Article

Multimodal Neural Network Analysis of Single-Night Sleep Stages for Screening Obstructive Sleep Apnea

Jayroop Ramesh *, Zahra Solatidehkordi , Assim Sagahyroon and Fadi Aloul

Department of Computer Science and Engineering, American University of Sharjah,
Sharjah 26666, United Arab Emirates; g00059068@aus.edu (Z.S.); asagahyroon@aus.edu (A.S.);
faloul@aus.edu (F.A.)
* Correspondence: b00057412@aus.edu

Abstract: Obstructive Sleep Apnea (OSA) is a prevalent chronic sleep-related breathing disorder characterized by partial or complete airway obstruction. The expensive, time-consuming, and labor-intensive nature of the gold-standard approach, polysomnography (PSG), and the lack of regular monitoring of patients' daily lives with existing solutions motivates the development of clinical support for enhanced prognosis. In this study, we utilize image representations of sleep stages and contextual patient-specific data, including medical history and stage durations, to investigate the use of wearable devices for OSA screening and comorbid conditions. For this purpose, we leverage the publicly available Wisconsin Sleep Cohort (WSC) dataset. Given that wearable devices are adept at detecting sleep stages (often using proprietary algorithms), and medical history data can be efficiently captured through simple binary (yes/no) responses, we seek to explore neural network models with this. Without needing access to the raw physiological signals and using epoch-wise sleep scores and demographic data, we attempt to validate the effectiveness of screening capabilities and assess the interplay between sleep stages, OSA, insomnia, and depression. Our findings reveal that sleep stage representations combined with demographic data enhance the precision of OSA screening, achieving F1 scores of up to 69.40. This approach holds potential for broader applications in population health management as a plausible alternative to traditional diagnostic approaches. However, we find that purely modality-agnostic sleep stages for a single night and routine lifestyle information by themselves may be insufficient for clinical utility, and further work accommodating individual variability and longitudinal data is needed for real-world applicability.

Keywords: deep learning; electronic health records; health informatics; hypnograms; machine learning; mental health; Obstructive Sleep Apnea

1. Introduction

Sleep disorders, which involve disturbances in sleep architecture, significantly affect psychophysical health and are linked to various health conditions. Obstructive Sleep Apnea (OSA) is a common sleep disorder, marked by the repeated narrowing or collapse of the throat while sleeping [1]. This condition affects 5–15% of the global population to a moderate extent and causes frequent sleep disturbances, occurring up to 15 times an hour [2]. OSA and other sleep disorders often result in increased sleep latency, fragmented sleep, and imbalances in sleep stages. Such conditions are associated with a wide range of pathophysiological states, including cardiovascular, respiratory, endocrine, metabolic, and neurological/psychiatric disorders [1]. In particular, the link between OSA and depression can be attributed to the concept of neuroplasticity, where sleep fragmentation and

chronic intermittent hypoxia trigger a series of physiological responses. These include activation of the sympathoadrenal system, oxidative stress, systemic inflammation, increased corticosteroid levels, and consequential cognitive decline and poor mental health [2]. Sleep disorders severely affect sleep quality, so even an average of 8 h of sleep does not feel completely sufficient for individuals. This has implications for general well-being, as it causes periodic episodes of depressive and anxiety and other mental health disorders [3].

Polysomnography (PSG), while accurate for sleep assessment, is limited in practice by the need for a clinical setting and its unique nature, which may not reflect normal sleep patterns. The discomfort of multiple sensors attached can also affect sleep quality, highlighting the need for more convenient monitoring methods. In contrast, the emergence of wearable monitoring technologies, such as the Apple iWatch 6, Fitbit Sense, and Garmin Vivo Active4, has marked a significant advancement in the detection and management of sleep disorders [4]. Our work leverages sleep stages scored by sleep technicians using the PSG, with the intention of empirically validating the utility of a single-night sleep architecture without additional physiological signals. We believe that this work can be extended to wearable devices, which provide access only to sleep stage information. These devices offer insights into night-to-night variability and overall sleep patterns, including the impact of patient activity levels and sleep stages on mental health and sleep disorder management. However, clinical and research applications of these devices are limited by proprietary algorithms and restricted access to raw sensor data, hindering the development of customized diagnostic models and the integration of wearables into clinical practice [5,6]. These limitations become particularly evident when dealing with complex conditions such as OSA, insomnia, and depression, which require a nuanced analysis of multimodal data. Our proposed deep learning framework addresses this challenge by combining data obtained from wearables with patient-specific medical history, effectively capturing the nuances of physiological and contextual information.

Furthermore, traditional diagnostic models do not account for individual differences, causing variability in the performance of the screening achieved. These models are trained with the intention of finding the largest separation between the classification outcomes and, as such, often have trouble with data that are on the periphery of multiple classes [7]. Addressing this gap, our study advocates the integration of multiple dimensions of data, reflective of the complex pathophysiology of comorbid conditions, to enable improved precision medicine. In this work, we propose a multimodal learning approach, in which we try to classify OSA, depression severity, and insomnia using the same set of independent characteristics while accounting for cofounders for each condition. This methodology promises a more nuanced understanding of the intricate relationship between sleep disorders and mental health.

For this purpose, we construct image representations of sleep stages that can be compared to the graphs most sleep tracking applications and devices can generate [8] to classify OSA. We implement the state-of-the-art Convolutional Bidirectional Long-Short-Term Memory (CNN-BiLSTM) model to process both the temporal sequence component (i.e., transitions) and the spatial content of the pixel colors (i.e., duration of each stage) [9]. This study expands beyond individual patient health to address broader aspects of population health management. It employs a noninvasive and cost-effective approach, using comprehensive patient data from electronic health records. These records provide detailed information on medical history, including prevalent comorbidities such as cardiovascular disorders, diabetes, obesity, and hypertension, as well as demographic variables such as age, race, and gender, which have been linked to OSA, insomnia, and depression.

Our focus is on elucidating the connection between fragmented sleep, its cognitive and mental health implications, and the transformative potential of machine learning in medical diagnostics and treatment.

The contributions of this study are listed below.

1. The development of deep learning models for the detection of OSA, insomnia, and depression using sleep stage data from wearable devices and patient-specific medical history.
2. The construction of hypnodensity and hypnogram visualizations to enrich quantitative data obtained from wearable devices.
3. The application of multimodal learning to analyze OSA in the context of its comorbidities.

The structure of this paper is as follows. Section 2 details the materials and methods, Section 3 reports the results, and Section 4 presents a discussion of the results.

2. Materials and Methods

The methodology of this study is structured as an end-to-end sequential pipeline (as shown in Figure 1, involving (i) preprocessing of datasets with hypnograms and hypnodensity graphs, (ii) processing of EHR data, and (iii) training and evaluation of deep learning models.
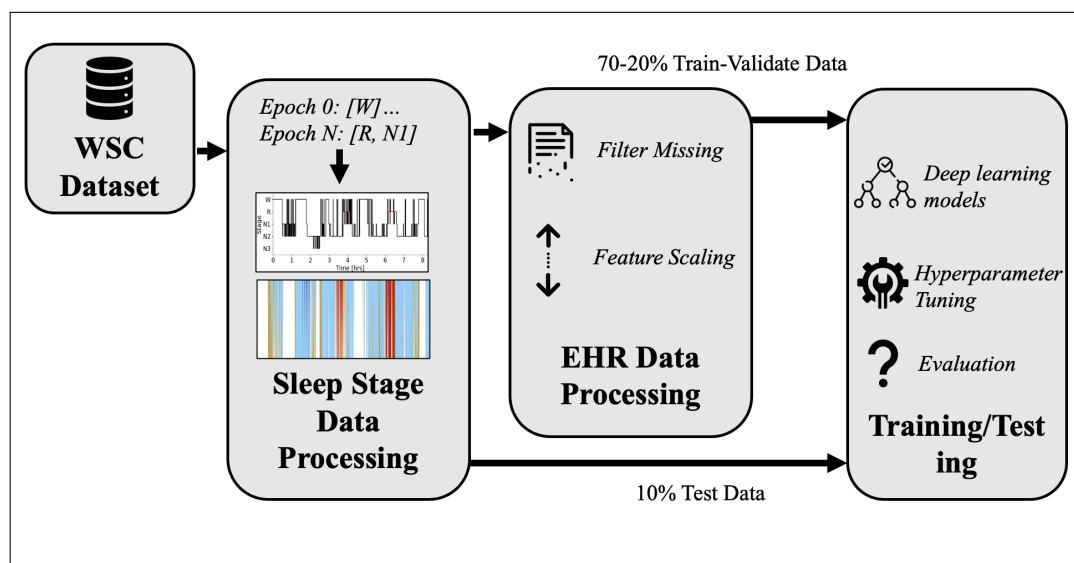

Figure 1. Flow diagram of the data pipeline from data preprocessing to analysis.

2.1. Dataset

The primary data source for this study is the Wisconsin Sleep Cohort (WSC) of the University of Wisconsin-Madison, which investigates the causes, consequences, and natural history of sleep disorders [10]. This data set consists of 2570 records from 1500 participants evaluated at four-year intervals, where each participant can have up to five records in the study. The dataset contains a wide range of electronic health records (EHR) that includes demographics and health history. The latter includes general health status, existing medical conditions, and sleep quality, collected via self-administered questionnaires. The exhaustive list of characteristics is provided in Table 1. Additional details can be found in the original work [10], and any specific pre-processing followed the pipeline outlined in [11,12]. We account for sleep parameters mainly corresponding to Rapid Eye Movement (REM) and non-REM sleep stages [13] to ensure unformity between modality and image representations.

Table 1. Categorized independent features available from WSC with their associated descriptions.

Feature	Description	Type
Demographics and General Habits		
Age	Measured in years	Numeric
Gender	Male or Female	Nominal
Body Mass Index	Measures height and weight ratio in kg/m^2	Numeric
Alcohol	Number of drinks weekly	Numeric
Caffeine	Number of drinks weekly	Numeric
Smoking	Indicates yes or no	Nominal
Health Status		
Excessive daytime sleepiness	Feeling of sleepiness and fatigue	Nominal
Epworth Sleepiness Scale	Likelihood of dozing off during day	Nominal
State Anxiety Score (STAI-1)	Quantifies current anxiety	Numeric
Trait Anxiety Score (STAI-2)	Quantifies general anxiety	Numeric
Cardiovascular Conditions	Indicates present or absent	Nominal
Diabetes Conditions	Indicates present or absent	Nominal
Thyroid Conditions	Indicates present or absent	Nominal
Arthritis Conditions	Indicates present or absent	Nominal
Asthma Conditions	Indicates present or absent	Nominal
Emphysema Conditions	Indicates present or absent	Nominal
Stroke Conditions	Indicates present or absent	Nominal
Medication Status		
Depression Medication	Indicates taking or not	Nominal
Anxiety Medication	Indicates taking or not	Nominal
Cholesterol Medication	Indicates taking or not	Nominal
Hypertension Medication	Indicates taking or not	Nominal
Diabetes Medication	Indicates taking or not	Nominal
Thyroid Medication	Indicates taking or not	Nominal
Asthma Medication	Indicates taking or not	Nominal
Narcotics Medication	Indicates taking or not	Nominal
Sedative Medication	Indicates taking or not	Nominal
Stimulant Medication	Indicates taking or not	Nominal
Antihistamines Medication	Indicates taking or not	Nominal
Androgen Medication	Indicates taking or not	Nominal
Decongestants Medication	Indicates taking or not	Nominal
Sleep Measures		
REM Latency	Time to reach first REM stage	Numeric
Wake After Sleep Onset	Wakefulness after first falling asleep	Numeric
Sleep Latency	Time to fall asleep	Numeric
REM Sleep Duration	Time in REM stage	Numeric
Total Sleep Duration	Time in bed asleep	Numeric
Sleep Efficiency	Percentage of time spent asleep while in bed	Numeric
NREM Sleep Duration	Time in NREM stage	Numeric
REM Sleep Percentage	Percentage of time spent asleep while in REM	Numeric
N1 Sleep Percentage	Percentage of time spent asleep while in N1	Numeric
N2 Sleep Percentage	Percentage of time spent asleep while in N2	Numeric
N34 Sleep Percentage	Percentage of time spent asleep while in N3 and N4	Numeric
Labels		
OSA	Indicates present or absent	Nominal
Insomnia	Indicates present or absent	Nominal
Depression	Indicates present or absent	Nominal

Objective measurements were obtained using an 18-channel polysomnography (PSG) system (Grass Instruments Model 78; Quincy, MA, USA). This system recorded sleep states using electroencephalography (EEG), electrooculography, and electromyography. Respiratory evaluations included nasal and oral airflow and oxyhemoglobin saturation, measured, respectively, by respiratory inductance plethysmography (Respitrace; Ambulatory Monitoring, Ardsley, NY, USA), thermocouples (ProTec, Hendersonville, TN and Validyne Engineering Corp pressure transducer, Northridge, CA, USA), and pulse oximetry (Ohmeda Biox 3740; Englewood, CO, USA). PSG recordings were analyzed every 30 s, scoring sleep stages and events indicative of apnea and hypopnea according to conventional standards. These events were defined by the discontinuation of airflow for 10 s and a discernible reduction in breathing, measured as the sum of chest and abdominal excursions accompanied by a decrease in oxyhemoglobin saturation of 4%. For each patient, these recordings were available alongside the EHR [10].

In our previous work, we outlined the feature processing steps for the same dataset, which was also followed for this work [11]. The Little Missing Completely Random Test was used to verify that the missing value pattern had no significant relationship with the rest of the data. Listwise deletion was used to remove entire records where the values for the clinical characteristics of interest were missing or had numeric values that were not plausible according to domain knowledge. Pearson's correlation coefficient for numerical variables, Kendall's Tau correlation coefficient for categorical variables, and mutual information and extremely random trees were also used as the feature selection method. Most features were found to have approximately similar levels of reasonable correlation with each of the target variables, except REM latency, which was consistently ranked higher. Thus, we used the subset of features described in Table 1, from the larger set of variables available in the WSC dataset. The features are readily collected from the combination of controllable factors on an individual basis (drinking AND physical health) and potentially chronic problems and medications. This suggests that individuals can have a noticeable level of agency in their own lives to curb the progression of adverse effects, provided the severity of comorbid conditions is reasonably low.

The binary labels used for OSA, depression, and insomnia classification use the Apnea–Hypopnea Index (AHI) with a cutoff at 5 and the Self-Rating Depression Scale score, which quantifies depression severity with a cutoff at 50, and the insomnia status was provided in the dataset [11]. After data cleansing to remove missing information, the dataset had 2564 records, split into 1456 patients with OSA and 1108 patients without OSA. For depression, there were 2020 patients with complete information, divided into 532 depressed patients and 2020 non-depressed patients. For insomnia, there were 2024 patients with complete information, divided into 993 patients with insomnia and 1031 patients without insomnia.

It is worth mentioning that relevant variables such as depression/anxiety medications and state/trait anxiety scores were not included in training models to classify depression, as this introduces obvious multicollinearity and could skew the learning process.

2.2. Sleep Stages

When sleep is analyzed in PSG, it is divided into discrete stages: wake, REM, non-REM (NREM) sleep stage 1 (N1), 2 (N2) and 3 (N3). Each stage is characterized by different criteria, as defined by consensus rules published in the American Academy of Sleep Medicine (AASM) Scoring Manual [14,15]. N1 (sleep onset) is characterized by a slowing of EEG, the disappearance of occipital alpha waves, decreased EMG, and slow rolling eye movements, while N2 is associated with spindles and K complexes. N3 is characterized by the dominance of slow, high-amplitude waves (>20%), while REM sleep is associated with

low voltage, desynchronized EEG with occasional saw tooth waves, low muscle tone and REM [16].

Generally, there is a progression from N1 to N3 and then to REM throughout the night, and this process repeats every 90 min. Each stage is associated with different physiological changes. For example, OSA is relatively less prevalent in N3 versus N2 because of the central control of breathing changes and is more severe in REM due to weakness in the upper airway muscles. Sleep continuity in depressed patients is often impaired, with more frequent wake-up periods, reduced sleep efficiency, and shortened REM [17].

Interpreting visual representations of sleep is a difficult task that requires domain knowledge and clinical experience. Sleep fragmentation and sleep stage distributions are particularly important considerations for clinician judgments, and automating this step with deep learning can help identify sleep-related conditions [18]. The rationale for considering the visual representation of the sleep stages is that clinicians often derive significant insights from visual representations of sleep architecture (as reported in Table 1).

Using the 30-second scored sleep stages, we constructed image representations that capture the entire night time spent in each stage (hypnodensity), as well as a sleep stage transition plot (hypnogram) [16,19]. Figure 2 shows the sleep patterns of a patient with WSC with high OSA and severity of depression, as well as the sleep patterns of a patient in the same cohort with low OSA and severity of depression. The rationale was that the information required to build both sets of features is readily available from commercial wearable fitness trackers or sleep-tracking apps.

The hypnogram provides a numeric indicator of wake and the sleep stages for the epoch score and often includes markings that include the time before the lights are turned off and the time after the lights are turned on. Labels and timestamps were provided in the dataset for removing these latter values, as well as for each epoch with the corresponding indicator. The *x*-axis shows the total duration of sleep throughout the night, and the *y*-axis shows the sleep stage label. On average, the participants in Figure 2 had a full night of sleep between 6 and 8 h, as in Patient 65492. It should be mentioned that some people woke up completely in the middle of the night and did not go back to sleep during the study, such as Patient 98255 in Figure 2.

The hypnodensity graph was introduced in [16] as a hypnogram that does not have a strict single sleep stage label but uses a membership function for each of the sleep stages, allowing more information to be conveyed about sleep trends. A hypnogram typically assigns a single sleep stage (e.g., REM, W, N1, N2 and N3) to each epoch; the cumulative probability of sleep refers to the fact that hypnodensity can show probabilities for multiple stages at the same time. This allows for the better capture of transitions and could help identify patterns during periods where sleep stages fluctuate. This is especially useful when an epoch has multiple labels (i.e., from different sleep scorers), and this uncertainty needs to be emphasized in the graphical representations. Essentially, instead of a numeric indicator as in the hypnogram, five probabilities are given that represent the likelihood of the sleep stage for the current epoch.

We adapt the code and steps provided in the original approach [16] (which utilized multiple models to predict fuzzy values for the multi-label probability distribution of the sleep stages) to produce a cumulative sleep stage probability throughout the night. This was performed with the intention of reducing oversampling and undersampling issues for certain fluctuating stages when strict single labels are used, especially during transitional states. The steps for graphical construction are as follows:

1. For each patient, a probability matrix is computed where each row corresponds to an epoch and each column corresponds to a specific sleep stage. This matrix is initialized

such that all elements are set to zero, and the stage for each epoch is then represented by setting the corresponding stage column to 1.

2. Once the epoch-wise probability matrix is computed for each patient, a cumulative sum is performed along the time axis (rows of the matrix).

3. The cumulative hypnodensity is then visualized as a polygon plot, where each polygon represents the cumulative probability of a sleep stage over time.
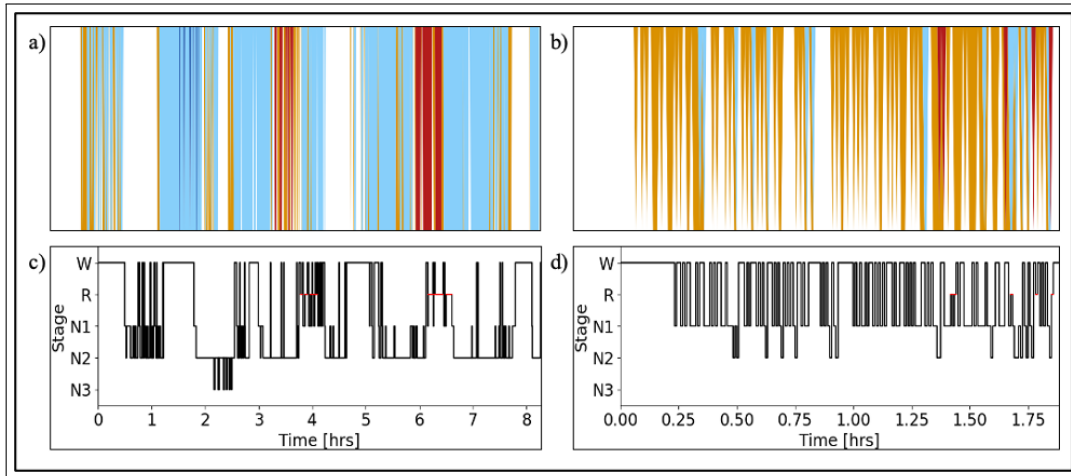

Figure 2. Examples of OSA-Depression labels transformed into hypnodensity graphs and hypnograms. (**a**) Hypodensity plot for Patient 65492 with high OSA and depression; (**b**) Hypnogram plot for Patient 65492 with high OSA and depression; (**c**) Hypodensity plot for Patient 98255 with low OSA and depression; (**d**) Hypnogram plot for Patient 98255 with low OSA and depression. Color scheme for stages: Wake: White; N1: Gold; N2: Light Blue; N2: Dark Blue; N3: Green; REM: Red.

2.3. Models

CNNs have become ubiquitous in the DL literature for the capability to automatically perform feature extraction effectively due to their inherent translation-invariance characteristics. More specifically, CNNs capture local conjunctions of features, namely sub-samples (either spatially or temporally) to reduce data dimensionality, detect semantic similarities, and utilize a shared-weight architecture to provide generalizability. We employ 1D CNNs, which are appropriate for one-dimensional inputs such as physiological signals. A standard CNN structure is composed of a set of layers, where each layer narrows down different aspects of the input data through local receptive fields. A series of simultaneous and then successive convolutional filters operate in tandem to extract relevant features and consolidate a learned representation of the input data in the form of feature maps/vectors. For a single convolution c_i of a signal $S_i^0 = [s_1, s_2, \ldots, s_n]$, where n is the 1500 sampling points (30-second waveform at a frequency of 50Hz), the following equation applies:

$$c_i^{lj} = h \left(b_j + \sum_{m=1}^{M} w_m^j x_{i+m-1}^j \right) \tag{1}$$

In Equation (1), l is the current layer index, h is the activation function, b is the bias of the jth feature map/vector, M is the kernel size, m is the filter index, and w_m^j is the weight of the corresponding feature map/vector.

Recurrent neural networks, particularly LSTMs, have proven their efficacy in many applications such as speech recognition, language modeling, and time series forecasting. Their ability to focus on sequential patterns in the data serves the purpose of capturing long-term temporal dependencies to a considerable extent. A single LSTM cell is made up of the cell state, the hidden stat,e and multiple gates, titled input i, output o, and

forget f gates. The gates themselves are composed of a sigmoid layer and a point-wise multiplication operation and regulate the flow of specific gradients to the cell state. The cell state essentially aggregates the information accumulated from the preceding time steps, whereas the hidden state encodes the information from the immediate previous timestep. The functional output of the LSTM cell C at any point in time can be defined as

$$C_t = f_t C_{t-1} + i_t c_t \tag{2}$$

In Equation (2), f_t is the activation of the forget gate, i_t is the activation of the input gate, and c_t is the input to the main cell. With the output gate activation o_t, the hidden-unit activations for a single cell are given by Equation (3):

$$h_t = o_t \tanh(c_t) \tag{3}$$

To address the unidirectional limitations of standard LSTM architectures in terms of being restricted to the previous context only, the solution proposed in [20] relating to the processing of data in both the forward and the backward directions is adopted. Let the forward layer consisting of T cells be indicated by h_t^f, and the backward layer consisting of the same T cells be denoted by h_t^b. The former processes the inputs in the fashion $[t_0, t_1, \dots t_T]$, while the latter processes in the opposite direction $[t_T, t_{T-1}, \dots t_0]$. The amalgamation of the outputs from both layers results in the vector $\hat{x}_T$ and is computed as follows:

$$
\begin{aligned}
h_t^f &= \tanh\left(W_{xh}^f x_t + W_{hh}^f h_{t-1} + b_h^f\right) \\
h_t^b &= \tanh\left(W_{xh}^b x_t + W_{hh}^b h_{t+1} + b_h^b\right) \\
y_t &= \left(W_{hy}^f h_t^f + W_{hy}^b h_t^b + b_y\right)
\end{aligned}
\tag{4}
$$

In Equation (4), W_x values are the input-to-hidden-layer weights, W_h values are the consecutive hidden-to-hidden-state weights, and b is the bias vector of the hidden state.

To emphasize any patterns found across the features, we employ the Luong-style soft-attention mechanism of the dot product proposed in [21]. Let the outputs of the BiLSTM block be $\hat{X}_i = [\hat{x}_1, \hat{x}_2 \dots \hat{x}_T]$ combined into a matrix A of size $N \times T$, where N is the size of the output vector and T is the number of time steps. Then, the weighted output vector h_{attn} of the attention mechanism is formulated as

$$
\begin{aligned}
\beta &= \mathrm{softmax}\left(w_{attn}^T A\right) \\
h_{attn} &= A\beta^T
\end{aligned}
\tag{5}
$$

In Equation (5), β is a weight vector calculated from the matrix A, and the output h_{attn} is calculated as the weighted sum of all outputs leaving the BiLSTM block.

The proposed model is shown in Figure 3 and consists of a single dilated convolutional layer with rectified linear unit (ReLU) on a time-distributed wrapper, a BiLSTM layer, and an attention mechanism followed by two fully connected layers with ReLU and sigmoid activations, respectively. In terms of regularization, batch normalization was added after the convolutional layer reduced covariance shift and a dropout layer preceded and succeeded the BiLSTMAttn block to mitigate overfitting. The model accepts the first input, with a dimensionality of [2552, 128, 128, 3, 1] indicating timesteps of 10 (3 s), with features of per-image signal sampling points. The CNN output is a X-dimensional feature vector that summarizes the spatial irregularities and patterns found in the raw input signal. This output is propagated through the BiLSTM network, where the temporal patterns are captured and fed into the attention mechanism to highlight the integral aspects of the

input. The model then accepts the second auxiliary input of sleep features and additional EHR measurements with a dimensionality of [37, 1] and concatenates this into the output of the attention mechanism to help improve the representation of the input. Finally, this concatenated vector passes through a sigmoid activation function for binary classification of the different classes.
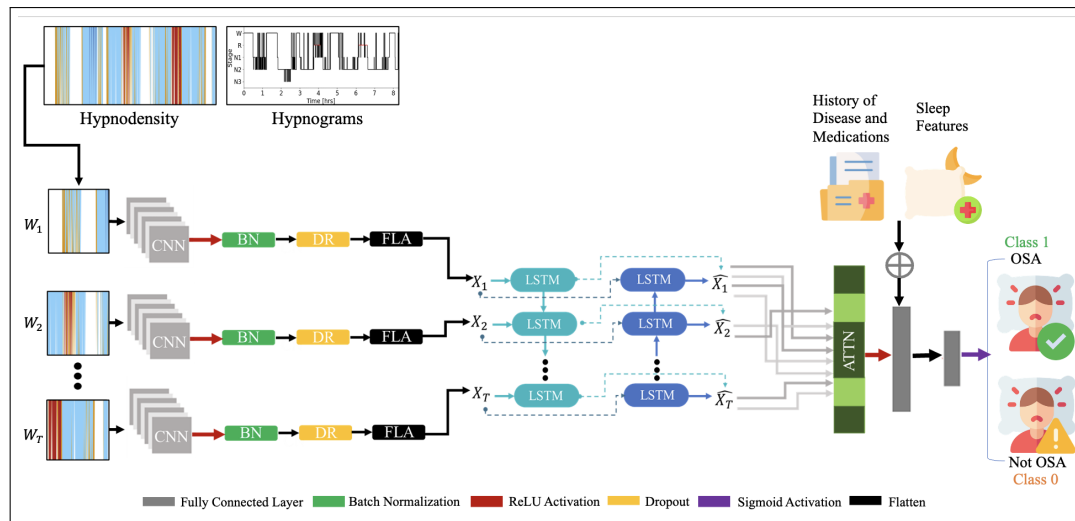

Figure 3. Architecture of the proposed CNN-BiLSTM-Attn model with sleep stage transition inputs with auxiliary data.

The CNN-BiLSTM architecture, combined with an attention mechanism, is hypothesized to be more applicable to hypnodensity and hypnogram data because it effectively utilizes spatial and temporal characteristics. CNNs are adept at capturing local patterns in the data, such as sharp transitions or microarousal, which are key to identifying sleep stages. By convolving over the input, CNNs extract hierarchical features while maintaining robustness to noise or slight temporal misalignments. This is especially important for hypnodensity images, where the transitions between stages or subtle changes in probabilities require detailed local feature extraction.

The BiLSTM component complements CNNs by capturing temporal dependencies and contextual trends within sleep data. Sleep staging has a strong temporal nature, as current stages are influenced by prior and subsequent epochs. BiLSTM processes sequences in both forward and backward directions, allowing the model to understand the broader context, such as recurring sleep cycles or transitions between sleep stages. The addition of an attention mechanism enhances this process by dynamically focusing on contextually important features, such as critical stage transitions or ambiguous epochs, making the model more robust to intra- and inter-rater disagreements. The attention mechanism also improves interpretability by highlighting which regions of the data were the most influential in the predictions.

Standard scaling was applied to the images to bring the pixel values to the [0, 1] range for faster convergence. The training of the model over 25 epochs was performed with the adaptive moment estimation (ADAM) optimizer with an initial learning rate of 1×10^{-5}, which was decreased by 10% if the performance in validation stagnated for 5 epochs [22]. We evaluated the trained model following a split 70%-20%-10% training–validation test separated at the patient level. A batch size of 16 and the binary cross entropy loss were utilized. Hyperparameters were selected empirically through grid search, with value ranges set as per [23] which uses similar models.

3. Results

The metrics of accuracy, sensitivity, specificity, and F1 score are used for quantitative evaluation. The results are reported in Table 2. The rationale for testing multiple feature combinations is to provide a systematic evaluation benchmark in the realm of machine learning for the purpose outlined in this work. The code is provided here: https://colab.research.google.com/drive/1t6n6Zy03TATbd-_TRfbJ1g8Tja_N0SN5?usp=sharing, accessed on 10 January 2024.

Table 2. Quantitative model performance metrics for classification of each condition with sleep stage architecture, sleep, and medical features.

Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
OSA Classification				
Hypnodensity	54.20	79.70	21.00	57.80
Hypnodensity + Sleep + Medical	63.10	74.20	48.70	69.40
Hypnograms	54.00	79.70	20.50	57.80
Hypnograms + Sleep + Medical	62.80	74.10	48.10	69.30
Insomnia Classification				
Hypnograms + Sleep + Medical	54.90	54.60	55.10	53.90
Depression Classification				
Hypnograms + Sleep + Medical	82.60	31.20	96.10	42.80

In OSA classification, models relying solely on a single feature type, either hypnodensity or hypnograms, achieved an F1 score of 57.80. The inclusion of sleep and medical data led to a noticeable improvement, with F1 scores increasing to 69.40 and 69.30, respectively. This enhancement indicates a more balanced model performance in terms of precision and recall.

In terms of accuracy, both hypnodensity (54.20%) and hypnograms (54.00%) perform roughly the same, i.e., once again only improving over random chance when additional contextual information is incorporated. This underscores the importance of adding patient characteristics and historical health records, allowing the assimilation of patient-specific nuances into the classification process. For insomnia and depression with the same features, scores of (54.90%) and (82.60%) are obtained, respectively. When using only image representations, the training process was unstable, likely because single-night sleep transitions offer insufficient information for these two conditions.

Our models achieved high sensitivity ($\approx$79%) but relatively lower specificity ($\approx$48%) for OSA classification. This suggests that the model effectively identifies true positives but struggles with false positives. Although high sensitivity is valuable for screening, especially for conditions such as OSA, where false negatives can lead to serious health risks, low specificity can result in unnecessary clinical evaluations. An approach to improve specificity is to incorporate stricter decision thresholds or ensemble models that combine predictions from multiple classifiers.

For insomnia, neither metric is remarkable ($\approx$55%), and for depression, there is a high specificity (96.10%), which is not ideal for screening potential at-risk patients as the sensitivity is only 31.20%).

4. Discussion

Although wearable devices provide valuable information on sleep stages, they typically do not provide the underlying physiological signal data, which are essential for a more detailed analysis. Our research, therefore, explored graphical representations that serve as a proxy for these underlying signals. Combining visual sleep stage patterns with medical histories and past health records, the models achieved a substantial improvement in the classification accuracy of OSA. This result indicates that data from wearables, while limited to sleep stages, gain significant value when combined with additional patient-centric information. The improved performance of the model when using this set of enriched features underscores the merit of this method, suggesting that it is a promising approach to the classification of sleep disorders.

Due to the multifaceted nature of OSA and its comorbidities, we also sought to examine its association with insomnia and depression (marginally better than random chance). For insomnia, models struggled to delineate insomniacs effectively, even among patients with higher-severity OSA. This ambiguity reflects the mixed findings within the literature. Some studies [11,24] report that nocturnal awakenings can be pertinent to chronic insomnia only and not related to OSA, some [7,25] suggest that they are not necessarily precursors to each other, and yet more [26,27] view insomnia as a comorbidity arising from OSA. In examining depression, as measured by the Zung questionnaire, our models similarly underperformed, mirroring the results of studies [28–30] that employ wearable technology. Although the initial data appeared promising, the link between depression and sleep disorders proved to be too weak to support the development of reliable machine learning models. Comparable machine learning efforts [31,32] have mirrored our results, which can be attributed to the significant variation in the severity of depression between individuals. In particular, the variance in the severity of depression symptoms and associated behaviors between subjects with mild to moderate depression presented a challenge. This variability led to difficulty in identifying clear patterns, raising concerns that the models may have learned spurious correlations rather than meaningful relationships. Such findings imply that the influence of depression on sleep architecture extends beyond the scope of data captured in a single night.

The limitations in distinguishing insomnia and depression highlight the inherent complexities of these conditions, as well as the subtlety and heterogeneity of the associated physiological and behavioral markers. Insomnia often reflects subjective complaints, such as difficulty in initiating or maintaining sleep, which do not always translate into measurable deviations in the sleep architecture [33]. Similarly, depression is characterized by various symptoms, such as mood disturbances and cognitive changes that may influence sleep, but are not directly detectable through wearable devices or PSG. This lack of specificity in the available data reduces the model's ability to generalize effectively across different patient profiles. Recent work also observes this phenomenon, and our work is in agreement with respect to the inconclusiveness of the interplay between sleep and these conditions [34,35]. To authentically capture their complex interactions, a longitudinal study design is necessary to observe long-term trends and patterns. To improve diagnostic precision, future studies can integrate additional features such as heart rate variability, respiratory rates, and accelerometer data, which can complement hypnodensity and hypnogram data by providing information on physiological states related to sleep quality. Leveraging transfer learning with large models pre-trained on similar biomedical or physiological datasets can provide a strong foundation and enable better generalization when applied to task-specific datasets. Data augmentation techniques, such as simulating variations in sleep stage transitions or durations, could also improve the robustness of the model by increasing the diversity of training data. In addition, multitask learning

approaches could improve the model's ability to detect overlapping conditions such as OSA, depression, and insomnia by leveraging shared representations across these disorders. These methods can significantly improve the generalizability of the model in real-world clinical applications.

Due to the subjectivity of the annotation task and the resulting inter-rater disagreements between experts, ambiguous or low-certainty epoch-by-epoch labeling may contain errors. The sleep staging methods in portable sleep trackers are optimized for epoch-by-epoch performance, which does not guarantee that clinically relevant characteristics will be captured throughout the night [36,37]. As such, fragmented sleep identification with respect to the reference PSG is poor, and there is no standardized approach for the validation of the wearable computation of the overnight statistics [38]. Thereby, the goal of our work was to combine multiple sources of data that may be readily available on wearables and study if deep learning approaches can prove useful in screening for multi-morbidity in limited data settings. We used the gold standard PSG data to perform our study, and this establishes a baseline for extensions to applications such as narcolepsy [39], emotional climate [40], and other sleep disorders. In terms of limitations, a lack of diversity in the collected data (single dataset) and the artificial structure of the experiments (PSG in controlled settings) may pose challenges while adapting to noisy real-world datasets. However, due to the scarcity of such data, we believe that this work affirms that single-instance, overnight, multimodal sleep stage information for a patient is insufficient to reveal signs of depression or insomnia but has the potential to detect the onset of OSA. This is in agreement with recent studies [41,42] in which sleep and circadian measures alone could not strongly discriminate depression outcomes, but several characteristics that overlap with our study (insomnia, excessive daytime sleepiness, time spent in N1) and others (snoring, inactivity at night, and lower morning activity) were shown to be common features between sleep and mental health problems. This emphasizes that sociodemographic, lifestyle, and genetic characteristics should be considered in addition to the metrics of existing wearable devices [36,38].

Recent studies consider sensing modalities such as EEG, electrocardiograms (ECG), or respiratory signals to classify OSA or depression [43]. We conducted this study to be modality-agnostic, such that only the epoch-wise sleep stage scores and routine demographic information are available, as proprietary algorithms used by wearables are difficult to access, and sought to validate the effectiveness of the hypnograms/hypnodensity graphs as a proxy nocturnal feature set. Our approach was found to underperform compared to methods using ECG [44], EEG [45], or a combination [46]. This is expected as we do not use information directly corresponding to individual physiological states, which better reflect the characteristics attributed to different conditions.

In WSC [10], it appears that the cohort included primarily healthy individuals in the community, which led the models to learn a poorer separation between positive and negative classes. Notably, there was significant data imbalance in the case of depressed and non-depressed patients, and it appears on average that none of the feature sets used showed a clear distinction between the two populations. This is likely due to having a sufficient sample size representative of the conditions and the lower levels of disease severity between populations. The dataset is also skewed toward the Caucasian demographic as well as limited by geography, limiting its generalization to other ethnicities in other locations. As found in [47], individuals with OSA and excessive daytime sleepiness had stronger associations with depression, suggesting that suffering from the effects of sleep disorders is more indicative of depression than simply having the disorder. In particular, the dataset overrepresents middle-aged and older adults, whose sleep characteristics may

vary differently from younger populations and may present a different relationship with depression or the other variables considered, especially in longitudinal studies [48].

Based on these findings, in future work, we will consider the National Sleep Research Resource (NSRR) as a source of public datasets focused on the standardized evaluation of sleep disorders. Some of the specific datasets of interest to the scope of our work include the Sleep Heart Health Study (SHHS) [49], Outcomes of Sleep Disorders in Older Men (MrOS Sleep Study) [50] and the Multiethnic Study of Atherosclerosis (MESA) [51]. This will allow cross-population validation, incorporating multicenter information and heterogeneity-aware analysis.

The findings of this study show the potential of wearable technologies in the diagnosis and management of sleep disorders. By offering a cost-effective and scalable alternative to traditional methods such as PSG, wearables can enhance the diagnostic accuracy for conditions such as OSA while reducing the reliance on clinical settings. Beyond passive monitoring, these devices can evolve into integral tools in personalized medicine, enabling early detection, long-term mental health monitoring, and proactive interventions. This advancement not only improves patient outcomes but also holds promise in reducing healthcare costs and broadening access to care.

To integrate these findings into clinical workflows, data from wearable devices should be incorporated into EHR systems to enable seamless access and analysis by clinicians. Decision support systems leveraging machine learning models can flag potential cases of OSA, insomnia, or depression for further evaluation. Collaborations between wearable technology developers and healthcare institutions should focus on creating standardized protocols for data integration, ensuring reliability and consistency in clinical applications.

5. Conclusions

Leveraging the transitional nature of sleep stages throughout the duration of sleep during a PSG test with image representations as a proxy and contextualizing them with additional patient-specific information could help mitigate the need for proprietary raw physiological signals. In summary, we have shown that sleep-related features derived from wearable devices, when supplemented with additional patient information and used within our constructed models, hold the potential to advance the diagnosis of sleep disorders and envision this work as a stepping stone to iterative development and validation in real-world settings. However, the complex interaction between these disorders and mental health conditions such as depression and insomnia requires further investigation with models that need to evolve to account for individual variability and longitudinal data.

Author Contributions: Conceptualization, A.S. and F.A.; data curation, Z.S.; investigation, J.R. and Z.S.; methodology, J.R.; project administration, A.S.; resources, F.A. and A.S.; software, J.R.; supervision, A.S. and F.A.; validation, J.R.; writing—original draft, J.R. and Z.S.; writing—review and editing, A.S. and F.A. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The WSC dataset adopted in this research is openly available in [National Sleep Research Resource] at https://doi.org/10.25822/js0k-yh52 (accessed on 1 January 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EHR	Electronic Health Records
CNN-BiLSTM	Convolutional Neural Network with Bidirectional Long Short-Term Memory
OSA	Obstructive Sleep Apnea
NREM	Non-Rapid Eye Movement Sleep
PSG	Polysomnography
REM	Rapid Eye Movement Sleep
WSC	Wisconsin Sleep Cohort

References

1. Lévy, P.; Kohler, M.; McNicholas, W.T.; Barbé, F.; McEvoy, R.D.; Somers, V.K.; Lavie, L.; Pépin, J.L. Obstructive sleep apnoea syndrome. *Nat. Rev. Dis. Prim.* **2015**, *1*, 15015. [CrossRef] [PubMed]
2. Saad, M.; Ray, L.B.; Bujaki, B.; Parvaresh, A.; Palamarchuk, I.; De Koninck, J.; Montplaisir, J.; Nielsen, T.A. Using heart rate profiles during sleep as a biomarker of depression. *BMC Psychiatry* **2019**, *19*, 168. [CrossRef] [PubMed]
3. Steiger, A.; Pawlowski, M. Depression and Sleep. *Int. J. Mol. Sci.* **2019**, *20*, 607. [CrossRef] [PubMed]
4. Grandner, M.A.; Lujan, M.R.; Ghani, S.B. Sleep-tracking technology in scientific research: Looking to the future. *Sleep* **2021**, *44*, zsab071. [CrossRef] [PubMed]
5. De Zambotti, M.; Goldstein, C.; Cook, J.; Menghini, L.; Altini, M.; Cheng, P.; Robillard, R. State of the science and recommendations for using wearable technology in sleep and circadian research. *Sleep* **2024**, *47*, zsad325. [CrossRef]
6. Jaiswal, S.J.; Pawelek, J.B.; Warshawsky, S.; Quer, G.; Trieu, M.; Pandit, J.A.; Owens, R.L. Using New Technologies and Wearables for Characterizing Sleep in Population-based Studies. *Curr. Sleep Med. Rep.* **2024**, *10*, 82–92. [CrossRef]
7. Lyall, L.M.; Sangha, N.; Zhu, X.; Lyall, D.M.; Ward, J.; Strawbridge, R.J.; Cullen, B.; Smith, D.J. Subjective and objective sleep and circadian parameters as predictors of depression-related outcomes: A machine learning approach in UK Biobank. *J. Affect. Disord.* **2023**, *335*, 83–94. [CrossRef]
8. de Zambotti, M.; Rosas, L.; Colrain, I.M.; Baker, F.C. The sleep of the ring: Comparison of the ŌURA sleep tracker against polysomnography. *Behav. Sleep Med.* **2019**, *17*, 124–136. [CrossRef]
9. Alkhodari, M.; Fraiwan, L. Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. *Comput. Methods Programs Biomed.* **2021**, *200*, 105940. [CrossRef]
10. Young, T.; Palta, M.; Dempsey, J.; Peppard, P.E.; Nieto, F.J.; Hla, K.M. Burden of sleep apnea: Rationale, design, and major findings of the Wisconsin Sleep Cohort study. *WMJ Off. Publ. State Med Soc. Wis.* **2009**, *108*, 246.
11. Ramesh, J.; Keeran, N.; Sagahyroon, A.; Aloul, F. Towards Validating the Effectiveness of Obstructive Sleep Apnea Classification from Electronic Health Records Using Machine Learning. *Healthcare* **2021**, *9*, 1450. [CrossRef] [PubMed]
12. Harlev, D.; Ravona-Springer, R.; Nuriel, Y.; Fruchter, E. Sleep monitoring using WatchPAT device to predict recurrence of major depression in patients at high risk for major depression disorder recurrence: A case report. *Front. Psychiatry* **2021**, *12*, 572660. [CrossRef] [PubMed]
13. Rykov, Y.; Thach, T.Q.; Bojic, I.; Christopoulos, G.; Car, J. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. *JMIR MHealth UHealth* **2021**, *9*, e24872. [CrossRef] [PubMed]
14. Kaufman, D.M.; Milstein, M.J. Chapter 17—Sleep Disorders. In *Kaufman's Clinical Neurology for Psychiatrists*, 17th ed.; Kaufman, D.M., Milstein, M.J., Eds.; W.B. Saunders: Philadelphia, PA, USA, 2013; pp. 365–396. [CrossRef]
15. Carskadon, M.A.; Rechtschaffen, A. Chapter 116—Monitoring and Staging Human Sleep. In *Principles and Practice of Sleep Medicine*, 4th ed.; Kryger, M.H., Roth, T., Dement, W.C., Eds.; W.B. Saunders: Philadelphia, PA, USA, 2005; pp. 1359–1377. [CrossRef]
16. Stephansen, J.B.; Olesen, A.N.; Olsen, M.; Ambati, A.; Leary, E.B.; Moore, H.E.; Carrillo, O.; Lin, L.; Han, F.; Yan, H.; et al. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. *Nat. Commun.* **2018**, *9*, 5229. [CrossRef]
17. Nutt, D.; Wilson, S.; Paterson, L. Sleep disorders as core symptoms of depression. *Dialogues Clin. Neurosci.* **2008**, *10*, 329–336. [CrossRef]
18. van der Woerd, C.; van Gorp, H.; Dujardin, S.; Sastry, M.; Garcia Caballero, H.; van Meulen, F.; van den Elzen, S.; Overeem, S.; Fonseca, P. Studying sleep: Towards the identification of hypnogram features that drive expert interpretation. *Sleep* **2024**, *47*, zsad306. [CrossRef]
19. Huang, W.C.; Lee, P.L.; Liu, Y.T.; Chiang, A.A.; Lai, F. Support vector machine prediction of Obstructive Sleep Apnea in a large-scale Chinese clinical sample. *Sleep* **2020**, *43*, zsz295. [CrossRef]
20. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]

21. Luong, M.T. Effective approaches to attention-based neural machine translation. *arXiv* **2015**, arXiv:1508.04025. [CrossRef]

22. Alkhodari, M.; Khandoker, A.H.; Jelinek, H.F.; Karlas, A.; Soulaidopoulos, S.; Arsenos, P.; Doundoulakis, I.; Gatzoulis, K.A.; Tsioufis, K.; Hadjileontiadis, L.J. Circadian assessment of heart failure using explainable deep learning and novel multi-parameter polar images. *Comput. Methods Programs Biomed.* **2024**, *248*, 108107. [CrossRef]

23. Alkhodari, M.; Hadjileontiadis, L.J.; Khandoker, A.H. Identification of Congenital Valvular Murmurs in Young Patients Using Deep Learning-Based Attention Transformers and Phonocardiograms. *IEEE J. Biomed. Health Informatics* **2024**, *28*, 1803–1814. [CrossRef] [PubMed]

24. Yan, B.; Zhao, B.; Jin, X.; Xi, W.; Yang, J.; Yang, L.; Ma, X. Sleep efficiency may predict depression in a large population-based study. *Front. Psychiatry* **2022**, *13*, 838907. [CrossRef] [PubMed]

25. Brock, M.S.; Mysliwiec, V. Comorbid insomnia and sleep apnea: A prevalent but overlooked disorder. *Sleep Breath.* **2018**, *22*, 1–3. [CrossRef]

26. Björnsdóttir, E.; Janson, C.; Sigurdsson, J.F.; Gehrman, P.; Perlis, M.; Juliusson, S.; Arnardottir, E.S.; Kuna, S.T.; Pack, A.I.; Gislason, T.; et al. Symptoms of insomnia among patients with Obstructive Sleep Apnea before and after two years of positive airway pressure treatment. *Sleep* **2013**, *36*, 1901–1909. [CrossRef]

27. Ong, J.C.; Crawford, M.R. Insomnia and Obstructive Sleep Apnea. *Sleep Med. Clin.* **2013**, *8*, 389–398. [CrossRef]

28. Luyster, F.S.; Buysse, D.J.; Strollo, P.J. Comorbid insomnia and Obstructive Sleep Apnea: Challenges for clinical practice and research. *J. Clin. Sleep Med.* **2010**, *6*, 196–204. [CrossRef]

29. Janssen, H.C.; Venekamp, L.N.; Peeters, G.A.; Pijpers, A.; Pevernagie, D.A. Management of insomnia in sleep-disordered breathing. *Eur. Respir. Rev.* **2019**, *28*, 190080. [CrossRef]

30. Mendes, M.S.; dos Santos, J.M. Insomnia as an expression of Obstructive Sleep Apnea syndrome–the effect of treatment with nocturnal ventilatory support. *Rev. Port. Pneumol. (Engl. Ed.)* **2015**, *21*, 203–208. [CrossRef]

31. Zhang, Y.; Folarin, A.A.; Sun, S.; Cummins, N.; Bendayan, R.; Ranjan, Y.; Rashid, Z.; Conde, P.; Stewart, C.; Laiou, P.; et al. Relationship between major depression symptom severity and sleep collected using a wristband wearable device: Multicenter longitudinal observational study. *JMIR MHealth UHealth* **2021**, *9*, e24604. [CrossRef]

32. Palagini, L.; Baglioni, C.; Ciapparelli, A.; Gemignani, A.; Riemann, D. REM sleep dysregulation in depression: State of the art. *Sleep Med. Rev.* **2013**, *17*, 377–390. [CrossRef]

33. Hein, M.; Wacquier, B.; Conenna, M.; Lanquart, J.P.; Point, C. Cardiovascular Outcome in Patients with Major Depression: Role of Obstructive Sleep Apnea Syndrome, Insomnia Disorder, and COMISA. *Life* **2024**, *14*, 644. [CrossRef] [PubMed]

34. Kim, B.; Kim, T.Y.; Choi, E.J.; Lee, M.; Kim, W.; Lee, S.A. Restless legs syndrome in patients with Obstructive Sleep Apnea: Association between apnea severity and symptoms of depression, insomnia, and daytime sleepiness. *Sleep Med.* **2024**, *117*, 40–45. [CrossRef] [PubMed]

35. Quevedo-Blasco, R.; Díaz-Román, A.; Quevedo-Blasco, V.J. Associations between Sleep, Depression, and Cognitive Performance in Adolescence. *Eur. J. Investig. Heal. Psychol. Educ.* **2023**, *13*, 501–511. [CrossRef] [PubMed]

36. Chinoy, E.D.; Cuellar, J.A.; Huwa, K.E.; Jameson, J.T.; Watson, C.H.; Bessman, S.C.; Hirsch, D.A.; Cooper, A.D.; Drummond, S.P.A.; Markwald, R.R. Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep* **2021**, *44*, zsaa291. [CrossRef]

37. Sridhar, N.; Shoeb, A.; Stephens, P.; Kharbouch, A.; Shimol, D.B.; Burkart, J.; Ghoreyshi, A.; Myers, L. Deep learning for automated sleep staging using instantaneous heart rate. *Npj Digit. Med.* **2020**, *3*, 106. [CrossRef]

38. Birrer, V.; Elgendi, M.; Lambercy, O.; Menon, C. Evaluating reliability in wearable devices for sleep staging. *Npj Digit. Med.* **2024**, *7*, 74. [CrossRef]

39. Christensen, J.A.E.; Carrillo, O.; Leary, E.B.; Peppard, P.E.; Young, T.; Sorensen, H.B.D.; Jennum, P.; Mignot, E. Sleep stage transitions during polysomnographic recordings as diagnostic features of type 1 narcolepsy. *Sleep Med.* **2015**, *16*, 1558–1566. [CrossRef]

40. Bubbico, G.; Di Iorio, A.; Lauriola, M.; Sepede, G.; Salice, S.; Spina, E.; Brondi, G.; Esposito, R.; Perrucci, M.G.; Tartaro, A. Subjective Cognitive Decline and Nighttime Sleep Alterations, a Longitudinal Analysis. *Front. Aging Neurosci.* **2019**, *11*, 142. [CrossRef]

41. Jiang, J.; Li, Z.; Li, H.; Yang, J.; Ma, X.; Yan, B. Sleep architecture and the incidence of depressive symptoms in middle-aged and older adults: A community-based study. *J. Affect. Disord.* **2024**, *352*, 222–228. [CrossRef]

42. Thakre, T.P.; Kulkarni, H.; Adams, K.S.; Mischel, R.; Hayes, R.; Pandurangi, A. Polysomnographic identification of anxiety and depression using deep learning. *J. Psychiatr. Res.* **2022**, *150*, 54–63. [CrossRef]

43. Rykov, Y.G.; Ng, K.P.; Patterson, M.D.; Gangwar, B.A.; Kandiah, N. Predicting the Severity of Mood and Neuropsychiatric Symptoms from Digital Biomarkers Using Wearable Physiological Data and Deep Learning. *Comput. Biol. Med.* **2024**, *180*, 108959. [CrossRef] [PubMed]

44. Sato, S.; Hiratsuka, T.; Hasegawa, K.; Watanabe, K.; Obara, Y.; Kariya, N.; Shinba, T.; Matsui, T. Screening for Major Depressive Disorder Using a Wearable Ultra-Short-Term HRV Monitor and Signal Quality Indices. *Sensors* **2023**, *23*, 3867. [CrossRef] [PubMed]

45. Markov, K.; Elgendi, M.; Menon, C. EEG-based Headset Sleep Wearable Devices. *Npj Biosensing* **2024**, *1*, 12. [CrossRef]

46. Moussa, M.M.; Alzaabi, Y.; Khandoker, A. ECG, EEG, Breathing Signals, and Machine Learning: Computer-Aided Detection of Obstructive Sleep Apnea Syndrome and Depression. In Proceedings of the 2022 Computing in Cardiology (CinC), Tampere, Finland, 4–7 September 2022 ; Volume 498, pp. 1–4. [CrossRef]

47. Lang, C.J.; Appleton, S.L.; Vakulin, A.; McEvoy, R.D.; Vincent, A.D.; Wittert, G.A.; Martin, S.A.; Grant, J.F.; Taylor, A.W.; Antic, N.; et al. Associations of Undiagnosed Obstructive Sleep Apnea and Excessive Daytime Sleepiness with Depression: An Australian Population Study. *J. Clin. Sleep Med.* **2017**, *13*, 575–582. [CrossRef]

48. Bi, K.; Chen, S. Sleep Profiles as a Longitudinal Predictor for Depression Magnitude and Variability Following the Onset of COVID-19. *J. Psychiatr. Res.* **2022**, *147*, 159–165. [CrossRef]

49. Quan, S.F.; Howard, B.V.; Iber, C.; Kiley, J.P.; Nieto, F.J.; O'Connor, G.T.; Rapoport, D.M.; Redline, S.; Robbins, J.; Samet, J.M.; et al. The Sleep Heart Health Study: Design, Rationale, and Methods. *Sleep* **1997**, *20*, 1077–1085. [CrossRef]

50. Vo, T.N.; Kats, A.M.; Langsetmo, L.; Taylor, B.C.; Schousboe, J.T.; Redline, S.; Kunisaki, K.M.; Stone, K.L.; Ensrud, K.E.; for the Osteoporotic Fractures in Men (MrOS) Study Research Group. Association of Sleep-Disordered Breathing with Total Healthcare Costs and Utilization in Older Men: The Outcomes of Sleep Disorders in Older Men (MrOS Sleep) Study. *Sleep* **2020**, *43*, zsz209. [CrossRef]

51. Dean, D.A., II; Wang, R.; Jacobs, D.R., Jr.; Duprez, D.; Punjabi, N.M.; Zee, P.C.; Shea, S.; Watson, K.; Redline, S. A Systematic Assessment of the Association of Polysomnographic Indices with Blood Pressure: The Multi-Ethnic Study of Atherosclerosis (MESA). *Sleep* **2015**, *38*, 587–596. [CrossRef]