This CVPR workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Deep Learning-Enabled Sleep Staging From Vital Signs and Activity Measured Using a Near-Infrared Video Camera

Jonathan Carter*[1,2] João Jorge[2] Bindia Venugopal[2] Oliver Gibson[2] Lionel Tarassenko[1]

[1]Institute of Biomedical Engineering, University of Oxford
[2]Oxehealth Ltd., Oxford

Abstract

Conventional sleep monitoring is time-consuming, expensive and uncomfortable, requiring a large number of contact sensors to be attached to the patient. Video data is commonly recorded as part of a sleep laboratory assessment. If accurate sleep staging could be achieved solely from video, this would overcome many of the problems of traditional methods. In this work we use heart rate, breathing rate and activity measures, all derived from a near-infrared video camera, to perform sleep stage classification. We use a deep transfer learning approach to overcome data scarcity, by using an existing contact-sensor dataset to learn effective representations from the heart and breathing rate time series. Using a dataset of 50 healthy volunteers, we achieve an accuracy of 73.4% and a Cohen's kappa of 0.61 in four-class sleep stage classification, establishing a new state-of-the-art for video-based sleep staging.

1. Introduction

The 'gold-standard' method for sleep monitoring is video polysomnography (vPSG, [35]), a test in which a number of physiological signals are recorded over a night's sleep. This typically requires an extensive set of contact sensors to be attached to the patient, which often leads to patient discomfort and an unrepresentative night's sleep [1]. After the recording, a human expert (sleep physiologist) must review the outputs from these sensors, manually annotating the record with sleep stages, leg movements and other events of interest from throughout the night. This process can often take multiple hours to complete. The cost of the study and the discomfort to the patient mean that longer-term studies are rare. Easier longitudinal monitoring would also benefit clinical research, from making trials for sleep-disorder treatments easier, to detecting the onset of dementia, by monitoring variations in slow-wave sleep [34].

Sleep stages are commonly annotated at 30-second intervals using the American Academy of Sleep Medicine (AASM) guidelines. These divide sleep into five discrete states: Wake, N1, N2, N3 and REM sleep. N1 to N3 account for varying degrees of non-rapid eye movement (NREM) sleep from near-wakefulness (N1) to deep slow-wave sleep (N3).

There is well-studied subjectivity between sleep physiologists in scoring sleep stages, which is often reported using the Cohen's kappa statistic (κ, [2]). The work of Danker-Hopfe *et al*. [4] showed that pairs of scorers agreed on 82% of epoch-by-epoch sleep stages ($\kappa = 0.76$) across a dataset of 72 sleep recordings.

A recent review has claimed that deep learning methods can score PSG records as accurately as a human scorer [26]. The SleepTransformer model [27] achieves a Cohen's κ of 0.83 in AASM sleep stage classification on the Sleep Heart Health Study (SHHS, [30]) dataset from an electroencephalogram (EEG) signal. However, the EEG is typically measured using electrodes which must be glued to the subject's scalp by a sleep physiologist.

To overcome this need, other work has investigated using cardio-respiratory and activity information obtained from wearable devices [28, 32, 45, 48]. Wearable sleep staging is possible from these inputs since they encode information about the underlying sleep state. For example, respiration tends to be more irregular during REM sleep [12], whilst low frequency power in the heart rate time series is typically lowest during N3 sleep [34]. Wearable-based methods have commonly simplified sleep staging into a four-class classification problem, by merging N1 and N2 into a 'light sleep' class. Doing so, Radha *et al*. [32] achieve a Cohen's κ of 0.65 in four-class sleep staging from wrist-worn photoplethysmography.

*jcarter@robots.ox.ac.uk

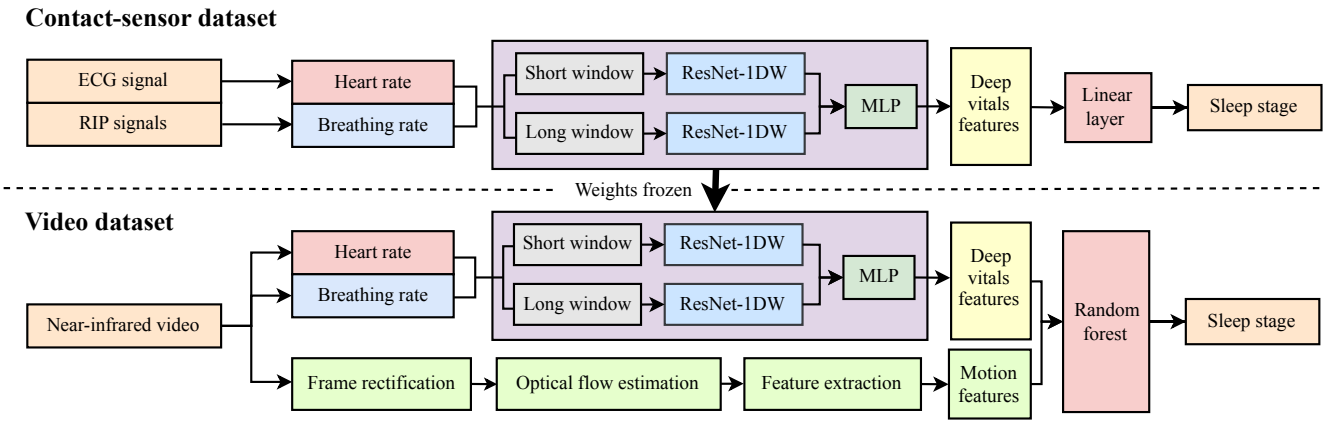

Figure 1. Overview of our approach. We first train a deep learning model on a contact sensor dataset to classify sleep stages from heart rate and breathing rate time series. We then use this trained model as a feature extractor to produce 'deep' features from the video-derived vital signs of our target dataset. To classify sleep stages from video data, we train a random forest classifier which uses these deep features in conjunction with motion features derived from optical flow.

Prior work has shown that both cardio-respiratory [38] and activity information [21] can be derived from video data during sleep, indicating that video could be used as a single source from which to classify sleep stages. As an entirely non-contact solution, this would further simplify and improve the sleep monitoring process for both patients and clinicians. Additionally, video remains an important modality for the diagnosis of conditions including periodic limb movement disorder, parasomnias and REM behaviour disorder [5, 36]. It may not be feasible to identify these conditions using alternative solutions such as wrist-worn wearables.

In this work, we use heart rate, breathing rate and activity signals, all derived from near-infrared video, to perform sleep stage classification. We use optical flow [11] to derive continuous measures of activity during sleep, which are made robust to camera position and bed orientation using image homography transformations [7].

To overcome data scarcity, we adopt a transfer learning approach (Fig. 1). We first train a deep neural network to classify sleep stages from heart rate and breathing rate time series derived from a substantially larger contact sensor dataset. The trained model is then applied to heart rate and breathing rate time series derived from near-infrared video to obtain 'deep' features. These deep features are used in combination with video-derived activity features as inputs to a classifier which is trained and evaluated on a video dataset of 50 healthy volunteers using 10-fold cross validation.

Using our approach, we present improved results over the existing state-of-the-art [40] for video-based four-class sleep staging in terms of both accuracy (73.4% vs. 67.9%) and Cohen's kappa (0.61 vs. 0.49).

2. Related work

Video actigraphy: Counting pixel differences above a threshold has proven to be a simple but effective activity signal for binary sleep–wake classification from both near-infrared (NIR, [33]) and depth images [17, 42].

Optical flow [3, 23] and spatio-temporal recursive search algorithms [9] have also proved effective at quantifying activity. These methods both compute motion vectors from the video feed and use average magnitudes as activity signals. The latter method has been used to detect periodic limb movements during sleep [10].

Using activity features derived from video, Long *et al.* [21] achieve an accuracy of 92.0% in sleep–wake detection from video using a dataset of 10 healthy infants.

Video-based sleep staging: To the best of our knowledge, only two prior works have attempted to classify sleep beyond a binary sleep-wake classification using a video camera. These works differ greatly in the information which they extract from the video to classify sleep stages.

Nochino *et al.* [24] use frame-differencing to derive an activity signal from the camera. From this signal, the authors derive a number of motion features for each sleep epoch and train a Support Vector Machine to perform four-class sleep staging, achieving an average accuracy of 40.5% and a Cohen's κ of 0.19 on a dataset of 6 participants.

van Meulen *et al.* [40] also classify sleep stages into four classes, but do so using camera-derived inter-pulse intervals (IPIs). The authors directly apply a deep learning model, previously trained on heart-rate variability (HRV) features derived from the electrocardiogram (ECG), to HRV features computed from a camera-derived IPI time series. This achieves an accuracy of 68% and a Cohen's κ of 0.49 across 46 healthy participants.

3. Methods

3.1. Data acquisition

In this work, we use video as the input data and polysomnogram recordings as the source of sleep stage labels collected from 50 healthy adult volunteers (18+) with no previous diagnoses of sleep disorders. We report age, weight, sex and Fitzpatrick skin type for the study population in Table 1. Fitzpatrick skin type [6] is a measure of skin pigmentation which classifies skin phototypes into one of six categories ranging from I (very pale) to VI (very dark).

Table 1. Sleep study population demographics.

Variable	Value
Age[1]	37±15 years
Weight[1]	70.4±12.9 kg
Sex	
Male	22 volunteers
Female	28 volunteers
Fitzpatrick skin type[2]	I: 13, II: 21, III: 12, IV: 3, VI: 1

[1]Mean (± std). [2]Skin type: No. volunteers.

Each recording took place overnight in one of two experiment rooms. Video data was captured in each room at 20 FPS using a single-channel 850 nm NIR video camera. Both rooms also contained LED illuminators centred at the same wavelength and placed adjacent to the cameras. Figure 2 shows the most common bed orientations and the fixed camera positions used in each room.

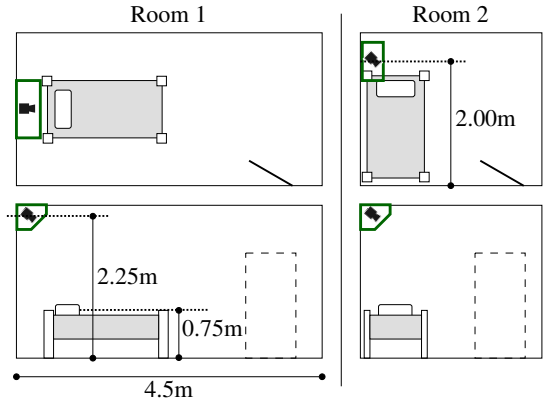

Figure 2. Room and camera layouts used for data acquisition.

Video polysomnography data was acquired using a SOMNOscreen™ Plus device, which recorded a typical vPSG montage which included respiratory inductance plethysmography (RIP), electrooculogram (EOG) and EEG signals and video data, using a separate camera set-up. Each recording was manually scored according to AASM guidelines in 30-second epochs by an experienced sleep physiologist.

3.2. Optical flow-based activity measurement

From the video camera, we derive continuous measures of subject activity using optical flow [11]. This process is illustrated in Figure 3. First, we apply an image homography transformation [7] which provides a consistent, virtual viewpoint of the bed region for different camera positions. This technique has previously been used to improve the robustness of sleep pose detection [22]. We then compute optical flow between transformed frames using the Dense Inverse Search algorithm [16]. From the 20 FPS video, we estimate the optical flow field between every 5th frame, then average over the previous four estimated fields, giving a field estimation frequency of 4 Hz.

From the optical flow field time series, we derive two simple activity signals by averaging the flow field magnitude in the upper third and lower two thirds of the rectified view, which correspond to approximate head and lower body regions. This gives a simple measure of the spatial scale of movements, since whole-body movements such as sleep posture changes will result in high activity for both signals.

3.3. Motion features

From our two video-derived activity signals, we compute a total of 20 time-series features for each epoch, listed in Table 2. These features broadly fall into two categories: integral features, which quantify the amount of motion that occurred within a time window centred on the epoch, and counter-based features, which quantify the elapsed time since a movement occurred. Our counter-based features are inspired by those of Nochino *et al*. [24].

Table 2. Motion features derived from the 1D optical flow activity signals for each sleep epoch. Features are computed from the upper and lower bed region signals separately.

Count	Feature description
14	**Integral features**
6	30-second Gaussian-weighted sums, time-shifted by (-30, 0.0, 30.0) seconds.
6	180-second Gaussian-weighted sums, time-shifted by (-180.0, 0.0, 180.0) seconds.
2	20-minute Gaussian weighted sums.
6	**Counter-based features**
4	Time elapsed since activity signal was above a threshold $\in \{0.1, 1.0\}$.
2	Time elapsed since activity signal was above a threshold (1.0) for 3 seconds.

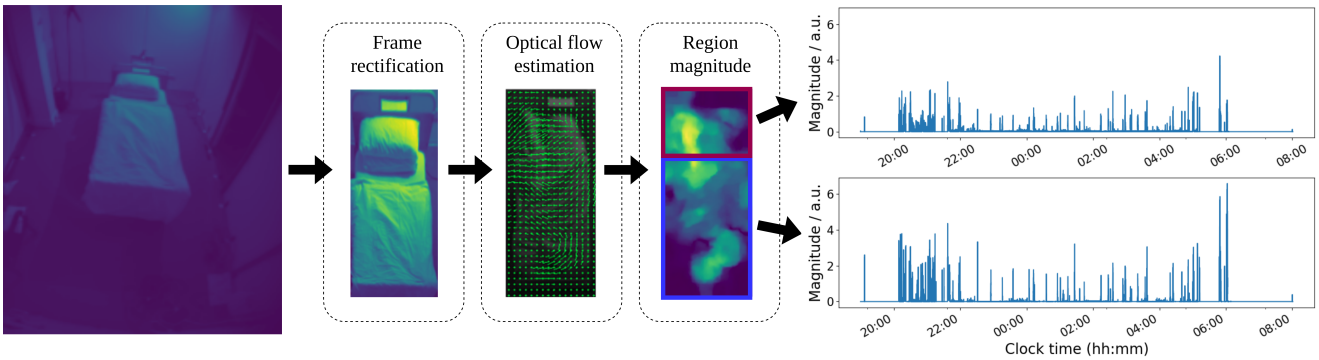

Figure 3. Extracting 1D activity signals from video using optical flow. A homography transformation is applied to each frame followed by cropping to the bed region, to obtain a consistent, virtual viewpoint. The magnitudes of optical flow vectors calculated between frames are averaged over the upper and lower bed regions to produce continuous activity signals for each region.

3.4. Video-based vital-sign monitoring

Many prior works have shown that both the heart rate [18, 29, 43, 47] and breathing rate [19, 23, 46] can be measured using video cameras. These methods have been validated in a number of clinical settings e.g. [14, 37, 44]. The recent study of van Gastel *et al*. [38] has also demonstrated the feasibility of continuous vital-sign monitoring during sleep.

In this work, we use components of Oxevision Vital Signs medical device software [13] to measure the heart rate and breathing rate from video data. For both vital signs, this produces 1 Hz estimates and binary Signal Quality Indices (SQIs, [20]), which indicate the software's assessment of the estimate's quality.

The software estimates both vital signs through a process of region of interest (ROI) selection, spatio-temporal filtering, cardio-respiratory signal extraction and data fusion. First, ROIs are identified by selecting areas in the image containing high variance that is periodic in nature [13]. These ROIs are then spatially smoothed with a median filter and band-pass filtered using pass-bands of 40–130 BPM and 8–39 BRPM for the heart and breathing rates respectively. The cardio-respiratory signals are extracted from the ROIs using principal component analysis as done in Villaroel *et al*. [44], with the heart and breathing rates identified from the dominant frequency components. SQIs are derived for each vital sign from heuristics including inter-beat interval consistency. These SQIs are typically low when the volunteer is active, e.g. out of bed, or absent from the room.

We define the coverage of each vital sign as the percentage of time for which the SQI is high when the volunteer is in bed. Using this definition, Figure 4 shows the coverage for each vital sign across recordings in the dataset. Coverage statistics computed across participants are given in Table 3.

Figure 5 shows an example of heart rate and breathing rate time-series estimated from near-infrared video over a night's recording along with AASM-annotated sleep stages. In this example, we can observe increased variance in the breathing rate around REM cycles, indicating how sleep state information can be encoded in the cardio-respiratory time series.

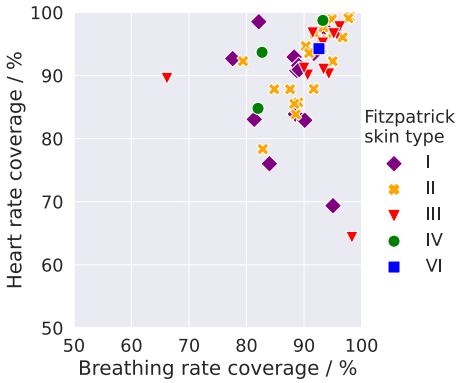

Figure 4. Coverage of video-derived vital-sign estimates.

Table 3. Coverage statistics for video-derived vital signs.

	Mean ± Std	Min	Max
BR coverage (%)	91.0±7.5	66.2	98.3
HR coverage (%)	90.0±6.0	64.3	99.3

3.5. Vital-sign representation learning

State-of-the-art EEG-based sleep staging performance has been achieved using supervised deep learning methods trained with large datasets, often containing thousands of PSG recordings [26]. These datasets e.g. SHHS [30], required multiple years to collect across several data collection centres.

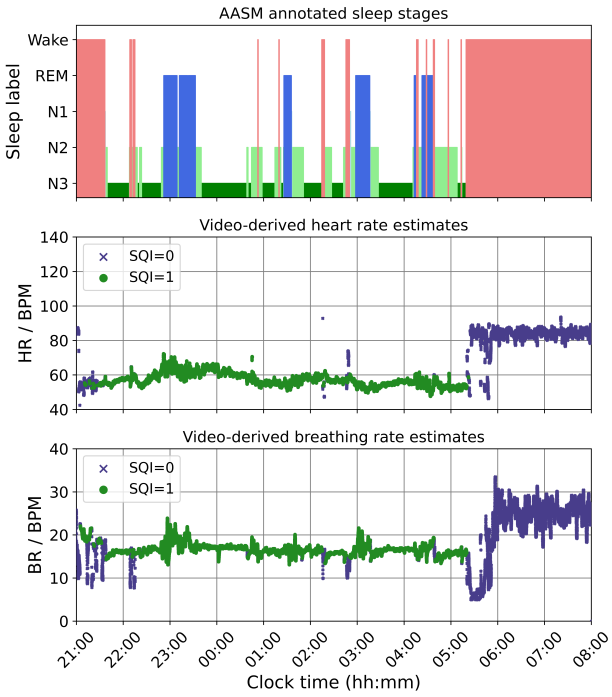

Figure 5. Example AASM sleep stages and video-derived heart rate and breathing estimates for a night's recording from the video dataset.

By training a model on a large, source dataset and fine-tuning on a smaller target dataset it is often possible to achieve greater performance than training directly on the target dataset. This approach has successfully been applied in the context of wearable sleep staging [32]. The authors first train a model to classify sleep on a large ECG dataset before fine-tuning it on a much smaller target photoplethysmography (PPG) dataset. This is shown to give improved performance over models trained solely on the target PPG dataset.

We use the SHHS dataset [30] to learn representations from heart rate and breathing rate time-series, acquired using wearable sensors, which can then be applied to similar time series obtained from video data. The SHHS dataset is substantially larger than our target video dataset, with polysomnogram recordings and accompanying AASM sleep stage labels from around 6000 participants.

Training inputs. Firstly, heart rate (HR) and breathing rate (BR) time-series for the SHHS dataset were estimated from the ECG and RIP signals respectively. An SQI was also calculated for each estimate. More information on our estimation process is given in Appendix A. Figure 6 shows synchronised AASM sleep stage labels and ECG-derived heart rate estimates for an example recording from the SHHS dataset. We see that there is a long period of low signal quality (SQI=0) at the end of the recording after the

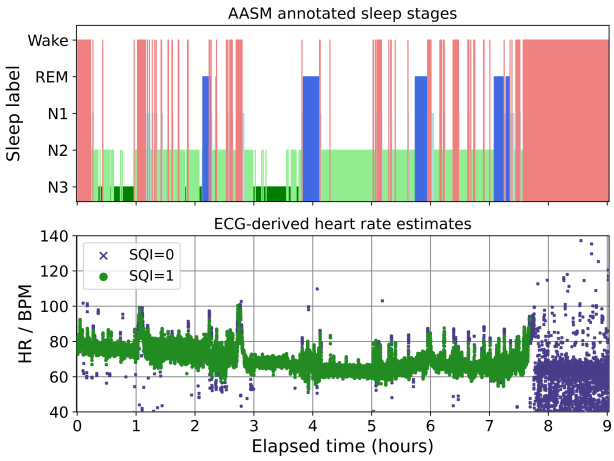

Figure 6. Example AASM sleep stages and ECG-derived heart rate estimates and SQI values from an SHHS polysomnogram recording.

participant has woken up. Periods of low signal quality are strongly correlated with wakefulness, since they are often induced by motion artefacts.

SQI Filtering. To improve performance, we set low quality sections of the time series to zero before passing them to the model. By zeroing out these sections, we still allow the model to know which periods are of low quality whilst preventing the model from overfitting to signal noise.

Model architecture. Figure 7 illustrates our neural network architecture. The model uses two input streams to classify each epoch: a high-frequency 5-minute HR/BR window sampled at 1 Hz and a lower frequency 50-minute window sampled at 0.1 Hz, both centred around the target epoch for classification. Each window of data is passed to a 1D ResNet [8] model, which transforms the two-channel input time series into a feature vector. Our ResNet architecture broadly follows the design of the original paper, except that 2D convolutions are replaced with their 1D equivalents. The feature vectors are concatenated and passed to a multi-layer perceptron (MLP), which mixes short-term and long-term information from the windows to produce a single feature vector. Finally, a linear classification layer is applied to the feature vector output of the MLP to obtain sleep stage probabilities.

Model training. The network was trained end-to-end using back-propagation, minimising the cross-entropy loss function against the annotated AASM (5-class) sleep stages. Further model implementation and training details are given in Appendix B. After training, we expect sleep stages to be linearly separable from the output features of the MLP, thus making them effective inputs to any alternative classification model. We use the trained model as a feature extractor by taking the outputs from the MLP layer.

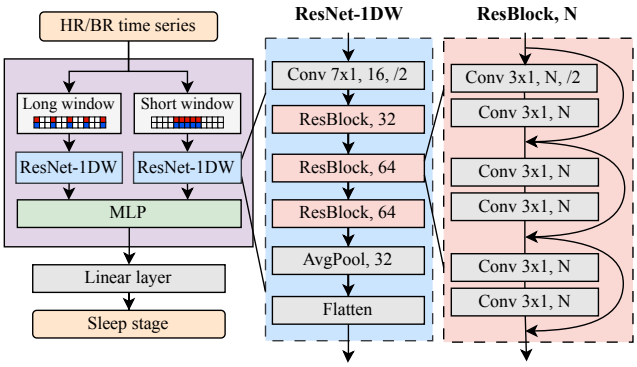

Figure 7. Sleep staging model architecture used to learn vital-sign representations from the SHHS dataset. When applying the model to video data, we use the outputs of the MLP (purple box) as input features to an alternative classifier which also incorporates motion information.

3.6. Transfer learning

From our video-derived heart rate and breathing rate time series, we use the trained feature extractor to produce 'deep' vital-sign features. These are used in conjunction with the motion features described in Section 3.3 as inputs to a random forest classifier which is trained to classify sleep stages on an epoch-by-epoch basis. A random forest classifier was chosen for its combination of low computational cost and robustness to overfitting.

3.7. Evaluation

The model was trained and evaluated using k-fold cross-validation, using 10 non-overlapping folds each containing 5 recordings. Performance on each fold was evaluated using a model trained on the remaining 9 folds. Results were then aggregated over all folds.

4. Results and Discussion

In this section, we report performance in terms of multi-class classification accuracy and Cohen's kappa between model output classifications and annotated sleep stages. A comparison with existing video-based four-class sleep staging algorithms is given in Table 4.

Table 4. Performance comparison between our deep transfer learning method and prior video-based four-class sleep staging work.

Method	N[1]	Cohen's κ	Accuracy (%)
Nochino *et al.* [24]	6	0.19±0.04	40.5±2.2
van Meulen *et al.* [40]	46	0.49±0.13	67.9±8.7
Ours	50	**0.61**±0.15	**73.4**±9.6

[1]Study population size. Mean (± std) across recordings.

Figure 8 shows the total test confusion matrix between scorer and model labels summed over all recordings.

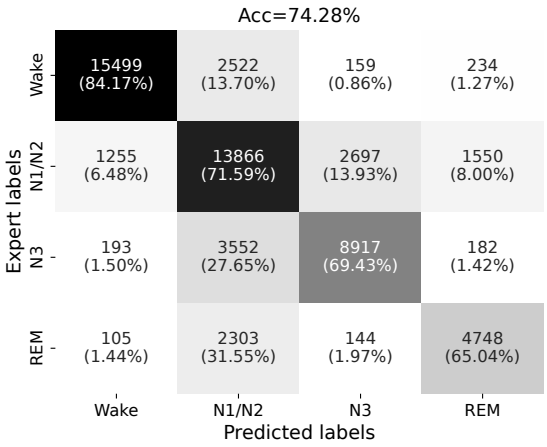

Figure 8. Four-class sleep staging confusion matrix across the dataset using our best performing method. Row-wise percentages are given in brackets, with values on the diagonal indicating class-wise sensitivities.

Performance with age. Figure 9 shows the variation in Cohen's kappa values with age. Here we observe a clear negative correlation between age and the Cohen's kappa statistic, findings which mirror those from prior ECG-based sleep staging work [31]. In older adults, common measures of autonomic function such as HRV are known to decrease with age [50], making sleep stages more difficult to distinguish when using cardio-respiratory inputs.

Figure 9. Scatter plot of Cohen's kappa values against age and Fitzpatrick skin type.

Example hypnogram output. Figure 10 shows an example four-class sleep hypnogram produced by our model.

This example corresponds to the median Cohen's kappa obtained across recordings. There is good visual agreement in the sleep architecture: the model correctly identifies all three REM cycles, sleep onsets and offsets, and most brief arousals during the night. Hypnograms corresponding to the minimum and maximum Cohen's kappa are given in Appendix D.

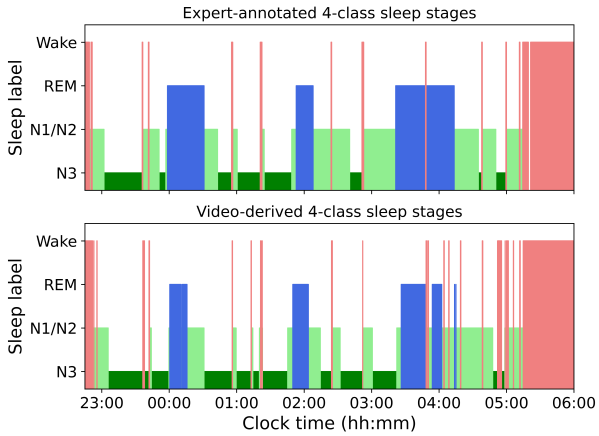

Figure 10. Model and scorer four-class hypnograms corresponding to the median Cohen's kappa statistic (0.63).

Performance with classification strategy. Table 5 indicates the performance of our model using alternative sleep stage classification strategies. The random forest classifier was re-trained for each strategy against the target labels which are derived from the AASM labels accordingly.

Table 5. Performance comparison across Sleep–Wake (SW), Wake–NREM–REM (3-class) and Wake–N1/N2–N3–REM (4-class) classification strategies.

	Cohen's κ	Acc. (%)	TPR[1] (%)	TNR[2] (%)
SW	0.72±0.24	91.6±6.9	94.2±5.3	81.0±25.0
3-class	0.68±0.18	84.9±7.1	-	-
4-class	0.61±0.15	73.4±9.6	-	-

[1]True positive rate i.e. sensitivity. [2]True negative rate i.e. specificity.

4.1. Ablation studies

Addition of motion and deep features. In Table 6 we assess the impact of including motion features and of using our deep vital-sign features over a simpler baseline of hand-designed features (see Appendix C). We observe that both the deep features and motion features are important to the overall performance.

SQI filtering. Table 7 indicates the performance improvement obtained from our SQI filtering approach. By zeroing out low quality sections of the time series in both training and testing we achieve an appreciable improvement in our results.

Table 6. Ablation results showing the effect of adding motion features (M) and of using deep features (DF) extracted from the vital signs compared with a simple feature engineering (FE) baseline.

Method	Cohen's κ	Accuracy (%)
FE	0.38±0.20	57.6±13.6
FE + M	0.56±0.17	70.2±10.2
DF	0.50±0.16	65.5±10.5
DF + M	0.61±0.15	73.4±9.6

Table 7. Ablation results showing the effect of SQI filtering.

Method	Cohen's κ	Accuracy (%)
Raw time-series	0.59±0.16	72.4±10.0
SQI filtering	0.61±0.15	73.4±9.6

5. Conclusions and Future Work

In this work, we have introduced a novel transfer learning approach to video-based sleep stage classification which uses non-contact measurements of the heart rate, breathing rate and activity measured using a near-infrared camera. This approach achieves state-of-the-art performance on video-based sleep staging on a healthy study population.

Performance is likely to be further improved through the use of an explicit sequence–sequence architecture such as a Transformer [41], which has led to superior performance in PSG-based sleep staging [27]. Using deep motion features, such as the intermediate outputs of a sleep pose detection model [22], may also improve classification performance over the relatively simple measures of activity used in this work.

In future work, we intend to investigate the performance of our approach across a broader demographic range and in a sleep-disordered population.

Acknowledgements

This work was supported by the EPSRC Centre for Doctoral Training in Autonomous Intelligent Machines and Systems [EP/S024050/1] and funded by Oxehealth Ltd.

We gratefully acknowledge the National Sleep Research Resource [49] for granting access to the SHHS dataset.

A. SHHS vital-sign estimation

Heart rate: We estimated the heart rate from the 128 Hz ECG signal available from each PSG recording. First, we applied the Pan-Tompkins algorithm [25] to produce a QRS detection time series. We then applied an FFT to 9-second

rolling windows, taking the frequency of the main spectral peak as our estimate. An SQI was calculated for each heart rate estimate given by:

$$\mathrm{SQI}_t^{\mathrm{ECG}} = \begin{cases} 1 & \text{if } |\mathrm{HR}_t^{\mathrm{ECG}} - \mathrm{HR}_t^{\mathrm{PPG}}| \leq 3\,\mathrm{bpm} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

i.e. our SQI is 1 when there is good agreement between our ECG-derived heart rate, $\mathrm{HR}_t^{\mathrm{ECG}}$, and a PPG-derived heart rate, $\mathrm{HR}_t^{\mathrm{PPG}}$, which is available from the SHHS dataset.

Breathing rate: We used a peak counting approach to estimate the breathing rate from both the abdomen and thorax RIP waveforms, using 30-second rolling windows. We then used the thorax-derived estimate as our breathing rate and used the disagreement between the two sources to derive a binary SQI:

$$\mathrm{SQI}_t^{\mathrm{RIP}} = \begin{cases} 1 & \text{if } |\mathrm{BR}_t^{\mathrm{Abd}} - \mathrm{BR}_t^{\mathrm{Thor}}| \leq 3\,\mathrm{brpm} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $\mathrm{BR}_t^{\mathrm{Abd}}$ and $\mathrm{BR}_t^{\mathrm{Thor}}$ are the time-series estimates derived from the abdomen and thorax respectively.

B. Further model implementation details

Each ResNet model produced a flattened 1D feature vector of length 2048. Our MLP therefore had an input dimension of 4096, given by the concatenation of the short- and long-window feature vectors. The MLP used a single hidden layer with a hidden dimension of length 100. The output of the MLP i.e. the 'deep' feature vector had length 32. This was chosen to be small and similar in size to the motion feature vector, to mitigate overfitting and encourage a balanced weighting of the two feature sources in the transfer learning phase.

To improve training stability, we normalised the input time series passed to the model by dividing by the mean and standard deviation of the heart and breathing rates calculated across the entire SHHS dataset. We used the Adam [15] optimiser with default hyper-parameters and employed early stopping after three epochs without an improvement in validation accuracy.

C. Feature engineering baseline

Our baseline consisted of 6 features extracted from the heart rate and breathing rate time series for each epoch: means over 3-minute rolling windows and standard deviations calculated using 3-minute and 10-minute rolling windows. Each time series was normalised by the 90th percentile value and filtered using a difference of Gaussians filter before computing features as done in the wearable sleep staging work of Walch *et al*. [45].

D. Additional model hypnograms

Figure 11 shows model and scorer four-class sleep hypnograms for the recording with the highest agreement as measured by Cohen's kappa.

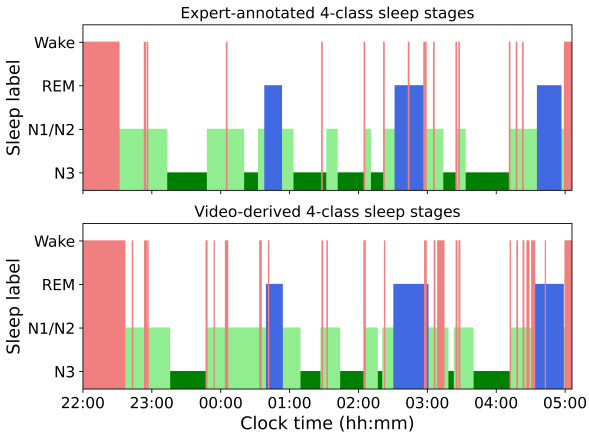

Figure 11. Four-class hypnograms corresponding to the highest Cohen's kappa statistic between model and scorer (0.82).

Figure 12 shows the hypnograms obtained for the recording with the lowest agreement. In their report, the sleep physiologist noted that there was a 'fuzzy signal on [many of the EEG channels] throughout, so sleep staging was difficult and not precise'. Measurement factors such as electrode placement and electrical interference are a common source of scorer uncertainty in conventional sleep staging [39].

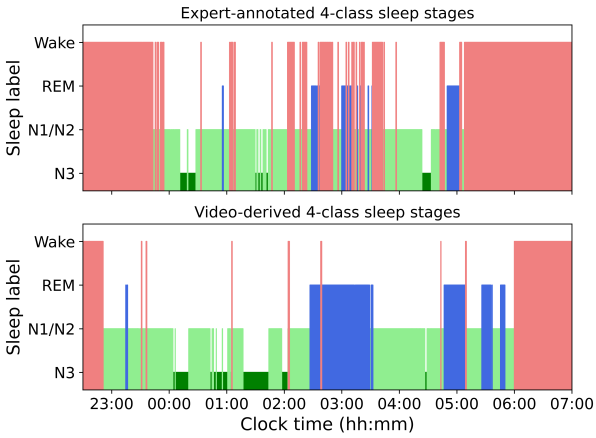

Figure 12. Four-class hypnograms corresponding to the lowest Cohen's kappa statistic between model and scorer (0.17).

References

[1] H. W. Agnew, W. B. Webb, and R. L. Williams. The first night effect: an EEG study of sleep. *Psychophysiology*, 2(3):263–266, Jan. 1966. 1

[2] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, Apr. 1960. 1

[3] Kris Cuppens, Lieven Lagae, Berten Ceulemans, Sabine Van Huffel, and Bart Vanrumste. Automatic video detection of body movement during sleep based on optical flow in pediatric patients with epilepsy. *Medical & Biological Engineering & Computing*, 48(9):923–931, Sept. 2010. 2

[4] Heidi Danker-Hopfe, Peter Anderer, Josef Zeitlhofer, Marion Boeck, Hans Dorn, Georg Gruber, Esther Heller, Erna Loretz, Doris Moser, Silvia Parapatics, Bernd Saletu, Andrea Schmidt, and Georg Dorffner. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. *Journal of Sleep Research*, 18(1):74–84, Mar. 2009. 1

[5] Lourdes M DelRosso, Caroline V Jackson, Kimberly Trotter, Oliviero Bruni, and Raffaele Ferri. Video-polysomnographic characterization of sleep movements in children with restless sleep disorder. *Sleep*, 42(4), Apr. 2019. 2

[6] T. B. Fitzpatrick. The validity and practicality of sun-reactive skin types I through VI. *Archives of Dermatology*, 124(6):869–871, June 1988. 3

[7] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2, 3

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. pages 770–778, 2016. 5

[9] Adrienne Heinrich, Xavier Aubert, and Gerard de Haan. Body movement analysis during sleep based on video motion estimation. In *2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013)*, pages 539–543, Oct. 2013. 2

[10] Adrienne Heinrich, Di Geng, Dmitry Znamenskiy, Jelte Peter Vink, and Gerard de Haan. Robust and Sensitive Video Motion Detection for Sleep Analysis. *IEEE Journal of Biomedical and Health Informatics*, 18(3):790–798, May 2014. Conference Name: IEEE Journal of Biomedical and Health Informatics. 2

[11] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1):185–203, Aug. 1981. 2, 3

[12] D. W. Hudgel, R. J. Martin, B. Johnson, and P. Hill. Mechanics of the respiratory system and breathing pattern during sleep in normal humans. *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology*, 56(1):133–137, Jan. 1984. 1

[13] Nicholas Dunkley Hutchinson and Simon Mark Chave Jones. Method and apparatus for monitoring of a human or animal subject field. US11563920B2, Jan. 2023. 4

[14] João Jorge, Mauricio Villarroel, Hamish Tomlinson, Oliver Gibson, Julie L. Darbyshire, Jody Ede, Mirae Harford, John Duncan Young, Lionel Tarassenko, and Peter Watkinson. Non-contact physiological monitoring of post-operative patients in the intensive care unit. *npj Digital Medicine*, 5(1):4, Dec. 2022. 4

[15] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 8

[16] Till Kroeger, Radu Timofte, Dengxin Dai, and Luc Van Gool. Fast Optical Flow using Dense Inverse Search, Mar. 2016. arXiv:1603.03590 [cs]. 3

[17] Bjorn Kruger, Anna Vogele, Marouane Lassiri, Lukas Herwartz, Thomas Terkatz, and Andreas Weber. Sleep detection using de-identified depth data. *Journal of Mobile Multimedia*, pages 327–342, Dec. 2014. 2

[18] Magdalena Lewandowska, Jacek Rumiński, Tomasz Kocejko, and Jedrzej Nowak. Measuring pulse rate with a webcam — A non-contact method for evaluating cardiac activity. In *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 405–410, Sept. 2011. 4

[19] Michael H. Li, Azadeh Yadollahi, and Babak Taati. A non-contact vision-based system for respiratory rate estimation. pages 2119–2122, Aug. 2014. ISSN: 1558-4615. 4

[20] Q. Li, R. G. Mark, and G. D. Clifford. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiological Measurement*, 29(1):15–32, Dec. 2007. Publisher: IOP Publishing. 4

[21] Xi Long, Renée Otte, Eric van der Sanden, Jan Werth, and Tao Tan. Video-Based Actigraphy for Monitoring Wake and Sleep in Healthy Infants: A Laboratory Study. *Sensors*, 19(5):1075, Jan. 2019. 2

[22] Sara Mahvash Mohammadi, Shirin Enshaeifar, Adrian Hilton, Derk-Jan Dijk, and Kevin Wells. Transfer Learning for Clinical Sleep Pose Detection Using a Single 2D IR Camera. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:290–299, 2021. 3, 7

[23] Kazuki Nakajima, Yoshiaki Matsumoto, and Toshiyo Tamura. Development of real-time image sequence analysis for evaluating posture change and respiratory rate of a subject in bed. *Physiological Measurement*, 22(3):N21–N28, Aug. 2001. 2, 4

[24] Teruaki Nochino, Yuko Ohno, Takafumi Kato, Masako Tanike, and Shima Okada. Sleep stage estimation method using a camera for home use. *Biomedical Engineering Letters*, 9(2):257–265, Apr. 2019. 2, 3, 6

[25] Jiapu Pan and Willis J. Tompkins. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236, Mar. 1985. 7

[26] Huy Phan and Kaare Mikkelsen. Automatic sleep staging of EEG signals: recent development, challenges, and future directions. *Physiological Measurement*, 2022. Publisher: IOP Publishing. 1, 4

[27] Huy Phan, Kaare B Mikkelsen, Oliver Chen, Philipp Koch, Alfred Mertins, and Maarten De Vos. SleepTransformer: Automatic Sleep Staging with Interpretability and Uncertainty Quantification. *IEEE Transactions on Biomedical Engineering*, pages 1–1, 2022. 1, 7

[28] Nicolò Pini, Ju Lynn Ong, Gizem Yilmaz, Nicholas I. Y. N. Chee, Zhao Siting, Animesh Awasthi, Siddharth Biju, Kishan Kishan, Amiya Patanaik, William P. Fifer, and Maristella Lucchini. An automated heart rate-based algorithm for sleep stage classification: Validation using conventional polysomnography and an innovative wearable electrocardiogram device. *Frontiers in Neuroscience*, 16:974192, Oct. 2022. 1

[29] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, Jan. 2011. Conference Name: IEEE Transactions on Biomedical Engineering. 4

[30] S. F. Quan, B. V. Howard, C. Iber, J. P. Kiley, F. J. Nieto, G. T. O'Connor, D. M. Rapoport, S. Redline, J. Robbins, J. M. Samet, and P. W. Wahl. The Sleep Heart Health Study: design, rationale, and methods. *Sleep*, 20(12):1077–1085, Dec. 1997. 1, 4, 5

[31] Mustafa Radha, Pedro Fonseca, Arnaud Moreau, Marco Ross, Andreas Cerny, Peter Anderer, Xi Long, and Ronald M. Aarts. Sleep stage classification from heart-rate variability using long short-term memory neural networks. *Scientific Reports*, 9(1):14149, Oct. 2019. 6

[32] Mustafa Radha, Pedro Fonseca, Arnaud Moreau, Marco Ross, Andreas Cerny, Peter Anderer, Xi Long, and Ronald M. Aarts. A deep transfer learning approach for wearable sleep stage classification with photoplethysmography. *npj Digital Medicine*, 4(1):1–11, Sept. 2021. 1, 5

[33] A. J. Schwichtenberg, Jeehyun Choe, Ashleigh Kellerman, Emily A. Abel, and Edward J. Delp. Pediatric Videosomnography: Can Signal/Video Processing Distinguish Sleep and Wake States? *Frontiers in Pediatrics*, 6:158, June 2018. 2

[34] Z. Shinar, A. Baharav, Y. Dagan, and S. Akselrod. Automatic detection of slow-wave-sleep using heart rate variability. In *Computers in Cardiology 2001. Vol.28 (Cat. No.01CH37287)*, pages 593–596, Sept. 2001. ISSN: 0276-6547. 1

[35] Ambra Stefani, David Gabelia, Thomas Mitterling, Werner Poewe, Birgit Högl, and Birgit Frauscher. A Prospective Video-Polysomnographic Analysis of Movements during Physiological Sleep in 100 Healthy Sleepers. *Sleep*, 38(9):1479–1487, Sept. 2015. 1

[36] Ambra Stefani and Birgit Högl. Sleep in Parkinson's disease. *Neuropsychopharmacology*, 45(1):121–128, Jan. 2020. 2

[37] Alexander Trumpp, Johannes Lohr, Daniel Wedekind, Martin Schmidt, Matthias Burghardt, Axel R. Heller, Hagen Malberg, and Sebastian Zaunseder. Camera-based photoplethysmography in an intraoperative setting. *BioMedical Engineering OnLine*, 17(1):33, Dec. 2018. 4

[38] Mark van Gastel, Sander Stuijk, Sebastiaan Overeem, Johannes P. van Dijk, Merel M. van Gilst, and Gerard de Haan. Camera-Based Vital Signs Monitoring During Sleep – A Proof of Concept Study. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1409–1418, May 2021. 2, 4

[39] Hans van Gorp, Iris A M Huijben, Pedro Fonseca, Ruud J G van Sloun, Sebastiaan Overeem, and Merel M van Gilst. Certainty about uncertainty in sleep staging: a theoretical framework. *Sleep*, 45(8):zsac134, Aug. 2022. 8

[40] Fokke B. van Meulen, Angela Grassi, Leonie van den Heuvel, Sebastiaan Overeem, Merel M. van Gilst, Johannes P. van Dijk, Henning Maass, Mark J. H. van Gastel, and Pedro Fonseca. Contactless Camera-Based Sleep Staging: The HealthBed Study. *Bioengineering*, 10(1):109, Jan. 2023. 2, 6

[41] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *arXiv:1706.03762 [cs]*, Dec. 2017. arXiv: 1706.03762. 7

[42] Christian Veauthier, Juliane Ryczewski, Sebastian Mansow-Model, Karen Otte, Bastian Kayser, Martin Glos, Christoph Schöbel, Friedemann Paul, Alexander U. Brandt, and Thomas Penzel. Contactless recording of sleep apnea and periodic leg movements by nocturnal 3-D-video and subsequent visual perceptive computing. *Scientific Reports*, 9(1):16812, Nov. 2019. 2

[43] Wim Verkruysse, Lars O. Svaasand, and J. Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics Express*, 16(26):21434–21445, Dec. 2008. 4

[44] Mauricio Villarroel, João Jorge, David Meredith, Sheera Sutherland, Chris Pugh, and Lionel Tarassenko. Non-contact vital-sign monitoring of patients undergoing haemodialysis treatment. *Scientific Reports*, 10(1):18529, Oct. 2020. Number: 1 Publisher: Nature Publishing Group. 4

[45] Olivia Walch, Yitong Huang, Daniel Forger, and Cathy Goldstein. Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. *Sleep*, 42(12):zsz180, Dec. 2019. 1, 8

[46] Wenjin Wang and Albertus C den Brinker. Algorithmic insights of camera-based respiratory motion extraction. *Physiological Measurement*, 43(7):075004, July 2022. 4

[47] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic Principles of Remote PPG. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, July 2017. 4

[48] Bernice M Wulterkens, Pedro Fonseca, Lieke W A Hermans, Marco Ross, Andreas Cerny, Peter Anderer, Xi Long, Johannes P van Dijk, Nele Vandenbussche, Sigrid Pillen, Merel M van Gilst, and Sebastiaan Overeem. It is All in the Wrist: Wearable Sleep Staging in a Clinical Population versus Reference Polysomnography. *Nature and Science of Sleep*, 13:885–897, June 2021. 1

[49] Guo-Qiang Zhang, Licong Cui, Remo Mueller, Shiqiang Tao, Matthew Kim, Michael Rueschman, Sara Mariani, Daniel Mobley, and Susan Redline. The National Sleep Research Resource: towards a sleep data commons. *Journal of the American Medical Informatics Association: JAMIA*, 25(10):1351–1358, Oct. 2018. 7

[50] Dan Ziegler, G. Laux, K. Dannehl, M. Spüler, H. Mühlen, P. Mayer, and F.a. Gries. Assessment of Cardiovascular Autonomic Function: Age-related Normal Ranges and Reproducibility of Spectral Analysis, Vector Analysis, and Standard Tests of Heart Rate Variation and Blood Pressure Responses. *Diabetic Medicine*, 9(2):166–175, 1992. 6