**结果：**

表格 1. 数据片段的预测混淆矩阵

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | Reference | | acu | TP | TN |
| AH | N |
| Estimated | AH | 25919 | 21469 | 0.878286 | 0.747678 | 0.899484 |
| N | 8747 | 192119 |  |  |  |

训练集一共包含13个被试合计超过34个小时的数据，总共包含248254个10秒长的数据片段。使用级联分类器对上述数据的预测结果的混淆矩阵如表格1所示。级联分类器的预测结果与多导睡眠图的人工标注结果对比显示对于上述248254个数据片段，级联分类器实现了87.8%的正确率，74.8%的正阳率以及89.9%的正阴率。

表格 2. 级联分类器对于事件的分类结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 虚警事件 | 检出事件 | 事件总数 | 精准率 | 召回率 |
| 569 | 1510 | 1786 | 0.726311 | 0.845465 |

表格2中统计了使用级联分类器定位出的事件结果与多导睡眠图人工标注的事件结果的对比。对于测试集中包含的13个被试人工标注出的总共1786次睡眠呼吸暂停事件，级联分类器模型检出了1510次事件，实现了84.5%的检出率，同时伴随着569次虚警事件，也就是72.6%的精准率。

(b)

(a)



图 1. (a) AHIest与AHIrefer三点关系图 (b) AHIest与AHIrefer的Bland-Altman图

图1(a)展示的是根据级联分类器预测出的AHIest与根据多导睡眠图人工标注出的AHIrefer的散点关系图。根据最小二乘法拟合出来的实线显示出了AHIest与AHIrefer之间明显的线性关系（皮尔逊相关指数0.98）。图1(b)展示的是AHIest与AHIrefer之间的Bland-Altman图，AHIest与AHIrefer的平均误差为-2.10次/小时，在95%的置信区间内误差范围为-5.65至1.46次每小时。

表格 3 级联分类器对SAHS严重程度预测结果

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 人工判断SAHS严重程度 | | | | |  | AHI cutoff(events/h) | | | |
|  |  | 正常 | 轻度 | 中度 | 重度 | 合计 |  | ≥5 | ≥15 | ≥30 | 平均值 |
| 模型预测严重程度 | 正常 | 0 | 0 | 0 | 0 | 0 | 正阳率(%) | 100.0 | 100.0 | 100.0 | 100.0 |
| 轻度 | 0 | 3 | 0 | 0 | 3 | 正阴率(%) | 100.0 | 75.0 | 75.0 | 83.3 |
| 中度 | 0 | 1 | 2 | 0 | 3 | 精准率(%) | 100.0 | 90.0 | 71.4 | 87.1 |
| 重度 | 0 | 0 | 2 | 5 | 7 | 正确率(%) | 100.0 | 92.3 | 84.6 | 92.3 |
| 合计 | 0 | 4 | 4 | 5 | 13 |  |  |  |  |  |

表格3总结了级联分类器对SAHS严重程度的预测结果，根据以往研究经验SAHS严重程度的阈值分别设置为5，15，30次/小时[1-5]，而AHI=5次/小时通常用来判断被试是否为SAHS阳性患者[6]，对于正常、轻度、中度与重度四种SAHS患者级联分类器模型达到的平均正阳率、正阴率、精准率和正确率为100.0，83.3，87.1，92.3。

图 2. (a) 患有轻度睡眠呼吸暂停综合征的患者的预测结果。(b) 患有中度睡眠呼吸暂停综合征的患者的预测结果。(c) 患有重度睡眠呼吸暂停综合征的患者的预测结果。



(b)

(c)

(a)

图2显示的分别是对一个轻度睡眠呼吸暂停综合征的患者、中度睡眠呼吸暂停综合征的患者、重度睡眠呼吸暂停综合征的患者的级联分类器模型事件预测结果与人工标注事件结果的对比。对于该轻度患者，级联分类器模型对于事件的预测正阳率与精准率分别达到了76.7%与78.0%，对于中度患者，模型的正阳率与精准率分别是89.8%与80.3%，对于该重度患者，模型的正阳率与精准率分别达到89.3%与86.8%。

**讨论：**

在这篇文章中，我们提出了一种级联分类器的模型可以实时精准地预测睡眠呼吸暂停与低通气事件。与传统的基于睡眠多导图的人工检测方法相比，级联分类器只使用口鼻流量通道信号与血氧通道信号即可对AH事件进行预测，并且通过事件检测器可以计算出AHI指数进而对睡眠呼吸暂停综合征的严重程度做出判断。

该种检测方法主要通过以下几个步骤对AH事件做出实时预测：1））将血氧通道信号主动提前23s钟的时间。2）对口鼻流量信号进行滤波处理，其中首先使用截止频率为0.05赫兹的三阶的巴特沃斯高通滤波器去除原始信号中由于被试移动而产生的极限漂移（图3.（a）），然后进行移动平滑滤波以消除原始信号中存在的由于被试触碰传感器而产生的高频干扰（图3.（b））。3）使用10s长度步长1s的滑动窗口将原始数据切割为10s的数据片段。4）从切割得到的数据片段中提取特征集合。5）使用级联分类器模型对数据片段进行预测6）使用事件检测器得到最终的AH事件预测结果与AHI指数计算。血氧通道信号往往在AH事件发生之后一段时间内才会发生欠饱和，为了更好的捕捉到血氧欠饱和这一特征在处理过程中对血氧通道信号进行主动提前，经过测试发现将血氧提前23s钟的时候得到的结果最好（图3）。



（a）

（b）

图 3. （a）主动对齐之前的流量通道信号与血氧通道信号；（b）主动对齐之后的流量通道信号与血氧通道信号

AH event

另一方面经过滤波处理之后的口鼻流量信号会更加近似于正弦信号，这会降低后续特征提取阶段的难度并且提高特征的准确度。如果不经过滤波处理，这些伴有外界噪声的数据片段可能会导致级联分类器做出错误分类。作为对比，我们分别将经过主动对齐与滤波处理得到的特征集合与未经过任何处理得到的特征集合输入到级联分类器中进行预测，未经过任何处理的特征集合只能达到85.5%的正确率与67.7%的正阳率，而经过主动对齐与滤波处理之后的特征集合可以达到87.8%的正确率与74.8%的正阳率。



（a）

（b）

图 4. 滤波前的原始信号与滤波后的信号

表格1所示为级联分类器得到的数据片段的预测结果与基于多导睡眠图人工标注的结果的混淆矩阵。级联分类器达到了74.8%的正阳率与89.9%的正阴率。需要注意的是级联分类器所犯的假阳性错误要多于假阴性错误数目，其中有一部分级联分类器预测中的假阳性错误符合2012年AASM[1]所推荐的睡眠呼吸暂停与低通气事件的定义，应该标注为AH事件，但是原始数据库中的人工标注并没有对这些事件进行标注（图4（a），（c），（e））。另一部分假阳性错误主要出现在阳性数据片段与阴性数据片段的边界处。而对于一部分数据库中标注的AH事件并没有达到2012年AASM所推荐的事件定义（图4（b），（d），（f）），所以这些数据片段被级联分类器错误预测成了阴性片段。

表格4所示为级联分类器估计AHI与人工标注Ahi的对比，因为级联分类器更倾向于犯假阳性错误，所以预测出的AHI指数通常高于人工标注的AHI指数。但是如表格3所示，这些错误并不会影响最终对于SAHS严重程度的判断。级联分类器对于SAHS严重程度的诊断结果的平均KAPPA系数达到0.83，意味着该种方法可以用来筛查SAHS的严重程度。

图 5. （a），（c），（e）为级联分类器预测结果中的假阳性错误；（b），（d），（f）为级联分类器预测结果中的假阴性错误



（a）

（b）

（c）

（d）

（e）

（f）

表格 4. 级联分类器模型估计的AHI与数据库人工标注AHI对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | No. of AH events | | AHI(events/h) | |
| Reference | Estimated | Reference | Estimated |
| Non-SAHS |  |  |  |  |
| Mild SAHS | 73.5±17.1 | 89.5±27.5 | 12.4±1.8 | 14.9±2.7 |
| Moderate SAHS | 133.2±23.3 | 160.0±38.3 | 23.0±3.6 | 27.6±6.6 |
| Severe SAHS | 191.8±70.5 | 216.2±71.0 | 43.8±16.3 | 40.7±14.0 |

表格5所示为与其他相关研究结果的对比，其他研究结果使用的大多是医院的内部数据库，可以看到我们的方法取得了与其他相关研究相当甚至更好的结果。Baile Xie等人也使用了UCD数据库，但是她们的研究结果并没有对SAHS严重程度的计算，这里只对数据片段的预测结果进行对比，可以看到我们的方法具有更高的正确率。

表格 5. 与现有的其他相关研究结果对比

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Related work | Method | Signal | AHI cutoff | acc | sen | spe |
| Our study | 级联随机森林 | Nasal flow and SpO2 | 5 | 100.0 | 100.0 | 100.0 |
| 15 | 92.3 | 100.0 | 75.0 |
| 30 | 84.6 | 100.0 | 75.0 |
| San Ho Choi et al. | Convolutional neural networks | Nasal pressure | 5 | 96.2 | 100.0 | 84.6 |
| 15 | 92.3 | 98.1 | 86.5 |
| 30 | 96.2 | 96.2 | 96.2 |
| Da Woon Jung et al. | Regression Model | SpO2 | 5 | 97.8 | 98.6 | 94.4 |
| 10 | 96.7 | 98.4 | 92.9 |
| 15 | 95.7 | 96.4 | 94.6 |
| 30 | 96.7 | 97.1 | 96.5 |
| Gonzalo C et al. | AdaBoost-Linear discriminant analysis | Nasal flow | 5 | 86.5 | 87.1 | 80.0 |
| 15 | 81.0 | 85.9 | 72.9 |
| 30 | 82.5 | 74.2 | 90.6 |

根据图1（a）可以看出我们的方法计算出的AHI指数与人工标注的AHI指数具有极高的相关性，并且在不同的被试之中显示出良好的一致性；在图1（b）中可以看到AHIest与AHIrefer的平均误差为-2.10次/小时，意味着我们的方法更倾向于对AHI指数估计过高，这是因为缺失了其他通道的信号所造成的。

我们也测试了级联分类器的训练与测试速度，使用合计超过24万的数据片段对级联分类器进行训练需要花费13.92秒钟的时间，而级联分类器对246970个数据片段的分类需要花费9.10秒钟的时间，也就是大概0.000037秒的时间可以分类出一个数据片段，说明我们的方法无论是训练时间还是响应速度都足以满足实时检测的需求。

与其他相关研究对比，我们的方法一方面可以预测出SAHS的严重程度，另外一方面可以定位到每一次SAHS事件的起止点。之前的相关研究有相当一部分只能做到对SAHS严重程度的判断，而不能够提供更多的具体SAHS事件信息的参考，这有可能在临床诊断中带来不便。在后来的研究中，有一部分学者开始对SAHS事件进行实时检测，如Da Woon Jung等人使用卷积神经网络的方法对SAHS事件进行实时检测，但是需要注意的是一方面卷积神经网路所要求的计算量巨大，所需要的训练时间要远远大于其他方法，可能会在临床检测的过程中带来不便；另外一方面卷积神经网络的可解释性不强，不能够提供更多的临床检测信息。而我们使用级联随机森林的分类方法对含有病理信息的特征集合进行训练与分类为整个检测过程提供了更好的可解释性，同时基于决策树的随机森林的分类方法对于计算量的需求大大降低，简化了整个检测过程的复杂性。

与此同时我们的研究中仍然存在着一些缺陷。首先对于血氧通道信号主动提前的时间我们是简单的根据结果而定的，对于这一问题应该蕴含着更多的研究空间；其次我们希望能够在未来能够进行在线测试以检验我们方法的实用性。

**2 材料和方法**

**2.1 数据库**

本文选取的数据库为Physionet上的公开数据库圣文森特大学医院/都柏林大学学院睡眠呼吸暂停综合征数据库（UCDDB），数据库中包含25名疑似患有睡眠呼吸障碍的被试的整夜睡眠多导图数据。其中包括脑电信号（C3-A2，C4-A1），左右眼电信号，颏下的肌电信号，使用热敏电阻测量的口鼻流量信号，使用应变片测量的胸部与腹部位移信号，使用手指脉搏血氧计记录的氧饱和度信号，使用麦克风记录的鼾声信号以及身体姿势信息共计十四个通道的信号。所有通道的信号都是使用德国的Jaeger-Toennies system进行记录的。包括睡眠分期与SAHS事件在内的数据库的标注文件是由一名经验丰富的睡眠技术专家进行标定的，其中包括每一次SAHS事件的发生与持续时间、类型等信息。根据以往研究经验SAHS严重程度的阈值分别设置为5，15，30次/小时[1-5]，而AHI=5次/小时通常用来判断被试是否为SAHS阳性患者[6]，我们选取了4名轻度SAHS患者、4名中度SAHS患者、5名重度SAHS患者合计13名患者来进行后续的训练与测试，选取的标准为。选取的被试的人口统计学与睡眠相关参数如表格6所示。

表格 6. 不同严重程度SAHS睡眠相关参数汇总（平均值±标准差）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Non-SAHS | Mild SAHS | Moderate SAHS | Severe SAHS |
| Age(years) |  | 49.00±1.63 | 57.50±7.19 | 46.60±5.46 |
| BMI(kg/m2) |  | 28.45±2.72 | 31.45±2.25 | 34.36±6.22 |
| AHI(events/h) |  | 12.40±1.81 | 25.59±3.58 | 43.75±16.34 |
| Study Duration(hours) |  | 7.03±0.17 | 6.63±0.43 | 6.84±0.82 |
| Epworth Sleepiness Score |  | 14.50±3.87 | 9.25±6.24 | 12.4±7.89 |

根据2012年AASM手册，睡眠呼吸暂停事件的定义为使用口鼻热敏传感器或者正压通气设备测得的流量信号峰值相对于上一次事件的偏移超过90%并且持续超过10s钟；睡眠呼吸低通气事件的定义为使用口鼻压力传感器测得的压力信号或者正压通气设备测得的流量信号峰值相对于上一次事件下降超过30%并且伴随着超过3%的血氧欠饱和或者觉醒且持续超过10s钟。所以一方面我们为了保证检测的准确度，另一方面为了简化信号采集与处理过程，选取了口鼻流量通道信号与血氧通道信号检测SAHS事件。

**2.2 方法设计**

级联分类器模型检测SAHS方法包含以下几个步骤：1）对原始信号进行滤波、分段等必要的预处理；2）对切割后的数据片段提取指定的特征集合；3）将特征集合输入到级联分类器中输出对于数据片段的预测结果（N或AH）；4）经过事件检测器计算出AHI指数对SAHS严重程度作出判断。

**2.3 信号预处理**

信号预处理阶段主要包括以下几个步骤：1）信号降采样。数据库中的原始信号采样率为128Hz，而成年人正常呼吸频率在16-20次每分钟，也就是0.27-0.33Hz，所以为了不影响检测精度同时降低检测过程计算量我们将两个通道的信号降采样到8Hz；2）随后为了去除原始信号中因为被试触碰或者移动造成的基线漂移与高频噪声干扰，我们使用了一个滑动平均滤波器与一个3阶巴特沃斯高通滤波器对原始口鼻流量信号进行滤波处理（图4），滑动平均滤波的过程如公式（1）所示，高通滤波器的截止频率设置在0.05Hz。

（1）

正常人的血氧饱和度水平在98%左右，而我们注意到数据库中的血氧通道信号存在过低的情况，这可能是由于传感器接触不良造成的，在预处理过程中我们将血氧低于80%的信号段切除，该部分数据大约占到了总数据的5.77%。3）滑动窗口切割。这里我们分别使用60s长度窗口与10秒长度窗口对原始数据进行切割，步长设置为1秒钟，切割后得到的数据片段标注为两种类型：AH与N。如果一个数据片段中包含有5秒钟以上的SAHS事件则标记为AH，其余的数据片段则标注为N。这样原始数据将被切割成为一段带有AH与N标记的数据片段序列**（图）**。最终我们将得到两组数据片段序列，一组60秒数据片段构成的序列与一组10秒数据片段构成的序列。4）血氧信号主动对齐。考虑到血氧信号总是在SAHS事件发生后一段时间才出现欠饱和的情况，所以为了更好地捕捉到血氧信号的特征我们将血样通道信号提前了23秒钟的时间（图3）。经过预处理之后我们总共获得了246970个数据片段，其中包括34666个AH片段与213588个N片段。

**2.4 特征提取**

**1）流量特征集合**

首先，我们提取出口鼻流量信号中的极值点作为呼气相与吸气相的转换点**（图）**，这样做的目的是可以更好的捕捉到被试每一次呼吸的潮气量变化。相邻的两个转换点之间的差值即可作为本次呼吸的潮气量的一个反映。处理完之后的流量信号转换为一系列的转换点序列（公式(2)）。

(2)

首先我们计算出每个数据片段中的平均潮气量、潮气量标准差、潮气量极差作为该数据片段的特征，随后我们每隔30秒钟的时间计算一次潮气量的基线，计算过程如公式3所示。

(3)

然后我们分别计算了每个数据片段中的呼吸暂停次数、呼吸低通气次数与正常呼吸次数，计算过程分别如公式4，5，6所示。

(4)

(5)

(6)

随后我们还考虑了流量信号的频域特征，正常人的呼吸频率在16-20次每分钟，而SAHS患者在受到SAHS事件影响之后呼吸频率会发生紊乱，所以我们计算了流量信号的频谱在0-4Hz频率段的频率谱峰度作为另外一个特征。

**2）血氧特征集合**

首先我们计算了典型的统计学特征例如血氧片段的标准差、极差，同时计算了数据片段内血氧信号的总体变化斜率作为血氧信号的一个变化特征。除此之外我们还计算了数据片段内血氧信号低于%92与%91水平的持续时间作为另外两个血氧特征。同时我们同样每个30秒钟的时间计算一次血氧的基线，计算过程如公式7、8所示。

(7)

(8)

随后我们分别计算了每个数据片段中的血氧欠饱和时长与血氧下降水平，计算过程分别如公式9-12所示。

(9)

(10)

(11)

(12)

最后，整个特征集合如表格7所示。

表格 7. 特征集合与特征定义

|  |  |  |
| --- | --- | --- |
| 特征编号 | 特征名称 | 特征定义 |
| 1 |  | 平均潮气量 |
| 2 |  | 潮气量标准差 |
| 3 |  | 潮气量极差 |
| 4 |  | 呼吸低通气次数、呼吸低通气次数与总呼吸次数所占比重 |
| 5 |  | 呼吸暂停次数、呼吸暂停次数与总呼吸次数所占比重 |
| 6 |  | 正常呼吸次数、正常呼吸次数与总呼吸次数所占比重 |
| 7 |  | 呼吸流量信号在感兴趣频段内的频谱峰度 |
| 8 |  | 血氧信号标准差 |
| 9 |  | 血氧信号极差 |
| 10 |  | 血氧欠饱和时长 |
| 11 |  | 血氧下降水平 |
| 12 |  | 血氧低于92、91水平的持续时长 |

**2.5 级联分类器的设计与训练**

采用滑动窗口切割信号来进行事件识别的方法有两个地方需要注意，一方面使用60秒或者更长的窗口切割原始信号并且对SAHS事件进行预测会存在分辨率不足的情况（图6（a））；另外一方面如果使用5秒甚至更短的窗口切割原始信号则不能够捕捉到相关的信号特征（图6（b）），所以综合以上两点考虑，我们决定使用一种级联分类器的设计方法解决这个问题。

级联分类器由两部分组成，第一部分我们将其称作是粗筛分类器，是由10棵CART决策树构成的随机森林，使用从60秒数据片段中提取得到的特征集合进行训练，可以筛除大部分与SAHS事件无关的正常数据片段；第二部分我们将其称作是精筛分类器，是由20棵CART决策树构成的随机森林，使用从10秒数据片段中提取得到的特征集合进行训练，可以在粗筛分类器的预测结果基础上进一步定位到SAHS事件的起止点，输出10秒数据片段的预测结果序列（图6（c））。

为了克服原始数据集中存在的不平衡的情况，我们修改了CART决策树对于不同类别样本的类权重，类权重的调整过程如公式13所示。

(13)

通过调整类权重使得分类器更加注重假阴性错误，可以帮助我们筛除掉大部分并不关心的正常数据片段而保留下绝大多数的SAHS数据片段。另外我们只使用4、8、9、10号特征构成的特征集合对粗筛分类器进行训练，这样做的目的是提高分类器的训练速度，同时对最终的预测结果几乎没有影响。为了防止过拟合我们限制每棵决策树的叶子节点上至少留下50个样本，分割内部结点所需的最小样本数50个同时最大深度不超过30。

在这里我们采用的是基于被试的两折交叉训练，也就是每次使用单个被试一半的数据作为训练集进行训练，余下一半的数据作为测试集；对整个数据库中的被试循环此过程直到得到所有被试的测试结果。

图 6. （a）采用长窗口切割数据并进行识别；（b）采用短窗口切割数据并进行识别；（c）级联分类器结构与预测过程



（a）

（b）

（c）

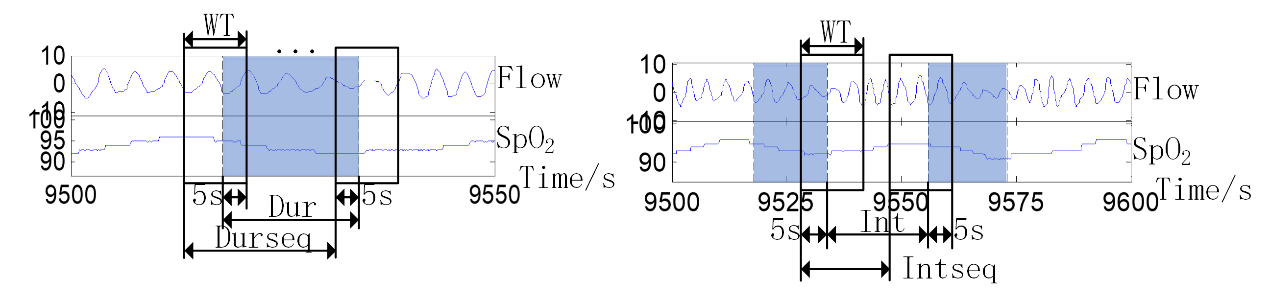
**2.6 事件检测器设计**

经过级联分类器后我们会得到10秒数据片段的预测结果序列，事件检测器会修正序列中的无效结果，输出最终的SAHS事件预测结果以及AHI。事件检测器的修正过程主要依赖以下两个标准：1）根据睡眠呼吸暂停与低通气事件的定义，一次SAHS事件的最短持续时长为10s，这对于预测结果中的SAHS序列长度有一定的约束（图7（a）），可以通过**公式**进行计算。

（14）

公式中Dur表示数据库标注中单次睡眠呼吸暂停与低通气事件的持续时长，满足Dur>10，WT表示所用滑动窗口长度，可以计算出预测结果中SAHS序列长度Durseq>WT。预测结果中不满足该标准的SAHS序列将被事件检测器重置为阴性序列。2）预测结果中两次相邻SAHS序列的间隔Intseq同样会存在一定的约束（图7（b））。容易计算得到*interval*>5，所以不满足该标准的N序列将被事件检测器重置为阳性序列。

图 7. （a）事件检测器修正假阳性事件标准；（b）事件检测器修正假阴性事件标准



（a）

（b）

经过级联分类器与事件检测器我们可以得到事件的总次数与每次事件的起点与终点，在评价过程中如果预测出的事件与人工标注中的事件存在重合则认为这次事件预测正确，这样认为的原因在于我们关心的重点是该种方法能不能定位到具体某次事件的发生位置，预测结果中定位到的事件能不能涵盖人工标注中的事件，所以具体到某一次事件的起止点的微小偏移认为是可以忽略的。

[1] R. Berry *et al.*, "Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine," (in eng), *J Clin Sleep Med,* vol. 8, no. 5, pp. 597-619, 2012.

[2] D. W. Jung *et al.*, "Real-Time Automatic Apneic Event Detection Using Nocturnal Pulse Oximetry," (in English), *Ieee Transactions on Biomedical Engineering,* Article vol. 65, no. 3, pp. 706-712, Mar 2018.

[3] G. C. Gutierrez-Tobal, D. Alvarez, F. del Campo, and R. Hornero, "Utility of AdaBoost to Detect Sleep Apnea-Hypopnea Syndrome From Single-Channel Airflow," (in English), *Ieee Transactions on Biomedical Engineering,* Article vol. 63, no. 3, pp. 636-646, Mar 2016.

[4] S. H. Choi *et al.*, "Real-time apnea-hypopnea event detection during sleep by convolutional neural networks," (in English), *Computers in Biology and Medicine,* Article vol. 100, pp. 123-131, Sep 2018.

[5] A. Qureshi, R. D. Ballard, and H. S. Nelson, "Obstructive sleep apnea," *Journal of Allergy and Clinical Immunology,* vol. 112, no. 4, pp. 643-651, 2003/10/01/ 2003.

[6] f. American academy of sleep medicine task, "Sleep-related breathing disorders in adults : recommendations for syndrome definition and measurement techniques in clinical research," *Sleep,* vol. 22, pp. 667-689, 1999 1999.