International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

SLEEP APENA DETECTION

Mr Krishnamoorthy. M[1], A GOKUL REDDY S[1], A HARI VIKASH S[2], CH JEEVAN KALYAN R[4]

[1]Asst.Professor, Dept. of Computer Science, Panimalar Engg College, Tamil Nadu
[2-4]Student, Dept. of Computer Science, Panimalar Engg College, Tamil Nadu

---

Abstract - Obstructive apnea (OSA) may be a common, but severely under-diagnosed disorder that affects the natural breathing cycle during roll in the hay periods of reduced respiration or no airflow in the least. As a first step towards the goal, we explore whether a limited subset of the physiological signs used in traditional OSA diagnosis, together with automatic classification, may be used to detect apnea occurrences in this study. We examine the effects of five data mining algorithms in classifying epochs of data from the PhysioNet Apnea-ECG and MIT-BIH datasets as interrupted or normal breathing. This research focuses on respiratory signals from the nose, abdomen, and chest, as well as oxygen saturation. We calculate the accuracy, sensitivity, specificity, and Kappa statistics of classification with data mining algorithms for any combination of these signals. With a collection of respiration data from both the chest and nose as input data, we reach an accuracy of 96.6 percent for Apnea-ECG, and an accuracy of more than 90 percent for other signal combinations. Surprisingly, these good results may also be obtained using the basic KNN approach. Because of noise, lesser size, and some class imbalance, the findings for MIT-BIH are lower.

Key Words: Obstructive Sleep Apnea, Data Mining, Evaluation.

1.INTRODUCTION

Sleep apnea is a serious sleep disorder that disrupts the natural respiration during sleep, resulting in periods of reduced or no airflow. Obstructive Sleep Apnea (OSA) occurs when the airway becomes physically obstructed, resulting in a reduction or complete cessation of air passage. When normal airflow is disrupted, oxygen saturation drops, and the brain forces an awakening in order to resume normal breathing. The sufferer is unlikely to recall the frequent awakenings. With such disrupted sleep, the person may never fall asleep, resulting in sleep disturbances and fatigue.

To be able to record data at residence without the assistance of personnel, the data must be simple to record, even for those without a high level of technical knowledge. Because it is important for patients to be comfortable while sleeping, the signals must be recorded without limiting patient movements during the night. As a result, we exclude EEG, ECG, EMG, and EOG, and instead concentrate on respiratory signal generated from the Respiratory Inductance Plethysmography (RIP) belt on the chest (C) and abdomen (A), as well as sensors that evaluate nasal airflow (N) and oxygen saturation (O). We evaluate four major data mining techniques to determine whether data mining can define iterations of data as either distorted or normal breathing using such signals and their configurations with sufficient accuracy, sensitivity, and specificity. We hope to discover some baselines by using these basic data mining algorithms and only slightly pre-processed data sets for training and validation, assuming that advanced data pre-processing and data mining algorithms will yield even better outcomes.

1.1 METHOD

The main objective of our research is to assess the performance obtained by employing one of the five candidate data mining algorithms on every possible combination of the four signal types: C,A,O, and N. This chapter defines the inputs used, the signal types, the parameterization of our classifiers, and the analysis methods.

1.2DATABASES

Our research is based on the PhysioNet databases Apnea-ECG and MITBIH. Those were the only two datasets with enough data for the signal types we're interested in, as well as annotations with sufficient level of detail. The recordings are derived from clinical PSG and are being used in a variety of related works. Two studies served as the foundation for the Apnea-ECG dataset. The first study (1993-1995) looked into he effect of OSA on arterial blood pressure in subjects with moderate and severe sleep apnea.

For our evaluation, we use Matlab versionR2017a (9.2.0.556344), which has reliable implementations of all five classifiers. Preliminary experiments are carried out to identify hyper-parameters that can be safely fixed during experimentation, either because they clearly yield superior performance or because varying the hyper-parameter has no effect on performance at all. In the latter case, we employ.

1.3. CLASSIFIERS

We evaluate classifiers with various algorithmic properties to see how these properties affect performance, including classifiers that have been used successfully in related work. It should be noted, however, that most related works perform extraction of features and signal processing on the signal prior to data mining, which we should not do. In related works, the second most prominent data mining methods are ANN and SVM. We include both methods in our evaluations because they produce good results, with an accuracy of more than 90% for epoch classification and

100% for differentiating between apneic and non-apneic subjects. Both are black-box methods with eager learners that can be increased to data sets where linear separation is not possible.

2. Classifier Parameterization

For our evaluation, we use Matlab versionR2017a (9.2.0.556344), which has reliable implementations of all five classifiers. Preliminary experiments are carried out to identify hyper-parameters that can be safely fixed during experimentation, either because they clearly yield superior performance or because varying the hyper-parameter has no effect on performance at all. In the latter case, we employ.

Classifier	Varied Parameter	Tested Values
ANN	Number of hidden nodes	10,20,30,50,100
SVM	Kernel Function Polynomial Order	Linear,Polynomial,radical basis/Gaussian. 2,3
DT	None	
KNN	Number of neighbours	1,5,10
RF	Number of users	10,50,100

Table 1: Varied Classifier Parameters.

3. EVALUATION METHOD

We assess how well the classifiers distinguish between normal breathing epochs and disrupted breathing epochs. To assess each classifier's generalization power, we employ a tenfold cross-validation scheme. A single partition serves as a test set, while the remaining nine partitions are used for training. We compare all possible combinations of signals available in a given database. For Apnea-ECG, this yields 15 signal combinations. Since Because MIT-BIH lacks records that capture respiration from both the chest and the abdomen at the same time, only 11 signal combinations are matched from this database.

For all combinations of classification models and signals, we use 3 different performance metrics: accuracy, sensitivity, and specificity, and compute the Kappa statistic. Furthermore, we obtain scheduling measurements for both the training and testing phases. The number of correctly classified objects is measured in terms of the total number of classified objects, whereas sensitivity measures the quantity of actual positives that are correctly classified as such, and specificity measures the proportion of actual negatives that are correctly identifier

III. EVALUATION

This section shows the classification performance with various signal combinations. Throughout this evaluation section, we refer to five major types of results, each of which is represented by its own table. Because results from all five classes are cited in most discussions, Section IV-A introduces all figures and tables containing our findings, namely Tables 1, 2, and 3. The results obtained with Apnea-ECG datasets differ significantly from the results obtained with MIT-BIH datasets. In Section IV-B, we investigate the cause of this discrepancy. In Section IV-C, we examine our results to see what kind of performance we can get with different signal combinations. In Section IV-E, we evaluate the training and classification times of all classifiers and connect the results to an online analytical context. Finally, Section IV-F summarizes the findings and discusses how they relate to our research questions.

A. TABLES AND FIGURES

This section describes the tables and figures that contain our findings, which are discussed and analyzed in the following sections. It displays the accuracy, sensitivity, and specificity values for the data mining method and parametrization that increase the overall accuracy for that signal combination for every signal combination. Please keep in mind that the y-axes for Apnea-ECG and MIT-BIH are not the same. They provide the findings for the classifier parametrization that yields the highest average accuracy among all signal combinations for such a database for each database. Each parameterized classifier is abbreviated by concatenating a classifier prefix (e.g., "RF") with a postfix containing the value of the relevant parameter that according table 1 Finally, Table 3 shows the training and testing times for each of our classifiers. The best performing parameterizations, i.e. the same classifiers that were used to generate the results, are shown

B.IMPACT OF DATA SET

In these results, there are two distinct patterns. First, when we use data from the Apnea ECG database, we get substantially superior results, with an accuracy of more than 90% for all signal combinations.

Signal(s)	Best performing classifier		Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy range (%)
Apnea-ECG:						
C	RF	(100 trees)	92.6	94.8	91.2	79.6 - 92.6
A	RF	(100 trees)	93.3	95.7	91.6	87.5 - 93.3
N	RF	(100 trees)	96.3	96.6	96.1	89.6 - 96.3
O	RF	(100 trees)	95.2	94.2	95.9	92.4 - 95.2
CA	RF	(100 trees)	93.5	96.0	91.8	87.1 - 93.5
CN	KNN, and RF	(100 trees) (5 neighbours)	96.6 (KNN) 96.6 (RF)	97.8 (KNN) 97.0 (RF)	95.6 (KNN) 96.4 (RF)	88.8 - 96.6
CO	RF	(100 trees)	95.8	96.3	95.5	92.3 - 95.8
AN	KNN	(1 neighbour)	96.3	95.2	97	89.1 - 96.3
AO	RF	(100 trees)	95.9	96.2	95.2	92 - 95.9
NO	SVM	(Gaussian/Radial Basis)	95.9	94.3	97	92.3 - 95.9
CAN	KNN	(5 neighbours)	96.3	95.6	97.6	88.7 - 96.3
CAO	RF	(100 trees)	95.7	96.1	95.4	92.9 - 95.7
CNO	KNN	(5 neighbours)	95.9	92.9	97.9	92.3 - 95.9
ANO	RF	(100 trees)	96.0	96.5	95.6	92.5 - 96.0
CANO	KNN	(1 neighbour)	95.9	93.5	97.5	92.2 - 95.9
MIT-BIH:						
C	KNN	(5 neighbours)	73.5	37.9	89.6	66.3 - 73.5
A	RF	(50 trees)	73.0	85.5	52.5	61.2 - 73.0
N	RF	(100 trees)	70.8	46.5	87.7	59.7 - 70.8
O	RF	(50 trees)	73.0	73.9	72.12	53.6 - 73.00
CN	RF	(100 trees)	75.4	41.7	90.7	66.7 - 75.4
CO	RF	(50 trees)	69.9	68.5	71.3	52.9 - 69.9
AN	KNN	(5 neighbours)	73.1	80.4	61	64 - 73.1
AO	RF	(100 trees)	71.7	72.3	71.2	53.1 - 71.7
NO	RF	(100 trees)	71.7	70.6	72.7	54.4 - 71.66
CNO	RF	(100 trees)	68.6	61.6	75.0	57.5 - 68.6
ANO	RF	(100 trees)	71.5	71.6	71.4	51.6 - 71.5

Table2: Comparision Of Signal Combinations for Apnea-ECG and MIT-BIH.

We must analyse critical factors such as data quality, class balance, and size to understand why the results differ so considerably between the databases. Without in-depth information of the physiological signal types and whether the data has been treated in any way before even being published, evaluating the data quality is impossible. The Apnea-ECG database was created as part of a contest sponsored by PhysioNet and Computers in Cardiology [7] with the goal of using datamining to detect sleep apnea using the ECG signal as the primary indicator. As a result, this database may be of higher quality than the MIT-BIH database.

C. COMPARING CLASSIFIERS

In both the Apnea-ECG and MIT-BIH databases, the classifiers produce similar results for the majority of signal combinations. This is evident in the fact that the majority of classifiers are linked by a horizontal line, indicating that there are few substantial distinctions between them. However, there is one crucial difference: RF100 yields a significantly higher accuracy than DT in Apnea ECG. RF furthermore ranks the highest in both databases.

D. TRAINING AND CLASSIFICATION TIMES

Although real-time analysis is not required for patient-friendly sleep monitoring, the classifiers chosen can be used in an on-line analysis and monitoring setting as well. As a result, comparing the strategies in terms of classification time is intriguing. It's worth noting that the classification time is affected by the inclusion of four eager learners and one slow learner (KNN).Table 3 shows that the training of all ten models complete within two minutes in all cases, which is well within what is acceptable. This results in higher training times for Apnea-ECG.

	Total time spent training ten models (seconds)					Total time spent testing on ten folds (seconds)				
Apnea-ECG: Signal(s)	**ANN-30**	**SVM-3**	**KNN-5**	**DT**	**RF-100**	**ANN-30**	**SVM-3**	**KNN-5**	**DT**	**RF-100**
C	7.533	4.362	0.555	5.516	62.679	0.085	0.233	1.297	0.018	2.065
A	5.049	2.529	0.139	3.964	51.339	0.076	0.121	0.861	0.017	2.031
N	5.822	2.421	0.100	5.202	49.638	0.077	0.124	0.864	0.017	1.999
O	14.291	1.669	0.095	1.574	27.699	0.079	0.086	0.866	0.017	1.969
CA	5.532	3.710	0.139	8.332	70.582	0.078	0.179	1.973	0.018	2.024
CN	6.521	3.372	0.112	9.653	66.610	0.075	0.182	1.968	0.019	2.101
CO	14.790	2.634	0.079	5.730	50.301	0.080	0.141	1.975	0.019	2.002
AN	5.303	3.206	0.078	7.746	68.407	0.079	0.162	2.002	0.019	2.081
AO	13.949	2.515	0.083	4.710	46.024	0.075	0.137	1.994	0.019	1.974
NO	12.540	2.435	0.094	6.306	51.850	0.078	0.130	2.054	0.018	2.001
CAN	5.755	4.690	0.088	11.467	87.372	0.082	0.240	3.215	0.020	2.089
CAO	14.273	4.047	0.090	8.456	64.972	0.080	0.200	3.195	0.195	2.018
CNO	12.507	3.849	0.094	11.103	72.580	0.080	0.201	3.202	0.020	2.081
ANO	12.645	4.105	0.101	9.106	67.123	0.080	0.216	3.255	0.020	2.037
CANO	13.510	5.481	0.139	12.950	79.924	0.082	0.309	4.339	0.020	2.013
MIT-BIH: Signal(s)	**ANN-50**	**SVM-3**	**KNN-5**	**DT**	**RF-50**	**ANN-50**	**SVM-3**	**KNN-5**	**DT**	**RF-50**
C	8.716	2.315	0.330	1.116	10.097	0.018	0.080	0.172	0.038	0.930
A	3.372	0.817	0.092	1.132	10.338	0.081	0.046	0.094	0.025	0.952
N	5.028	17.264	0.088	5.816	61.769	0.081	0.620	1.161	0.025	1.333
O	2.779	0.429	0.079	0.795	8.927	0.075	0.031	0.051	0.013	0.918
CN	3.744	0.763	0.077	1.762	15.574	0.077	0.050	0.164	0.014	0.940
CO	2.467	0.111	0.073	0.339	2.313	0.071	0.011	0.017	0.010	0.845
AN	3.290	0.953	0.084	2.113	14.474	0.079	0.059	0.171	0.014	0.946
AO	2.785	0.393	0.078	1.206	8.036	0.076	0.031	0.066	0.013	0.904
NO	2.850	0.524	0.091	2.960	10.490	0.081	0.037	0.083	0.014	0.914
CNO	2.407	0.108	0.074	0.255	2.671	0.070	0.011	0.018	0.013	0.854
ANO	2.780	0.437	0.081	1.714	9.848	0.071	0.036	0.096	0.014	0.906

TABLE3: Training and testing times for the best-performing classifier parametrizations in seconds.

The amount of time the classifiers spend classifying individual epochs of data is referred to as testing time. As a result, determining their suitability for online analysis is a critical requirement. In MIT-BIH, we notice that RF has the fastest test times for all signal combinations, and in Apnea-ECG, it has the fastest test times for practically all one- and two-signal combinations. KNN has the fastest testing times for three and four signals in Apnea-ECG. The impact of the number of included signals on testing time, as well as the impact of the number of included epochs on testing time, is both examined.

3. CONCLUSIONS

The research provided in this publication is a first step toward our long-term aim of reducing the number of people who aren't identified with OSA and shortening the time it takes for them to get a clinical diagnosis. The primary concept is to allow people to collect physiological signals while sleeping at home and then analyses the data via data mining and determine whether a visit to the MD is recommended. The findings described in this publication and in permitted us to participate in a big clinical trial at the University of Oslo that is now collecting sleep monitoring data from a population of roughly 50 patients with various degrees of OSA using the Nox T3 and the Flow sensor. This This data collection and preparation is still being worked on, and a full analysis of which data mining techniques should be utilized to obtain good classification performance will be done in the future.

REFERENCES

[1] Z. Abedi, N. Naghavi, and F. Rezaeitalab. Detection and classification of sleep apnea using genetic algorithms and SVM-based classification of thoracic respiratory effort and oximetric signal features. Computational Intelligence, 33(4):1005–1018, 2017. COIN-OA-07-15-2237.R2.

[2] H.M. Al-Angari and A.V.Sahakian. Automated Recognisation of obstructive sleep apnea syndromeusing support vendor machine classifier.IEEE Transactions on Information Technology in BioMedicine,16(3):463-468.

[3] L. Almazaydeh, K. Elleithy, and M. Faezipour. Detection of obstructive sleep apnea through ECG signal features. In Electro/Information Technology (EIT), 2012 IEEE International Conference on, pages 1–6. IEEE,2012.

[4] Apnea-ECG. https://www.physionet.org/ physiobank/database/apnea-ECG/additional-information.txt. Accessed: September 2018.

[5] Bitalino. http://bitalino.com/en/. Accesse ed: September 2018.