | 0    |

做出服用降压药和不服用降压药的箱线图，显然服用降压药的血压药的样本收缩压和舒张压高。因为是高血压患者服用降压药。由于特征较多，此处不再一一画箱线图



#### 四. 建立模型与评估

我们使用最简单的线性回归分析。

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, \dots, n$$

由于我们的目标是预测血压值，所以使用均方误差的平方根来评估预测效果

$$\text{RMSD} = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}.$$

Y 是收缩压和舒张压，X 考虑四种选择，一种只用 PPG 信号，另一种使用所有变量，其中人的 ID 需要先进行 One-Hot 编码。第 3 种使用除了人的 ID 以外的变量。最后是仅使用性别、年龄、身高、体重、是否使用降压药（即不使用 PPG 信号和人的 ID）。各自的测试均方误差的平方根如下表所示。

|     | 仅用 PPG<br>信号 | 使用所<br>有变量 | 使用除<br>了人的 ID 以<br>外的变量 | 不 使用<br>PPG 信号和人<br>的 ID |
|-----|--------------|------------|-------------------------|--------------------------|
| Sbp | 14.58        | 7.18       | 13.74                   | 16.80                    |
| Dbp | 8.96         | 4.34       | 8.42                    | 10.32                    |

而 Sbp 的样本标准差为 17.67，Dbp 的样本标准差为 11.22。

由于我们的目标是预测血压值，所以我们关注均方误差的平方根。从表中可以看到，仅使用性别、年龄、身高、体重、是否使用降压药的预测效果特别差，测试均方误差的平方根和 Sbp Dbp 的样本标准差十分相近。这说明使用线性模型时，我们无法从性别、年龄、身高、体重、是否使用降压药中获取多少对预测血压有价值的信息。仅用 PPG 信号的预测效果不够好，测试均方误差的平方根和 Sbp Dbp 的样本标准差相近，但是还是具有一定的预测效果。加上性别、年龄、身高、体重、是否使用降压药这些变量后效果提升不明显。使用所有变量能得到相对满意的结果。我们可以看出使用人的 id 的信息对预测效果提升明显。这比较好理解，因为每个人的血压有个基准，方差相对总体分布而言更小。利用人的 id 这一信息相当于利用了测试样本以往的血压值这个信息。

所以使用 PPG 信号对于血压有一定的预测效果，但是结果不够好。测试样本以往的血压值这个信息对预测效果来说非常重要。虽然在有此信息后预测误差比较小，但是更有实际应用价值的情况可能是并不知道待预测的样本以往的血压值，只给定训练样本的 PPG 信号特征和性别、年龄、身高、体重、是否使用降压药等变量，然后对病人的血压进行预测。

继续探索的话可考虑更复杂的模型，如果效果仍旧不理想那么可能存在两种情况。一种是我们提取的 PPG 信号特征包含的信息不够多，可考虑重新提取。另一种情况是 PPG 信号本身提供的信息就不够多，所以无法帮助我们获得非常满意的预测结果。