IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 3, April 2024

Review Paper on Sleep Apnea Detection from Single-Lead ECG: A Comprehensive Analysis of Machine Learning and Deep Learning Algorithms

Aaryan Dhage[1], Shubham Bornare[2], Siddhi Karve[3], Siony Chaudhari[4], Prof. V. M. Dilpak[5]

Students, Department of Artificial Intelligence and Machine Learning[1,2,3,4]

Professor, Department of Artificial Intelligence and Machine Learning[5]

All India Shri Shivaji Memorial Society Polytechnic Pune, Maharashtra, India

Abstract*: Sleep apnea is a common condition that is characterized by sleep-disordered breathing. Worldwide the number of apnea cases has increased and there has been a growing number of patients suffering from apnea complications. Unfortunately, many cases remain undetected, because expensive and inconvenient examination methods are formidable barriers with regard to the diagnostics. Furthermore, treatment monitoring depends on the same methods which also underpin the initial diagnosis; hence issues related to the examination methods cause difficulties with managing sleep apnea as well. Computer-Aided Diagnosis (CAD) systems could be a tool to increase the efficiency and efficacy of diagnosis. To investigate this hypothesis, we designed a deep learning model that classifies beat-to-beat interval traces, medically known as RR intervals, into apnea versus non-apnea. The RR intervals were extracted from Electrocardiogram (ECG) signals contained in the Apnea-ECG benchmark Database. Before feeding the RR intervals to the classification algorithm, the signal was band-pass filtered with an Ornstein–Uhlenbeck third-order Gaussian process.*

Keywords: Sleep Apnea Detection, Electrocardiogram (ECG), RR (R waves in ECG), Long short term memory (LSTM).

I. INTRODUCTION

Sleep is a fundamental human activity which is characterized by reduced or suspended consciousness. Hence, the ability to avoid or correct distur bances, such as sleep disordered breathing, is reduced . Sleep apnea is a common cause for sleep-disordered breathing. In the middle-aged workforce about 2% of women and 4% of men were apnea patients in 1993. In 2003, about 4% of the US population had sleep apnea [3]. The world-wide prevalence was estimated to be 6% in 2008. It is predicted thatthis upward trend will continue. Without diagnosis and adequate treatment patients might be exposed to an increased risk of cardiovascular diseases ,such as stroke and hypertension. Apnea might also disturb recreational activities and by doing so cause mental suffering and in some cases clinical depression. Apnea is also linked to narcolepsy, insomnia, and obesity. Studies show that patients with apnea have a higher chance of being involved in a road traffic accident. The disease is also a risk factor for complications during operations under anesthesia. Finally, patients with untreated apnea have a significantly higher mortality risk when compared to a control group with the same age, sex and Body Mass Index (BMI).Current diagnostic methods depend on Polysomnography (PSG). The measurements include ECG, Electroencephalogram (EEG), Electrooculogram (EOG), Electromyogram (EMG), respiratory effort, airflow and oxygen saturation (SaO2). To capture these signals, the patient must sleep with intrusive measurement equipment in a clinical environment. The process requires supervision by medical specialists. The PSG process makes apnea diagnosis expensive and inconvenient. To improve this situation new methods are required which are less intrusive and more cost effective, but equally accurate. Mobile technology and advanced physiological signal measurement methods might be able to address the intrusiveness and cost issues. One promising measurement technology is single lead ECG for signal acquisition and mobile soft processing for beat-to-beat (RR) interval extraction.As such, that measurement setup has a significantly lower complexity when compared with PSG. Furthermore, it is notably cheaper to communicate and process the resulting RR interval signals, when compared with

DOI: 10.48175/IJARSCT-17223

the multitude of physiological signals measured during PSG. However, major issues remain with the diagnosis support quality provided by these systems. One critical component to ensure diagnosis support quality are the algorithms which extract the relevant information or provide decision support. The setup was designed with We achieved remarkable results with our system using a benchmark dataset from the MIT-BIH Polysomnographic Database. Through 10-fold cross-validation, we attained an accuracy of 99.80%, a sensitivity of 99.85%, and a specificity of 99.73%. This demonstrates the effectiveness of our system, even with a simpler data acquisition setup.

Additionally, we made a noteworthy design improvement. We discovered that applying low- and high-pass filtering to the RR interval signal boosted the classification accuracy by more than 20%. This preprocessing step enhances the detection quality for various CAD systems by allowing deep learning algorithms to focus on Heart Rate Variability (HRV).

To support these claims, we outline our design of an apnea detection algorithm. The next section introduces the medical background of sleep apnea. Section 3 details the methods used to construct the test setup. Thereafter,we present the results achieved while testing the proposed diagnosis support system. In the Discussion section, we relate our work to other studies done on similar topics. Having this extended scope allows us to show how the RR interval filtering might help to improve the classification accuracy for other detection tasks. The conclusion summarizes the work and puts forward the highlights of the study.

II. BACKGROUND

During apnea the patient ceases to breath for 10 s or more. Obstructive Sleep Apnea (OSA) and Central Sleep Apnea (CSA) are the two main causes for the pauses in breathing. The pauses usually occur during during rapid eye movement sleep. An OSA event occurs when the airway is blocked completely. The blockage might be due to fatty tissue, musculus geniohyoideus, or musculus genioglossus. In contrast, a CSA event is characterized by a lack of respiratory effort, i.e. there is a problem with respiration control.OSA is diagnosed more often than CSA. There are several therapies for sleep apnea, such as Positive Airway Pressure (PAP) and Palato Pharyngo Plasty (PPP). In general, these therapies are more effective when sleep apnea is detected early. In current clinical practice, polysomnograms, which result from PSG sleep studies, are used to evaluate an index score. The score value determines the apnea severity. An important component of these index scores is the airflow signal and blood oxygen content. However, measuring 83 these signals is intrusive and inconvenient for the patient. To reduce the in convenience, apnea detection methods were developed using respiratory and 85 single-lead ECG signals [24, 19]. In response, PhysioNet held a competition called CinC Challenge 2000 , which provided ECG data with 87 minute-by-minute labeling. After the challenge, the training dataset,88 with 35 recordings, was made publicly available by PhysioNet. Over the 89 years, the dataset was used to design apnea detection algorithms and it is now considered a benchmark that can be used to compare individual method performances. Digital biomarkers fail to capture all sleep apnea induced morphological changes because transient abnormalities appear randomly, and long-term abnormalities are difficult to quantify. Deep neural networks can refine the information even further and provide medical decision support which can help to diagnose sleep apnea.

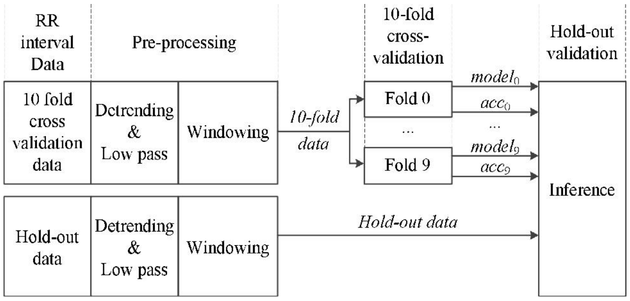

Figure 1: Block diagram for training and validating the deep learning model

The research provided precedents of employing Convolutional Neural Network (CNN) to detect disease using ECG signals. In apnea detection tasks, directly feeding original ECG signals to deep neural networks is adopted by some researchers but the high ECG data rate limits the network depth. As such, the RR interval signal is derived from the ECG extracting the beat-to-beat record of RR-intervals and is, as a time series, irregularly sampled. Studies show that there is a physiologic link between the breathing rate and thebeat-to-beat variations of the human heart. Hence, it is possible to detect sleep disordered breathing based on RR interval signals. The next section describes the methods we have used to detect apnea induced sleep disordered breathing based on RR interval signals.

III. METHODS AND MATERIALS

Based on lessons learned from the aforementioned literature we formulate the hypothesis that: the ECG alone is a promising signal to use for sleep apnea detection. In addition, an adequate feature selection is preponderant for the classifier accuracy, and the SVM algorithm revealed its suitability to cope with apneic ECG signals. With those notions in mind, we developed a system to detect sleep apnea in which feature selection and classifiers were benchmarket. The flow of the proposed model is depicted in Figure 2 including: pre-process, feature extraction, classification, and feature selection. These architecture is explained in detail in the sections below.

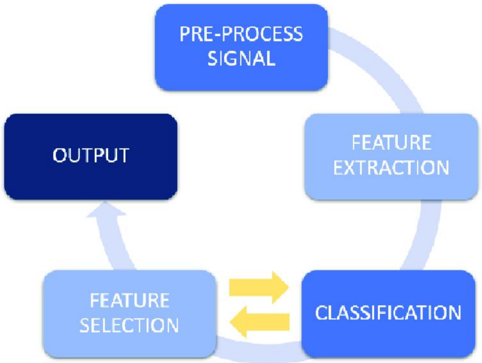

FIGURE 2. Proposed system activity model.

For the development of this study, we used GNU Octave, which is software compatible with MATLAB and its packages.

A. DATABASE

PhysioNet Apnea-ECG Database v1.0.0 was used in this study. This dataset consists of 70 recordings (i.e., A01 to A20, B01 to B05, C01 to C05 and X01 to X35) which is split into train and test set. Training set consist of A01 to A10, B01 to B05 and C01 and C10 while X01 to X35 are used as test set.

ECG signals are digitized in 100 Hz. The sleep apneaannotation is available in each 1-min of data that was labeled by experts. Apnea index (AI), hypopnea index (HI) and apnea-hypopnea index (AHI) are the main indexes which are studied in sleep apnea analysis. Age, Sex, and physiological characteristics such as height and weight may be effective in sleep apnea analysis. Eight recordings in test set and five recordings in train set are from females. Minimum, maximum, and mean values of different subject characteristics in train and test sets are summarized.

B. Pre-Processing and Data Preparation

In this paper, we addressed the challenge of noise in ECG signals caused by baseline wander, muscle artifacts, and power line interference. To mitigate this noise, we applied a bandpass finite impulse response (FIR) filter.

The ECG signal was divided into 1-minute intervals. Within each interval, we utilized the Hamilton-Tompkins R-peak detection method to identify R-peaks. Subsequently, we calculated the amplitudes of the R-peaks and the intervals between consecutive R-peaks (R-R intervals).

To further analyze the signal, we computed the Power Spectral Density (PSD) of both the ECG R-peak amplitudes and the R-R intervals using the Welch method. This approach helped us better understand the frequency content and characteristics of the ECG signal, enabling more accurate analysis and interpretation.

DOI: 10.48175/IJARSCT-17223

C. FEATURE EXTRACTION

In this study, we conducted feature extraction on both the HRV and EDR signals, aligning with the one-minute segments indicated in the database annotations. Initially, we extracted 18 features from the HRV signal and 2 features from the EDR signal, totaling 20 features.

Later, to expand our analysis and incorporate findings from existing literature, we included more features. This extended our study to encompass a total of 50 features from the HRV signal and 34 features from the EDR signal, resulting in 84 features overall.

The additional features were derived from the 256-point FFT power spectral density, with 32 points considered for each HRV and EDR signal. This comprehensive approach allowed for a more in-depth analysis of the results and the performance of various classifiers.

D. CLASSIFICATION

With the classification in mind, all records extracted fromone-minute segments were labeled as 0 or 1 representing non-apnea or apnea event respectively. The database containing 17401 records in which 46.33% are related toapnea moments, whereas non-apnea moments are observable

in 53.67%. Then, database was segmented into three different vectors for training, testing, and validation purposes.The k-fold cross-evaluation method was adopted with k=10,in order to improve the training of the classifiers. Finally, sensitivity, specificity and accuracy were calculated as follows:

(1) sensitivity =TP/TP + FN

(2) specificity =TN/TN + FP

(3) accuracy =TP + TN/P + N

where P: Positive. N: Negative. TP: True Positive. TN: True Negative. FP: False Positve. FN: False Negative.

In the classification phase, five classifiers (ANN, SVM, LDA, PLS, and aNBC) were implemented and its performance were comparatively evaluated. All algorithms were implemented following its default settings except the ANN and the SVM that were configured for our experiments.

1) ARTIFICIAL NEURAL NETWORK (ANN)

The ANN was implemented with both 20 and 84 input neurons (congruently with the 20 and the 84 features extracted respectively). The hyperbolic tangent sigmoid transfer function, i.e. tansig was used as a transfer function between the input layer and the hidden layer. Then, the linear transfer function i.e. purelin was used as a transfer function between the hidden layer and the output layer.

2) SUPPORTVECTOR MACHINE (SVM)

$$K(xi, xj) = e{-}\gamma\ kxi{-}xjk,\ \gamma \geq 0$$

In which γ determines the variance i.e. the similarity measure between two points. A large value means a small variance (two points are similar when they are close to each other). On the contrary, a lower value means a large variance (two points are similar even if are distant to each other) . On the other hand, aiming at to obtain a better overall fit model [38], we tuned the SVM soft margin, namely the C parameter. Based on the experimental results the models performed best with the C parameter equal to 512.

3) LINEAR DISCRIMINANT ANALYSIS (LDA)

The LDA was introduced by [39] for dimensionality reduction. On the one hand its simple to implement since is based on generalized eigenvalue decomposition. In addition, its easy to adapt for discriminating non-linearly separable classes. In other words, the LDA aims to identify a low-dimensional linear subspace whereon instances of multiple classes; at least two, are best separable .depicts a two class-separation using the LDA by means of axes maximization.

E. Classification

Different machine learning algorithms have been applied for sleep apnea detection. However, the main problem in this field is the lack of a fair and unified comparison between different algorithms. In this study, different well-known machine learn- ing algorithms are compared. More information about these machine learning algorithms is available in [40].

1) Linear Discriminant Analysis (LDA): LDA is a simple classifier that uses a linear decision boundary, generated by fitting Gaussian densities to each class using Bayes' rule [16]. The main hyperparameters that need to be predefined before training LDA are the solver algorithm and a tolerance para- meter that defines the absolute threshold for a singular value to be considered significant.

2) Quadratic Discriminant Analysis (QDA): QDA is similar to LDA but uses a quadratic boundary decision [17]. The solver for QDA is usually set to SVD, and therefore tolerance is the main hyperparameter that needs to be defined before training.

3) Logistic Regression (LR): LR uses a logistic function for modeling dependent binary variables [20]. Solver algorithm,tolerance for stopping criteria, regularization strength (1/C), and penalty norm are its main hyperparameters.

4) Gaussian Naïve Bayes (GNB): GNB is a probabilistic method based on applying Bayes' theorem with naïve independence assumption between features. Its main hyperparameter is the portion of the largest variance of all features that has to be added to variances for stability [41].

5) Gaussian Process (GP): GP classification is a nonpara- metric approach that uses a kernel function to assign class labels. Kernel, optimizer algorithm, and the maximum number of iterations for approximation posterior in Newton's method are its main hyperparameters.

6) SVM: SVM uses a kernel function for mapping data into a higher dimension. Lagrange equations are used to solve the problem in higher dimensions [21]. The type of the kernel and the kernel parameters are its main hyperparameters.

7) K-Nearest Neighborhood (KNN): KNN classifies data samples based on their nearest neighbors. There are different criteria for the distance metric in the KNN method, including Euclidean distance and Mahalanobis distance . Distance metric, the metric parameters, number of nearest neighbors, and leaf size are its main hyperparameters.

8) Decision Tree (DT): DT is a nonparametric classifier that predicts the class labels by learning simple decision rules inferred from the extracted features [18]. Interpretability is the main advantage of this method. Max depth of trees, the strategy used to choose the split, and the criterion measuring the quality of splits are the main hyperparameters of DT.

9) Random Forest (RF): RF is an extended version of DT. RF uses several DTs for combinational learning by subsampling the input data with replacement (bootstrapping) [20]. This method largely prevents over-fitting during training, although improving performance in this method and reducing over-fitting depends on the type of dataset. Maximum depth, the minimum number of samples required to split an internal node, split criterion, and the number of estimators (trees) are the main hyperparameters to be set.

10) Extra Tree (ET): ET is another evolved model based on DTs that is similar to RF but unlike RF that uses bootstrapping methods, uses the whole original sample. In addition, unlike RF which chooses the optimum cut points to split nodes, ET chooses them randomly. In terms of computational complexity and time required to implement the algorithms, ET is faster and more efficient than RF. Maximum depth, the minimum number of samples required to split an internal node, split criterion, and the number of estimators (trees) are the main hyperparameters to be set.

V. EXPERIMENTS AND DISCUSSION

The study used the PhysioNet Apnea-ECG Database v1.0.0 to detect sleep apnea, which had imbalanced data with fewer apnea episodes. To ensure reliable model evaluation, they applied stratified cross-validation. Machine learning and deep learning algorithms were implemented in Python using Google Colaboratory and libraries like Keras and scikit-learn. Feature extraction from ECG signals was done using hrvanalysis, scipy, and entropy libraries.

HRV features, especially frequency-domain features like HF power and LF/HF ratio, were found crucial for sleep apnea detection. Time-domain features such as mean and kurtosis, along with nonlinear features like SD2, were also important. Among conventional machine learning algorithms, MLP and SVM performed best, while deep learning algorithms showed superior performance. Hybrid architectures like ZFNet-BiLSTM had the highest accuracy and specificity. Deep neural networks were favored for their automatic feature extraction ability. CNN-based networks generally outperformed DRNNs, possibly due to short ECG segment analysis. Models like ZFNet-GRU and AlexNet-GRU showed the highest sensitivity, which is crucial for minimizing false negatives in apnea detection.

VI. RESULTS ANALYSIS

Our computational experiments were based on the above mentioned classifiers. Firstly, 20 extracted features were applied to train and simulate the model. Secondly, 64 additional features obtained via PSD/FFT points were added to the initial features set (i.e., 84 features in total). All five classifiers were trained using 8507 records of features, with 2836 records used as training set and 5671 records used to evaluate the performance. The data provided to the classifiers for training and testing were divided using the k-fold cross-validation method with k=10.

VII. CONCLUSION

This study presents an ECG-based model for minute-based analysis of sleep apnea. The main goal is to implement an efficient and precise alternative method to the classical PSG, based on a single signal, the ECG. In addition, a benchmark with five classifiers are implemented, namely: ANN, SVM, LDA, PLS, and a NBC. As expected and according with the presented results,it can be concluded that different classifiers have different behaviors to solve the same problem. Additionally, it is shown that the model proposed in this study is suitable, feasible and accurate in the detection of sleep apnea with an ECGsignal. Our findings highlighted the ANN using 20 features as the most accurate model with an accuracy of 82.12%, a sensitivity of 88.41% and a specificity of 72.29%. Moreover, the experimental results revealed that is crucial deter- mining the most relevant features with the ambition to enhance the accuracy of the model. Indeed, a same classifier may present contrasting performances as observed on the lower accuracy obtained when classifiers were evaluated with 84 features. Future work may include the introduction of feature selection in order to determine an optimized characteristic set for the detection of sleep apnea; improving sensitivity so that all apnea moments are detected; comparing and calculating the performance of the different methods applied in the study, including evaluating the computational costs of classifiers; and simulating the same study in real patients to examine the viability of the method presented here and its implementation.

REFERENCES

[1]. V. M. Altevogt and H. R. Colten, Sleep Disorders and Sleep Depriva- tion: An Unmet Public Health Problem. Washington, DC, USA: National Academies Press (U.S.), 2006.

[2]. K. Feng, H. Qin, S. Wu, W. Pan, and G. Liu, "A sleep apnea detection method based on unsupervised feature learning and single-lead electro- cardiogram," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–12, 2021.

[3]. A B. Neikrug and S. Ancoli-Israel, "Sleep disorders in the older adult— A mini-review," Gerontology, vol. 56, no. 2, pp. 181–189, 2010.

[4]. Q. Shen, H. Qin, K. Wei, and G. Liu, "Multiscale deep neural network for obstructive sleep apnea detection using RR interval from single-lead ECG signal," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–13, 2021.

[5]. M. M. Lyons, N. Y. Bhatt, A. I. Pack, and U. J. Magalang, "Global burden of sleep-disordered breathing and its implications," Respirology, vol. 25, no. 7, pp. 690–702, Jul. 2020.

[6]. A Pinho, N. Pombo, B. M. C. Silva, K. Bousson, and N. Garcia, "Towards an accurate sleep apnea detection based on ECG signal: The quintessential of a wise feature selection," Appl. Soft Comput., vol. Oct. 2019, Art. no. 105568.

[7]. W. Conwell et al., "Prevalence, clinical features, and CPAP adherence in REM-related sleep-disordered breathing: A cross-sectional analysis of a large clinical population," Sleep Breathing, vol. 16, no. 2, pp. 519–526, Jun. 2012.

[8]. G. Cybenko, "Approximation by superpositions of a sigmoidal function," Math. Control, Signals Syst., vol. 2, no. 4, pp. 303–314, 1989.

[9]. I. Ahmad, M. Basheri, M. J. Iqbal, and A. Raheem, "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection," IEEE Access, vol. 6, pp. 33789–33795, 2018.

[10]. Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, "Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network," Inf. Fusion, vol. 53, pp. 174–182, Jan. 2020.

[11]. K. Jankowsky and U. Schroeders, “Validation and generalizability of machine learning prediction models on attrition in longitudinal studies,” Int. J. Behav. Develop., pp. 1–8, 2021, doi: 10.1177/01650254221075034.

[12]. M. Forouzanfar, F. C. Baker, I. M. Colrain, A. Goldstone, and M. Zambotti, “Automatic analysis of pre-ejection period during sleep using impedance cardiogram,” Psychophysiology, vol. 56, no. 7, Jul. 2019, Art. no. e13355.

[13]. M. Forouzanfar, F. C. Baker, M. de Zambotti, C. McCall, L. Giovangrandi, and G. T. A. Kovacs, “Toward a better noninvasive assessment of preejection period: A novel automatic algorithm for B-point detection and correction on thoracic impedance cardiogram,” Psychophysiology, vol. 55, no. 8, Aug. 2018, Art. no. e13072.

[14]. F. C. Baker et al., “Changes in heart rate and blood pressure during nocturnal hot flashes associated with and without awakenings,” Sleep, vol. 42, no. 11, p. zsz175, Oct. 2019.

[15]. H. Li and Y. Guan, “DeepSleep convolutional neural network allows accurate and fast detection of sleep arousal,” Commun. Biol., vol. 4, no. 1, pp. 1–11, Dec. 2021.

[16]. M. Bahrami and M. Forouzanfar, “Deep learning forecasts the occurrence of sleep apnea from single-lead ECG,” Cardiovascular Eng

DOI: 10.48175/IJARSCT-17223