The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

How Effectively Can ECG Signals Alone Be Used to Detect Sleep Apnea Events in Patients?

Sam I Anil

Cambridge Centre for International Research (CCIR)

This research was conducted as part of the CCIR Program, focusing on artificial intelligence in healthcare: biomedical sensors, signal processing, and machine learning.

Abstract

Sleep apnea is a common disorder characterized by pauses in breathing during sleep, which can lead to serious health complications if left undiagnosed. This study investigates the effectiveness of using only ECG signals to detect sleep apnea events, aiming to evaluate an alternative, non-invasive diagnostic approach. The research utilizes the Apnea-ECG Database, which includes ECG data from 35 patients, where features such as heart rate, RR interval, and frequency domain metrics were extracted from filtered ECG signals. Supervised machine learning models, including Naive Bayes (NB), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), were employed, with hyperparameter tuning and cross-validation applied to optimize performance. The NB model achieved a weighted average F1-score of 90%, indicating its effectiveness in identifying apnea events. These findings indicate that ECG-based detection may serve as a promising tool for detecting sleep apnea, with future research focusing on expanding the dataset, integrating additional physiological features, and validating the approach in clinical environments.

Keywords: Physiological Signals, Machine Learning, ECG Signals, Sleep Apnea, Classification

Introduction

Sleep apnea is a common and potentially serious disorder where breathing repeatedly stops during sleep, which can result in lower oxygen levels and disrupted sleep for many people (National Heart, Lung, and Blood Institute, 2022). Detecting sleep apnea early can be important because it can help prevent related health problems like heart disease, memory issues, and a reduced quality of life (Punjabi, 2008). Commonly, sleep apnea has been diagnosed using polysomnography (PSG), a detailed and costly test done in an overnight sleep lab (Wang et al., 2019). Even though PSG is effective, its cost, complexity, and limited availability have driven the need for more accessible and affordable ways to detect sleep apnea (Wang et al., 2019). Finding a simpler way to diagnose sleep apnea could make it more accessible to many people who need it.

Recent developments suggest that electrocardiogram (ECG) signals alone may be used to detect sleep apnea (Yeh et al., 2022). By looking at patterns in ECG signals, like the ECG-derived respiratory (EDR) signals and heart rate variability (HRV), researchers have made significant progress in identifying apnea events with good accuracy (Yeh et al., 2022). This approach simplifies the diagnostic process since ECG is already commonly used in many healthcare settings, making it easier to detect sleep apnea without using complex and costly tests like PSG (Yeh et al., 2022).

Research in the detection of sleep apnea using ECG signals has shown some promising results. For example, one study utilized a combination of respiratory and ECG signals to improve the accuracy of sleep apnea detection (Sheta et al., 2021).

The researchers used advanced processing techniques to analyze these signals, leading to better detection accuracy (Sheta et al., 2021). An-other study focused on using filter bank decomposition and convolutional neural networks (CNNs) to evaluate the contribution of different sub bands of ECG in sleep apnea detection (Yeh et al., 2022). This method effectively filtered out the unnecessary frequencies, and then they normalized the signals, significantly increasing the sleep apnea detection performance (Yeh et al., 2022). These studies highlight the potential of combining other advanced signal processing techniques with ECG analysis to improve the detection of sleep apnea.

Further research has investigated machine learning and deep learning models to improve sleep apnea detection from ECG signals. A comprehensive review compared many machine learning models, such as support vector machines (SVM) and random forests, finding that these models could effectively diagnose sleep apnea with high accuracy (Zhu et al., 2022). Another study applied a deep learning approach, combining CNNs and long short-term memory (LSTM) networks, to automatically detect sleep apnea from single-lead ECG signals (Zarei et al., 2022). This approach avoided traditional feature extraction methods, allowing for end-to-end learning and achieving high detection accuracy (Zarei et al., 2022). These studies emphasize the increasing role of ma-chine learning and deep learning in developing more accurate and automated methods for sleep apnea detection.

In this study, the aim is to establish the effectiveness of using ECG signals alone to detect sleep apnea events in patients. By focusing on various features of ECG signals, including those derived from time and frequency domains, this research paper aims to improve the reliability and accuracy of sleep apnea detection (Penzel et al., 2000).

The Apnea-ECG Database from PhysioNet is the main resource for this research because it provides detailed ECG recordings from 35 subjects who participated in sleep studies (Penzel et al., 2000). If proven effective, using only ECG signals to detect sleep apnea could be simpler compared to traditional methods, making the diagnosis process easier for patients (Wang et al., 2019). The use of ECG data as a detection tool for sleep apnea provides a different method compared to current complex methods of detection, making the process more affordable and accessible (Sheta et al., 2021). The significance of this approach process is its potential to obtain earlier detection and treatment for patients, particularly for individuals who might otherwise remain undiagnosed with sleep apnea.

Materials and Methods

Dataset Description

This study utilizes the Apnea-ECG Database, a publicly available dataset on PhysioNet contributed by Dr. Thomas Penzel from Phillips-University in Marburg, Germany, and curated by George Moody and Roger Mark (Penzel et al., 2000). The dataset showcases sleep apnea research and has been employed in numerous studies to evaluate machine learning models for detecting sleep apnea events based solely on ECG signals (Penzel et al., 2000).

The database contains 70 records, split into two groups: a learning set of 35 records and a test set of 35 records. The learning set includes subjects labeled as a01–a20 (sleep apnea), b01–b05 (borderline cases), and c01–c10 (healthy), while the test set, identified as x01–x35, does not have apnea labels (Penzel et al., 2000). Recordings typically span from 7 to 10 hours of continuous ECG signals and were obtained at a sample rate of 100 samples per second (Penzel et al., 2000). Each ECG signal is recorded at a sampling frequency of 100 Hz, with segments indicating sleep apnea events, allowing the application of signal processing techniques for further analysis (Penzel et al., 2000). Each file in the dataset includes machine-generated QRS annotations as well as expert-labeled apnea annotations, which help in accurately identifying sleep apnea events across time segments (Penzel et al., 2000).

Patient Demographics

The dataset does not provide specific demographic details such as age, gender, BMI, or medical history for each patient (Penzel et al., 2000). However, the population likely includes subjects characteristic of sleep apnea research, where the condition is more prevalent among middle-aged and older adults, particularly males and individuals with higher body mass indexes (BMIs) (Ling, 2023; Penzel et al., 2000).

These inferred characteristics help relate the findings of this study to real-world applications. Additionally, eight recordings (a01–a04, b01, c01–c03) contain respiratory signals, such as chest and abdominal respiratory effort (Resp C and Resp A), oronasal airflow (Resp N), and oxygen saturation (SpO2) (Penzel et al., 2000). This analysis focuses only on ECG data, even though other signals could provide a more complete picture of the physiological changes during apnea events (Penzel et al., 2000).

Data Selection and Features

Out of the total 70 subjects, 35 were selected for this study: 20 diagnosed with sleep apnea (the ’a’ group), 5 borderline patients (the ’b’ group), and 10 healthy individuals (the ’c’ group) (Penzel et al., 2000). The 35 test set patients labeled as ’x’ were excluded due to the lack of apnea labels, which would limit the analysis. The ECG signals were digitized at 16-bit precision and set to 200 A/D units per millivolt (Penzel et al., 2000).

Preprocessing

The preprocessing was structured to prepare the ECG signals for feature extraction and machine learning analysis. Initially, the ECG signals were loaded from the dataset using the WFDB library. The signals were preprocessed through several steps:

Normalization

Normalization ensures that the signals are scaled properly, preventing any particular feature from dominating due to the differing ranges in amplitude. The ECG signals were normalized by subtracting their mean and dividing by their standard deviation to ensure a consistent range across different signals (Haykin, 2009). The formula is shown below:

$$X - \mu \ \sigma$$

$$X_{\text{norm}} = \underline{\qquad}$$

where X is the original ECG signal, μ is the mean of the signal, and σ is the standard deviation of the signal (Haykin, 2009).

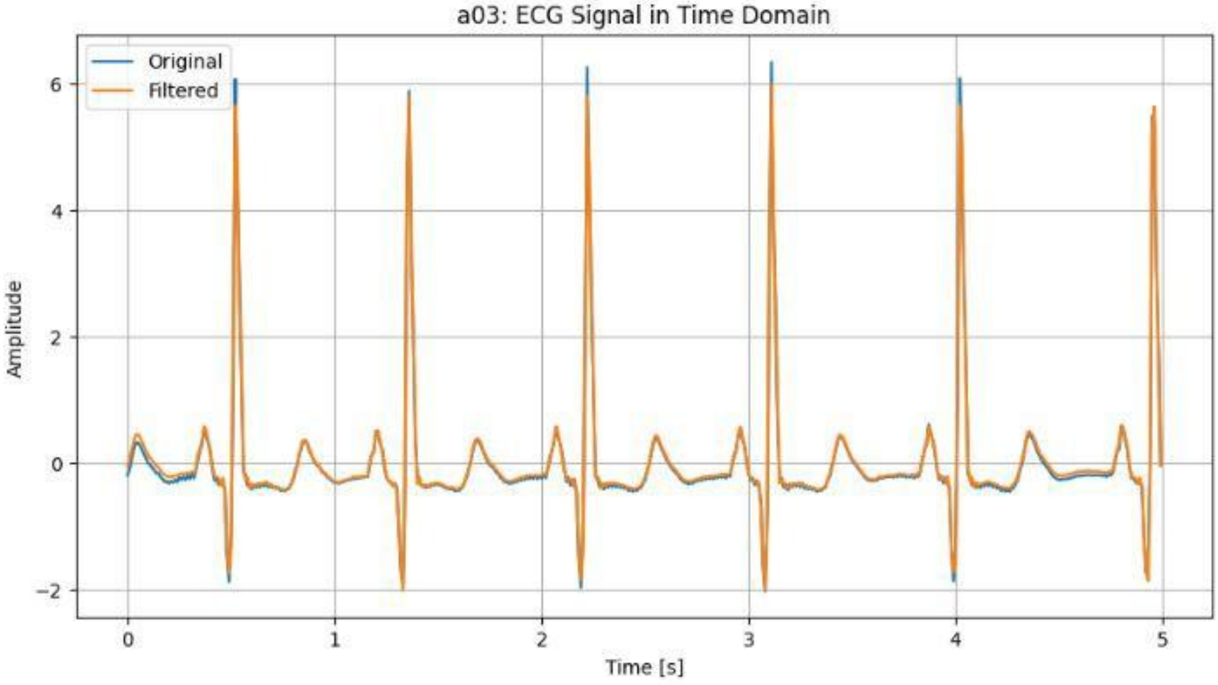
Filtering

A band-pass Butterworth filter was applied to the ECG signals with a low cut-off frequency of 0.5 Hz and a high cut-off frequency of 40 Hz. This step removed noise and retained the relevant frequency components for sleep apnea detection. The Butterworth filter is designed to have a frequency response as flat as possible in the passband (Butterworth Filter - an Overview — ScienceDirect Topics, n.d.). The band-pass filter was designed with the following characteristics:

- Low cutoff frequency: 0.5 Hz
- High cutoff frequency: 40 Hz

Figure 1

Comparison of the ECG signal before and after filtering



- Filter order: 2

Frequency Domain

To further examine the filtering process, Welch's method was used to calculate the power spectral density (PSD) of both the raw and filtered ECG signals. Welch's method estimates the power of a signal at different frequencies, providing insights into the distribution of power across frequencies (Solomon Jr, 1991).

Feature Extraction

After preprocessing, various time and frequency domain features were extracted from the ECG signals to enable the classification of apnea events. These features were further analyzed and visualized to provide an understanding of the ECG data.

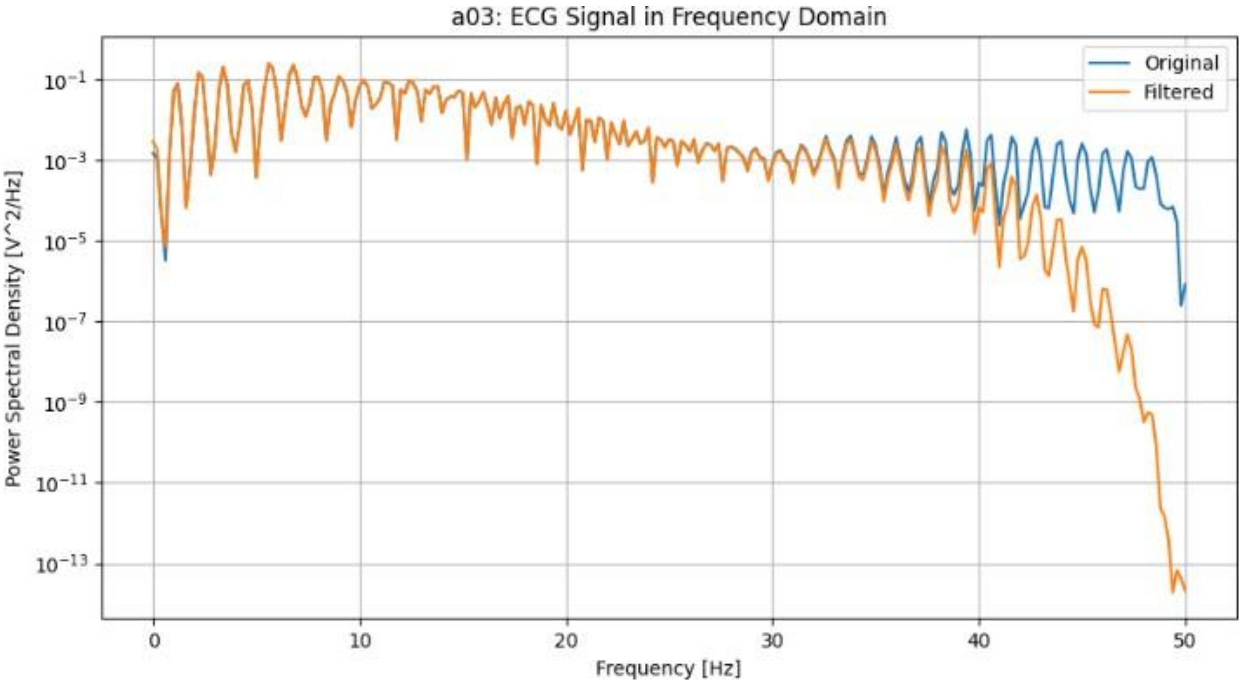
Time-Domain Features

Mean, Standard Deviation, Max, and Min

These basic statistical features describe the overall distribution of the ECG signal in the time domain. They help describe the variability and spread of the signal. As shown in Figure 3, the box plot illustrates the maximum time-domain feature, while the same figure also depicts the mean time-domain feature. These patterns highlight how apnea patients tend to show greater variability and higher peak values, which may reflect physiological instability during disordered breathing.

Figure 2

Power Spectral Density of the ECG signal before and after filtering



RR Interval and Heart Rate (HR)

The RR interval, representing the time between con-secutive R-peaks, was calculated by detecting the R-peaks in the ECG. From the RR intervals, the heart rate (HR) was calculated. These features can be useful in identifying irregular heartbeats, often associated with apnea events (Sulistyo et al., 2018). Figure 3 also shows a box plot of the RR interval time-domain feature.

Frequency-Domain Features

Area Under the Curve

The area under the PSD curve, specifically between the frequencies of 0.5 Hz and 40 Hz, was calculated to capture the overall energy in the lower frequency band. This range aligns with typical respiratory-related changes observed in the ECG during sleep apnea (Solomon,1991). Figure 3 shows a box plot of the area under the curve frequency domain feature. Analysis of the frequency-domain features revealed that apnea patients generally exhibited higher energy in the lower frequency bands compared to healthy individuals, reflecting stronger respiratory-related variations in the ECG during apnea events. The maximum frequency and amplitude were also observed to be more variable in apnea patients, suggesting irregular cardiac dynamics associated with disordered breathing.

Max Frequency and Max Amplitude

The maximum frequency and corresponding am-plitude were extracted from the PSD. These features indicate the dominant frequency components of the ECG signal and their strength, which may reflect changes associated with sleep apnea (Yeh et al., 2022). Figure 3 shows box plots of both the max frequency feature and the max amplitude feature.

Figure 3 illustrates all of the features visualized in box plots, with separate plots for the two classes: healthy and apnea.

Machine Learning Model

The analysis employed multiple machine learning models to classify ECG signals as either being of sleep apnea or healthy. Machine learning is particularly suitable for this task due to its abil-ity to handle complex data patterns and identify relationships in high-dimensional data (Machine Learning – ScienceDirect Topics, n.d.).

After preprocessing the ECG signals and extracting both time-domain and frequency-domain features, these were used as inputs to various supervised machine learning algorithms. The primary objective was to compare the performance of different classifiers and optimize the model for accurate prediction.

Feature Matrix and Label Encoding

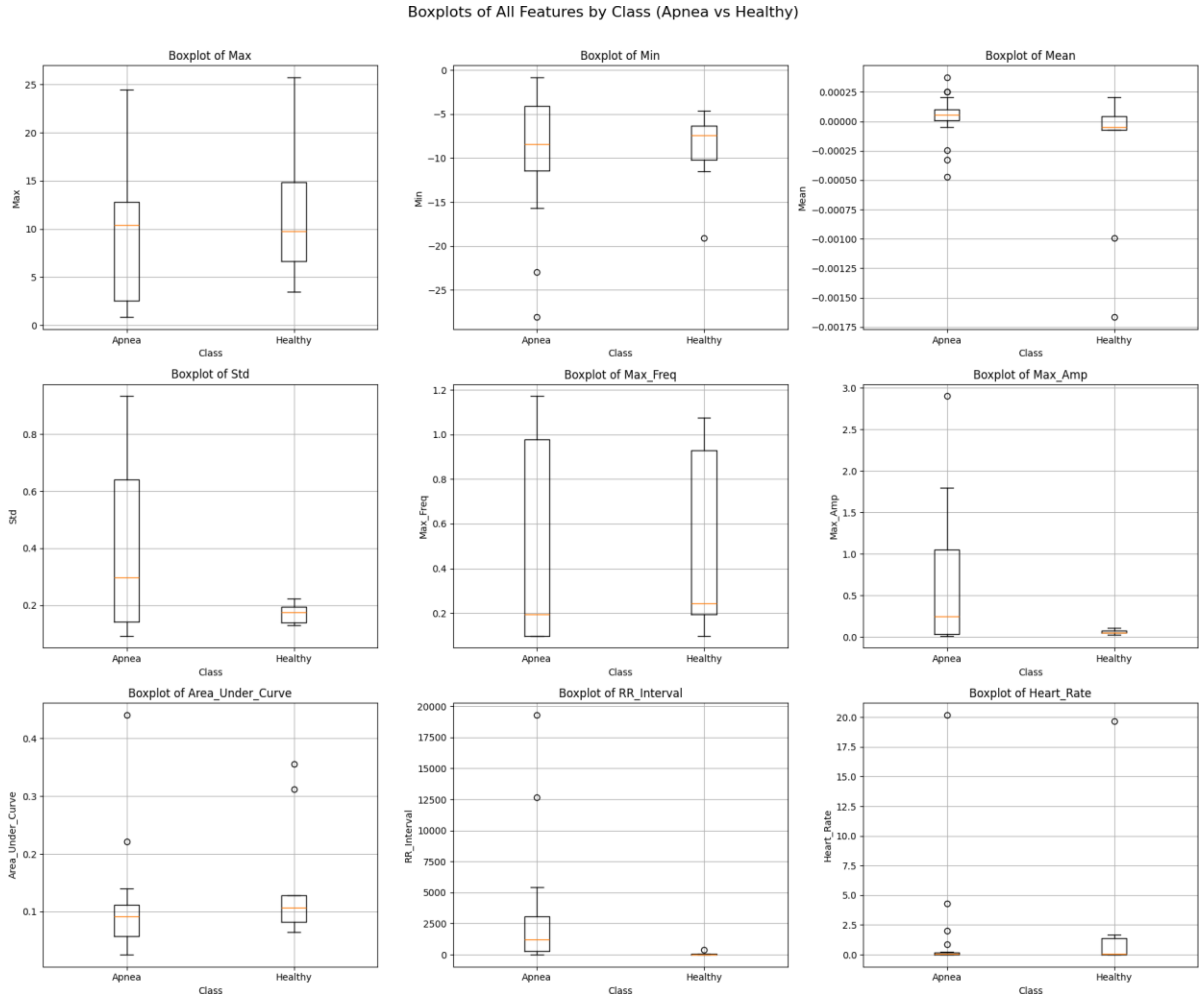
The features extracted from the ECG signals were consolidated into a feature matrix, with each row representing a patient and columns corresponding to features such as the mean, standard deviation, maximum amplitude, and RR intervals. The labels were manually encoded, where ’Healthy’ was represented as 0 and ’Apnea’ as 1, to facilitate binary classification. This binary encoding is a common approach in classification tasks as it simplifies the modeling process (Keylabs, 2024).

Data Normalization and Splitting.

To ensure consistency in the range of feature values, the features were normalized using the StandardScaler. The dataset, comprising 35 patients (25 with apnea and 10 healthy), was then split into training and test sets in an 80:20 ratio using stratified sampling to maintain the class distribution across both sets. Stratification was crucial due to the class imbalance, ensuring both the ’Healthy’ and ’Apnea’ categories were represented proportionally (Cao et al., 2024).

Figure 3

Box plots visualizing the extracted features for healthy and apnea classes



Model Selection and Cross-Validation

Several machine learning models were selected for evaluation, including Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Decision Trees (CART), Gaussian Naive Bayes (NB), and Support Vector Ma-chine (SVM). These models were chosen based on their popularity and effectiveness in classification tasks (Mitchell, 1997; Haykin, 2009; Hastie, 2009). To identify the best-performing model, each classifier was evaluated using 5-fold cross-validation with hyperparameter tuning via Grid-SearchCV on the training dataset. This method ensured robust model evaluation by training and testing the models across different subsets of the data, minimizing the risk of overfitting. The performance of each model was measured using accuracy as the evaluation metric (Arlot & Celisse, 2010). The cross-validation process ensured that the most robust model was selected based on performance metrics relevant to classification accuracy and F1-score.

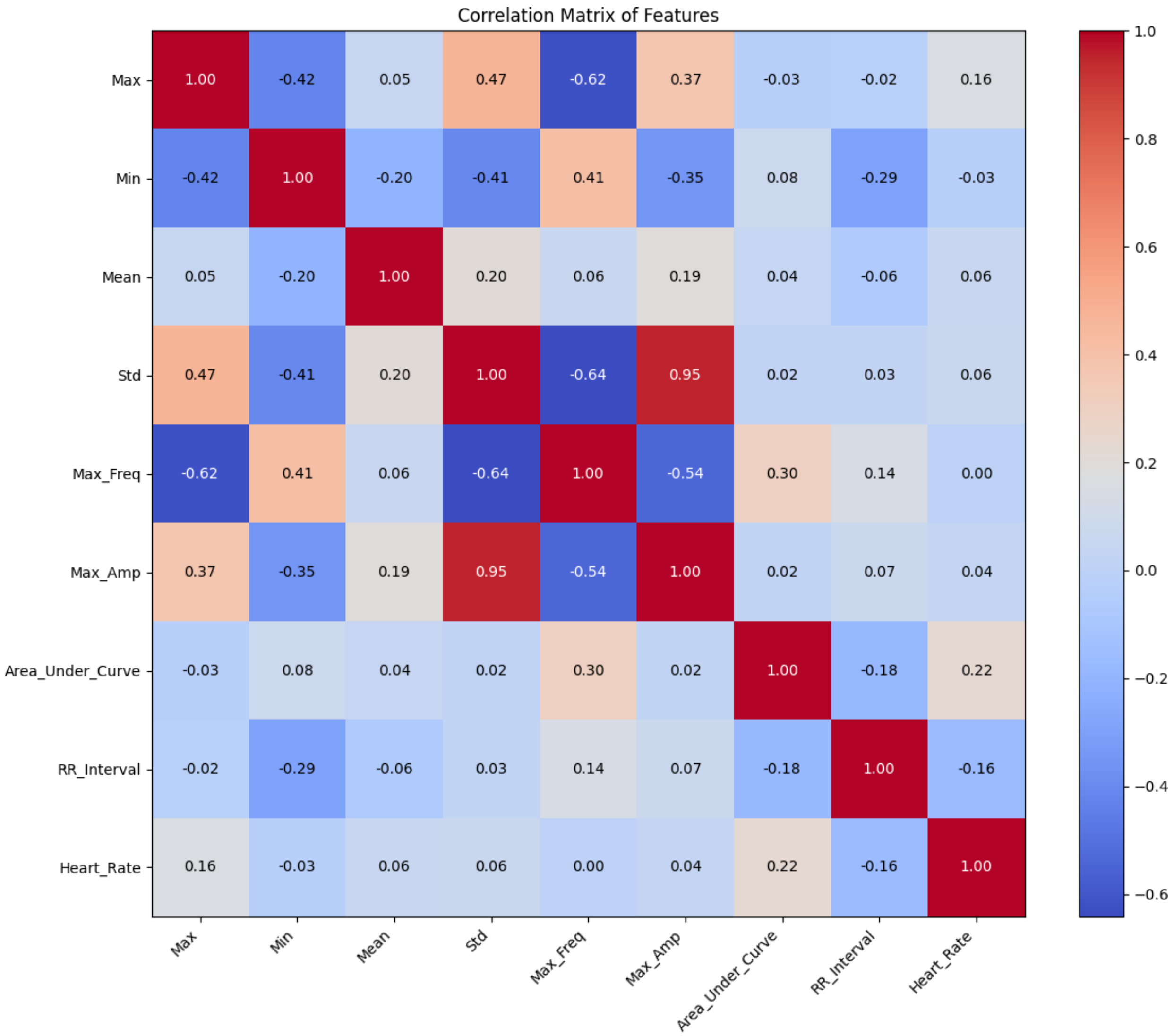
The results of the cross-validation process revealed the following accuracies for each model:

- Support Vector Machine (SVM): 86%
- Logistic Regression (LR): 70%
- Linear Discriminant Analysis (LDA): 86%
- K-Nearest Neighbors (KNN): 84%
- Decision Trees (CART): 79%
- Naive Bayes (NB): 90%

The weighted average of F1-score results from the cross-validation process was compared across the different models. While Naive Bayes achieved the highest weighted average of F1-score at 90%, both SVM and LDA demonstrated robust performances with a weighted average of F1-score of 86%. KNN followed closely with a weighted average F1-score of 84%. Logistic regression had a lower performance with a weighted average F1-score of 70%, indicating its limitations in this context. The performance results suggest that SVM, LDA, and KNN are strong candidates for further analysis.

Figure 4

Correlation Matrix of Features



	Max	Min	Mean	Std	Max_Freq	Max_Amp	Area_Under_Curve	RR_Interval	Heart_Rate
Max	1.00	-0.42	0.05	0.47	-0.62	0.37	-0.03	-0.02	0.16
Min	-0.42	1.00	-0.20	-0.41	0.41	-0.35	0.08	-0.29	-0.03
Mean	0.05	-0.20	1.00	0.20	0.06	0.19	0.04	-0.06	0.06
Std	0.47	-0.41	0.20	1.00	-0.64	0.95	0.02	0.03	0.06
Max_Freq	-0.62	0.41	0.06	-0.64	1.00	-0.54	0.30	0.14	0.00
Max_Amp	0.37	-0.35	0.19	0.95	-0.54	1.00	0.02	0.07	0.04
Area_Under_Curve	-0.03	0.08	0.04	0.02	0.30	0.02	1.00	-0.18	0.22
RR_Interval	-0.02	-0.29	-0.06	0.03	0.14	0.07	-0.18	1.00	-0.16
Heart_Rate	0.16	-0.03	0.06	0.06	0.00	0.04	0.22	-0.16	1.00

Correlation matrix of extracted ECG features, indicating relationships between time-domain and frequency-domain metrics relevant to sleep apnea detection. The strongest correlations highlight which features tend to vary together, helping explain why some models performed better with this feature combination.

Best Model Selection and Evaluation

Based on the cross-validation results, the Naive Bayes (NB) model was identified as the best-performing model, achieving a weighted average F1-score of 90%. The weighted average of the F1-score was selected over accuracy as it provided the most accurate assessment of model performance, while accuracy alone showed bias and was not as reliable. The NB model was trained on the entire training set and tested on the test set. The model's predictions were compared against the true labels, and its performance was evaluated using the confusion matrix, classification report, and weighted average of the F1-score. Achieving a weighted average F1-score of 90% is promising, especially in the context of medical applications where reliable diagnosis is critical.

Naive Bayes Algorithm

The Naive Bayes (NB) algorithm is a probabilistic classifier based on Bayes' theorem, which ap-plies the ”naive” assumption that the features used for classification are conditionally independent given the class label. This means that the presence or absence of a particular feature does not affect the presence of any other feature. Despite this simplifying assumption, Naive Bayes can perform surprisingly well in practice, especially in cases where the feature independence assumption holds reasonably true (Hastie, 2009; Haykin, 2009).

The fundamental equation behind the Naive Bayes algorithm is derived from Bayes' theorem:

$$P(C/X) = \frac{P(X/C) \cdot P(C)}{P(X)} \tag{1}$$

Where:

- $P(C/X)$ is the posterior probability of class C given the features X.

- $P(X/C)$ is the likelihood, which is the probability of observing the features X given class C.

- $P(C)$ is the prior probability of class C.

- $P(X)$ is the evidence or the total probability of the features.

In the Naive Bayes classifier, the likelihood $P(X/C)$ is calculated as the product of the individual probabilities of each feature given the class label:

$$P(X/C) = P(x_1/C) \cdot P(x_2/C) \cdot \ldots \cdot P(x_n/C) \tag{2}$$

Here, $x_1, x_2, \ldots, x_n$ are the individual features.

The classification rule then involves selecting the class C that maximizes the posterior probability:

$$\hat{C} = \arg\max_{C} P(C/X) \tag{3}$$

Given the Naive assumption, this can be simplified to:

$$\hat{C} = \arg\max_{C} \left(P(C) \cdot \prod_{i=1}^{n} P(x_i/C) \right) \tag{4}$$

Naive Bayes is particularly effective for text classification, spam detection, and, as demonstrated in this study, medical diagnoses, including sleep apnea detection from ECG signals. Its strength lies in its efficiency and performance even with small datasets, as evidenced by its highest weighted average of F1-score in this research (90%) (Hastie, 2009; Haykin, 2009).

Results

This section presents the results of the classification of ECG signals for the detection of sleep apnea. After preprocessing, feature extraction, and model training, the performance of the machine learning models was evaluated using the test set. The evaluation was based on standard classification metrics such as precision, recall, F1-score, and accuracy, along with confusion matrixes to visualize the classification outcomes.

Class Distribution

The class distribution in the training set is as follows:

- Class 0 (Healthy): 10
- Class 1 (Apnea): 25
-

Preprocessing and Feature Extraction Results

The ECG signals were preprocessed by applying a bandpass filter with a lowcut of 0.5 Hz and a highcut of 40 Hz. The filtered signals were then normalized using the StandardScaler. The extracted time-domain features included maximum amplitude, minimum amplitude, mean, and standard deviation, while frequency-domain features included the area under the curve. Addition-ally, RR intervals and heart rates were calculated. These features formed the feature matrix, with each row corresponding to a patient and each column representing a specific feature.

Model Evaluation and Comparison

Six classifiers were evaluated: Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Decision Trees (CART), Naive Bayes (NB), and Support Vector Ma-chine (SVM). Each model was assessed using 5-fold cross-validation with hyperparameter tuning via GridSearchCV. The classification performance metrics, including accuracy, precision, recall, and F1-score, are summarized in Table 1.

From Table 1, Logistic Regression (LR) achieved an accuracy of 0.70, while Naive Bayes (NB) achieved the highest accuracy of 0.90. The support vector machine (SVM) and linear discriminant analysis (LDA) both achieved an accuracy of 0.86. K-Nearest Neighbors (KNN) had an accuracy of 0.84, and Decision Trees (CART) achieved an accuracy of 0.79. These results indicate that NB, SVM, and LDA consistently demonstrated stronger predictive capability compared to the other models.

Model Performance Tables

Table 1 summarizes the weighted average F1-scores for each model.

Table 1

Weighted Average F1-Scores for Each Model

Model	Weighted Avg F1-Score
Logistic Regression (LR)	0.70
Linear Discriminant Analysis (LDA)	0.86
K-Nearest Neighbors (KNN)	0.84
Decision Trees (CART)	0.79
Naive Bayes (NB)	0.90
Support Vector Machine (SVM)	0.86

Confusion Matrix for Naive Bayes (NB)

The confusion matrix for the Naive Bayes (NB) model is shown in Table 2. The confusion matrix provides insight into the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts for the classification task. This breakdown helps illustrate how well the model distinguishes apnea from healthy cases.

Table 2

Confusion Matrix for Naive Bayes (NB)

	Predicted Healthy (0)	Predicted Apnea (1)
Actual Healthy (0)	TN: 4	FP: 1
Actual Apnea (1)	FN: 0	TP: 5

Naive Bayes Model Performance Metrics

The detailed performance metrics for the Naive Bayes (NB) model are presented in Table 3. This table includes precision, recall, F1-score, and support for both classes. The results indicate that the model was able to correctly detect apnea cases with perfect recall (1.00), meaning no apnea samples were misclassified.

Table 3

Performance Metrics for Naive Bayes (NB)

Metric	Healthy (0)	Apnea (1)	Macro Average	Weighted Average
Precision	1.00	0.83	0.92	0.92
Recall	0.80	1.00	0.90	0.90
F1-Score	0.89	0.91	0.90	0.90

Overall, the Naive Bayes (NB) model showed strong performance, achieving an accuracy of 0.90. The SVM, KNN, and LDA models also performed well, with accuracies of 0.86, while Decision Trees yielded good performance as well, but with slightly lower accuracy compared to the best models. These findings suggest that several models are viable for detecting sleep apnea using ECG data, although NB demonstrated the most balanced performance across all evaluated metrics.

Discussion

Effectiveness of ECG Signals in Sleep Apnea Detection

Overall, the findings indicate that ECG signals can effectively support the detection of sleep apnea events in patients. The results show that among the machine learning models applied, the Naive Bayes (NB) model achieved the highest performance, with a weighted average of F1-score of 90%, followed closely by the Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA), both with weighted averages of F1-scores of 86%. These findings suggest that ECG-based detection could serve as a promising and cost-effective alternative to traditional polysomnography (PSG) for sleep apnea diagnosis. However, it is important to note that these results should be interpreted with caution due to the small dataset, potentially leading to overfitting and inflated performance metrics. Specifically, the success of models like Naive Bayes and others might be attributed to the specific characteristics of the dataset rather than generalizable patterns, emphasizing the necessity for validation with larger, more diverse patient populations to confirm these findings. Nonetheless, the results provide valuable insights and indicate that ECG signals hold significant promise for non-invasive apnea detection.

Interpretation of Results

Figures 1 and 3 present the preprocessing steps and feature extraction outcomes, which were crucial in enhancing model performance. The normalization of ECG signals effectively reduced amplitude variations, ensuring consistency across the dataset. As shown in Figure 3, the box plots of the extracted features (mean, standard deviation, max, min, RR interval, heart rate, area under the curve, max frequency, and max amplitude) highlight the variability between healthy and apnea classes. The filtering process significantly improved the accuracy of capturing features such as the RR interval and heart rate, both of which displayed notable differences between the two classes. These features served as critical indicators for classifying sleep apnea events, with the reduced variability within each class further demonstrating the effectiveness of the preprocessing steps.

The correlation matrix, shown in the correlation matrix image (Figure 4), highlights significant relationships between the extracted features, revealing that RR intervals and specific frequency-domain components are particularly indicative of apnea events. This clustering of related features suggests that dimensionality reduction techniques could help simplify the model and improve interpretability while retaining relevant information.

The Naive Bayes (NB) model, which demonstrated the highest weighted average of F1-score, benefited from the probabilistic nature of the algorithm in handling the extracted ECG features. SVM and LDA also performed well, showcasing the potential of linear and discriminant analysis-based methods in classifying apnea events. Although the K-Nearest Neighbors (KNN) model and Decision Tree (CART) models achieved reasonably high weighted averages of F1-scores, they are more susceptible to overfitting in small datasets, as evidenced by slight performance drops during cross-validation. Logistic Regression (LR) performed the lowest with a 70% weighted average of F1-score, likely due to the linear decision boundary limitations, which may not capture the complexity of the ECG features. Taken together, these performance patterns indicate that models capable of capturing nonlinear relationships or leveraging probabilistic assumptions tend to be more effective for ECG-based apnea detection.

Future Directions

While this study demonstrates the potential of ECG signals for detecting sleep apnea, several aspects for future research have been identified to enhance the model's accuracy and generalizability. First, future studies should aim to expand the dataset. The current dataset includes 35 patients (25 with sleep apnea and 10 healthy controls). Increasing the sample size and incorporating more diverse patient populations will be critical for validating the model's performance in broader clinical settings. This expansion will help improve the statistical power of the findings and en-sure that the models generalize well to different demographic groups. Increasing the sample size and incorporating a broader demographic range would allow more robust validation and improve applicability in clinical settings.

Additionally, incorporating demographic information such as age, gender, and BMI is also essential, as these factors are known to influence sleep apnea prevalence and severity (Punjabi, 2008; Ling, 2023). Including these variables as features could provide deeper insights into their interactions with ECG-based detection methods and allow the model to be more personalized and effective across varied populations.

While this study focuses solely on ECG signals, integrating additional physiological signals, such as respiratory effort and oxygen saturation, could further enhance the detection of sleep apnea events. Such multi-signal approaches would provide a more comprehensive view of the patient's physiological state and improve the robustness of the classification model (Penzel, Schöbel, & Fietze, 2018).

To address concerns about overfitting, future work should explore more sophisticated validation techniques, such as stratified cross-validation. Additionally, incorporating feature reduction methods like principal component analysis (PCA) or recursive feature elimination (RFE) could improve model efficiency by reducing complexity while maintaining classification performance (Jaadi, 2019; Smolic, 2024). Applying these approaches can simplify models, improve interpretability, and reduce the risk of overfitting in small datasets.

Lastly, refining the feature extraction methods and exploring alternative machine learning algorithms will be crucial in enhancing accuracy and generalizability. By addressing these aspects, the development of a more accessible, reliable, and cost-effective ECG-based diagnostic tool for sleep apnea detection can be realized, improving patient care and outcomes.

Limitations of the Study

Several limitations must be acknowledged in this study. The dataset comprised only 35 patients (25 with sleep apnea and 10 healthy patients), which may limit the generalizability of the results. This relatively small dataset also restricts the statistical power of the findings, potentially inflating the reported model performance (Cao, Chen, & Katz, 2024). Expanding the dataset with more diverse patients, in terms of size and demographics, is critical for validating the model in clinical settings.

Additionally, demographic information such as age, gender, and BMI, which influence sleep apnea prevalence, was not included (Punjabi, 2008; Ling, 2023). Incorporating these variables could provide insights into their interaction with ECG-based sleep apnea detection and im-prove the model's ability to generalize across different populations.

Another limitation is the focus solely on ECG signals. While this simplifies the detection process, other physiological signals, such as respiratory patterns and oxygen saturation, are highly relevant to sleep apnea detection (Penzel, Schöbel, & Fietze, 2018). Future research should explore the integration of these complementary signals to improve diagnostic accuracy and robustness.

Lastly, the machine learning models used in this study exhibited signs of possible overfitting. Although cross-validation techniques were applied, the small sample size may have affected the reliability of these models. More sophisticated validation techniques, such as stratified cross-validation or bootstrapping, should be considered to better evaluate the generalizability of the models in future studies (Ying, 2019).

Conclusion

In conclusion, this study demonstrates the feasibility of using ECG signals to detect sleep apnea events, with the Naive Bayes (NB) model showing particularly strong performance. The general pipeline of this project began with signal acquisition, utilizing the Apnea-ECG Database, which provided a robust online dataset containing ECG recordings from diverse patients.

Following the data acquisition, preprocessing steps, such as filtering, normalization, and segmentation of ECG signals, were performed to enhance the quality of the data and minimize noise. This preprocessing was closely integrated with feature extraction, where significant time-domain and frequency-domain metrics were derived, including heart rate, RR intervals, and specific frequency components. These features are crucial for effectively training machine learning models.

The final phase involved employing various supervised machine learning algorithms, such as Naive Bayes, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), to classify apnea events based on the extracted features. The use of hyperparameter tuning and cross-validation further optimized the model performance, leading to the identification of the Naive Bayes model as the best performer.

Although the findings emphasize the possible benefits of ECG-based approaches, continued refinement of feature extraction methods, expanded datasets, and integration of additional physiological signals remain essential for improving diagnostic accuracy and clinical applicability. By advancing these areas, ECG-based tools may become more accessible, reliable, and useful for the early detection of sleep apnea, ultimately enhancing patient care and benefiting healthcare providers alike.

Acknowledgment

The author would like to thank Dr. Shadi Ghiasi for her invaluable guidance and mentorship throughout this research project.

This research was conducted as part of the CCIR Program, focusing on artificial intelligence in healthcare: biomedical sensors, signal processing, and machine learning.

The author thanks PhysioNet for providing access to the Apnea-ECG Database, which was instrumental in this study.

References

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection.

Butterworth Filter - an overview — ScienceDirect Topics. (n.d.). Www.sciencedirect.com. https://www.sciencedirect.com/topics/engineering/butterworth-filter

Cao, Y., Chen, R. C., & Katz, A. J. (2024). Why is a small sample size not enough?. *The Oncologist, 29*(9), 761-763.

Hastie, T. (2009). The elements of statistical learning: data mining, inference, and prediction.

Haykin, S. (2009). *Neural networks and learning machines, 3/E*. Pearson Education India.

Jaadi, Z. (2019). A step by step explanation of principal component analysis. *Towards Data Science, 1-9.*

Keylabs. (2024, September 2). *Types of Classification Models: Binary, Multiclass&Multilabel*. Keylabs: Latest News and Updates. https://keylabs.ai/blog/types-of-classification-models-binary-multiclass-and-multilabel/

Ling, V. (2023). Sleep apnea statistics and facts you should know. *National Council on Ageing.*

Machine Learning - an overview — ScienceDirect Topics. (n.d.). Www.sciencedirect.com. https://www.sciencedirect.com/topics/earth-and-planetary-sciences/machine-learning

Mitchell, T. M. (1997). Does machine learning really work?. *AI magazine, 18*(3), 11-11.

Penzel, T., Moody, G. B., Mark, R. G., Goldberger, A. L., & Peter, J. H. (2000, September). The apnea-ECG database. In *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)* (pp. 255-258). IEEE.

Penzel, T., Schöbel, C., & Fietze, I. (2018). New technology to assess sleep apnea: wearables, smartphones, and accessories. *F1000Research, 7*, 413.

Punjabi, N. M. (2008). The epidemiology of adult obstructive sleep apnea. *Proceedings of the American Thoracic Society, 5*(2), 136-143.

Sheta, A., Turabieh, H., Thaher, T., Too, J., Mafarja, M., Hossain, M. S., & Surani, S. R. (2021). Diagnosis of obstructive sleep apnea from ECG signals using machine learning and deep learning classifiers. *Applied Sciences, 11*(14), 6622.

National Heart, Lung, and Blood Institute. (2022, March 24). *Sleep apnea - what is sleep apnea?*

Www.nhlbi.nih.gov. https://www.nhlbi.nih.gov/health/sleep-apnea

Smolic, H. (2024, September 25). *Feature selection*. Graphite Note. https://graphite-note.com/feature-selection/

Solomon Jr, O. M. (1991). PSD computations using Welch's method. *NASA STI/Recon Techni-cal Report N, 92*, 23584.

Sulistyo, B., Surantha, N., & Isa, S. M. (2018). Sleep apnea identification using HRV features of ECG signals. *Int. J. Electr. Comput. Eng, 8*(5), 3940.

Wang, T., Lu, C., & Shen, G. (2019). Detection of Sleep Apnea from Single-Lead ECG Sig-nal Using a Time Window Artificial Neural Network. *BioMed research international, 2019*(1), 9768072.

Yeh, C. Y., Chang, H. Y., Hu, J. Y., & Lin, C. C. (2022). Contribution of different subbands of ECG in sleep apnea detection evaluated using filter bank decomposition and a convolu-tional neural network. *Sensors, 22*(2), 510.

Ying, X. (2019, February). An overview of overfitting and its solutions. In *Journal of physics: Conference series* (Vol. 1168, p. 022022). IOP Publishing.

Zarei, A., Beheshti, H., & Asl, B. M. (2022). Detection of sleep apnea using deep neural net-works and single-lead ECG signals. *Biomedical Signal Processing and Control, 71*, 103125.

Zhu, J., Zhou, A., Gong, Q., Zhou, Y., Huang, J., & Chen, Z. (2022). Detection of sleep ap-nea from electrocardiogram and pulse oximetry signals using random forest. *Applied Science 12(9), 4218.*