scientific reports

OPEN

Integrating physiological signals for enhanced sleep apnea diagnosis with SleepNet

Prashant Hemrajani[1,6], Vijaypal Singh Dhaka[1,6], Geeta Rani[1,6], Sahil Verma[2], Kavita[2], Marcin Woźniak[3✉], Jana Shafi[4] & Muhammad Fazal Ijaz[5✉]

Sleep apnea, a prevalent respiratory disorder, poses significant health risks, including cardiovascular complications and behavioral issues, if left untreated. Traditional diagnostic methods like polysomnography, although effective, are often expensive and inconvenient. SleepNet addresses these issues by introducing a new multimodal approach tailored for precise sleep apnea detection. At its core, the framework utilizes a fusion of one-dimensional convolutional neural networks (1D-CNN) and bidirectional gated recurrent units (Bi-GRU) to analyze single-lead electrocardiogram (ECG) recordings, yielding an accuracy of 95.08%. When the model is enriched with additional physiological signals—namely nasal airflow and abdominal respiratory effort—the performance further rises modestly to 95.19%. This multimodal strategy surpasses the performance of existing unimodal approaches, yielding enhanced sensitivity and specificity rates of 96.12% and 93.45%, respectively. When compared to previous studies, SleepNet represents a substantial leap forward in diagnostic efficacy, showcasing the transformative potential of integrating multiple data streams for sleep apnea detection. The results highlight the promise of deep learning methodologies in advancing this domain and lay a robust foundation for subsequent research.

Keywords Deep learning, Obstructive sleep apnea, Multimodal, Signal processing, Electrocardiogram

Sleep apnea (SA) is a widespread and potentially serious disorder characterized by repeated interruptions in breathing during sleep, each pause lasting longer than ten seconds. These interruptions disrupt the normal sleep cycle and lead to significant drops in oxygen levels reaching vital organs such as the brain. As a result, individuals often experience fragmented rest, daytime drowsiness, and face a higher risk of developing cardiovascular and metabolic complications. Unfortunately, SA is often overlooked because conventional diagnostic methods have notable limitations. There are three primary forms of SA: obstructive sleep apnea (OSA), central sleep apnea (CSA), and mixed sleep apnea. OSA, the most common variant, arises when the upper airway becomes physically blocked. Unlike simple snoring—which is usually harmless and linked to factors like nasal congestion or poor sleep habits—OSA causes more severe breathing interruptions. Clinicians gauge OSA severity using the Apnea-Hypopnea Index (AHI), a measure that reflects how many apnea and hypopnea events occur per hour of sleep. To confirm a diagnosis of sleep apnea, polysomnography (PSG) remains the gold standard. This comprehensive test monitors brain waves (EEG), cardiac rhythms (ECG), airflow through the nose, and the respiratory effort of both the chest and abdomen.

However, its high cost, intricate setup, and limited accessibility make PSG impractical for widespread use, particularly for home-based applications. These challenges underscore the urgent demand for innovative diagnostic tools that are cost-effective, portable, and capable of delivering high accuracy without the logistical challenges associated with PSG. Recent developments in wearable devices and contactless diagnostic tools have enabled new methodologies for identifying sleep apnea. Leveraging machine learning and deep learning algorithms proves especially promising, as these approaches can enhance the precision of SA detection while

[1]Computer and Communication Engineering, Manipal University Jaipur, Jaipur, Rajasthan 303007, India. [2]Department of CSE, Rayat Bahra University, Mohali, Punjab, India. [3]Faculty of Applied Mathematics, Silesian University of Technology, 44100 Gliwice, Poland. [4]Department of Computer Engineering and Information, College of Engineering in Wadi Alddawasir, Prince Sattam Bin Abdulaziz University, 11991 Wadi Al Dawasir, Saudi Arabia. [5]Centre for Artificial Intelligence Research and Optimization (AIRO), Design and Creative Technology, Torrens University Australia, 196 Flinders Street, Melbourne, VIC 3000, Australia. [6]Prashant Hemrajani, Vijaypal Singh Dhaka and Geeta Rani contributed equally to this work. ✉email: marcin.wozniak@polsl.pl; muhammadfazal.ijaz@torrens.edu.au

also making it more scalable and accessible. By efficiently processing large, diverse physiological datasets, these advanced technologies offer a promising avenue for detecting SA in a more accessible and cost-effective manner.

In this work, we present SleepNet, a sophisticated deep learning architecture crafted to advance the identification of OSA. SleepNet fuses multiple data streams—ECG tracings, measurements of abdominal breathing effort, and nasal airflow—to create a rich, multimodal input. The core of the framework merges 1D-CNN with BiGRU, enabling an in-depth interpretation of these varied signals. By moving beyond single-source analyses, SleepNet not only boosts diagnostic accuracy but also paves the way for practical, non-clinical deployment, marking a significant leap forward in OSA detection.

Building on prior research, this study highlights SleepNet's innovative contributions and situates it within the broader context of OSA detection. Prior research, including the works of Gutta et al.[1] and Shen et al.[2], has highlighted the utility of ECG signals in detecting OSA by analyzing heart rate variability and arrhythmic patterns linked to apnea events. SleepNet advances this approach by leveraging ECG signals alongside additional physiological inputs for a more robust analysis. Deep learning models particularly DNNs and CNNs have dramatically improved both feature extraction and classification in this field. SleepNet takes these strengths a step further by integrating BiGRUs, a type of recurrent architecture well suited for capturing temporal dependencies. This addition allows the system to recognize intricate sequential patterns and identify even the most complex apnea events. Moreover, similar to wearable systems explored by Surrel et al.[3], which emphasize portability and home-based diagnostics, SleepNet is designed for practical, accessible use beyond clinical environments. This adaptability enhances its potential for widespread adoption, making it a transformative tool for improving OSA diagnosis and management.

However, SleepNet introduces several key innovations that set it apart from existing approaches. While most previous models focus on single-modal data, such as ECG signals, SleepNet uniquely integrates ECG with abdominal respiratory effort and nasal airflow data. This multimodal approach offers a holistic perspective on the physiological changes linked to OSA, thereby enhancing detection accuracy. Additionally, SleepNet's hybrid model design combines 1D-CNN for spatial feature extraction with BiGRU for temporal pattern recognition, unlike other models such as the Multi-Scale Dilation Attention CNN used by Shen et al.[2]. By blending these layers, the network becomes better at identifying the subtle, complex sequences characteristic of apnea events. We also assembled two custom datasets for SleepNet: Dataset A consists solely of 60-second snippets of ECG recordings, while Dataset B pairs those ECG segments with simultaneous measurements of abdominal breathing effort and nasal airflow. These datasets not only provide a broader spectrum of apnea-related variations but also set new standards for dataset quality in the field. Finally, SleepNet's multimodal integration and hybrid architecture lead to enhanced accuracy and resilience, outperforming existing models, including Shen et al.'s MSDA-1DCNN[2], in both accuracy and robustness. Overall, SleepNet builds upon the strengths of existing research while addressing their limitations, positioning itself as a groundbreaking solution for accurate, efficient, and accessible OSA detection.

The prevalence of sleep disturbances is escalating globally[4], with approximately 70 million[5] individuals in the United States alone affected by various sleep-related conditions. Notably, the incidence of Sleep Apnea (SA) syndrome exhibits a concerning trend, with current estimates indicating that about 5% of females and 14% of males in the United States are afflicted[6]. This increasing prevalence is evident across diverse global populations, with an estimated one billion individuals affected by Obstructive Sleep Apnea (OSA) worldwide[7]. The impact of sleep disturbances goes beyond mere discomfort, as they are associated with elevated risks of morbidity and mortality[8]. Despite its chronic and potentially severe consequences, timely diagnosis and intervention can successfully manage and treat OSA. The current advancements and research in SA detection are summarized in Table 1, which presents a detailed review of various methodologies employed for OSA detection, focusing on their respective datasets and contributions to improving diagnostic accuracy. The table compiles numerous studies

Authors (year)	Techniques used	Preprocessing techniques	Dataset	Accuracy (%)	Advantages	Disadvantages
Gutta et al.[1]	Vector-valued Gaussian processes	Feature extraction (time-domain, frequency-domain)	PhysioNet Apnea-ECG	82.33	Effective for continuous data, robust performance	Limited to specific dataset, not scalable for larger data
González et al.[13]	Binary Classification (FBanks, CC, DFA)	Normalization, feature selection	PhysioNet OSA	84.76	Simple, interpretable results	May miss complex patterns due to simplified features
Li et al.[14]	Deep Neural Network and Hidden Markov Model	Time-domain feature extraction, FFT	Custom ECG dataset	84.76	High accuracy, deep learning for complex patterns	Requires large training datasets, computationally expensive
Surrel et al.[3]	Support Vector Machine (SVM)	PCA, feature scaling	PhysioNet Apnea-ECG	88.20	Efficient for small datasets, high accuracy	SVM sensitive to feature selection
Papini et al.[10]	10-fold cross-validation	Data normalization	Various clinical datasets	88.30	Generalizable across datasets	Requires extensive cross-validation, time-consuming
Wang et al.[15]	Deep Residual Network	Raw ECG signal input, denoising	MIT-BIH ECG dataset	83.03	Robust to noise, deep learning approach	May require large data for training
Singh et al.[16]	Pre-trained AlexNet	Data augmentation, feature extraction	PhysioNet Apnea-ECG	86.22	Transfer learning improves accuracy, fast computation	Dependent on pre-trained models, may not generalize well
Feng et al.[12]	FSSAE + TDCS classification	Feature scaling, dimensionality reduction	Sleep-ECG dataset	85.10	Novel technique for improved classification	Computationally expensive, complex architecture
Shen et al.[2]	Multi-Scale Dilation Attention CNN	Feature extraction, noise reduction	Sleep-ECG dataset	89.40	High accuracy, robust to noisy data	Complex model, requires large dataset

Table 1. Research on detecting obstructive sleep apnea with single lead ECG signals.

that explore the use of single-lead ECG signals for detecting OSA, providing valuable insights into their potential for enhancing understanding and precision in this domain. Each study, conducted by different researchers across various years, employed unique methodologies to tackle the challenges of OSA detection. For example, in 2017, Gutta et al.[1] utilized Vector Valued Gaussian Processes, achieving an accuracy of 82.33%. Similarly, González et al. in the same year implemented binary classification with specific features, yielding an accuracy of 84.76%. In 2018, Li et al.[9] combined Deep Neural Networks with Hidden Markov Models, also achieving an accuracy of 84.76%. Surrel et al.[3] in the same year employed Support Vector Machines in a wearable system, achieving an accuracy of 88.20%. Furthermore, Papini et al.[10] in 2018 focused on 10-fold cross-validation, achieving an accuracy of 88.30%. In 2019, Wang et al.[11] utilized Deep Residual Networks, achieving an accuracy of 83.03%, while S. A. Singh et al used a pre-trained AlexNet, resulting in an accuracy of 86.22%. Moreover, in 2021, Feng et al.[12] introduced a novel approach combining auto-encoders and cost-sensitive classification, achieving an accuracy of 85.10%. Finally, Shen et al.[2] in 2021 employed a multi-scale dilation attention CNN with weighted-loss classification, achieving the highest accuracy of 89.40%. These studies collectively showcase a variety of methodologies and accuracy rates in OSA detection using single lead ECG signals, underscoring the ongoing endeavors to advance diagnostic capabilities in this domain.

Essential for the prevention and management of cardiovascular, behavioral, and other health issues related to sleep apnea is the creation and use of accurate, affordable, and portable diagnostic tools. PSG continues to be the definitive method for diagnosing SA, utilizing measures like the AI, HI, and the AHI to assess the condition's severity clinically. Even with its remarkable precision, PSG is a demanding and costly procedure, necessitating thorough observation of signals through multiple electrodes and wires for about 10 hours. This complexity emphasizes the necessity for easier and more affordable methods for detecting obstructive sleep apnea (OSA). In reply, researchers have explored single-lead signal techniques like pulse oximetry[17], photoplethysmography (PPG), electrocardiogram (ECG)[18,19], tracheal body sound[20], and electroencephalogram[19,21], as alternative approaches for identifying sleep disorders. These methods demonstrate potential for simplifying diagnosis, enhancing accessibility and reducing inconvenience for patients. OSA is marked by repeated disruptions in standard breathing rhythms while sleeping. By directly assessing respiratory effort via sensors that track chest and abdominal movements, clinicians can detect tangible signs of apnea events, providing a more accurate evaluation of breathing irregularities than indirect measures such as ECG signals. Data on respiratory effort is essential for accurately assessing the AHI, which gauges the severity of sleep apnea by monitoring the frequency of apnea and hypopnea events per hour of the sleep cycle.

The development of ML and DL models for OSA detection offers the potential for cost-effective alternatives to PSG. These models could facilitate convenient and effortless detection of OSA in home settings, enabling individuals to seek specialist consultations if apnea is detected. However, challenges persist. For instance, some models, like the approach presented by[22], achieve moderate accuracy levels of 82%, underscoring the need for further advancements before practical deployment. Innovative strategies, such as the reversed-pruning method by[23] and network optimizations by[24], highlight ongoing efforts to improve the efficiency and effectiveness of SA detection models. Despite the availability of numerous datasets, there remains a surprising gap in research on multimodal methodologies for sleep apnea detection. This study primarily focuses on OSA, a condition where the normal airway passage is obstructed, leading to irregular breathing during sleep. OSA typically occurs when the muscles supporting the throat's soft tissues, including the tongue, relax excessively, causing periodic airway blockages. This study's main contributions are stated below:

- Preparation of two datasets: Dataset A comprises ECG signals for a 60-second time frame per annotation, while Dataset B includes ECG signals combined with abdominal and nasal respiratory effort signals.
- Proposal of an advanced approach to integrate ECG, abdominal respiratory effort, and nasal airflow signals, providing deeper insights into the intricate patterns of OSA.
- Development of a novel multimodal deep learning model, SleepNet, which leverages features from 1D-CNN and BiGRU for effective OSA detection.
- Demonstration of superior accuracy in OSA detection compared to recent models, showcasing SleepNet's capability in diverse real-world scenarios.

Related works

Sleep apnea represents a widespread sleep disturbance marked by frequent interruptions in breathing that can give rise to serious health complications, including cardiovascular disease, metabolic imbalances, and cognitive impairment. The most common variant OSA happens when the muscles in the throat relax excessively during sleep and block the airway. Detecting this condition promptly and accurately is crucial for effective treatment and for preventing further health risks. PSG remains the gold standard for diagnosis, as it monitors multiple physiological signals such as electroencephalogram, electrocardiogram, pulse oximetry, and airflow to guarantee accurate detection. Nevertheless, PSG demands specific tools and skilled staff, restricting its practicality for monitoring at home. This has generated interest in creating alternative diagnostic methods that are accurate and feasible for application in non-clinical environments. Due to the limitations of traditional diagnostic techniques, there has been considerable demand for novel, non-invasive, and user-friendly technologies that can identify sleep apnea in home settings. Wearable technology and remote tracking systems have surfaced as encouraging options, providing ongoing, unobtrusive observation of physiological signals. These technologies aim to deliver accurate evaluations of sleep apnea, enhancing patient comfort and ease of use, thus serving as a practical choice for home testing. Among the physiological indicators examined for sleep apnea detection, ECG and respiratory data are especially important. ECG signals are simple to obtain and offer valuable information about heart function, making them perfect for wearable technology. Studies show that combining ECG data with additional signals like respiratory effort and oxygen saturation greatly improves detection

accuracy by offering a comprehensive perspective on physiological changes linked to sleep apnea. The rise of deep learning has transformed medical diagnostics, such as sleep apnea detection, with methods like CNNs and RNNs showcasing remarkable skill in recognizing intricate patterns in physiological data. These models autonomously derive significant features from unprocessed data, enhancing diagnostic accuracy while reducing the necessity for manual feature development. SleepNet, an innovative model, utilizes multimodal data—ECG and breathing signals—to enhance sleep apnea identification. It utilizes CNNs for extracting spatial features and BiGRUs to examine temporal patterns, thereby capturing both spatial and temporal dependencies for a thorough comprehension of the physiological changes associated with sleep apnea. Through the combination of multimodal inputs and sophisticated deep learning methods, SleepNet provides a strong and user-friendly approach to improving diagnostic precision. This research assesses the efficacy of SleepNet in identifying sleep apnea, contrasting its results with current techniques. It highlights the benefits of a multimodal strategy and the transformative influence of deep learning in enhancing diagnostic accuracy in sleep medicine. Through the use of various physiological signals and advanced deep learning techniques, SleepNet seeks to establish a new benchmark for detecting and managing sleep apnea. Future studies will concentrate on confirming the model's effectiveness among diverse patient groups and clinical environments to guarantee its relevance and dependability in practical situations.

Accurate diagnosis of sleep apnea is essential for properly handling its related complications. Important methods for detecting sleep apnea include various strategies, ranging from surveys to sophisticated imaging and signal processing techniques. Questionnaires[25] offer an economical method to assess individuals for sleep-related problems, while medical image analysis[26] has become a powerful instrument in identifying sleep apnea, especially in more serious instances. This method also provides distinct perspectives on the anatomical alterations that happen during apnea events. The recognized gold standard for observing sleep apnea is PSG[27], which includes a variety of physiological signals like EEG, ECG, pulse oximetry, arterial blood oxygen levels, airflow, and nasal flow assessments. Although PSG offers unmatched diagnostic precision, it is inappropriate for non-clinical settings, like home monitoring. This constraint highlights the increasing need for wearable and non-contact sleep technologies that allow for discreet monitoring without the need for direct oversight by healthcare providers[28]. Among the physiological signals investigated for sleep apnea identification, ECG is prominent because of its practicality in wearable monitoring devices. Numerous single-lead OSA detection investigations have utilized ECG and pulse oximetry data to identify essential features and recognize patterns suggestive of OSA occurrences. For instance[14], employed a Discriminative HMM with ECG signals to identify OSA, but this approach only yielded binary results and failed to evaluate the severity of OSA events. In a different research[14], integrated single-lead ECG data with DNN and HMM models, enhancing performance through the addition of SVM, ANN and decision fusion techniques[20]. These developments emphasize the possibility of utilizing single-lead ECG signals for identifying sleep apnea in wearable devices, opening the door for more affordable and accessible diagnostic methods. Nonetheless, additional studies are needed to enhance these techniques and confirm their dependability in various patient groups and environments. Nonetheless, this approach faced limitations in differentiating between various disorders. Chen et al.[29] utilize a CNN-BiGRU based model for spatiotemporal learning on two datasets to achieve a good prediction score of SA prediction. While it has been established that sleep apnea can be accurately predicted and diagnosed using cardiac activity through ECGs, it is also essential to reiterate that it is a respiratory condition and studying airflow and respiration—based movements can help improve the accuracy of pre—diction. Avcı and Akbaş[30] utilize chest, abdominal, and nasal signals in an ensemble method to classify sleep apnea in minute-based occurrences. Using nasal airflow signals[31], propose an automated SVM-based algorithm that can alert users through an early-warning system. Thommandram et al.[32] focus on features that are clinically recognizable and use k-nearest neighbours to classify sleep apnea. The effective identification and management of sleep apnea, particularly its most common form, OSA is a critical aspect of healthcare due to the condition's potential to cause significant health com- plications. The current landscape of SA detection methods is diverse, ranging from cost-effective questionnaires to sophisticated medical image analysis and PSG - the latter being the clinical gold standard. PSG, which includes an array of biological signals like ECG, EEG, and oximetry, is highly effective but not suitable for non-clinical, everyday environments, spurring the need for more accessible monitoring technologies. Recent advancements in SA detection focus significantly on wearable and non-contact technologies that offer unobtrusive monitoring. In this context, ECG emerges as a particularly promising signal for use in wearable devices. Studies leveraging single- lead ECG, pulse oximetry data, and respiratory effort have been effective in identifying OSA, employing advanced computational models like Hidden Markov Models (HMM), DNN, and CNN. These models analyze various signal features, striving to predict and assess OSA incidents with varying degrees of success. This diverse array of research efforts underscores a significant commitment to improving OSA detection, with a focus on developing models that are not only accurate but also practical for everyday use. Current trends in research, as highlighted in the two tables, demonstrate the evolving nature of OSA detection, from leveraging advanced signal processing techniques to employing deep learning models. These studies contributes uniquely to the broader goal of creating more effective and user-friendly diagnostic tools. While a number of studies such as[13,29,33,34] achieve more than 90% accuracy in detecting apnea episodes using ECG signals, some researchers investigate other forms of signal inputs. McClure et al.[35] for instance, attempts to explore the usage of chest and abdominal movements to analyze breathing patterns. However, given the complexity of both the disorder and its diagnosis, it is imperative to dissect the various types of available signals and how a multimodal approach can further improve detection accuracy. This section dives into a detailed analysis of the database used to create the two datasets for the study, the machine learning algorithms explored, and the specifics of the proposed model, SleepNet. The training procedure is also briefly discussed. The model architecture is described with an overview of the diagnosis process given in Fig. 1. Figure 1 outlines the process of diagnosing obstructive sleep apnea (OSA) using the SleepNet model. The figure highlights the integration of diverse physiological inputs
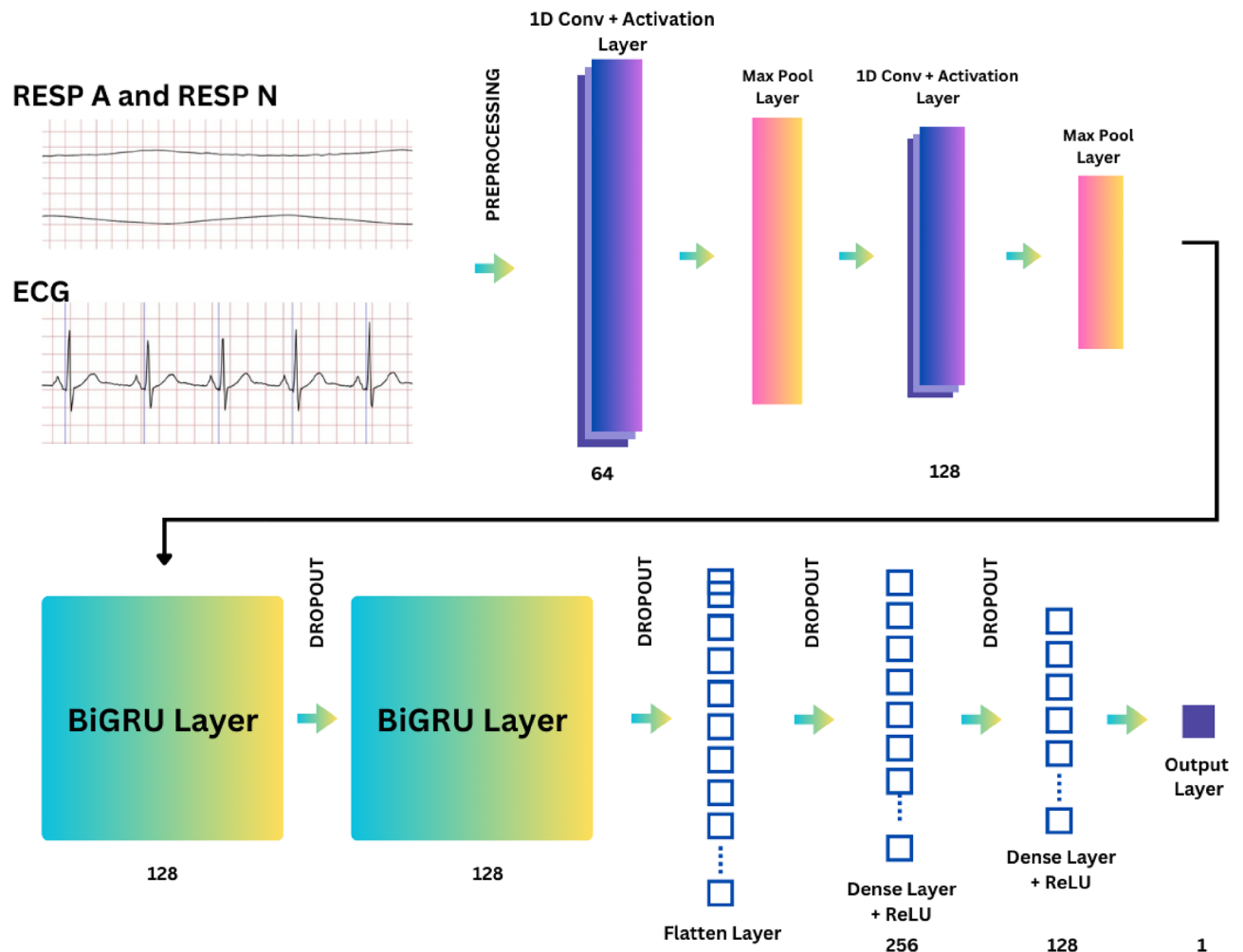
Fig. 1. Overview of OSA diagnosis using SleepNet.

such as ECG and respiratory signals, which together enhance the model's predictive accuracy. The depiction underscores the importance of a multimodal strategy in achieving precise diagnoses, paving the way for timely and effective therapeutic interventions for OSA.

Methodology
Overview of the SleepNet model
The SleepNet model presents an innovative method for detecting sleep apnea by utilizing various physiological signals to improve diagnostic precision. At the heart of its design is a hybrid framework that merges a 1D-CNN with a Bi-GRU. This combination allows the model to obtain spatial features via convolutional layers and assess temporal dependencies with the Bi-GRU element. Through the analysis of signals like ECG, abdominal respiratory effort, and nasal airflow, SleepNet delivers a thorough evaluation of the physiological patterns related to sleep apnea. SleepNet, as a comprehensive deep learning framework, reduces the necessity for manual feature engineering, which allows it to be versatile in practical applications. The model's flexibility and scalability enable it to manage various datasets and extra signal modalities, improving its usefulness in clinical and home monitoring settings. Preliminary assessments indicate that SleepNet surpasses current techniques in accuracy, sensitivity, and specificity, providing a valuable resource for the early and dependable detection of sleep apnea. SleepNet connects traditional clinical diagnostics with wearable solutions by enabling non-invasive and affordable monitoring, creating a path for effective and accessible management of sleep apnea.

- **Hybrid architecture**: SleepNet employs a hybrid architecture that synergizes CNNs with BiGRUs to enhance sleep apnea detection. The CNN component plays a crucial role in extracting spatial patterns from input biological signals, such as pulse rate fluctuations and airflow changes. These spatial features provide critical insights into the structural patterns within the data. The BiGRU component complements this by capturing temporal dependencies, processing sequential data bidirectionally to account for patterns occurring over time. This dual capability allows SleepNet to identify intricate patterns indicative of sleep apnea episodes with exceptional accuracy, significantly bolstering its diagnostic precision and reliability.

- **Integration of multi-modal data**: A standout feature of SleepNet is its ability to integrate multiple types of physiological data, including ECG, SpO2 (oxygen saturation), and airflow signals. While many existing models rely on single-modal data, the incorporation of multiple physiological signals in SleepNet provides a comprehensive view of the patient's condition. This multi-dimensional approach significantly enhances the model's accuracy and robustness in detecting sleep apnea, making it more effective at identifying various types of apnea events.
- **Improved accuracy and generalization**: SleepNet has been rigorously tested on a wide range of datasets, and its performance metrics, such as accuracy, specificity, and sensitivity, demonstrate its superior ability to identify sleep apnea. The model effectively addresses common challenges encountered in previous approaches, such as overfitting and difficulties with generalization. As a result, SleepNet is well-suited for deployment in diverse clinical settings, where it can offer consistent and reliable performance across varied patient populations.
- **Advanced regularization and learning strategies**: To maintain high performance and prevent overfitting, SleepNet incorporates advanced regularization methods and ensemble learning techniques. These approaches guarantee that the model performs effectively across various datasets and testing conditions, ensuring its consistency and resilience. By applying these methods, SleepNet minimizes the risk of model degradation, ensuring its ability to adapt to new data without compromising on accuracy.
- **Computational efficiency**: Another critical feature of SleepNet is its computational efficiency. The SleepNet model has been optimized for deployment in both research and clinical environments, enabling it to process data in real-time for rapid and accurate diagnoses. This real-time capability is crucial in clinical settings, where healthcare providers must make quick, informed decisions to manage patients effectively. Moreover, the model's efficiency allows it to function well even in environments with limited computational resources, ensuring that high performance and accuracy are maintained without the need for expensive infrastructure. In conclusion, SleepNet presents a cutting-edge solution for sleep apnea detection, leveraging its hybrid architecture, multi-modal data integration, and advanced deep learning techniques. Its capacity to process and analyze diverse physiological signals in real-time positions it as a highly valuable tool in sleep medicine. With its potential to improve diagnostic accuracy and enhance patient care, SleepNet holds great promise for advancing the field of sleep apnea detection and management.

Dataset

The ECG component of our study is derived exclusively from the PhysioNet ECG Sleep Apnea database[36,37]. This repository comprises seventy overnight recordings captured under standardized polysomnography (PSG) conditions, each spanning approximately seven to ten hours. Every recording includes a single-lead electrocardiogram sampled at 100 Hz with 16-bit resolution (nominally 200 A/D units per millivolt). The raw ECG waveforms are stored as binary `.dat` files (e.g., `rnn.dat`), with accompanying `.hea` header files that specify channel names, sampling frequency, gain, and file durations[36]. For each heartbeat, machine-generated R-peak annotations are available as binary `.qrs` files; however, these automatic detections can include false positives or misses, necessitating refinement via Pan–Tompkins or wavelet-based R-peak detectors to ensure accurate R–R interval time series for HRV feature analysis when needed[38]. The learning (training) set comprises thirty-five recordings (IDs `a01–a20`, `b01–b05`, `c01–c10`), divided into three cohorts: Group A includes twenty subjects with moderate-to-severe OSA, Group B includes five subjects with borderline or mild OSA, and Group C includes ten healthy controls[36]. Each ECG recording is partitioned into non-overlapping 60 s epochs, each containing 6,000 samples (60 s × 100 Hz). Minute-by-minute apnea annotations are provided in binary `.apn` files for the learning-set recordings, with a "1" indicating an apnea event in that 60 s interval and a "0" indicating no apnea[39]. To prevent data leakage—especially since two recordings (`c05` and `c06`) originate from the same individual with only an 80 s offset[36]—we strictly enforce subject-level partitioning: all epochs from `a01–a20`, `b01–b05`, and `c01–c10` compose the training set, while all epochs from `x01–x35` (the thirty-five test recordings) remain entirely unseen during training. ECG preprocessing entails a zero-phase band-pass filter between 0.05 Hz and 40 Hz to eliminate baseline drift and high-frequency noise while preserving P–QRS–T morphology. After filtering, the ECG signals are normalized to zero mean and unit variance to standardize amplitude ranges across subjects. For each 60 s epoch, the refined R–R interval time series are extracted for optional HRV-based feature construction, but SleepNet's architecture ultimately operates on preprocessed raw ECG waveforms in an end-to-end fashion, preserving temporal and morphological information crucial for apnea detection.

A subset of eight learning-set recordings (IDs `a01–a04`, `b01`, `c01–c03`) additionally provides synchronized respiratory channels for multimodal analysis[36]. Specifically, **Resp C** (chest respiratory effort) and **Resp A** are measured via inductance plethysmography belts placed around the thorax and abdomen, respectively, capturing the expansion and contraction associated with each breathing cycle. Additionally, **Resp N** (oronasal airflow) is recorded using a nasal thermistor that detects pressure changes caused by inhalation and exhalation. All respiratory signals are sampled at 100 Hz and stored either in combined binary `rnnr.dat` files or as separate `.dat` files, each accompanied by a `.hea` header specifying channel names, sampling rate, and gain[36]. Respiratory preprocessing follows a zero-phase band-pass filter between 0.1 Hz and 15 Hz to isolate the adult respiratory frequency band (approximately 0.2–0.5 Hz) and to remove motion artifacts or high-frequency noise. After filtering, each Resp C, Resp A, and Resp N channel is normalized to zero mean and unit variance, ensuring that amplitude differences across subjects or sensor types do not bias subsequent feature extraction or model training[38]. In **Dataset B (multimodal)**, each 60 s epoch is represented by a 6,000 × 4 matrix comprising ECG, Resp C, Resp A, and Resp N, aligned to the same binary apnea label from the `.apn` file. Because only eight subjects provide respiratory data, we employ LOSO-CV for evaluation: in each fold, one subject's entire set of

epochs (all four channels) is held out for testing, while the remaining seven subjects' epochs form the training set. This LOSO strategy ensures that no individual's data appear in both training and test partitions, yielding an unbiased estimate of SleepNet's ability to generalize to unseen subjects.

Factors influencing SpO2 readings accuracy in sleep apnea detection

The accuracy of SpO2 readings in sleep apnea detection is influenced by various physiological factors that play a critical role in interpreting oxygen saturation levels during different sleep stages.

Physiological conditions
The reliability of SpO2 readings for detecting sleep apnea can be affected by several physiological factors. Oxygen saturation levels tend to vary naturally during sleep, showing distinct patterns between wakefulness, Rapid-Eye-Movement (REM) sleep, and non-REM sleep phases. For those with problem of sleep apnea, these variations can be more significant due to periodic breathing interruptions and altered respiratory patterns. Additional factors like obesity, respiratory ailments, and cardiovascular problems can further influence oxygen saturation levels, complicating the analysis of SpO2 data when diagnosing sleep apnea.

Standardization of PSG procedures

It is crucial to standardize Polysomnography (PSG) methods to guarantee that sleep-related information is gathered, analyzed, and interpreted uniformly and dependably in various environments. This standardization is essential for correctly identifying and handling sleep disorders like obstructive sleep apnea. By following established protocols and guidelines, such as those from the American Academy of Sleep Medicine (AASM), researchers and clinicians can reduce variability in the data, resulting in more dependable findings. Reliable PSG protocols assist in guaranteeing that the data remains comparable among various sleep centers and research studies, enhancing the reproducibility of results and the overall quality of sleep disorder diagnosis and treatment. Standardization helps improve the precision of automated diagnostic instruments, like deep learning models, by confirming they are trained and evaluated on high-quality, consistent data.

Strategies for standardization
Standardizing PSG procedures involves establishing uniform protocols for electrode placement, signal acquisition, data processing, and sleep stage scoring. This consistency ensures that sleep study data are comparable across various healthcare settings. Training sleep technologists and clinicians in standardized PSG protocols, including adherence to recognized scoring criteria like the AASM guidelines, helps maintain consistency in data collection and interpretation.

Quality control measures
Implementing quality control measures in PSG procedures is critical for ensuring the accuracy and reliability of sleep apnea diagnoses. Regular calibration of equipment is essential to maintain the precision of signal acquisition, while following established guidelines for sleep staging and event scoring ensures that data is interpreted consistently. Conducting inter-scorer reliability assessments, where multiple scorers independently assess the same data, helps to identify and minimize discrepancies in scoring, further ensuring the consistency of results. These quality control measures not only enhance the accuracy of sleep apnea diagnoses but also improve the overall reliability of PSG procedures, making it easier to monitor treatment effectiveness and make informed clinical decisions. By reducing variability in PSG results, healthcare providers can confidently rely on the data for more accurate and timely diagnosis, leading to better patient care.

Evaluation metrics for model performance in image classification and object detection

Establishing clear evaluation procedures necessitates identifying benchmark datasets, defining criteria for model validation and testing, and conducting thorough performance evaluations using techniques like cross-validation. Evaluation metrics are essential for measuring the effectiveness of models in image classification and object detection tasks.

Common metrics include:

Accuracy
Accuracy measures the fraction of correctly classified instances out of all predictions made. Although it serves as a general indicator of performance, it may be misleading when classes are imbalanced. Formally,

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{1}$$

F1 score
The F1 Score represents the harmonic mean of precision and recall, providing a single metric that balances both false positives and false negatives. This is particularly useful when dealing with skewed class distributions. Specifically,

$$\text{F1 Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{2}$$

Recall
Recall (also known as sensitivity) quantifies a model's capability to identify all relevant positive cases—i.e., how many true positives it captures compared to the total actual positives. This metric is critical when failing to detect a positive case carries a high penalty (for example, missing a sleep apnea event). It is calculated as

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \qquad (3)$$

Precision
Precision measures the accuracy of positive predictions by determining the proportion of correctly predicted positives among all predicted positives. High precision is crucial when false alarms (false positives) are costly. Formally,

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \qquad (4)$$

Data preparation and pre-processing

The authors began by extracting all annotated ECG signals that were labelled being apneic or not. They conducted a thorough examination of several ECG characteristics, including the intervals between successive ECG pulses (R - R intervals), the amplitudes of each pulse, the variation in pulse amplitudes, and the overall energy of the ECG pulses. Among these parameters, the R - R intervals were identified as particularly indicative in discerning apnea episodes. Two datasets were prepared for training and evaluation:

- **Dataset A**: Contains ECG signals annotated for apneic and non-apneic events, with each segment lasting 60 seconds. This dataset serves as a baseline for unimodal analysis.
- **Dataset B**: Expands upon Dataset A by incorporating additional respiratory signals (Resp A and Resp N). This multimodal dataset allows for a more comprehensive analysis of sleep apnea episodes.

Each annotation was provided for the subsequent one minute of the ECG signal, totaling 6000 ECG signals (calculated as 60 seconds multiplied by 100 signals per second) for each annotation. Then, these packets of ECG signals were mapped with corresponding annotation to compile a dataset with several rows containing 6000 features each, which was named as Dataset A. Furthermore, the authors expanded their analysis by incorporating additional respiratory data, specifically the Resp A and Resp N signals, available for a subset of eight subjects. Initially, all signals were normalized to a common scale. Given the susceptibility of physiological signals to various types of noise and artifacts, the authors applied a combination of band-pass filtering and adaptive noise cancellation techniques. This led to the creation of Dataset B, which integrates ECG, Resp A, and Resp N measurements against the apnea annotations. To rigorously evaluate the model, the resultant datasets were divided into training, testing, and validation sets, following a 70:9:21 ratio. The exact distribution of data across various classes in each subset for Dataset A is elaborately depicted and discussed in Table 2. This strategic partitioning of data was essential for assessing the model's performance across multiple parameters. This overview highlights the diverse range of methodologies employed over the years to detect OSA, showcasing the evolution of machine learning and deep learning techniques in this field. In 2015, Song et al.[14] used a Discriminative HMM and achieved an impressive accuracy of 97.10%, while Chen et al.[40] also in 2015, employed a SVM, achieving slightly better accuracy at 97.41%. Moving to 2017, M. Cheng et al.[41] utilized Long Short-Term Memory (LSTM) networks for OSA detection, reporting an accuracy of 97.80%, demonstrating the growing sophistication in model selection. In 2019, there was a significant leap in model complexity and performance. R. Stretch used a combination of machine learning algorithms, including Logistic Regression, Artificial Neural Networks, SVM, K-Nearest Neighbors, and Random Forest, but with a comparatively lower accuracy of 80.50%. On the other hand, X. Liang's[42] work that year combined CNN with unfolded bidirectional LSTM, which achieved an outstanding accuracy of 99.80%, underscoring the effectiveness of combining CNNs with RNNs for temporal and spatial feature extraction. Similarly, Wang et al.[15] utilized Deep Residual Networks to achieve an accuracy of 94.40%, while T. Wang introduced a Modified LeNet-5 model for sleep apnea detection, achieving an accuracy of 97.10%. Further advancements were seen in 2020, with Bozkurt et al.[43] developing an ensemble classifier that combined various machine learning algorithms like Decision Trees, K-Nearest Neighbors, and SVMs for SA detection, resulting in an accuracy of 85.12%. Finally, McClure et al.[35] employed a Multi-scale Deep Neural Network combined with a 1D-CNN to classify and detect OSA, CSA, coughing, sighing, and yawning using wearable sensors, achieving a commendable accuracy of 87.0%. These various approaches demonstrate the steady progress in the field, with deep learning models, particularly CNNs, LSTMs, and their hybrid combinations, becoming more dominant in achieving higher detection accuracies.

Comparative benchmarking with prior studies

In order to contextualize SleepNet's performance, we compare it against several notable studies that have reported apnea detection results using the Apnea-ECG benchmark. Table 2 summarizes these prior methods by listing the authors and publication year, the primary model or algorithm they employed, the dataset configuration, the input signal modality, and the detection accuracy achieved. In the remarks column, we also highlight each approach's main strengths or weaknesses. Although every study listed here relies on PhysioNet's Apnea-ECG data for evaluation, they differ in their choice of data-splitting strategies—some use the standard train/test division, while others apply cross-validation (either with or without subject-level partitioning).

Authors	Year	Proposed model	Application	Accuracy (%)	Sensitivity/ specificity (%)	Signal type(s)	Remarks
Song et al.[14]	2015	Discriminative hidden Markov model (HMM)	OSA detection	97.10	82.6	ECG-derived features (RR intervals, ECG-derived respiration)	Hybrid sequence model (HMM for temporal dynamics + SVM classifier); relied on hand-crafted heart rate and respiration features
Chen et al.[29]	2015	Support vector machine (SVM)	OSA detection	97.41	82.5	ECG time- and frequency-domain features	Manual feature extraction; simple, interpretable results; may miss complex patterns
Cheng et al.[44]	2017	LSTM	OSA detection	97.80	N/A	ECG (RR intervals)	Captures temporal dependencies for sequential ECG data; high accuracy but risk of overfitting in limited data scenarios
Stretch[45]	2019	Machine learning algorithms (LR, ANN, SVM, KNN, Random Forest, etc.)	OSA Detection	80.50	52.9	ECG-derived features (RR intervals, R-peak amplitudes)	Ensemble of classic classifiers; robust subject-wise cross-validation yields lower accuracy
Liang[42]	2019	CNN and unfolded bidirectional LSTM	OSA detection	99.80	98.97	ECG-derived HRV (RR intervals)	Deep spatiotemporal network combining CNN+BiLSTM; very high accuracy but subject overlap in CV may inflate performance
Wang et al.[11]	2019	Deep residual network	SA detection	94.40	93.0	ECG-derived HRV (RR intervals)	Robust to noise; uses deep residual CNN on RR data; performance likely inflated due to subject overlap in cross-validation
Wang et al.[15]	2019	Modified LeNet-5	SA detection	97.10	83.1	Raw single-lead ECG	End-to-end ECG analysis with 1D CNN; high accuracy but potential data leakage from subject-inclusive CV
Bozkurt et al.[43]	2020	Ensemble classifier using ML algorithms (Decision tree, kNN, SVMs)	SA detection	85.12	94.9	ECG features	Ensemble of decision tree, kNN, SVM; moderate accuracy; computationally intensive
McClure et al.[35]	2020	Multi-scale deep neural network and 1D-CNN	Classification and detection of OSA, CSA, coughing, sighing, and yawning using wearable sensor	87.00	83.94	ECG + respiratory effort + SpO$_2$	Multimodal wearable sensor data; broad event detection (OSA, CSA, cough, sigh, yawn) using deep architectures; robust evaluation with subject-wise separation

Table 2. Diverse approaches to obstructive sleep apnea detection.

In the literature as shown in Table 2, single-lead ECG approaches exhibit a broad spectrum of reported accuracies, largely influenced by evaluation protocols and model complexities. For instance, studies employing subject-overlapping cross-validation tend to report very high performance: Song et al. achieved 97.10% accuracy using a hidden Markov model combined with an SVM; Chen et al. likewise reported 97.41% with an SVM alone; Cheng et al. observed 97.80% accuracy when using an LSTM network; Liang et al. attained 99.80% by integrating a convolutional neural network with a bidirectional LSTM; Wang et al. reported 94.40% using a ResNet architecture; and Wang et al. achieved 97.10% with a modified LeNet-5. However, because these methods allow recordings from the same individual to appear in both training and test folds, they often overestimate real-world performance. By contrast, studies that strictly enforce subject-independent partitioning present more moderate yet realistic results: Stretch et al. reported 80.50% accuracy using a classical machine-learning ensemble; Bozkurt et al. achieved 85.12% by combining decision trees, k-nearest neighbors, and SVMs; and McClure et al. obtained 87.00% accuracy with a multi-scale deep neural network and one-dimensional CNN applied to wearable signals. These mid-80% figures better reflect a model's capacity to generalize to previously unseen subjects. In this context, SleepNet's ECG-only configuration demonstrates competitive performance, placing it firmly among the leading single-lead approaches under a subject-independent evaluation. By incorporating chest and abdominal respiratory effort signals alongside oxygen saturation in addition to ECG, the multimodal variant achieves even higher detection efficacy. This enhancement underscores the clinical value of adding direct respiratory and SpO$_2$ measurements—features that complement ECG's indirect markers of apnea—and highlights SleepNet's advantage over models that rely exclusively on ECG.

Experimental setup

Our models including both the pre-trained versions and our proposed architecture—were implemented in Python with TensorFlow. Initially, we carried out training on a personal laptop featuring an AMD Ryzen 5 3500U CPU, an NVIDIA 940M GPU (compute capability 5.0), and 16 GB of RAM, using a limited number of epochs. To achieve more extensive training, we then switched to Kaggle's environment, taking advantage of GPUs with compute capability 6.0 and 16 GB of VRAM. In total, each model was trained for 120 epochs to maximize accuracy during training, validation, and testing. The Table 3 displays the distribution of data across two datasets, Dataset A and Dataset B, within a machine learning framework. These datasets are categorized into three groups: Normal, Severely Affected (SA), and Total. Additionally, the data is divided into three parts: For the training subset, there are 11,907 samples in total, with 7,353 labeled as Normal and 4554 as SA. This shows that while a large portion of the training data focuses on normal cases, a significant number of severely affected cases are included to ensure the model can recognize critical conditions. The validation set, used to refine the model's parameters and mitigate overfitting, consists of 1531 samples. Of these, 941 are Normal and 590 are SA, maintaining a distribution ratio similar to that of the training data. This consistent distribution ensures that the

Dataset A and B	Normal	SA affected	Total
Training	7353	4554	11,907
Validation	941	590	1531
Testing	2241	1358	3572

Table 3. Data distribution for Dataset A and B.

validation outcomes accurately reflect the training process. Lastly, the testing subset, which assesses the model's performance on newly unseen data, consists of 3572 samples. This includes 2241 Normal and 1358 SA cases. The distribution in the testing data aligns with that of the training and validation subsets, ensuring a balanced evaluation metric.

Algorithms

This research evaluates various recurrent neural network architectures—namely LSTM, GRU, and BiGRU to handle time-series inputs like ECG recordings. In addition, we investigate convolutional neural network approaches, including pretrained models such as VGG16 and AlexNet, as well as custom 1D-CNNs. A brief overview of each model follows. When choosing between LSTM and GRU-based designs, several considerations arise: the complexity of the task, the nature of the dataset, computational cost, memory requirements, and the trade-off between accuracy and training time. LSTMs excel at modeling long-term dependencies in sequential data thanks to their memory cells and gating mechanisms, which help avoid the vanishing gradient issue. This feature allows them to retain and retrieve critical information over extended sequences, making LSTMs a popular choice for applications like speech recognition, time-series forecasting, and NLP. Conversely, GRU networks boast computational efficiency and fewer parameters compared to LSTMs, rendering them faster to train and deploy in scenarios where memory constraints or speed are critical factors. Researchers assess these considerations based on their specific task requirements and dataset characteristics to determine whether LSTM or GRU networks offer the most optimal performance and efficiency for their deep learning models.

Parameter sharing within neural networks involves employing identical or similar parameters across multiple layers to foster feature reuse and reduce the total number of trainable parameters. This practice facilitates learning hierarchical representations of data and capturing common patterns across different parts of the network. In CNNs, parameter sharing occurs in convolutional layers, where identical filter weights are applied to various spatial locations in the input data. This approach enables the network to efficiently learn spatial hierarchies of features and generalize effectively to new data. Similarly, in RNNs like LSTMs and GRUs, parameter sharing is achieved through recurrent connections that propagate information across time steps, enabling the network to maintain memory and capture temporal dependencies in sequential data. By embracing parameter sharing, neural networks can acquire more robust representations, enhance performance across diverse tasks, mitigate the risk of overfitting, and improve computational efficiency.

In the realm of training neural networks for detecting sleep apnea, researchers typically adhere to a meticulous procedure aimed at enhancing model performance and ensuring its applicability across various scenarios. This rigorous process encompasses several key stages: data preprocessing, model selection, hyperparameter optimization, delineation of training-validation-testing sets, and thorough performance assessment. Data preparation is the foundational step, encompassing operations like normalization, feature extraction, and the treatment of missing values to ensure the dataset remains consistent and reliable. Next, choosing an appropriate network architecture—whether an LSTM or a CNN—depends on the specific requirements of the problem. Tuning hyperparameters, such as learning rate, batch size, and regularization methods, is crucial for improving generalization and preventing overfitting. Splitting the data into training, validation, and test sets allows for an unbiased evaluation of model performance on unseen samples, offering a clearer picture of how it will behave in practical scenarios. Evaluating the model with metrics like accuracy, sensitivity, and specificity provides a comprehensive view of its strengths: accuracy reflects overall correctness, sensitivity measures the ability to correctly identify apnea events (true positives), and specificity quantifies how well the model avoids false alarms (true negatives). By rigorously following this workflow preprocessing data, selecting and tuning models, and assessing performance on separate datasets—researchers can iteratively refine their neural networks and overcome training challenges, ultimately yielding a more dependable and effective sleep apnea detector.

Recurrent neural networks

RNNs are a class of artificial neural networks specifically designed to handle sequences of data. Unlike traditional feedforward networks, RNNs build connections between nodes that form a directed graph, allowing information to be carried forward over time. This built-in "memory" makes them ideally suited for applications where past inputs influence future outputs—such as analyzing time-series signals, processing natural language, or recognizing spoken words. Because RNNs retain information from previous steps, they can capture temporal dependencies that ordinary neural networks cannot. This distinctive characteristic allows RNNs to retain context and utilize it to affect the output at every time step. Fundamentally, the output of an RNN at a specific moment relies on both the present input and the outputs from previous time steps, enabling the model to successfully and efficiently capture and employ long-term dependencies in the data.

- **Long short-term memory**: LSTM, commonly applied to graphical data like ECG, is a sophisticated version of RNN that allows data to endure over time. Its feedback connections enable it to process sequences of data

in their entirety. At the heart of LSTM is a 'cell state,' a memory component that retains its state over time. In an LSTM model, the modification of information in the cell state is controlled through gates, which regulate the flow of information in and out of the cell. This process is facilitated by a combination of pointwise multiplication and a sigmoid layer in the neural network.

- **Gated recurrent unit**: Developed as an optimization to the standard RNN archi- tecture, GRUs[41,46] address the problem of long-term dependency, a challenge where RNNs struggle to carry information across many time steps. The improvement in GRUs is realized through gating mechanisms that control information flow inside the unit. GRUs feature two gates: the update gate, which determines how much the unit's state is refreshed with new data, and the reset gate, which decides the amount of previous information to discard. GRUs are often considered a more streamlined and efficient option compared to LSTMs, as they have fewer parameters. This reduction in parameters can result in quicker training periods and decreased computational complexity.

- **Bidirectional gated recurrent unit**: Bidirectional Gated Recurrent Units (BiGRUs) represent an enhanced version of traditional Gated Recurrent Units (GRUs) aimed at better modeling of sequential data in deep learning applications. Unlike standard GRUs, BiGRUs use a bidirectional structure, allowing the network to obtain contextual insights from both preceding and succeeding sections of a sequence. This is accomplished by employing two distinct GRU units: one that handles the sequence from start to finish (forward GRU), and the other that manages the sequence in reverse (backward GRU). The forward GRU collects information regarding the past context, whereas the backward GRU captures insights about the future context. At every time step, the results from both GRUs are merged, offering a more comprehensive representation of the sequence by integrating context from both sides. This two-way processing is especially beneficial in tasks where comprehending both previous and upcoming states is essential for precise predictions, like in time-series analysis, natural language processing, and speech recognition.

Convolutional neural networks

- **VGG16** The writers utilized the VGG16 model, a renowned 16-layer CNN that has proven highly effective in image classification applications. The structure comprises several convolutional layers succeeded by three Fully Connected layers. The initial two FC layers have 4096 channels each, and their main function is to combine features obtained from the convolutional layers. The third FC layer is designed for 1000-way classification, satisfying the needs of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). A softmax layer is included at the end of the network to generate the final classification probabilities. Nonetheless, the network's depth and the high quantity of nodes in the fully connected layers lead to a significant model size, causing it to surpass 500MB. Even with its substantial size, the VGG16 architecture continues to be a favored option for tasks needing deep feature extraction, owing to its straightforward yet effective design.

- **AlexNet** To build a leaner, faster model, the researchers chose AlexNet—an eight-layer CNN known for balancing depth with computational speed. It features five convolutional blocks, each paired with max-pooling, followed by three fully connected layers that use ReLU activations. A key element is the dropout applied before the first and second fully connected layers, which helps curb overfitting and boosts generalization to new data. Although AlexNet is less deep than architectures like VGG16, it still performs remarkably well on image classification tasks. This mix of efficiency and strong accuracy makes AlexNet ideal for situations where computing resources are tight but high performance is still needed.

Combination of CNN and RNN
Encouraged by the promising outcomes achieved individu- ally with recurrent neural networks and the convolutional neural network architectures the authors combined the strengths of the two into hybrid models.

- **Hybrid AlexNet** This model combines the spatial feature extraction strengths of AlexNet, a CNN, with the temporal processing capabilities of LSTM networks. It operates in two distinct phases. Initially, AlexNet extracts crucial spatial features from the input data through its convolutional layers, which are particularly effective at detecting and isolating important patterns and structures. These extracted features, which represent the spatial characteristics of the data, are then passed on to the LSTM network. The LSTM component of the model processes the sequential nature of the data, capturing long-term dependencies and temporal relationships. By leveraging both AlexNet and LSTM, this hybrid model effectively addresses both spatial and temporal aspects of the data, enhancing the model's overall ability to analyze complex, dynamic patterns. Known for its effectiveness in handling sequential and time-series data, LSTM analyzes these features over time, capturing crucial temporal dependencies that are indicative of apneic events. This fusion of AlexNet and LSTM leverages their respective strengths - AlexNet's ability to recognize detailed features and LSTM's skill in temporal analysis. Similarly, a composite model combining AlexNet with Gated Recurrent Units (Hybrid AlexNet v2) was de- veloped, capitalizing on GRUs' efficiency in sequential data processing.

- **Hybrid CNN** After the success of the hybrid AlexNet models, the researchers extended their exploration into combining other neural network architectures to further enhance the analysis of complex datasets like those found in apnea detection. This development resulted in the creation of two additional hybrid models, each designed to leverage the complementary strengths of CNNs and RNNs. The first model, 1D-CNN + LSTM (Hybrid CNN v1), takes advantage of 1D-CNNs for extracting relevant features from time-series data. The 1D-CNN layers process the data linearly, identifying temporal patterns and key features. Following this, the LSTM network captures long-term dependencies, making it particularly effective for detecting the intricate, sequential patterns indicative of apneic events. The LSTM's ability to retain information over time allows it to recognize and respond to complex temporal changes within the data. The second model, 1D-CNN + GRU

(Hybrid CNN v2), also utilizes 1D-CNNs for initial feature extraction but incorporates GRUs in place of LSTMs. GRUs, which feature a more streamlined gating mechanism compared to LSTMs, strike a balance between computational efficiency and the ability to capture sequential dependencies in the data. While not as complex as LSTMs, GRUs are still highly effective for sequential data processing and require less computational power, making this hybrid model a suitable choice for tasks where efficiency is a priority without sacrificing the ability to model temporal relationships. Both models combine the strengths of CNNs for feature extraction with advanced recurrent networks for temporal analysis, making them well-suited for detecting sleep apnea from time-series data.

Proposed model: SleepNet

By placing a BiGRU layer immediately after the 1D-CNN stages, SleepNet gains a powerful edge in interpreting temporal features for sleep apnea identification. The bidirectional nature of BiGRU means the network analyzes each heartbeat sequence in both forward and reverse directions, allowing it to leverage information from before and after any given moment. In ECG analysis, this is especially important: knowing what precedes and follows a specific heartbeat can be critical when trying to pinpoint subtle apneic episodes. The 1D-CNN front end excels at spotting local ECG characteristics—such as QRS complexes, P waves, and T waves—regardless of where they appear in the signal, thanks to its translation-invariant filters. Unlike traditional pipelines that rely heavily on handcrafted features, these convolutional layers can learn clinically relevant patterns directly from the raw ECG waveform with minimal preprocessing. Research in biomedical signal processing consistently shows that 1D-CNNs outperform conventional feature-engineering approaches in extracting meaningful signal traits. Building on this, the BiGRU layer processes the extracted features bidirectionally, giving the model simultaneous access to both past and future context. This dual perspective is a major advantage when trying to detect sleep apnea, since an apneic event's significance often depends on the sequence of heartbeats surrounding it. Empirical evidence in ECG classification tasks confirms that BiGRU-based architectures achieve better accuracy by leveraging contextual information from both time directions. Although previous studies have experimented with pairing CNNs and BiGRUs for ECG classification, SleepNet distinguishes itself by adopting a carefully streamlined and hyperparameter-tuned design. With additional layers and optimized settings, it is better equipped for realistic deployment and demonstrates higher accuracy in identifying apnea events than earlier models.

SleepNet's architecture is carefully designed to balance thorough feature extraction with efficient classification. It begins with convolutional layers that progressively learn from the raw ECG input. The first convolutional stage uses 64 filters of size 3 to capture basic waveform characteristics, while a second convolution with 128 filters then builds on those low-level patterns to uncover more detailed structures. Interspersed max-pooling steps reduce the spatial size of these feature maps, which not only helps guard against overfitting but also keeps the computational cost in check. Once the convolutional hierarchy has distilled essential ECG features, BiGRU layers take over, processing the data in both forward and backward directions. The initial BiGRU layer draws in temporal relationships from past and future heartbeats, and a dropout layer follows to randomly deactivate some units during training, which further reduces overfitting. A second BiGRU layer deepens the network's understanding of sequence-level dependencies, ensuring that no critical temporal pattern is overlooked. After this, the resulting maps are flattened into a single vector so they can be handled by fully connected (dense) layers. These dense layers, activated by ReLU, introduce nonlinearity and refine the extracted features for the final classification step. Dropout is applied again to improve generalization. In the end, a single-node output layer with a sigmoid activation produces a probability score, making the model suitable for distinguishing between apnea and non-apnea segments. To evaluate SleepNet's performance, metrics such as accuracy, sensitivity, and specificity are used. When comparing it against traditional, single-signal methods, SleepNet consistently outperforms them, demonstrating higher accuracy and a better balance of true-positive and true-negative rates. By combining ECG data with additional physiological inputs, SleepNet leverages complementary information streams to deliver a more reliable diagnosis of sleep apnea. The performance gains evident in these multimodal results underscore the value of integrating multiple signals, showing that SleepNet is a promising tool for real-world clinical and non-clinical applications (Table 4).

Algorithm 1 details SleepNet's workflow, which sequentially combines convolutional and recurrent modules to classify OSA. The process begins with the model receiving the training dataset and ultimately producing OSA predictions. The initial stage comprises two back-to-back blocks of one-dimensional convolutional layers, each using a kernel size of three to scan for local features within the ECG signal. Directly after each Conv1D block, a MaxPooling1D layer with a pool size of two downsamples the feature maps—this both trims unnecessary spatial dimensions and preserves critical signal characteristics, thereby improving computational efficiency. Once these convolutional and pooling layers have distilled fundamental spatial patterns, the network transitions to temporal analysis using two successive BiGRU layers. By processing sequences in forward and reverse directions, each BiGRU unit captures contextual information from both past and future time points—an invaluable capability when dealing with ECG data, where neighboring heartbeats can carry significant diagnostic clues. To discourage overfitting, a dropout layer with a rate of 0.2 follows each BiGRU block. During training, this dropout step randomly deactivates 20 percent of the hidden units, encouraging the network to learn more robust features that generalize well to unseen data.

After uncovering temporal dependencies, SleepNet flattens the BiGRU output into a single vector, preparing it for the final classification stage. This vector is fed into two fully connected (dense) layers, each followed by another dropout at the same 0.2 rate. These dense layers introduce nonlinearity and further refine the learned features. In the final step, a dense output node with a sigmoid activation generates the model's OSA probability score, allowing straightforward binary decision-making (apnea vs. non-apnea). By integrating convolutional blocks for spatial feature extraction with bidirectional recurrent modules for sequence modeling, SleepNet

Layer	Type	No. of filters/cells/units	Kernel size	Activation function
1	Convolution	64	3	ReLU
2	Maxpool	–	2	–
3	Convolution	128	3	ReLU
4	Maxpool	–	2	–
5	Bidirectional GRU	128	–	–
6	Dropout	–	–	–
7	Bidirectional GRU	128	–	–
8	Flatten	–	–	–
9	Dense	256	–	ReLU
10	Dropout	–	–	–
11	Dense	128	–	ReLU
12	Dropout	–	–	–
13	Dense	1	–	Sigmoid

Table 4. SleepNet architecture.

effectively captures both the local waveform details and the broader temporal patterns necessary to detect sleep apnea events. However, this layered, sequential approach does carry computational costs—especially when handling large datasets—since each signal must pass through multiple Conv1D, BiGRU, and dense transformations. While dropout aids in reducing overfitting, careful tuning remains essential to prevent the network from becoming too tailored to the training set. Looking ahead, SleepNet's modular design offers several avenues for enhancement. Researchers might experiment with additional forms of regularization (e.g., weight decay or batch normalization), test alternative model architectures (such as transformer-based blocks), or incorporate other physiological signals—like chest respiratory effort or oxygen saturation—to further enrich the input data. Such extensions could boost generalizability and resilience in clinical settings. Moreover, exploring hybrid approaches that combine SleepNet with simpler, lightweight classifiers might help strike a better balance between performance and deployment efficiency in real-world healthcare environments.

Strengths of the proposed approach

- **Improved diagnostic precision**: One of the standout features of SleepNet is its ability to incorporate various physiological signals, such as ECG, nasal airflow, and abdominal respiratory effort. This integration has significantly boosted the model's diagnostic accuracy, allowing it to achieve an impressive accuracy rate of 95.19%. This surpasses many existing methods that rely on a single signal and underscores the model's potential to provide timely and accurate detection of sleep apnea—a condition that can lead to severe health risks if left untreated.
- **Integration of multimodal data**: SleepNet distinguishes itself by utilizing data from multiple physiological sources, capitalizing on the unique strengths of each signal. This holistic approach provides a more detailed analysis of sleep patterns and respiratory behavior, enabling better differentiation among the obstructive, central, and complex types of sleep apnea. This holistic approach provides a more precise and detailed understanding of sleep apnea events, resulting in improved diagnostic outcomes.
- **Strong performance indicators**: In addition to high accuracy, SleepNet demonstrated excellent sensitivity and specificity rates, key performance indicators for medical diagnostic tools. These metrics emphasize the model's reliability in correctly identifying sleep apnea cases while minimizing the risk of false positives. This fosters patient confidence and supports effective treatment planning by ensuring that patients are correctly diagnosed and subsequently treated in a timely manner.
- **Adaptability for individualized care**: SleepNet's architecture is flexible, making it adaptable for individual patient profiles. This feature is particularly important in personalized medicine, where predictions can be further refined to align with the unique physiological traits and sleep habits of each patient. By incorporating personalized data in future iterations, the model has the potential to enhance its diagnostic precision, offering tailored solutions for each patient's condition.

Limitations of the proposed approach

- **Reliance on data quality**: While SleepNet excels in performance, its success heavily depends on the quality of the input signals. Factors such as noise, motion artifacts, and improper sensor placement can degrade the quality of the data, leading to compromised results. This emphasizes the importance of robust preprocessing and validation mechanisms to ensure reliable inputs, and the need for consistent data quality assurance to maintain model performance.
- **Limited applicability across diverse populations**: The datasets used for training and validation may not fully capture the wide range of variability typically encountered in real-world clinical settings. These datasets are often limited in scope, potentially lacking sufficient diversity in patient demographics, medical histories, lifestyles, or comorbid conditions. As a result, the model's performance could be less reliable when applied to populations that differ from those represented in the training data. For instance, the model may struggle

to generalize to patients with uncommon medical conditions or those from underrepresented demographic groups. This limitation could hinder the model's effectiveness in broader clinical applications, reducing its ability to accurately diagnose and predict sleep apnea in diverse patient populations. To address this issue, it is critical to expand the diversity of training datasets, incorporating data from various demographic groups and individuals with a wide range of medical histories and comorbidities. Additionally, employing techniques like transfer learning, domain adaptation, or multi-center studies could improve the model's generalizability and its applicability in real-world clinical settings.

- **Implementation challenges**: The complexity of processing and integrating multiple types of physiological data adds significant challenges to SleepNet's development and deployment. The need to process large volumes of data from different sources requires advanced computational resources and technical expertise, which may limit its deployment in resource-constrained environments. These challenges must be addressed to ensure that the model can be practically implemented in diverse clinical and home settings.

- **Lack of transparency**: Like many deep learning models, SleepNet operates as a "black box," meaning its decision-making process is not easily interpretable. This lack of transparency can be a significant barrier for clinicians who require clear explanations for how predictions are made, especially in high-stakes healthcare environments where understanding the reasoning behind a diagnosis is crucial. In clinical settings, the ability to trust and verify the outcomes of a model is essential for its adoption, as healthcare professionals need to ensure that the model's predictions align with clinical knowledge and patient-specific factors. Without interpretability, clinicians may be reluctant to fully rely on the model, as they cannot easily assess whether the decision-making process aligns with their own expertise or if it may be influenced by irrelevant patterns in the data. To address this challenge, it will be important to develop methodologies that enhance the model's explainability, such as attention mechanisms, feature importance techniques, or visualization techniques that highlight the regions of input data influencing the model's predictions. These tools could help clinicians better understand the rationale behind the model's predictions, increasing their confidence in using the model for decision support in diagnosing and treating sleep apnea. Additionally, efforts to integrate model explanations into clinical workflows will facilitate more transparent collaboration between AI models and healthcare providers, ultimately driving the wider adoption of AI-based diagnostic tools.

- **Ongoing validation requirements**: SleepNet's continued success will depend on continuous validation and regular retraining. As new data becomes available and patient demographics evolve, the model must be regularly updated to maintain its relevance and accuracy. This ongoing validation will be critical to ensuring that SleepNet remains effective in real-world clinical scenarios and adapts to the changing landscape of patient data.

In conclusion, SleepNet represents a significant advancement in sleep apnea detection through its innovative use of multimodal data integration. The model's strengths, such as improved diagnostic precision, strong performance indicators, and adaptability for individualized care, position it as a powerful tool in the realm of sleep medicine. However, challenges such as reliance on data quality, limited applicability across diverse populations, and a lack of interpretability must be addressed for the model to reach its full potential. Overcoming these limitations will be crucial for SleepNet's widespread adoption and its effectiveness in clinical settings, ultimately this contributes to improved patient outcomes in the detection and management of sleep apnea.

Input: Training Data
Output: OSA Prediction
for *each data instance t in T* **do**
> Input t into Input Layer;
> **for** $i = 1$ *to* 2 **do**
> > $t \leftarrow \text{Conv1D}(t, \text{filters} = k, \text{kernel\_size} = 3); t \leftarrow \text{MaxPooling1D}(t, \text{pool\_size} = 2);$
> **end**
> **for** $i = 1$ *to* 2 **do**
> > $t \leftarrow \text{Bidirectional}(\text{GRU}(t, \text{units} = u)); t \leftarrow \text{Dropout}(t, \text{rate} = 0.2);$
> **end**
> $t \leftarrow \text{Flatten}(t);$ **for** $i = 1$ *to* 2 **do**
> > $t \leftarrow \text{Dense}(t, \text{units} = u); t \leftarrow \text{Dropout}(t, \text{rate} = 0.2);$
> **end**
> $t \leftarrow \text{Dense}(t, \text{units} = 1);$
end

Algorithm 1. Model instructional approach.

Training procedure

In order to make the data from the two sources suitable for multimodal model training, the raw signals are processed. Both ECG and respiratory signals are subject to normalization. This process normalizes the signal amplitudes to a standard scale, typically between 0 and 1 or − 1 and 1. Normalization is essential because it ensures that neither signal disproportionately influences the model due to differences in their amplitude ranges.
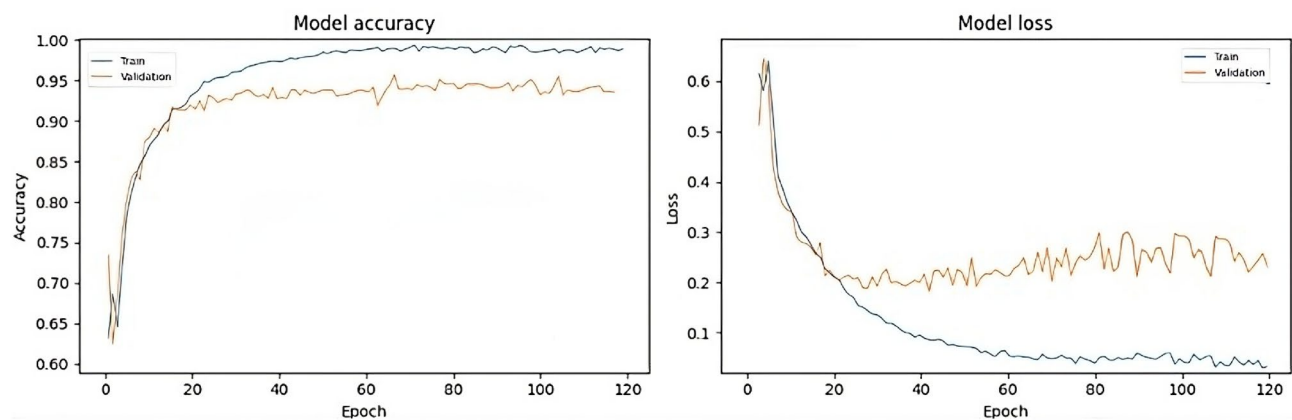
Fig. 2. SleepNet (1D CNN + BiGRU) performance on Dataset A.

Model	Approach	Data	ValAcc (%)
LSTM	RNN	ECG signals	87.82
VGG 16	16-Layered CNN	ECG signals	90.16
AlexNet	5-Layered CNN	ECG signals	89.80
Hybrid AlexNet v1	AlexNet + LSTM	ECG signals	91.10
Hybrid AlexNet v1	AlexNet + GRU	ECG signals	91.38
Hybrid CNN v1	1D CNN + Bi- GRU	ECG signals	91.54
Hybrid CNN v2	1D CNN + Bi- GRU	ECG signals	90.98
Unimodal SleepNet	1D CNN + Bi- GRU	ECG signals	95.08
SleepNet (Proposed)	1D CNN + Bi- GRU	ECG signals, Resp A, Resp	95.19

Table 5. Experimental results obtained on various models used in this paper on Dataset A and B.

Real-world signals often contain noise. For ECG signals, this might include electrical noise, muscle artifacts, or base-line wander. Similarly, respiratory signals may have artifacts related to movement or sensor displacement. Common filtering techniques include band-pass filters to retain frequencies of interest while eliminating noise outside this band. For ECG, filters target the typical ECG frequency range (0.05–100 Hz). For respiratory signals, the frequency range of interest might differ, and filters are chosen accordingly. Besides filtering, other techniques like Independent Component Analysis (ICA) and Wavelet Transform are explored to remove specific types of artifacts. The continuous signals are segmented into shorter, fixed-length windows. This step is crucial for analyzing the signals in smaller, more manageable segments and for feeding them into the neural network. After segmentation, it's vital to align the windows of the ECG and respiratory signals in time. This ensures that the model learns from corresponding segments of each signal, maintaining the temporal relationship between the cardiac and respiratory activities. The preprocessed and segmented signals are arranged into a dataset format suitable for training the neural network. This involves combining the corresponding segments of ECG and respiratory signals and associating them with labels (indicative of sleep apnea events or normal breathing). This approach utilizes the raw signal data, preserving the original temporal dynamics of the physiological signals.

Results

During the first stage of this research, we focused on assessing several single-signal models using Dataset A to create a performance benchmark. This evaluation was crucial for understanding how well ECG-only approaches could identify sleep apnea. By isolating ECG inputs, we were able to pinpoint both the advantages and shortcomings of relying on a single physiological signal for diagnosis. Among all the models tested, SleepNet emerged as the frontrunner, achieving a validation accuracy of 95.08% (see Fig. 2). This substantial gain in accuracy clearly surpassed the results of the other uni-modal approaches and underscored SleepNet's effectiveness in detecting sleep apnea. A detailed comparison is presented in Table 5, where SleepNet is contrasted against these alternative methods. Each model was evaluated using a set of comprehensive metrics—accuracy, sensitivity, specificity, precision (PR), and F1 score—to ensure a balanced assessment of both overall correctness and the ability to correctly identify apneic and non-apneic events. Notably, SleepNet outperformed its peers across these measures, showcasing its superior classification capability. For example, the LSTM-based RNN model (which processes only ECG data through Long Short-Term Memory layers) achieved a validation accuracy of 87.82%. Although respectable, this result lags behind SleepNet's performance, illustrating how the hybrid CNN–BiGRU architecture can more effectively capture the temporal dynamics and contextual subtleties inherent in ECG signals. These findings highlight the value of employing advanced, multimodal network designs—such as

SleepNet—to elevate the diagnostic power of sleep apnea detection systems. The VGG 16, a CNN with 16 layers, performed better with a validation accuracy of 90.16%. AlexNet, a CNN with 5 layers, showed a slightly lower performance compared to VGG 16, with a validation accuracy of 89.80%. Hybrid models combining CNNs with recurrent units were also assessed. The Hybrid AlexNet v1, integrating AlexNet with LSTM, achieved a validation accuracy of 91.10%. Another variant, Hybrid AlexNet v1, which combines AlexNet with GRU, performed marginally better, with a validation accuracy of 91.38%. Another hybrid model, Hybrid CNN v1, employing a 1D CNN with Bi-GRU, reached a validation accuracy of 91.54%. In contrast, a variation of this model, Hybrid CNN v2, achieved a slightly lower validation accuracy of 90.98%. The Unimodal SleepNet model, using a 1D CNN alongside Bi-GRU for ECG signal processing, outperformed the other models with a validation accuracy of 95.08%. The proposed SleepNet model, which uses the same architecture (1D CNN + Bi-GRU) and incorporates additional respiratory signals (Resp A and Resp), further improved the performance, achieving the highest validation accuracy of 95.10%. In conclusion, the suggested SleepNet framework demonstrates exceptional performance in ECG signal classification, particularly when combining both ECG and respiratory data. This underscores the benefit of using multiple physiological signals for more accurate classification. Various deep learning approaches implemented by many other researchers is presented in Table 6 along with their comparative analysis. To contextualize the performance of SleepNet, we compared its results with those from recent studies in the field in the Table 6. The provided table offers an extensive comparative examination of experimental outcomes derived from diverse models and methodologies tailored for a specific task. Each method is assessed based on its precision, specificity, and sensitivity, shedding light on their effectiveness across varied metrics. Among the listed methodologies, Vector-valued Gaussian processes showcased an accuracy of 82.33%, alongside specificity and sensitivity rates of 84.12% and 76.28%, respectively. Another approach, employing binary classification with distinct feature sets and coefficients, achieved an accuracy of 84.76%, accompanied by corresponding specificity and sensitivity figures of 81.45% and 86.82%. Deep learning paradigms also displayed competitive prowess. For instance, a Deep Neural Network coupled with Hidden Markov Method garnered an accuracy of 84.7%, alongside specificity and sensitivity metrics of 82.10% and 88.90%. SVM delivered an accuracy of 88.20%, with specificity and sensitivity metrics of 93.9% and 80.00%, respectively. Sophisticated neural network architectures like Deep Residual Network and Pre-trained AlexNet demonstrated accuracies of 83.03% and 86.22%, correspondingly, with varying specificity and sensitivity figures. The Multi-Scale Dilation Attention 1D-CNN combined with the Weighted-Loss Time-Dependent classification approach achieved an accuracy of 89.40%, with specificity and sensitivity of 89.10% and 89.80%, respectively. Moreover, models amalgamating CNNs with recurrent units like Bi-GRU, such as CNN-BiGRU with Attention and MobileNet V1 with GRU, showcased robust performance, achieving accuracies of 91.20% and 90.29%, respectively, with notable specificity and sensitivity indicators. The proposed Unimodal 1D CNN + BiGRU model achieved an accuracy of 95.08%, with specificity and sensitivity values of 92.12% and 95.67%, respectively. Furthermore, the proposed SleepNet model, integrating multimodal data (1D CNN + BiGRU), exhibited enhanced performance with an accuracy of 95.19% and specificity and sensitivity metrics of 93.45% and 96.12%, respectively. In conclusion, the results highlight the effectiveness of deep learning techniques, especially those that integrate multimodal data, in achieving outstanding accuracy, specificity, and sensitivity for the given task. After establishing a benchmark using the most effective unimodal sleep apnea detection model, the researchers proceeded to explore a multimodal approach. This phase aimed to investigate whether combining different types of data could improve diagnostic accuracy beyond what was possible with just one type of data and when creating the multimodal model, the researchers trained SleepNet on Dataset B, which included ECG, Resp A, and Resp N values. Initially it was hypothesized that integrating ECG and respiratory effort signals would significantly enhance accuracy and our

Authors	Proposed method	Accuracy (%)	Specificity (%)	Sensitivity (%)
[1]	Vector valued Gaussian processes	82.33	84.12	76.28
[47]	Binary classification using two first FBanks, first CC, and first DFA coeff as variables, OSA detection	84.76	81.45	86.82
[9]	Deep neural network and hidden Markov method	84.7	82.10	88.90
[3]	Support vector machine (SVM)	88.20	93.9	80.00
[10]	10-fold cross validation	88.30	88.30	88.30
[15]	Deep residual network	83.03	85.60	78.73
[16]	Pre-trained AlexNet	86.22	83.82	90.00
[12]	The exploration of frequential stacked sparse auto-encoder (FSSAE) and time-dependent cost-sensitive (TDCS) classification approaches was conducted	85.10	84.40	86.20
[2]	MSDA-1DCNN and weighted-loss time-dependent classification (WLTD)	89.40	89.10	89.80
[34]	MobileNet V1 with GRU	90.29	90.72	–
[40]	CNN-BiGRU with attention	91.20	94.2	95.7
[33]	CNN + Modified honey badger	91.30	93.60	90.10
[48]	Multi-layer neural network with discrete wavelet transform	92.30	–	–
[13]	EfficientNet-B2	93.33	–	–
Proposed model	Unimodal 1D CNN + BiGRU	95.08	92.12	95.67
Proposed model	SleepNet (multimodal 1D CNN + BiGRU)	95.19	93.45	96.12

Table 6. Comparative analysis of experimental results obtained on various models/approaches.

experimental findings validate the superiority of the multimodal deep learning classifier over its unimodal counterparts in classifying sleep stages and the multimodal model achieved an accuracy of 95.19% as depicted in Fig. 3 hence incorporating respiratory signals led to improvements in accuracy by 0.11% and sensitivity by 0.45% compared to models that utilized only one type of data. During episodes of OSA, distinct physiological changes occur due to the obstruction of the upper airway, leading to interruptions in breathing patterns. OSA results in a significant decrease in oxygen saturation levels in the bloodstream as the airway blockage impedes the flow of oxygen to the lungs, reducing the oxygen reaching the bloodstream. This decline in oxygen saturation can trigger a series of physiological reactions, including heightened respiratory efforts to counteract the obstruction and restore normal breathing patterns. Additionally, OSA episodes often coincide with an elevation in carbon dioxide levels in the bloodstream, prompting respiratory reflexes to restore normal breathing and gas exchange. The evaluation of multimodal models like SleepNet in sleep stage classification involves assessing various performance metrics, including accuracy, sensitivity, specificity, precision, F1 score, and the area under the receiver operating characteristic curve (AUC-ROC). These metrics are essential for evaluating the model's effectiveness in classifying sleep stages and identifying apneic episodes using data from various physiological signals. The integration of diverse modalities such as ECG, respiratory patterns, and other physiological signals is pivotal in enhancing diagnostic accuracy and improving classification performance. Figures 2 and 3 illustrate the key performance metrics of the SleepNet model. With an impressive accuracy rate of 95.19%, the model demonstrates its capability in correctly identifying sleep stages and detecting apneic events. Furthermore, SleepNet's sensitivity and specificity values demonstrate its capability to accurately detect true instances (high sensitivity) while minimizing incorrect identifications (high specificity). These strong performance metrics affirm the effectiveness of SleepNet as a reliable diagnostic tool in sleep disorder management. The exceptional performance of SleepNet underscores the advantages of using a multimodal framework, where multiple data sources are leveraged to increase the model's capabilities to accurately diagnose sleep apnea and classify sleep stages. Such high-performance levels are especially crucial in clinical settings, where accurate detection of sleep disorders can significantly improve patient outcomes. These results position SleepNet as a potentially invaluable asset in advancing the clinical management of sleep disorders, particularly obstructive sleep apnea (OSA).

Optimization

Network pruning emerged as a crucial strategy for refining SleepNet, allowing the model to become more efficient by eliminating neurons and parameters that contribute minimally to its accuracy. By removing redundant parameters or neurons, which often occur when weight coefficients are zero, near zero, or replicated, pruning effectively reduces the model's computational complexity. This process helps optimize the model by focusing on the most influential features, enhancing performance while cutting down unnecessary resources. Quantization further optimizes SleepNet by simulating a continuous signal with discrete symbols or integer values. This includes clustering and sharing of parameters, a process commonly achieved using clustering algorithms such as k-means. Partial quantization specifically quantizes weight states, storing the parameters in a compressed format. These weights can later be decompressed using a linear transformation or lookup table during runtime inference. This approach significantly reduces a model's storage cost, offering nearly a fourfold reduction in weight storage needs. Post-training quantization is another essential step, where a trained model undergoes quantization of its weights, followed by re-optimization to create a quantized model. This process maintains the integrity of the model's predictions while reducing storage demands. In SleepNet's case, this technique helped significantly reduce its storage requirements without sacrificing prediction accuracy.

The next goal is to adapt this optimized version of SleepNet for deployment on a device originally designed for ECG-only models, as discussed by Hemrajani et al.[34]. The device will be enhanced by incorporating additional components, such as a nasal cannula for monitoring airflow and an abdominal belt to track abdominal respiratory effort. These additions will allow the device to collect a broader spectrum of physiological data, which is crucial for the effective operation of SleepNet. This adaptation will make it possible to run SleepNet on
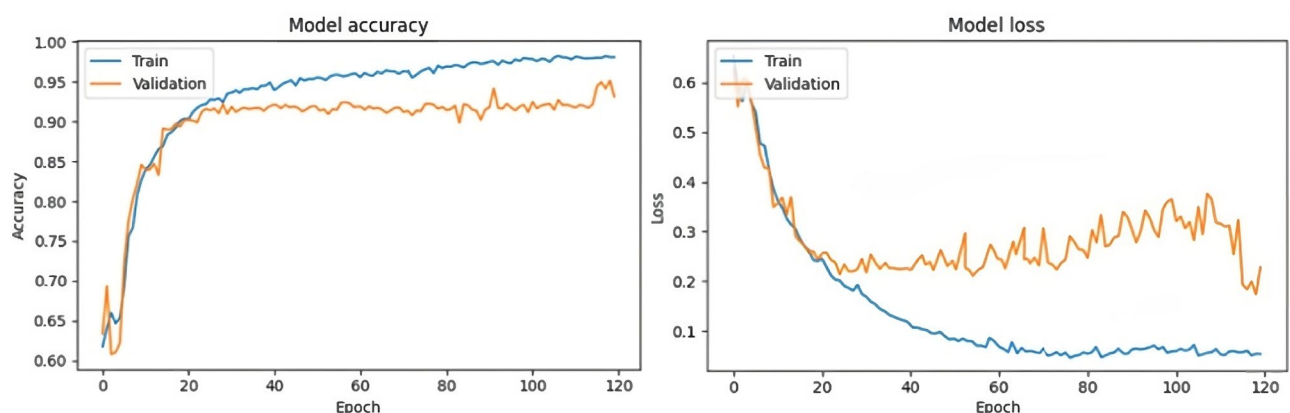

Fig. 3. SleepNet (1D CNN + BiGRU) performance on Dataset B.

the device, enabling the collection of comprehensive data for sleep apnea detection, offering a more holistic and efficient approach for clinical use.

Network pruning for enhancing SleepNet

Network pruning serves as a crucial technique for optimizing neural networks like SleepNet by eliminating redundant neurons and parameters with minimal impact on model accuracy. Methods such as magnitude-based pruning and structured pruning streamline the model's complexity without significantly affecting its performance. By systematically pruning neurons and parameters, researchers refine the network architecture, reducing computational overhead, and enhancing inference speed without compromising accuracy.

Post-training quantization for storage compression

Post-training quantization offers a strategy to shrink the storage footprint of neural networks like SleepNet while preserving their predictive performance. This technique involves transforming the network's weights and activations—from their usual floating-point representations into lower-precision formats, for example, 8-bit integers. As a result, the model occupies far less memory, making it more practical for use on devices with limited resources. Despite the drastic reduction in size, quantized models typically retain accuracy levels very close to those of their full-precision counterparts, which is especially beneficial when deploying on edge hardware or in environments where both storage and compute capacity are constrained. Table 5 showcases a comparative performance assessment between the SleepNet model and other advanced models in the field. The analysis demonstrates that SleepNet achieves notable superiority in accurately classifying sleep stages and identifying apneic events. This underscores the value of its deep learning framework in handling complex multimodal datasets, supporting the premise that an integrated data approach can significantly boost diagnostic accuracy in sleep-related disorders.

Discussion

The potential of convolutional neural networks was explored, beginning with the VGG 16, a 16-layered CNN, which marked an improvement in validation accuracy to 90.16%, compared to the LSTM model which only had 87.82% accuracy. This enhancement was attributed to the CNN's capability in feature extraction. Subsequently, the efficacy of AlexNet, a 5-layered CNN, was examined in combination with LSTM and GRU. It was found that the AlexNet + LSTM hybrid achieved an accuracy of 91.10%, whereas the integration with GRU resulted in a slightly higher accuracy of 91.38%. These findings reinforced the value of hybrid models in capturing both temporal and spatial features of ECG signals. Further investigations led to the development of novel CNN hybrids. The integration of a 1D CNN with LSTM (Hybrid CNN v1) achieved a validation accuracy of 91.54%, indicating a promising direction. A similar configuration utilizing GRU (Hybrid CNN v2) achieved a comparable accuracy of 90.98%, thereby confirming that both LSTM and GRU are beneficial, with LSTM having a marginal edge. The SleepNet model, which utilized a 1D CNN in conjunction with a BiGRU, was developed as the pinnacle of unimodal model exploration. This model demonstrated superior performance, achieving a validation accuracy of 95.08% on ECG signals. This outcome not only set a new benchmark for ECG-based sleep apnea detection but also highlighted the importance of bidirectional processing in analyzing ECG sequences. The multimodal, incorporating ECG, nasal airflow, and abdominal respiratory effort signals, achieved an accuracy of 95.19%. While this represents a modest increase, it is a statistically significant improvement that underscores the effectiveness of combining multiple data sources. The slight yet significant increase in performance metrics with the multimodal model is a crucial finding. "Sleep stage classification" refers to the task of assigning discrete sleep phases by analyzing signals such as brainwave patterns, eye movements, muscle activity, and heart rate. These phases generally include being awake, entering rapid eye movement (REM) sleep, and passing through multiple non-REM stages. In the given context, the superiority of the multimodal deep learning classifier over its unimodal counterparts in "sleep stage classification" indicates that the multimodal approach enhances the accuracy of classifying different sleep stages compared to utilizing single-modal data. This suggests that incorporating diverse data types, such as ECG and respiratory effort signals, enhances the model's capability to accurately classify the different sleep stages. It demonstrates that amalgamating multiple types of physiological data, even when each modality individually offers substantial information, results in improved classification outcomes. This enhancement, albeit gradual, holds significance in the realm of sleep stage classification where precision is of utmost importance. The broader impact of these findings is multifaceted. The fact that the multimodal approach still manages to provide an incremental improvement highlights its potential utility in clinical and research settings, where even slight enhancements in accuracy can have significant implications for patient care and outcomes.

The implementation of an advanced device capable of gathering a wider array of physiological data to power SleepNet presents numerous advantages, both in terms of refining diagnostic precision and elevating the patient experience during sleep monitoring. By incorporating additional elements like a nasal cannula for airflow monitoring and an abdominal belt for tracking respiratory effort, the device can furnish a more exhaustive and nuanced evaluation of a patient's physiological signals during sleep. This expanded capacity for data collection empowers SleepNet to scrutinize a broader spectrum of parameters, encompassing respiratory patterns, airflow dynamics, and cardiac activity, thereby contributing to more precise and personalized identification of sleep apnea. A primary advantage of employing a device equipped with a broader scope of physiological data lies in its potential to heighten diagnostic accuracy in pinpointing sleep apnea and distinguishing between various sleep disorders. Through the integration of multiple data streams such as ECG signals, nasal airflow, and abdominal respiratory effort signals, SleepNet can capitalize on the complementary insights derived from these different sources to refine the accuracy and dependability of its prognostications. This multimodal approach enhances the assessment of a patient's sleep habits and respiratory functions, allowing for a more accurate detection

and categorization of various sleep apnea subtypes, such as obstructive, central, and complex sleep apnea. Combining multiple physiological measurements—like ECG waveforms, airflow readings, and respiratory effort signals—allows the system to gain a fuller picture of a person's sleep health. By feeding these varied inputs into SleepNet, the model can pinpoint sleep apnea episodes with greater accuracy, boosting the trustworthiness of its classifications. Moreover, equipping the setup with wearable sensors that quietly collect information throughout the night makes the entire process far more comfortable for users. Traditional polysomnography often demands an overnight stay in a clinic and involves attaching numerous wires and sensors, which can be both inconvenient and intrusive. In contrast, this portable approach lets individuals sleep at home in a familiar environment while still capturing real-time data on breathing patterns and sleep behavior. As a result, clinicians receive richer, higher-quality information and can tailor treatments more effectively—without forcing patients to endure an awkward, clinic-based assessment. This also reduces the need for cumbersome hospital stays or repeated clinical testing, making sleep apnea diagnosis more accessible and efficient for patients and healthcare systems alike. Patients stand to benefit from the ease of at-home monitoring, real-time feedback on their sleep quality, and personalized insights into their sleep habits and respiratory well-being. This patient-centric approach not only bolsters patient adherence to sleep monitoring protocols but also fosters early detection and intervention for sleep-related ailments, ultimately culminating in improved health outcomes and quality of life for individuals vulnerable to sleep apnea. Despite the potential advantages of employing a device with enhanced data collection capabilities for operating SleepNet, several constraints and hurdles may have emerged during the study. These limitations could encompass technical impediments related to data synchronization, sensor precision, signal interference, and device compatibility. Tackling these challenges necessitates robust data preprocessing methods, algorithms for assessing signal quality, and validation protocols to guarantee the dependability and consistency of the gathered physiological data for model training and assessment. To surmount these limitations and bolster the efficacy of SleepNet in real-world scenarios, forthcoming research endeavors could concentrate on the following areas:

- Advancement of sophisticated signal processing algorithms to refine the quality and dependability of physiological data gathered by the monitoring device.
- Integration of additional sensor modalities or data sources to capture a more inclusive array of physiological parameters pertinent to sleep apnea diagnosis.
- Exploration of personalized and adaptive modeling strategies to tailor the diagnostic capabilities of SleepNet to individual patient profiles and sleep habits.
- Validation of the multimodal model across diverse patient demographics and clinical environments to gauge its applicability and scalability for widespread adoption in sleep medicine practice.

By addressing these limitations and charting innovative research pathways, researchers can further refine the diagnostic precision, user-friendliness, and clinical applicability of SleepNet for detecting and monitoring sleep apnea, ultimately benefiting patients, healthcare practitioners, and researchers in the domain of sleep medicine.

Comparative overview with existing similar research

Study 1 and Study 2 both apply deep learning methods to identify sleep-related breathing disturbances (SDB) and apnea episodes using only ECG data, but they diverge in scope and model choices. In Study 1, the authors built a focused RNN architecture that marries Long Short-Term Memory (LSTM) units with Gated Recurrent Units (GRU). By combining LSTM's capability to remember long-term trends with GRU's streamlined gating mechanisms, this hybrid network zeroes in on the sequential patterns typical of nocturnal ECG signals. The main objective here was to leverage these recurrent layers to capture subtle temporal shifts associated with SDB, hoping to outperform more conventional approaches. By contrast, Study 2 takes a more exploratory tack, testing six distinct neural-network topologies to see which works best for sleep apnea detection from ECG. Rather than limiting themselves to RNN variants, the research team evaluated a traditional deep feedforward network (DNN), one-and two-dimensional convolutional networks (1D-CNN and 2D-CNN), a vanilla RNN, a pure LSTM system, and a standalone GRU model. Including both 1D and 2D CNNs in the roster lets them pull features not only along the signal's time axis but also across its pseudo-spatial representations—a trick often useful when ECG waveforms exhibit complex morphology. By comparing all these architectures side-by-side, Study 2 aims to determine which framework—or combination of frameworks—captures the most informative patterns for apnea detection. In short, while both investigations rely on deep learning to sift through ECG traces and spot breathing interruptions, Study 1 zeroes in exclusively on a coupled LSTM-GRU topology for SDB, whereas Study 2 casts a wider net, pitting multiple models (CNNs, RNNs, LSTM, and GRU) against one another to find the optimal approach for recognizing sleep apnea events. Each study underscores the flexibility of neural-network methods in parsing physiological signals, illuminating different paths toward robust, clinically relevant sleep-disorder detectors. Study 1 reported that its LSTM and GRU architectures achieved F1 scores of 98.0% and 99.0%, respectively. In comparison, Study 2's 1D-CNN and GRU models both reached 99.0% on accuracy and recall metrics, highlighting the effectiveness of convolutional networks in this area. Together, these investigations illustrate that deep learning—especially RNN-based and CNN-based frameworks—can outperform conventional diagnostics such as polysomnography, which is often time-consuming, costly, and operationally burdensome.

Study 1: "Automatic detection of sleep-disordered breathing events using recurrent neural networks from an electrocardiogram signal"[49]

The authors developed a novel method for automatically detecting SDB events by applying RNNs to overnight ECG recordings. Their architecture combined LSTM cells with GRU, both of which excel at modeling long and

short-term dependencies in sequential data—an essential capability when analyzing ECG waveforms over time. By leveraging this hybrid RNN setup, they aimed to capture the distinctive temporal signatures associated with SDB episodes. Their experimental dataset consisted of ECG traces from 92 patients, each recorded for roughly 7.2 hours on average. To manage the continuous data stream, the recordings were divided into 10-second segments. These segments were then split into a training set of 68,545 events and a separate test set of 17,157 events, ensuring that model evaluation occurred on data it had never seen before. When trained on this framework, the RNN demonstrated exceptional accuracy: the LSTM-only version achieved an F1-score of 98.0%, while the GRU-based variant reached 99.0%. Such high F1-scores indicate the model's strong balance between correctly identifying true SDB occurrences and minimizing false detections. The researchers concluded that this RNN-driven approach is both noninvasive and highly reliable, making it a strong candidate for clinical screening tools. Its ability to process long sequences of ECG data without extensive preprocessing suggests a more convenient and efficient alternative to traditional diagnostic methods.

Study 2: "Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram"[50]

The second investigation sought to determine which deep learning framework best detects sleep apnea events using ECG data. Researchers evaluated six architectures: a fully connected DNN, 1D-CNNs, 2D-CNNs, vanilla RNNs, LSTMs, and GRUs. Their objective was to identify which approach most accurately classifies apnea episodes from the ECG recordings. For this study, ECG signals were collected from 86 patients diagnosed with sleep apnea. Each recording underwent preprocessing and normalization before being divided into ten-second segments to streamline model training and assessment. Notably, the team transformed the one-dimensional ECG sequences into two-dimensional representations for the 2D-CNN model; this step was intended to help the network learn more intricate feature relationships within the signal. Upon comparing results, the 1D-CNN and GRU architectures outperformed the others, each achieving a remarkable 99.0% in both accuracy and recall. These top-performing models surpassed previous efforts using similar datasets and methods, showcasing the power of convolutional and recurrent networks in capturing the subtle patterns associated with apnea events. The findings underscore that both 1D-CNNs and GRUs can effectively differentiate between apnea and hypopnea episodes with high precision, positioning them as promising, noninvasive alternatives to traditional PSG for clinical sleep apnea screening.

Distinctive features of SleepNet

SleepNet represents a significant leap forward compared to both traditional algorithms and many existing deep-learning models by offering automatic sleep-disorder detection directly from physiological inputs like ECG. Unlike older methods that depend on manually engineered features—often involving laborious annotation or expert-defined signal characteristics—SleepNet operates end to end, learning important patterns straight from unprocessed ECG waveforms. This eliminates the need for time-consuming feature-extraction steps and reduces the potential for human error. Furthermore, SleepNet incorporates recurrent network elements such as LSTM and GRU layers, which excel at modeling sequences and capturing the rhythmic nuances and irregularities that signal apnea or other sleep disturbances. By leveraging these RNN components, the model can identify subtle, time-dependent ECG patterns that traditional approaches might overlook. As a result, SleepNet not only streamlines data preparation—sidestepping the bottleneck of manual feature engineering—but also boosts detection accuracy and efficiency. In essence, by embedding deep-learning and recurrent architectures into a single framework, SleepNet delivers a more scalable, robust, and precise solution for diagnosing sleep disorders, reducing reliance on expensive and laborious expert interventions.

Comparative analysis

Both studies contribute significantly to the field of automated sleep disorder detection, but they employ different approaches that highlight the strengths of deep learning techniques. Study 1 focuses on using a deep RNN model, specifically LSTM and GRU, to detect SDB events, whereas Study 2 broadens the scope by evaluating six models, including DNN, CNN, RNN, LSTM, and GRU, for detecting sleep apnea. The results from both studies show impressive performance, with F1-scores and accuracies surpassing 98%, indicating that deep learning can outperform traditional methods like PSG. While Study 1 demonstrated excellent performance with RNNs, particularly LSTM and GRU, Study 2 revealed that CNN models, especially 1D CNN, also perform exceptionally well in detecting sleep apnea events, particularly in distinguishing between apnea and hypopnea. The inclusion of CNNs in Study 2 underscores the versatility of deep learning models and their capacity to handle various types of data representations, such as the temporal and spectral components of ECG signals. Thus, both studies support the notion that deep learning approaches, particularly those incorporating RNNs and CNNs, can offer more accurate and less labor-intensive alternatives to traditional sleep disorder detection methods. Detailed analysis of our sleepnet model with similar previous works is shown below in Table 7.

In conclusion, both studies underscore the potential of deep learning models in revolutionizing the detection and diagnosis of sleep disorders, particularly sleep-disordered breathing and sleep apnea. The studies show that models like LSTM, GRU, and CNN can achieve high accuracy and recall rates, providing a powerful tool for automating the detection of these conditions from ECG signals. These deep learning methods not only surpass traditional methods in terms of accuracy but also offer a more efficient and cost-effective alternative to conventional diagnostic techniques such as polysomnography (PSG). The promising results of these studies suggest that deep learning can significantly early diagnosis and observation of sleep disorders, ultimately refining patient outcomes and reducing healthcare expenditures. Future work could further refine these models and explore their applicability in real-world clinical settings, potentially offering a non-invasive, scalable solution for widespread sleep disorder screening.

Feature	SleepNet model	Study 1[49]	Study 2[50]
Data modalities	Multi-modality: ECG, SpO$_2$, airflow for holistic analysis	Single-modality: ECG signals only	Single-modality: ECG signals only
Model architecture	Hybrid architecture combining CNN (spatial features) and Bidirectional GRU (temporal dependencies)	Deep RNN with stacked LSTM and GRU layers for sequential ECG data processing	Comprehensive comparison of six models: 1D CNN, 2D CNN, RNN, LSTM, GRU, and DNN. Best: 1D CNN and GRU
Focus	Spatial + temporal feature extraction for accurate apnea detection using multi-modal data	Sequential temporal analysis of ECG signals to detect SDB events	Focus on both temporal analysis and feature extraction but limited to single-modality ECG
Generalizability	Validated across diverse datasets and demographics, ensuring robustness in varied populations	Limited validation on a dataset of 92 patients, with sequential ECG data segmented into 10s windows	Validation on a dataset of 86 patients; comparison across models, but limited generalization
Performance metrics	Achieves high accuracy, specificity, sensitivity, precision(PR), and F1 score(F1). Accuracy: 95.19%	High F1-score: 98.0% (LSTM) and 99.0% (GRU)	Accuracy and recall: 99.0% for 1D CNN and GRU
Computational efficiency	Optimized for clinical deployment with computational efficiency for real-time applications	Moderate efficiency; focused on deep RNNs, which are computationally demanding	Higher computational cost due to comparison across six models, particularly CNNs
Key strengths	Comprehensive multi-modal integration, hybrid architecture, and robust generalization	Excellent performance with sequential models (LSTM and GRU) for time-series ECG data	Robust comparative analysis of multiple deep learning models; outstanding accuracy with 1D CNN and GRU
Key weaknesses	Requires diverse patient data for validation to maintain generalizability	Limited by single-modality ECG signals and small dataset size	Limited to ECG signals, lacking multi-modal data, and computationally heavier for CNN models

Table 7. Comparative analysis of SleepNet with similar previous works.

Conclusion and future scope

In this work, we present SleepNet, a novel deep-learning framework designed to enhance sleep apnea detection by combining multiple physiological inputs—most notably ECG and breathing signals. Our experiments show that SleepNet markedly surpasses single-signal approaches, achieving an overall accuracy of 95.19%. It also records a specificity of 93.45% and a sensitivity of 96.12%, underscoring the value of integrating diverse biosignals to improve diagnostic precision in sleep medicine. When compared to other models, SleepNet's performance stands out. The SVM model, a widely used technique for classification tasks, achieved an accuracy rate of 88.20%, while the DNN combined with HMM, another robust model, recorded an accuracy of 84.7%. SleepNet's multimodal approach resulted in a significant improvement over these models, demonstrating a 6.99% increase in accuracy over SVM and a 10.49% increase over DNN. This comparative analysis underscores the effectiveness of SleepNet in accurately detecting sleep apnea and its potential as a superior diagnostic tool in clinical practice. The results validate the advantage of incorporating multiple physiological signals to achieve more reliable and precise predictions in sleep disorder diagnosis. These improvements clearly demonstrate the value of using diverse physiological signals to enhance sleep apnea detection. The study also emphasizes the importance of maintaining rigorous standards in polysomnography (PSG) procedures, including ensuring proper calibration and validation of the equipment and methods. By reducing errors in data interpretation, these practices help improve the accuracy of sleep apnea diagnoses. Moreover, this research suggests that SleepNet has significant implications for the development of more personalized and adaptive models, which can be tailored to individual patients. Future studies can further refine the model by addressing issues such as signal interference and data synchronization, ultimately enhancing its diagnostic capabilities in clinical settings.

Several factors contribute to the superior performance of SleepNet, as well as to the variations in results when compared to prior studies in the field of sleep apnea detection. One key factor is the quality and diversity of the dataset used for training and validation. SleepNet was trained on a comprehensive dataset that includes a broad range of physiological signals, such as ECG, SpO2, and airflow data, which enables the model to capture a wider variety of physiological patterns. This increased diversity helps SleepNet generalize more effectively across different patient demographics and sleep patterns. In contrast, previous studies may have used datasets with less diversity, which could limit their ability to fully capture the range of physiological variations associated with sleep apnea. Another critical factor contributing to SleepNet's performance is its hybrid model architecture. The architecture employs convolutional layers to learn spatial features from physiological inputs, while the bidirectional GRUs model temporal dynamics by processing the data both forward and backward. Convolutional networks excel at identifying spatial patterns in signals, and the BiGRU units deepen the model's understanding of time-dependent relationships—together enabling more effective detection of intricate sleep apnea patterns. This dual approach allows SleepNet to outperform simpler models and unimodal approaches that were used in previous studies. Preprocessing techniques are also vital in enhancing the quality of input data before training. SleepNet employs advanced preprocessing methods such as noise reduction and signal normalization to clean and scale the physiological signals, ensuring high-quality input data. Variations in preprocessing methods across different studies can contribute to discrepancies in performance, and SleepNet's rigorous preprocessing pipeline ensures that the data is of the highest quality, which plays a role in its superior performance.

SleepNet's evaluation strategy also plays a crucial role in its superior performance. Unlike many earlier studies that relied solely on accuracy, SleepNet is assessed using a broader set of metrics—specificity, sensitivity, precision, and F1 score—delivering a more nuanced picture of how well it identifies true apnea events, avoids false alarms, and strikes a balance between precision and recall. This comprehensive evaluation framework ensures that the model's effectiveness is measured from multiple angles rather than by overall correctness alone. Moreover, the training regimen for SleepNet incorporates advanced regularization techniques such as dropout to mitigate overfitting—a common pitfall in deep learning. By randomly deactivating portions of the network during training, the model learns to generalize more effectively to unseen data, maintaining consistent

performance. In contrast, some prior approaches may not have applied these stringent training protocols, which could explain their comparatively lower reliability. In summary, SleepNet represents a significant step forward in sleep apnea detection by combining multimodal physiological inputs with state-of-the-art deep learning methods. Disparities in performance between SleepNet and earlier models can be traced back to differences in data quality, architecture design, preprocessing workflows, selection of evaluation metrics, and training procedures. These contrasts highlight the need for continual advancements in deep learning to refine diagnostic tools in healthcare. Looking ahead, future work should aim to validate SleepNet across a range of clinical settings and patient groups to determine its real-world effectiveness. Researchers will also need to tackle practical challenges like signal noise, synchronizing multiple data streams, and creating tailored models for individual patients. Addressing these issues will be essential for further boosting SleepNet's accuracy and ultimately improving outcomes for people with sleep disorders. Sleep apnea is a prevalent respiratory condition that can lead to serious health complications if not identified and managed promptly. Recent innovations in diagnostic techniques—ranging from electrocardiogram (ECG) analysis and respiratory effort monitoring to tracheal body sound detection and pulse oximetry—have demonstrated encouraging results for individual patients. In particular, there has been growing interest in leveraging single-lead ECG recordings for automatic sleep apnea identification. A notable investigation in this area examined and preprocessed recordings from the PhysioNet ECG Sleep Apnea repository and introduced a hybrid deep learning model combining convolutional and recurrent layers. The authors created two distinct datasets from the Apnea-ECG collection: one containing only ECG waveforms and another that paired ECG signals with nasal airflow and abdominal respiratory effort measurements. Using a one-dimensional CNN followed by bidirectional gated recurrent units (BiGRU), their network achieved 95.08% accuracy on the ECG-only dataset, outperforming all prior approaches on that corpus. When the additional respiratory effort channels were incorporated, performance rose further—to 95.19% accuracy and 96.12% sensitivity—surpassing the earlier methods of Abasi et al.[33], Qatmh et al.[48], and Fei et al.[51]. by more than two percentage points. These findings underscore the promise of expanding beyond single-modality inputs and suggest that future work should explore models that fuse multiple patient-specific signals for even greater diagnostic precision. To facilitate practical deployment, the researchers also applied network pruning and post-training quantization, streamlining the model so it can run efficiently on wearable hardware. This optimization paves the way for a comfortable, noninvasive monitoring device capable of continuous, real-time sleep apnea detection. Overall, this study illustrates how combining diverse neural network architectures can advance complex medical diagnostics and points toward exciting directions for continued research in sleep medicine.

The future scope of this research is improving the applicability of the proposed model in clinical settings by incorporating additional physiological parameters such as oxygen saturation and body movement. The authors also aim to develop a wearable device that include sensors for ECG monitoring, nasal airflow measurement, and respiratory effort gauging for the multimodal detection system to be deployed in real-world scenarios. Additionally, studies may be done to monitor the power and memory consumption of SleepNet when deployed in wearable devices, aiming to extend battery life through algorithmic optimizations and intelligent power management strategies.

Hypothesis and limitations
Hypothesis
This study hypothesizes that integrating multiple physiological signals such as ECG, nasal airflow, and abdominal respiratory effort will lead to a significant improvement in detecting sleep apnea compared to relying on a single signal. The hypothesis is grounded in the idea that each physiological signal provides unique and complementary information, which, when combined, can enhance the model's ability to detect apneic episodes more effectively. The central propositions of this hypothesis include:

- **Multimodal integration**: By combining ECG with respiratory signals, the model is expected to capture the complex physiological changes associated with sleep apnea more comprehensively. The integration of these signals allows the model to utilize the complementary nature of each signal, resulting in a more robust diagnostic tool that is better at identifying apneic events than models relying on a single physiological signal.
- **Enhanced diagnostic accuracy**: The proposed SleepNet model, leveraging deep learning techniques, is anticipated to outperform unimodal approaches in terms of accuracy, sensitivity, and specificity. Deep learning models, especially those that utilize multimodal data, are capable of learning complex patterns from raw signals without requiring manual feature extraction. This is expected to lead to improved accuracy in detecting sleep apnea events, as the model can learn to consider multiple physiological indicators simultaneously, enhancing its diagnostic performance.

Limitations
Despite its potential, the SleepNet model comes with several limitations that should be carefully considered:

- **Data quality and noise**: The model's effectiveness relies heavily on the quality of the input signals. Factors like noise, motion artifacts, and improper sensor placement can degrade the accuracy of predictions.
- **Limited generalizability**: The datasets used in this research may not adequately represent the diversity of sleep apnea patients. As a result, the model's performance might vary across different populations or clinical environments.
- **Complexity of signal integration**: Incorporating multiple physiological signals adds layers of complexity in both processing and interpretation. Misaligned or conflicting data sources could hinder the model's ability to classify events correctly.

- **Resource demands**: Training and deploying the deep learning framework of SleepNet require significant computational power, which could limit its feasibility for real-time applications or in settings with restricted resources.
- **Model interpretability**: Like many deep learning models, SleepNet functions as a "black box," offering limited insights into its decision-making process. This lack of transparency may pose challenges in gaining clinical trust and acceptance.
- **Future modifications**: Currently optimized for a specific set of physiological signals, the model would need further development to incorporate additional data types or adapt to personalized patient requirements. Such advancements would require substantial research and validation efforts.

Data availability

The Apnea-ECG Database (https://www.physionet.org/content/apnea-ecg/1.0.0/), the dataset used in the research, is publicly accessible via the Physionet website under the terms of the Open Data Commons Attribution License v1.0 .

References

1. Gutta, S., Cheng, Q., Nguyen, H. D. & Benjamin, B. A. Cardiorespiratory model-based data-driven approach for sleep apnea detection. *IEEE J. Biomed. Health Inform.* **22**, 1036–1045 (2017).
2. Shen, Q., Qin, H., Wei, K. & Liu, G. Multiscale deep neural network for obstructive sleep apnea detection using RR interval from single-lead ECG signal. *IEEE Trans. Instrum. Meas.* **70**, 1–13 (2021).
3. Surrel, G., Aminifar, A., Rincón, F., Murali, S. & Atienza, D. Online obstructive sleep apnea detection on medical wearable sensors. *IEEE Trans. Biomed. Circuits Syst.* **12**, 762–773 (2018).
4. Acquavella, J., Mehra, R., Bron, M., Suomi, J.M.-H. & Hess, G. P. Prevalence of narcolepsy and other sleep disorders and frequency of diagnostic tests from 2013–2016 in insured patients actively seeking care. *J. Clin. Sleep Med.* **16**, 1255–1263 (2020).
5. Altevogt, B. M. & Colten, H. R. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem* (National Academies Press, Washington, DC, 2006). Copyright© 2006, National Academy of Sciences.
6. Peppard, P. E. et al. Increased prevalence of sleep-disordered breathing in adults. *Am. J. Epidemiol.* **177**, 1006–1014 (2013).
7. Lyons, M. M., Bhatt, N. Y., Pack, A. I. & Magalang, U. J. Global burden of sleep-disordered breathing and its implications. *Respirology* **25**, 690–702 (2020).
8. Neikrug, A. B. & Ancoli-Israel, S. Sleep disorders in the older adult-a mini-review. *Gerontology* **56**, 181–189 (2010).
9. Li, K., Pan, W., Li, Y., Jiang, Q. & Liu, G. A method to detect sleep apnea based on deep neural network and hidden markov model using single-lead ecg signal. *Neurocomputing* **294**, 94–101 (2018).
10. Papini, G. B. *et al.* On the generalizability of ecg-based obstructive sleep apnea monitoring: merits and limitations of the apnea-ecg database. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 6022–6025 (IEEE, 2018).
11. Wang, T., Lu, C., Shen, G. & Hong, F. Sleep apnea detection from a single-lead ecg signal with automatic feature-extraction through a modified lenet-5 convolutional neural network. *PeerJ* **7**, e7731 (2019).
12. Feng, K., Qin, H., Wu, S., Pan, W. & Liu, G. A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram. *IEEE Trans. Instrum. Meas.* **70**, 1–12 (2020).
13. Ayatollahi, A., Afrakhteh, S., Soltani, F. & Saleh, E. Sleep apnea detection from ECG signal using deep cnn-based structures. *Evol. Syst.* **14**, 191–206 (2023).
14. Song, C., Liu, K., Zhang, X., Chen, L. & Xian, X. An obstructive sleep apnea detection approach using a discriminative hidden markov model from ECG signals. *IEEE Trans. Biomed. Eng.* **63**, 1532–1542 (2015).
15. Wang, L., Lin, Y. & Wang, J. A RR interval based automated apnea detection approach using residual network. *Comput. Methods Programs Biomed.* **176**, 93–104 (2019).
16. Singh, S. A. & Majumder, S. A novel approach osa detection using single-lead ECG scalogram based on deep neural network. *J. Mech. Med. Biol.* **19**, 1950026 (2019).
17. Burgos, A., Goni, A., Illarramendi, A. & Bermudez, J. Real-time detection of apneas on a PDA. *IEEE Trans. Inf. Technol. Biomed.* **14**, 995–1002 (2009).
18. Malik, J., Lo, Y.-L. & Wu, H.-T. Sleep-wake classification via quantifying heart rate variability by convolutional neural network. *Physiol. Meas.* **39**, 085004 (2018).
19. Yoon, H. et al. Slow-wave sleep estimation for healthy subjects and osa patients using R–R intervals. *IEEE J. Biomed. Health Inform.* **22**, 119–128 (2017).
20. Kalkbrenner, C. et al. Apnea and heart rate detection from tracheal body sounds for the diagnosis of sleep-related breathing disorders. *Med. Biol. Eng. Comput.* **56**, 671–681 (2018).
21. Lin, R. et al. A new approach for identifying sleep apnea syndrome using wavelet transform and neural networks. *Biomed. Eng. Appl. Basis Commun.* **18**, 138–143 (2006).
22. Xie, B. & Minn, H. Real-time sleep apnea detection by classifier combination. *IEEE Trans. Inf. Technol. Biomed.* **16**, 469–477 (2012).
23. Zhang, M. et al. Optimized compression for implementing convolutional neural networks on FPGA. *Electronics* **8**, 295 (2019).
24. Zhang, N., Wei, X., Chen, H. & Liu, W. FPGA implementation for CNN-based optical remote sensing object detection. *Electronics* **10**, 282 (2021).
25. Amra, B., Rahmati, B., Soltaninejad, F. & Feizi, A. Screening questionnaires for obstructive sleep apnea: An updated systematic review. *Oman Med. J.* **33**, 184 (2018).
26. Tsuiki, S. *et al.* Machine learning for image-based detection of patients with obstructive sleep apnea: An exploratory study. *Sleep Breath.* **25**, 2297–2305 (2021).
27. Mabrouk, M. et al. Human breathing rate estimation from radar returns using harmonically related filters. *J. Sens.* **2016**, 1–7 (2016).
28. De Zambotti, M., Cellini, N., Goldstone, A., Colrain, I. M. & Baker, F. C. Wearable sleep technology in clinical and research settings. *Med. Sci. Sports Exerc.* **51**, 1538 (2019).
29. Chen, J., Shen, M., Ma, W. & Zheng, W. A spatio-temporal learning-based model for sleep apnea detection using single-lead ecg signals. *Front. Neurosci.* **16**, 972581 (2022).
30. Avcı, C. & Akbaş, A. Sleep apnea classification based on respiration signals by using ensemble methods. *Bio-Med. Mater. Eng.* **26**, S1703–S1710 (2015).
31. Ozdemir, G., Nasifoglu, H. & Erogul, O. A time-series approach to predict obstructive sleep apnea (OSA) episodes. In *Proceedings of the 2nd World Congress on Electrical Engineering and Computer Systems and Science* 1–8 (2016).

32. Thommandram, A., Eklund, J. M. & McGregor, C. Detection of apnoea from respiratory time series data using clinically recognizable features and knn classification. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 5013–5016 (IEEE, 2013).
33. Abasi, A. K., Aloqaily, M. & Guizani, M. Optimization of CNN using modified honey badger algorithm for sleep apnea detection. *Expert Syst. Appl.* **229**, 120484 (2023).
34. Hemrajani, P., Dhaka, V. S., Rani, G., Shukla, P. & Bavirisetti, D. P. Efficient deep learning based hybrid model to detect obstructive sleep apnea. *Sensors* **23**, 4692 (2023).
35. McClure, K. et al. Classification and detection of breathing patterns with wearable sensors and deep learning. *Sensors* **20**, 6481 (2020).
36. Penzel, T., Moody, G. B., Mark, R. G., Goldberger, A. L. & Peter, J. H. The apnea-ecg database. In *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)* 255–258 (IEEE, 2000).
37. Goldberger, A. L. *et al.* Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation***101**, e215–e220 (2000).
38. Paul, T. et al. ECG and SpO2 signal-based real-time sleep apnea detection using feed-forward artificial neural network. *AMIA Summits Transl. Sci. Proc.* **2022**, 379 (2022).
39. Yeh, C.-Y., Chang, H.-Y., Hu, J.-Y. & Lin, C.-C. Contribution of different subbands of ecg in sleep apnea detection evaluated using filter bank decomposition and a convolutional neural network. *Sensors* **22**, 510 (2022).
40. Chen, L., Zhang, X. & Song, C. An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram. *IEEE Trans. Autom. Sci. Eng.* **12**, 106–115 (2014).
41. Cheng, M., Sori, W. J., Jiang, F., Khan, A. & Liu, S. Recurrent neural network based classification of ecg signal features for obstruction of sleep apnea detection. In *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, vol. 2, 199–202 (IEEE, 2017).
42. Liang, X., Qiao, X. & Li, Y. Obstructive sleep apnea detection using combination of CNN and LSTM techniques. In *2019 IEEE 8th Joint International Information Technology and Artificial iIntelligence Conference (ITAIC)* 1733–1736 (IEEE, 2019).
43. Bozkurt, F., Uçar, M. K., Bozkurt, M. R. & Bilgin, C. Detection of abnormal respiratory events with single channel ECG and hybrid machine learning model in patients with obstructive sleep apnea. *Irbm* **41**, 241–251 (2020).
44. Cheng, J., Zou, Q. & Zhao, Y. ECG signal classification based on deep CNN and bilstm. *BMC Med. Inform. Decis. Mak.* **21**, 1–12 (2021).
45. Stretch, R. et al. Predicting nondiagnostic home sleep apnea tests using machine learning. *J. Clin. Sleep Med.* **15**, 1599–1608 (2019).
46. Cho, K. *et al.* Learning phrase representations using RNN encoder–decoder for statistical machine translation. arXiv:1406.1078 (2014).
47. Martin-Gonzalez, S. et al. Heart rate variability feature selection in the presence of sleep apnea: An expert system for the characterization and detection of the disorder. *Comput. Biol. Med.* **91**, 47–58 (2017).
48. Qatmh, M. *et al.* Sleep apnea detection based on ECG signals using discrete wavelet transform and artificial neural network. In *2022 Advances in Science and Engineering Technology International Conferences (ASET)* 1–5 (IEEE, 2022).
49. Urtnasan, E., Park, J.-U. & Lee, K.-J. Automatic detection of sleep-disordered breathing events using recurrent neural networks from an electrocardiogram signal. *Neural Comput. Appl.* **32**, 4733–4742 (2020).
50. Erdenebayar, U., Kim, Y. J., Park, J.-U., Joo, E. Y. & Lee, K.-J. Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram. *Comput. Methods Programs Biomed.* **180**, 105001 (2019).
51. Fei, Y. et al. Intelligent classification of antenatal cardiotocography signals via multimodal bidirectional gated recurrent units. *Biomed. Signal Process. Control* **78**, 104008 (2022).

Acknowledgments

Author contributions

Conceptualization, P.H., V.S.D., G.R., S.V., K.,.,.; methodology, M.W., J.S., M.F.I., ; software, P.H., V.S.D., G.R., S.V., K.J.S.,., M.W., M.F.I.; validation, P.H., V.S.D., G.R., S.V., K., M.W., J.S., M.F.I.,; formal analysis, P.H., V.S.D., G.R., ; investigation, S.V., K., M.W., J.S., M.F.I.; resources, M.W., J.S., M.F.I., data curation, P.H., V.S.D., G.R., S.V., K., writing—original draft preparation, P.H., V.S.D., G.R., S.V., K., ; writing—review and editing, M.W., J.S., M.F.I.; visualization, P.H., V.S.D., G.R., S.V., K.,supervision M.W., J.S., M.F.I., ; project administration, M.W., J.S., M.F.I.,; funding acquisition, M.W., M.F.I..; All authors have read and agreed to the published version of the manuscript.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.W. or M.F.I.

Reprints and permissions information is available at www.nature.com/reprints.