Contactless Oxygen Monitoring with Gated Transformer

Hao He[1*] **Yuan Yuan**[1*] **Ying-Cong Chen**[2*] **Peng Cao**[1] **Dina Katabi**[1]
[1]MIT CSAIL [2]HKUST (GZ) and HKUST
{haohe, miayuan, pengcao, dk}@mit.edu, yingcong.ian.chen@gmail.com

Abstract

With the increasing popularity of telehealth, it becomes critical to ensure that basic physiological signals can be monitored accurately at home, with minimal patient overhead. In this paper, we propose a contactless approach for monitoring patients' blood oxygen at home, simply by analyzing the radio signals in the room, without any wearable devices. We extract the patients' respiration from the radio signals that bounce off their bodies and devise a novel neural network that infers a patient's oxygen estimates from their breathing signal. Our model, called *Gated BERT-UNet*, is designed to adapt to the patient's medical indices (e.g., gender, sleep stages). It has multiple predictive heads and selects the most suitable head via a gate controlled by the person's physiological indices. Extensive empirical results show that our model achieves high accuracy on both medical and radio datasets.

1 Introduction

Remote health monitoring and telehealth are increasingly popular because they reduce costs and facilitate access to healthcare, particularly for people in remote locations [2]. Further, remote health monitoring can track the long-term physiological state of a patient or an older person who lives alone at home, and enable family and professional caregivers to provide timely help [6, 42, 13, 47]. Delivering such services, however, depends on the availability of solutions that continuously measure people's physiological signals at home, with minimal overhead to patients.

Oxygen saturation is an important physiological signal whose at-home monitoring would benefit very old adults and individuals at high risk for low blood oxygen [29]. Oxygen saturation refers to the amount of oxygen in the blood – that is the fraction of oxygen-saturated hemoglobin relative to the total blood hemoglobin. Normal oxygen levels range from 94% to 100%. Lower oxygen can be dangerous, and if severe, lead to brain and lung failure [10, 21].

Today, measuring oxygen saturation requires the person to wear a pulse oximeter on their finger, and actively measure themselves. While pulse oximeters are very helpful, they can be impractical in some at-home monitoring scenarios. In particular, old people in their late 80's and 90's are at high risk for low blood oxygen [32], and should regularly monitor their oxygen. Many of them however may suffer from dementia or cognitive impairment that prevent them from measuring themselves. COVID and pneumonia patients recovering at home can suffer from delirium [17], which can affect their reasoning and ability to measure their oxygen levels. Additionally, blood oxygen tends to drop during sleep, making it particularly important to track oxygen overnight [33, 15]. Yet people cannot actively measure themselves while asleep.

The above use cases motivate us to try to complement pulse oximetry with a new approach that can work passively and continuously, assessing blood oxygen throughout the night without requiring the

[*]Equal contribution, order determined via a random coin flip

Workshop on Learning from Time Series for Health, 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

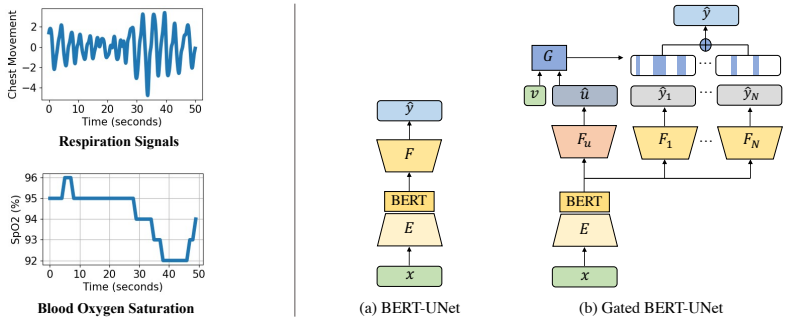


Figure 1: Left: Examples of respiration signals and corresponding blood oxygen saturation. Right: Illustrations of the proposed models: (a) The backbone BERT-UNet model, which contains an encoder E and a predictor F as the UNet structure, and a BERT module at the bottleneck of the UNet. (b) The Gated BERT-UNet, where a gate G, controlled by the accessible variable v and the predicted inaccessible variable $\hat{u}$, is used to select among multiple predictive heads.

person to wear a sensor or actively measure themselves. Prior attempts at estimating blood oxygen passively, without a wearable sensor, rely on cameras [27, 43, 44, 39, 3, 16]. For example, [27] proposes a convolution neural network that analyzes a video of the person's palm to estimate his/her blood oxygen. While not requiring wearable sensors, it cannot work continuously since the user cannot keep their hand in front of the camera for a long time. It also cannot operate in dark settings and thus cannot monitor one's oxygen during sleep. To bypass the limitations imposed by cameras, we propose a different sensing modality, *radio-frequency (RF) signals*.

We propose to monitor oxygen saturation by analyzing the radio signals that bounce off a person's body. Recent research has demonstrated the feasibility of monitoring breathing, heart rate, and even sleep stages by transmitting a very low power RF signal and analyzing its reflections off a person's body [1, 50]. The medical literature shows an inherent dependence and dynamic interaction between the breathing signal and oxygen saturation [18, 4, 34]. Building on these advances, we use RF signals to track a person's breathing signal and train a neural network to infer oxygen from respiration. Such a design can measure a person's oxygen without any physical contact or wearable devices. Thus, it does not burden the patient or interfere with their sleep. Using breathing as a mediator has an additional side benefit. It would be very hard to collect a large dataset of radio signals and the corresponding oxygen levels. Luckily however there are multiple large medical datasets that contain continuous breathing signals paired with oxygen measurements. This allows us to train a model to infer oxygen from breathing and test it directly on breathing extracted from radio signals.

While designing our model, motivated by personalized medicine, we aim for a model that can adapt to the patient's medical indices (e.g., gender, disease diagnoses). We observe that many medical indices are binary or categorical. To leverage such variables, we propose *Gated BERT-UNet*, a new transformer model that has multiple predictive heads. It selects the most suitable head for each person via a gate controlled by the person's categorical indices. We evaluate our model on medical and RF datasets. Experiments show that our model's average absolute error in predicting oxygen saturation is 1.3%, which is significantly lower than the state-of-the-art (SOTA) camera-based models [27].

2 Our System and Model (Gated BERT-UNet)

The proposed oxygen estimation solution operates in two steps: (1) Extracting breathing signals from RF signals and (2) Estimating SpO2 from breathing signals. Please refer to Appendix B for the details of step one. We focus on step two in this section.

We formulate the oxygen saturation prediction from respiration signals as a sequence-to-sequence regression task. Figure 1 left shows an example of respiration signals and corresponding blood oxygen saturation. The model takes the breathing signal $x \in \mathbb{R}^{1 \times f_b T}$ over a T-second interval and predicts the oxygen time series $y \in \mathbb{R}^{1 \times f_o T}$ over the same period, where f_b and f_o are the sampling frequencies of respiration and oxygen, respectively. In this study, the default sampling rates for respiration and oxygen are $f_b = 10Hz$ and $f_o = 1Hz$.

Table 1: Performances on medical datasets. (* indicates using extra physiological variables.)

Model	SHHS			MESA			MrOS			Overall		
	Corr$^{\uparrow}$	MAE$^{\downarrow}$	RMSE$^{\downarrow}$	Corr$^{\uparrow}$	MAE$^{\downarrow}$	RMSE$^{\downarrow}$	Corr$^{\uparrow}$	MAE$^{\downarrow}$	RMSE$^{\downarrow}$	Corr$^{\uparrow}$	MAE$^{\downarrow}$	RMSE$^{\downarrow}$
CNN	0.47	1.72	1.81	0.46	1.62	1.72	0.49	1.77	1.82	0.47	1.73	1.78
CNN-RNN	0.48	1.70	1.80	0.50	1.56	1.65	0.52	1.76	1.84	0.49	1.67	1.77
BERT-UNet	0.51	1.67	1.77	0.52	1.55	1.66	0.54	1.75	1.84	0.51	1.64	1.76
BERT-UNet + VarAug*	**0.52**	1.65	1.76	0.52	1.51	1.62	**0.55**	1.68	1.78	0.52	1.61	1.72
Gated BERT-UNet*	**0.52**	**1.61**	**1.72**	**0.53**	**1.50**	**1.61**	**0.55**	**1.65**	**1.75**	**0.53**	**1.58**	**1.70**

As shown in Figure 1(a), our backbone model is a combination of a BERT module [9] and UNet [38]. The *BERT-UNet* backbone consists of an encoder E and an oxygen predictor F. The encoder has convolutional layers and bidirectional attention modules while the decoder is fully convolutional and has skip-links from the encoder. The backbone is simply trained by L_1 loss and correlation loss. Please refer to Appendix C for the details of model architecture and loss functions.

Our full model, *Gated BERT-UNet* augments BERT-UNet with multiple predictive heads to adapt the model's prediction to different medical indices. It is important for boosting the model's performance since the relationship between breathing and oxygen saturation has been shown related to many medical indices. For example, gender is an influential factor as men and women have differences in their oxygen transport systems [36]. Another example is that sleep stage affects a person's resting oxygen levels [8].

To do the adaptation, our Gated BERT-UNet selects the most suitable head for a person via a gate controlled by the subject's categorical indices. Specifically, as shown in Figure 1(b), it has a gate function $G(v, u) : \mathcal{V} \times \mathcal{U} \rightarrow \{1, 2, \cdots, N\}$ where $v \in \mathcal{V}$ and $u \in \mathcal{U}$ are accessible/inaccessible variables, N is the number of gate statuses. We use the term *accessible variable* for variables easily available during inference time (e.g., gender, race) and the term *inaccessible variable* for information that is not available during inference, but available during training, like a person's sleep stages. The prediction for inaccessible variable is learned concurrently with the main task under full supervision.

The model has N heads $\{F_i\}_{i=1}^{N}$ that adapt the prediction $\hat{y}_i = F_i(E(x))$ to the gate status. It also has an extra predictor F_u to infer inaccessible variables u. During testing time, based on the accessible variables v and estimated inaccessible variables $\hat{u}$, we evaluate the gate status $s = G(v, \hat{u})$. Please see Appendix D for the details of (1) the loss function used to train our full model; (2) the exact construction of the gate function $G(v, u)$.

3 Experiments and Results

Dataset and Metrics. We conduct experiments on three medical datasets: SHHS [51], MrOS [5] and MESA [7], and a self-collected RF dataset. We consider three evaluation metrics: correlation (Corr), mean-absolute error (MAE) and rooted mean-squared error (RMSE). Please refer to Appendix E for more details on datasets and metrics.

Baselines. We compare our model with the following baselines: (a) *CNN*, a fully convolutional model; (b) *CNN-RNN*, an augmentation of CNN with RNN units in the bottleneck; (c) *BERT-UNet + VarAug*, which takes BERT-UNet as the backbone and uses medical variables as extra inputs or outputs. Appendix F explains more details of the baselines.

Training and Evaluation Protocols. Since the RF dataset is too small for training, we train models on breathing signals from the medical datasets, and test them directly on the respiration signals extracted from the RF dataset as well as medical datasets' test sets. More details are in Appendix G.

3.1 Results on Medical Datasets

Quantitative. The results on medical datasets are shown in Table 1. Since there is no past work that predicts oxygen from breathing or radio signals, all models in the table refer to variants of our neural network. The table shows that all variants achieve relatively low prediction errors with an average MAE that ranges from 1.58 to 1.73 percent, and an average RMSE that ranges from 1.70 to 1.78 percent. Such a relatively low RMSE shows our model rarely has predictions that largely deviate from the ground truth. All variants also achieve reasonable high correlations ranging from 0.47 to 0.53. Such a correlation level indicates our model's prediction capture the dynamics of the ground

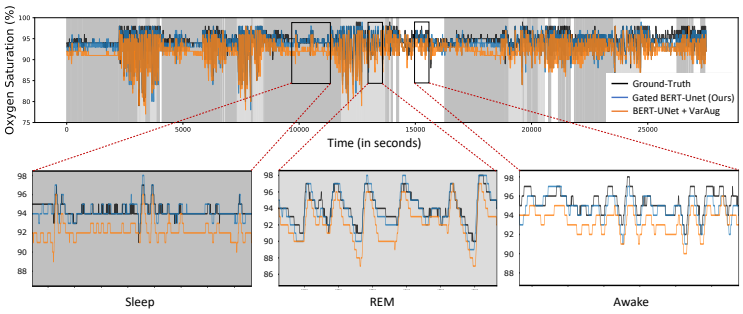


Figure 2: An illustrative example of the oxygen level predicted by the BERT-UNet + VarAug (Orange Curve), and our Gated BERT-UNet model (Blue Curve). The background color indicates sleep stages. The 'dark grey', 'light grey' and 'white' corresponds to non-REM 'Sleep', 'REM' and 'Awake'.

Table 2: Results on the RF dataset.

Model	Corr↑	MAE↓	RMSE↓
CNN	0.45	1.65	1.73
CNN-RNN	0.49	1.85	1.93
BERT-UNet	0.48	1.49	1.58
BERT-UNet + VarAug	0.49	1.49	1.59
Gated BERT-UNet	**0.52**	**1.32**	**1.54**

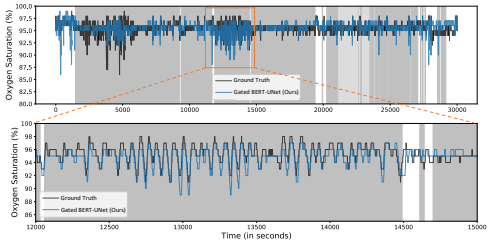


Figure 3: Example of SpO2 prediction results on RF dataset.

truth SpO2 which is also visually shown in Figure 2. These quantitative results highlight that our system can be useful for continuous monitoring of patients' oxygen at home.

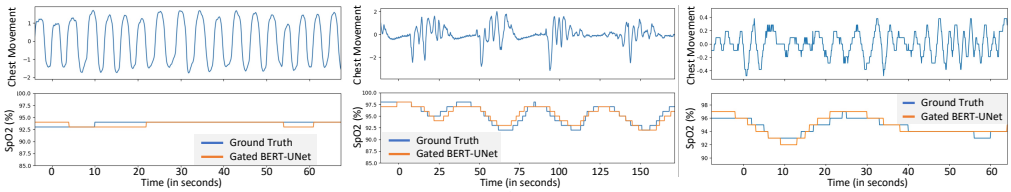
The upper rows in the table show the variants that do not leverage side variables. The table shows that the BERT-UNet model consistently outperforms the CNN model and the CNN-RNN model on all datasets, in all metrics. This indicates that BERT-UNet is a preferable architecture for this task.

The bottom rows in the table show the results of models leveraging accessible and inaccessible medical variables. The table shows that BERT-UNet + VarAug and Gated BERT-UNet outperform models that do not leverage physiological indices, demonstrating the benefit of leveraging such variables. In addition, Gated BERT-UNet outperforms VarAug on all three datasets, demonstrating that a gated multi-head approach works best for such categorical side variables.

Qualitative. Figure 2 visualizes the predicted oxygen saturation of the Gated BERT-UNet model and the VarAug model on a male subject in the SHHS dataset. As the ground-truth oxygen saturation are integers, we round the predicted oxygen values. The background color indicates different sleep stages. The 'dark grey', 'light grey' and 'white' correspond to non-REM 'Sleep', 'REM' and 'Awake' stage, respectively. We observe that Gated BERT-UNet consistently outperforms VarAug over the whole night. The small panel focuses on different sleep stages. In general, different sleep stages tend to show different behavior and hence the importance of using a gated model. Specifically, oxygen is typically more stable during non-REM 'Sleep' than during REM and Awake stages. The figures show that Gated BERT-UNet can track the ups and downs in oxygen and is significantly more accurate than VarAug. This experiment demonstrates that the way we incorporate the sleep stages into the model improves performance across different sleep stages.

3.2 Results on RF Dataset

The results on the RF dataset are shown in Table 2. All model variants have low prediction error and high correlation. Among the models that do not leverage physiological variables, CNN-RNN and Bert-UNet perform better than the vanilla CNN, which shows the importance of modeling temporal information. When using physiological variables, the performance of BERT-UNet+VarAug is similar to that of BERT-UNet, while Gated BERT-UNet is better than all other variants. This demonstrates that the gating design leverages auxiliary variables better. Overall, the MAEs, RMSEs and correlations on the RF dataset are comparable to those on the medical datasets. This indicates that our model is directly applicable to respiration signals from RF. We have also visualized the prediction results in Figure 3. As shown, our model can accurately track the fluctuation of ground truth SpO2.



(a) (b) (c)

Figure 4: Examples of breathing signals and corresponding predicted oxygen saturation. (a) is an example of normal breathing signals. (b)(c) are examples of different abnormal breathing signals. In each figure, the first row shows the breathing signals while the second row shows the ground truth oxygen saturation (in blue) and the oxygen saturation predicted by our model (in orange).

Table 3: Prior SOTA oxygen saturation prediction methods [11, 27] v.s. our *Gated BERT-UNet*.

Method	Corr$^{\uparrow}$	MAE$^{\downarrow}$	RMSE$^{\downarrow}$	Contactless
Ding et al. 2018	0.26	2.43	2.85	No
Mathew et al. 2021	0.46	1.97	2.16	Yes
Ours	**0.52**	**1.32**	**1.54**	Yes

3.3 Visualization of Breathing-Oxygen Patterns

We demonstrate several visual results on different patterns of breathing signals and the corresponding ground truth/predicted oxygen saturation, as shown in Figure 4. Figure 4(a) shows a normal breathing pattern, which leads to constant oxygen saturation. In contrast, Figure 4(b,c) present two different abnormal breathing signals and the resulting fluctuated oxygen predictions. These figures show the diversity of the oxygen and breathing patterns as well as the complexity of their relationship. The model however is able to capture this relationship for highly diverse patterns.

3.4 Comparison with Past Works

We compare our approach with two recent deep-learning camera-based SpO2 monitoring methods. The first method [11] asks the user to press his/her finger against a smartphone camera, and uses a CNN to estimate SpO2 from the video. The second method [27] uses a CNN to estimate oxygen from a video of the person's hand. The input of these systems is different from ours (camera vs. RF); so to compare them we follow the setup in [27]. In their setting, test samples are 180-240 seconds and vary between normal breathing to no or minimal breathing. Similarly, we divide the RF test dataset into non-overlapping 240-second segments containing both regular breathing and shallow breathing (i.e., apnea or hypopnea) and compute metrics on them. Table 3 reports the output of the three models. The results of the baselines are taken from [27], and the results of our RF-based model are computed as described. The results show that our model improves the correlation and reduces the MAE and RMSE in comparison to past work.

More results are in the appendix: Section H analyzes (predicted) oxygen value distributions in different race groups; Section I has more visualization on our model's performance on different datasets, patients with different relevant diseases.

4 Concluding Remarks

This paper introduces the new task of inferring oxygen saturation from radio signals. It develops a new gated transformer architecture to deliver this application and adapt deep models to auxiliary categorical variables. We note the work have some limitations. First, the paper focuses on special use cases (e.g., at-home oxygen monitoring in very old adults or during sleep), but is not suitable for other use cases (e.g., measuring oxygen to optimize performance during exercise). Second, the results in the paper provide an initial proof of concept that the shape and dynamics of the breathing signal include sufficient clues to infer a useful estimate of oxygen saturation. However, before this system can be used in clinical care, one needs clinical studies to quantify the performance for different disease conditions.

References

[1] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015.

[2] Mohammed Al-Khafajiy, Thar Baker, Carl Chalmers, Muhammad Asim, Hoshang Kolivand, Muhammad Fahim, and Atif Waraich. Remote health monitoring of elderly through wearable sensors. *Multimedia Tools and Applications*, 78(17):24681–24706, 2019.

[3] Ufuk Bal. Non-contact estimation of heart rate and oxygen saturation using ambient light. *Biomedical optics express*, 6(1):86–97, 2015.

[4] N.A. Bergman. Cyclic variations in blood oxygenation with the respiratory cycle. *Anesthesiology*, 1961.

[5] Terri Blackwell, Kristine Yaffe, Sonia Ancoli-Israel, Susan Redline, Kristine E Ensrud, Marcia L Stefanick, Alison Laffan, Katie L Stone, and Osteoporotic Fractures in Men Study Group. Associations between sleep architecture and sleep-disordered breathing and cognition in older community-dwelling men: the osteoporotic fractures in men sleep study. *Journal of the American Geriatrics Society*, 59(12):2217–2225, 2011.

[6] BG Celler, W Earnshaw, ED Ilsar, L Betbeder-Matibet, MF Harris, R Clark, T Hesketh, and NH Lovell. Remote monitoring of health status of the elderly at home. a multidisciplinary project on aging at the university of new south wales. *International journal of bio-medical computing*, 40(2):147–155, 1995.

[7] Xiaoli Chen, Rui Wang, Phyllis Zee, Pamela L Lutsey, Sogol Javaheri, Carmela Alcántara, Chandra L Jackson, Michelle A Williams, and Susan Redline. Racial/ethnic differences in sleep disturbances: the multi-ethnic study of atherosclerosis (mesa). *Sleep*, 38(6), 2015.

[8] Eunkyung Choi, Doo-Heum Park, Jae-hak Yu, Seung-Ho Ryu, and Ji-Hyeon Ha. The severity of sleep disordered breathing induces different decrease in the oxygen saturation during rapid eye movement and non-rapid eye movement sleep. *Psychiatry investigation*, 13(6):652, 2016.

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[10] Genaro Díaz-Regañón, Eduardo Miñambres, Marisol Holanda, Segundo González-Herrera, Francisco López-Espadas, and Carlos Garrido-Díaz. Usefulness of venous oxygen saturation in the jugular bulb for the diagnosis of brain death: report of 118 patients. *Intensive care medicine*, 28(12):1724–1728, 2002.

[11] Xinyi Ding, Damoun Nassehi, and Eric C Larson. Measuring oxygen saturation with smartphone cameras using convolutional neural networks. *IEEE journal of biomedical and health informatics*, 23(6):2603–2610, 2018.

[12] Lijie Fan, Tianhong Li, Rongyao Fang, Rumen Hristov, Yuan Yuan, and Dina Katabi. Learning longterm representations for person re-identification using radio signals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[13] Lijie Fan, Tianhong Li, Yuan Yuan, and Dina Katabi. In-home daily-life captioning using radio signals. In *ECCV*, pages 105–123. Springer, 2020.

[14] John R Feiner, John W Severinghaus, and Philip E Bickler. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. *Anesthesia & Analgesia*, 2007.

[15] Robert E Gries and Lee J Brooks. Normal oxyhemoglobin saturation during sleep: how low does it go? *Chest*, 110(6):1489–1492, 1996.

[16] Alessandro R Guazzi, Mauricio Villarroel, Joao Jorge, Jonathan Daly, Matthew C Frise, Peter A Robbins, and Lionel Tarassenko. Non-contact measurement of oxygen saturation with an rgb camera. *Biomedical optics express*, 6(9):3320–3338, 2015.

[17] Dong Han, Chenyang Wang, Xiaojing Feng, and Jing Wu. Delirium during recovery in patients with severe covid-19: Two case reports. *Frontiers in Medicine*, 7, 2020.

[18] PG Hanson, KH Lin, and MB McIlroy. Influence of breathing pattern on oxygen exchange during hypoxia and exercise. In *J Appl Physiol*, June 1975.

[19] Chen-Yu Hsu, Yuchen Liu, Zachary Kabelac, Rumen Hristov, Dina Katabi, and Christine Liu. Extracting gait velocity and stride length from surrounding radio signals. In *Proc. of the 2017 CHI Conf. on Human Factors in Computing Systems*, 2017.

[20] Dani Kiyasseh, Tingting Zhu, and David A Clifton. Clops: Continual learning of physiological signals. *arXiv preprint arXiv:2004.09578*, 2020.

[21] SE Lapinsky, M Aubin, S Mehta, P Boiteau, and AS Slutsky. Safety and efficacy of a sustained inflation for alveolar recruitment in adults with respiratory failure. *Intensive care medicine*, 25(11):1297–1301, 1999.

[22] Albert Lecube, Gabriel Sampol, Patricia Lloberes, Odile Romero, Jordi Mesa, Cristina Hernández, and Rafael Simó. Diabetes is an independent risk factor for severe nocturnal hypoxemia in obese patients. a case-control study. *PloS one*, 4(3):e4692, 2009.

[23] Tianhong Li, Peng Cao, Yuan Yuan, Lijie Fan, Yuzhe Yang, Rogerio S Feris, Piotr Indyk, and Dina Katabi. Targeted supervised contrastive learning for long-tailed recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6918–6928, 2022.

[24] Tianhong Li, Lijie Fan, Yuan Yuan, Hao He, Yonglong Tian, Rogerio Feris, Piotr Indyk, and Dina Katabi. Addressing feature suppression in unsupervised visual representations. *arXiv e-prints*, pages arXiv–2012, 2020.

[25] Tianhong Li, Lijie Fan, Yuan Yuan, and Dina Katabi. Unsupervised learning for human sensing using radio signals. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2022.

[26] Yifan Liu, Bohan Zhuang, Chunhua Shen, Hao Chen, and Wei Yin. Auxiliary learning for deep multi-task learning. *arXiv preprint arXiv:1909.02214*, 2019.

[27] Joshua Mathew, Xin Tian, Min Wu, and Chau-Wai Wong. Remote blood oxygen estimation from videos using neural networks. *arXiv preprint arXiv:2107.05087*, 2021.

[28] Taylor Mordan, Nicolas Thome, Gilles Henaff, and Matthieu Cord. Revisiting multi-task learning with rock: a deep residual auxiliary block for visual detection. In *NeurIPS*, 2018.

[29] Mark Moss, Mark Franks, Pamela Briggs, David Kennedy, and Andrew Scholey. Compromised arterial oxygen saturation in elderly asthma sufferers results in selective cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, 27(2):139–150, 2005.

[30] Murugappan Murugappan, Nagarajan Ramachandran, Yaacob Sazali, et al. Classification of human emotion from eeg using discrete wavelet transform. *Journal of biomedical science and engineering*, 3(04):390, 2010.

[31] Shashi Narayan, Nikos Papasarantopoulos, Shay B Cohen, and Mirella Lapata. Neural extractive summarization with side information. *arXiv preprint arXiv:1704.04530*, 2017.

[32] NLM. Aging changes in the lungs. https://medlineplus.gov/ency/article/004011.htm, 2020. Accessed: 2020-10-01.

[33] David T Palma, George M Philips, Miguel R Arguedas, Susan M Harding, and Michael B Fallon. Oxygen desaturation during sleep in hepatopulmonary syndrome. *Hepatology*, 47(4):1257–1263, 2008.

[34] K. J. Parkins, L. M. Poets, C. F.and O'Brien, V. A. Stebbens, and D. P. Southall. Effect of exposure to 15% oxygen on breathing patterns and oxygen saturation in infants: interventional study. *BMJ (Clinical research ed.)*, 1998.

[35] Tauhidur Rahman, Alexander T Adams, Ruth Vinisha Ravichandran, Mi Zhang, Shwetak N Patel, Julie A Kientz, and Tanzeem Choudhury. Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar. In *UbiComp*, pages 39–50, 2015.

[36] Tony Reybrouck and Robert Fagard. Gender differences in the oxygen transport system during maximal exercise in hypertensive subjects. *Chest*, 115(3):788–792, 1999.

[37] Beanbonyka Rim, Nak-Jun Sung, Sedong Min, and Min Hong. Deep learning in physiological signal data: A survey. *Sensors*, 20(4):969, 2020.

[38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 2015.

[39] Dangdang Shao, Chenbin Liu, Francis Tsow, Yuting Yang, Zijian Du, Rafael Iriya, Hui Yu, and Nongjian Tao. Noncontact monitoring of blood oxygen saturation using camera and dual-wavelength imaging system. *IEEE Transactions on Biomedical Engineering*, 2015.

[40] Xiaoyong Shen, Aaron Hertzmann, Jiaya Jia, Sylvain Paris, Brian Price, Eli Shechtman, and Ian Sachs. Automatic portrait segmentation for image stylization. In *Computer Graphics Forum*, volume 35, pages 93–102, 2016.

[41] Michael W Sjoding, Robert P Dickson, Theodore J Iwashyna, Steven E Gay, and Thomas S Valley. Racial bias in pulse oximetry measurement. *New England Journal of Medicine*, 383(25):2477–2478, 2020.

[42] Yonglong Tian, Guang-He Lee, Hao He, Chen-Yu Hsu, and Dina Katabi. Rf-based fall monitoring using convolutional neural networks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–24, 2018.

[43] Mark Van Gastel, Sander Stuijk, and Gerard De Haan. New principle for measuring arterial blood oxygenation, enabling motion-robust remote monitoring. *Scientific reports*, 6(1):1–16, 2016.

[44] Mark van Gastel, Wim Verkruysse, and Gerard de Haan. Data-driven calibration estimation for robust remote pulse-oximetry. *Applied Sciences*, 9(18):3857, 2019.

[45] Yangtao Wang, Xi Shen, Shell Xu Hu, Yuan Yuan, James L Crowley, and Dominique Vaufreydaz. Self-supervised transformers for unsupervised object discovery using normalized cut. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14543–14553, 2022.

[46] Yangtao Wang, Xi Shen, Yuan Yuan, Yuming Du, Maomao Li, Shell Xu Hu, James L Crowley, and Dominique Vaufreydaz. Tokencut: Segmenting objects in images and videos with self-supervised transformer and normalized cut. *arXiv preprint arXiv:2209.00383*, 2022.

[47] Yuzhe Yang, Yuan Yuan, Guo Zhang, Hao Wang, Ying-Cong Chen, Yingcheng Liu, Christopher G Tarolli, Daniel Crepeau, Jan Bukartyk, Mithri R Junna, et al. Artificial intelligence-enabled detection and assessment of parkinson's disease using nocturnal breathing signals. *Nature medicine*, pages 1–9, 2022.

[48] Yuan Yuan, Xiaodan Liang, Xiaolong Wang, Dit-Yan Yeung, and Abhinav Gupta. Temporal dynamic graph lstm for action-driven video object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1801–1810, 2017.

[49] Yuan Yuan, Yueming Lyu, Xi Shen, Ivor W Tsang, and Dit-Yan Yeung. Marginalized average attentional network for weakly-supervised learning. *arXiv preprint arXiv:1905.08586*, 2019.

[50] Shichao Yue, Hao He, Hao Wang, Hariharan Rahul, and Dina Katabi. Extracting multi-person respiration from entangled rf signals. In *UbiComp*, 2018.

[51] 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*, 2018.

[52] Mingmin Zhao, Fadel Adib, and Dina Katabi. Emotion recognition using wireless signals. In *MobiCom*, 2016.

[53] Mingmin Zhao, Shichao Yue, Dina Katabi, Tommi S Jaakkola, and Matt T Bianchi. Learning sleep stages from radio signals: A conditional adversarial architecture. In *ICML*, 2017.

A Related Work

Monitoring Oxygen Saturation. The most accurate measurements of oxygen saturation are invasive and require arterial blood samples. The non-invasive and widely-used method for measuring oxygen saturation (SpO2) uses a pulse oximeter, a small device worn on the finger. To enable remote SpO2 measurements, past work has investigated the use of cameras [43, 44, 39, 3, 16]. However those methods have limitations like susceptibility to noise, sensitivity to motion, and a need for ambient light. Recently, deep learning has been considered to aid SpO2 monitoring using cameras. [11] tried to monitor SpO2 using smartphones. But their solution requires the fingertip to be pressed against the camera, and hence cannot provide continuous overnight measurements. A more recent work [27] has estimated SpO2 in a contactless way with regular RGB cameras. Their method first extracts the region of interest from the video of the person's palm, then uses a CNN model to estimate SpO2. While this approach is contactless, it still requires the user to keep their hand in front of the video camera for the duration of the monitoring, which is not practical for continuous or overnight monitoring. Our work differs from all of these prior works in that we predict oxygen values from breathing or radio waves, which allows for continuous oxygen sensing in a contactless and passive manner.

Contactless Health Sensing with Radio Signals. The past decade has seen a rapid growth in research on passive sensing using RF signals. Early work has demonstrated the possibility of sensing one's breathing and heart rate using radio signals [1]. Later, researchers have shown that by analyzing the RF signals that bounce off the human body, they can monitor a variety of health metrics including sleep, gait, falls, and even human emotions [53, 19, 42, 52, 12, 25, 47, 24]. We build on this work to enable SpO2 monitoring with radio waves.

Adaptation of ML Models to Medical Indices. Prior deep learning models [46, 45, 23, 49, 48] do not adapt to a person's medical indices. For example, the literature has models that infer sleep stages from respiration [53], detect arrhythmia from ECG [20], and classify emotion from EEG signals [30]. A recent survey [37] collected 147 papers about learning with physiological signals. None of the deep models therein adapt to a person's medical indices. The deep learning literature includes a few approaches for leveraging auxiliary variables. If the variable is available at inference time, typically it is taken as an extra input [31, 40]. Variables accessible only during training are typically used as extra supervisors to regularize the model via multi-task learning [26, 28]. We propose a gating mechanism to handle categorical variables and show that it performs better.

B Extracting Breathing Signals from RF Signals

We leverage past work on extracting breathing from the RF signals. Specifically, our system is equipped with a multi-antenna Frequency-Modulated Continuous-Wave (FMCW) radio, which is commonly used in passive health monitoring [35, 50, 13]. The radio transmits a very low power RF signal and captures its reflections from the environment. We process these reflections using the algorithm in [50] to infer the subject's breathing signal. Past work shows that breathing signals extracted in this manner are highly accurate. Specifically, their correlation with an FDA-approved breathing belt on the person ranges from 91% to 99%, depending on the distance from the radio and the distance between people [50].

C Backbone Model: BERT-UNet

Our backbone model a combination of a BERT module [9] and UNet [38]. As shown in Figure 1(a), our *BERT-UNet* model consists of an encoder $E(\cdot;\theta_e)$ and an oxygen predictor $F(\cdot;\theta_f)$. The encoder is composed of a fully convolutional network (FCN) followed by a bidirectional-transformer (BERT) module [9]. The FCN extracts local features from the raw respiration signals, then the BERT module captures long-term temporal dependencies based on those features. The predictor F is composed of several deconvolutional layers, which up-sample the extracted features to the same time resolution of oxygen saturation. Formally, we have $E : \mathbb{R}^{1\times f_b T} \to \mathbb{R}^{n\times \alpha f_b T}$ and $F : \mathbb{R}^{n\times \alpha f_b T} \to \mathbb{R}^{1\times f_o T}$ where n is the dimension of the respiration feature and α is the down-sampling factor ($\alpha = 1/240$ in our experiments).

The model is trained with a combination of the L_1 loss and the correlation loss given below:

$$\mathcal{L}(\hat{y}, y) = \frac{\|\hat{y}-y\|_1}{f_o T} - \lambda \frac{\sum_i (\hat{y}^i - \mu_{\hat{y}})(y^i - \mu_y)}{\sqrt{\sum_i (\hat{y}^i - \mu_{\hat{y}})^2 \sum_i (y^i - \mu_y)^2}}. \quad (1)$$

Here $\hat{y} = F(E(x;\theta_e);\theta_f)$ is the model prediction, y is the ground truth oxygen, μ_y and $\mu_{\hat{y}}$ are the mean values of y and $\hat{y}$, and λ is a hyper-parameter to balance the two loss terms. We choose the L_1 loss over other regression loss functions, since it is more robust to outliers and empirically has better performance. We also use the correlation loss to help in matching the fluctuations in the predicted oxygen with the fluctuations of the ground truth.

Architecture Specifications. The encoder has nine 1-D convolutional layers (Conv-BatchNorm-RReLU) that shrink the features' temporal dimension by 240 times. It is then followed by several bi-directional multi-head self-attention layers (BERT) [9] to aggregate the temporal information at the bottleneck. We use 8 layers, 6 heads with hidden-size of 256, intermediate-size of 512 for self-attention, and the max position embeddings is 2400. The decoder contains 7 layers of 1-D de-convolutional layers (DeConv-Norm-RReLU). We also use a skip connection [38] by concatenating the convolutional layers in the encoder to the de-convolutional layers in the predictor. Figure 5 illustrates the overall network architecture.

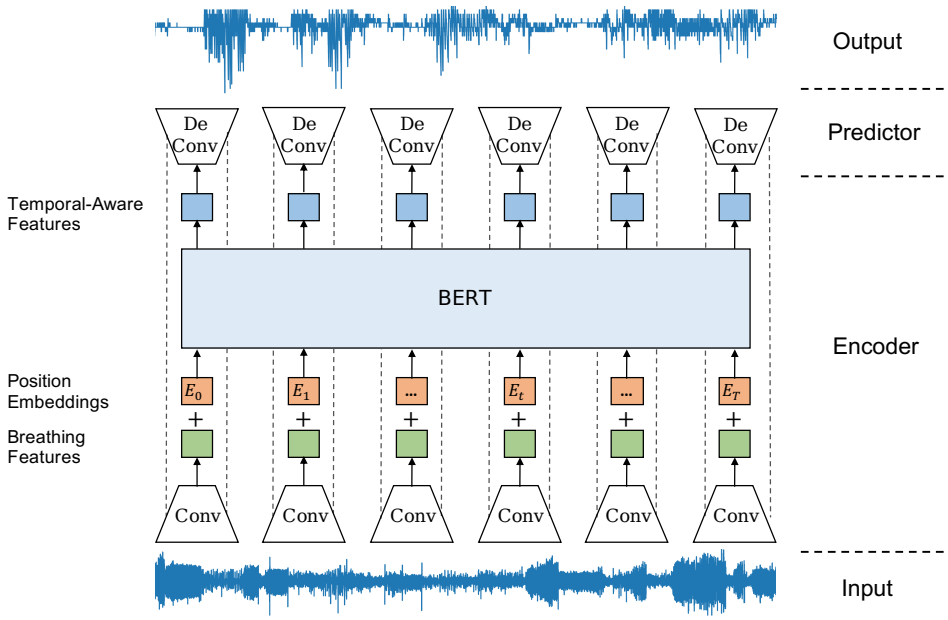


Figure 5: Network architecture for the backbone BERT-UNet model.

D Full Model: Gated BERT-UNet

In medical applications, there are useful side variables. Adapting to such variables will likely improve performance and make the results more personalized. In many cases, the relevant variables are binary or categorical. For example, the relevant variables for oxygen saturation include gender, whether the person is a smoker, whether they have asthma, etc. The categorical nature of these variables induces **discontinuity** in the learned function over the physiological indices. Specifically, consider oxygen saturation $y = g_s(x)$ as a function of respiration x and gender s (0 for male and 1 for female). $g_0(x)$ and $g_1(x)$ can be two different functions since men and women have differences in their oxygen transport systems [36].

To handle such discontinuity for better leveraging the categorical variables, we propose *Gated BERT-UNet*, a new model that augments BERT-UNet with multiple predictive heads. It selects the most suitable head for a person via a gate controlled by the subject's categorical indices. The model supports both variables available at the time of inference (e.g., gender), as well as dense categorical variables concurrently learned from the input signals (e.g., sleep stages).

Figure 1(b) illustrates the model. It has a gate function $G(v, u) : \mathcal{V} \times \mathcal{U} \to \{1, 2, \cdots, N\}$ where $v \in \mathcal{V}$ and $u \in \mathcal{U}$ are accessible/inaccessible variables, N is the number of gate statuses. We use the term *accessible variable* for variables easily available during inference time, e.g., gender, and the term *inaccessible variable* for information that is not available during inference, but available during training, like a person's sleep stages. Inaccessible variables are typically dense time series (e.g., sleep stages). Their prediction are learned concurrently with the main task under full supervision. The construction of the gate function $G(v, u)$ is described in the next sub-section.

The model has N heads $\{F_i\}_{i=1}^N$ that adapt the prediction $\hat{y}_i = F_i(E(x))$ to the gate status. It also has an extra predictor F_u to infer inaccessible variables u. During testing time, based on the accessible variables v and estimated inaccessible variables $\hat{u}$, we evaluate the gate status $s = G(v, \hat{u})$.

In the case of oxygen prediction, $\hat{y}_i$ and the gate status s are time series. As shown in Figure 1(b), the final prediction at each time step, is the gated combination of every head's output, i.e. $\forall t = 1, \ldots, f_o T, \hat{y}^t = \sum_{i=1}^N \mathbf{1}[s^t = i]\hat{y}_i^t$. We train Gated BERT-UNet (GBU) with the following loss,

$$\mathcal{L}_{\texttt{GBU}}(\hat{y}, \hat{u}, y, u) = \mathcal{L}(\hat{y}, y) + \frac{\lambda_u}{f_o T} \sum_{t=1}^{f_o T} \mathcal{L}_{\texttt{CE}}(\hat{u}^t, u^t), \tag{2}$$

where $\mathcal{L}$ is the main loss defined in Equation 1, $\mathcal{L}_{\texttt{CE}}$ is the cross-entropy loss to train the branch for predicting inaccessible variables and λ_u is a balancing hyperparameter.

Mapping Variables to Heads. The number of heads in a Gated BERT-UNet model puts an upper bound on the number of possible gate states. For example, if the model has 6 heads, the gate can take only 1 of 6 states. Typically, we have many more variable states than gate states. To find a proper mapping $G(v, u)$ from variable state to gate state, we rely on gradient similarity. For example, if we want to check whether male smokers should be in the same group as female smokers, we take a pretrained backbone BERT-UNet and compute its averaged gradient (w.r.t the loss function) over all male smokers and all female smokers in the dataset. Then we check the cosine similarity between the two gradients. If the gradients are similar, which means the two categories move the loss function in the same direction, we can use the same predictor for them. On the other hand, if the gradients are vastly different, it is preferable to separate such categories and assign them to different gate statuses.

E Datasets and Metrics

Medical Datasets. We leverage three public medical datasets: Sleep Heart Health Study (*SHHS*) [51], Multi-Ethnic Study of Atherosclerosis (*MESA*) [7], and Osteoporotic Fractures in Men Study (*MrOS*) [5]. The datasets were collected during sleep studies. For each subject, they include the respiration signals throughout the night along with the corresponding blood oxygen time series. The breathing signals are collected using a breathing belt around the chest or abdomen, and the oxygen is measured using a pulse oximeter. The datasets also contain side variables including sleep stages, which for every time instance assign to the subject one of the following: Awake, Rapid Eye Movement (REM) or non-REM stage.

We note that the subjects in these studies have an age range between 40 and 95, and some of them suffer from a variety of diseases such as chronic bronchitis, cardiovascular diseases, and diabetes. This allows for a wider range of oxygen variability beyond the typical range of healthy individuals.

RF Datasets. We collected a dataset of RF signals paired with SpO2 measurements. The data is collected from two sleep labs. In total, there are 400 hours of data from 49 overnight recordings of 32 subjects. The dataset contains subjects of different genders and races. Some subjects are healthy volunteers while others are patients with sleep problems. Thus, the ground truth oxygen saturation distribution is wider than normal ranges. A radio device is installed in the room to collect RF signals, as shown in Fig. 6. The radio signals are synchronized with the SpO2 measurements and processed with the algorithm in [50] to extract the person's breathing signals.

Figure 6: The radio device to collect RF signals.

Metrics. Let $\hat{y}$, y denote the predicted and ground truth oxygen saturation. Following the previous work [27], we use three standard metrics for evaluation: (1) Correlation: $\frac{\sum_t (y^t - \mu_y)(\hat{y}^t - \mu_{\hat{y}})}{\sqrt{\sum_t (y^t - \mu_y)^2 \sum_t (\hat{y}^t - \mu_{\hat{y}})^2}}$, where μ_y and $\mu_{\hat{y}}$ are the averaged oxygen saturations; (2) MAE: $\frac{1}{T}\sum_{t=1}^{T} |y^t - \hat{y}^t|$; and (3) RMSE: $\sqrt{\frac{1}{T}\sum_{t=1}^{T}(y^t - \hat{y}^t)^2}$. The models take a night of breathing as input and predict the corresponding oxygen levels. During the evaluation, we divide the model's prediction and ground truth oxygen into 240-second segments to compute the metrics. We note that those metrics are sensitive to the segment's length. For a fair comparison, we follow [27] and use 240-second intervals.

F Baselines.

We compare the following neural network architectures for our backbone model: (a) *CNN* is a fully convolutional model composed of eight 1-D convolutional layers and seven 1-D deconvolutional layers; (b) *CNN-RNN* augments the CNN model with a recurrent unit (one layer LSTM) in the bottleneck to better captures the long-term temporal relationships of the data. (c) *BERT-UNet* further makes two improvements on the CNN-RNN model. First, it uses an attention module to replace recurrent unit for temporal modelling. Second, it adds skip links between encoding convolutional layers and decoding deconvolutional layers at the same temporal scale to better capture the signal's local information.

To evaluate our design for incorporating side information, we compare the following models: (a) *BERT-UNet + VarAug*, which uses BERT-UNet as backbone and takes accessible variables as extra inputs and inaccessible variables as auxiliary tasks. (b) Our *Gated BERT-UNet* model, which uses a multi-head model gated by physiological variables.

G Training and Evaluation Protocols

Train/Valid/Test Spilt. Due to the limited data amount in the RF dataset, the RF data is all hold out for testing. In our study, all models are trained on the medical datasets and evaluated on both the medical datasets and the RF dataset. Collectively, the medical datasets have about 48,000 hours of data from 5,765 subjects in total. We randomly split subjects, 70% for training and validation, and 30% for testing, and fix the splits in all experiments. We train each model on the union of the training sets from the three medical datasets, and tested on each test set.

Side Variable Specifications. For models that incorporate side variables, we use *gender* as the accessible variable and *sleep stages* as the inaccessible variable. In Gated BERT-UNet, we use gradient similarity to map gate status as described in the method section, which results in the following 6 categories: (male, awake), (male, REM sleep), (male, non-REM sleep), (female, awake), (female, REM sleep), (female, non-REM sleep). The sleep stages themselves are learned from the input since they are an inaccessible variable. In the baseline VarAug, the gender variable is provided as an additional input and the sleep stages are used as an auxiliary task in a multitask model.

Implementation Details. We implement all models in PyTorch. All experiments are carried out on a NVIDIA TITAN Xp GPU with 12 GB memory. The number of parameters for the Gated BERT-UNet model is 26,821,113 and the model size is 107.28MB. In the training process, we use the Adam optimizer with a learning rate of 2×10^{-4}, and train the model for 500 epochs. Due to the varying input length, we set the batch size for all models to 1 (i.e., one night of respiration signal and the corresponding oxygen time series).

H Results for Different Skin Colors

Since pulse oximeters rely on measuring light absorbance through the finger, they are known to be affected by skin color and tend to overestimate blood oxygen saturation in subjects with dark skin [14, 41]. A large study that looked at tens of thousands of white and black COVID patients found that the "reliance on pulse oximetry to triage patients and adjust supplemental oxygen levels may place black patients at increased risk for hypoxaemia" [41]. In contrast, breathing and RF signals have no intrinsic bias against skin color. Figure 7 shows the distributions of the oximetry-based ground-truth oxygen and the Gated BERT-UNet prediction for different races, for the union of all

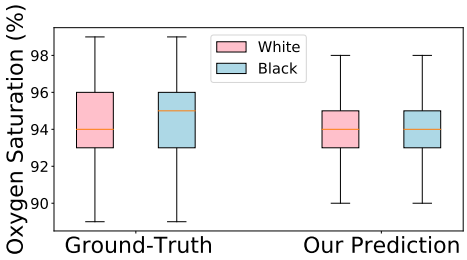


Figure 7: Box plot compares the ground-truth oxygen saturation measured by pulse oximeter and our predicted oxygen saturation for two races.

datasets. The ground-truth measurements from oximetry show a clear discrepancy between black and white subjects. In particular, black subjects have higher average blood oxygen. This is compatible with past findings that pulse oximeters overestimate blood oxygen in dark-skinned subjects [14]. In contrast, the breathing-based oxygen prediction corrects or reduces this bias, and shows more similar oxygen distributions for the two races.

I Extra Results

I.1 Example of Same Breathing Rate but Different Oxygen Level

One might wonder whether our model predicts oxygen merely by the breathing frequency. We show it is not the case. Our model understands that the same breathing frequency does not translate to the same oxygen level. Figure 8 shows two people breathing at the same frequency of 14 BPM, but the model correctly realizes that one of them has high and stable oxygen of 98 while the other has a relatively low oxygen level of 93. The model even follows the dynamic change in their oxygen levels from one second to the next. This is possible because the model analyzes the full details of the breathing signal and its dynamics, which is a dense 1D input, not just one variable like the rhythm. To see that, consider again the example in Figure 8. While the two people breathe at the same frequency, the person whose breathing is plotted in blue suddenly starts taking deeper breaths. This is an indication that the person is low on oxygen and is trying to increase his oxygen intake. When the person takes such deep breaths, the oxygen level increases. In contrast, the person whose oxygen is plotted in orange is not struggling with low oxygen and his breathing is steady.

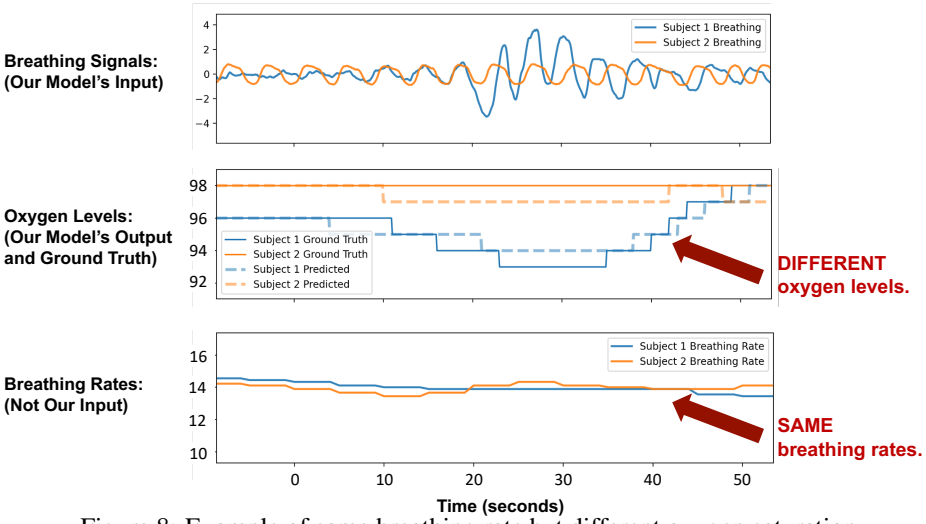


Figure 8: Example of same breathing rate but different oxygen saturation.

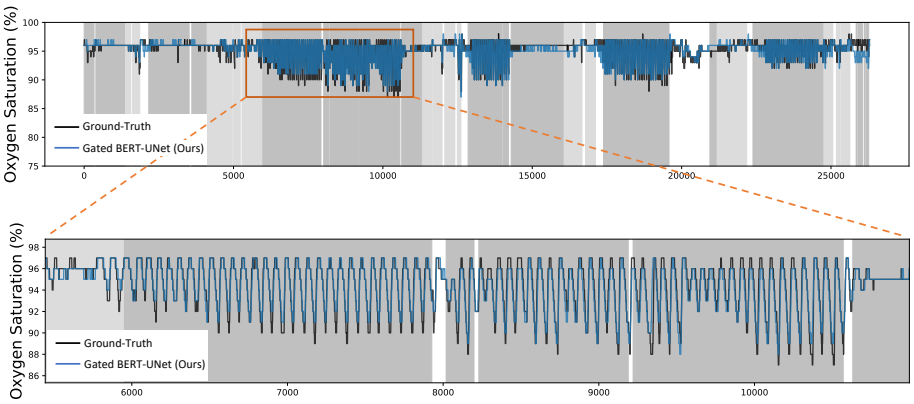


Figure 9: *MESA* Example 1.

I.2 Qualitative Results in Different Datasets

We include more visualizations of the breathing signals and the corresponding oxygen saturation predicted by the *Gated BERT-UNet* model for the medical datasets: MESA (Figure 9 and Figure 10), MrOS (Figure 11 and Figure 12). We also visualize the model's prediction on unhealthy subjects with various diseases including chronic obstructive pulmonary disease (COPD), asthma, diabetes. In the plots, The background color indicates different sleep stages. The 'dark grey', 'light grey' and 'white' correspond to 'Non-REM sleep', 'REM' and 'Awake', respectively.

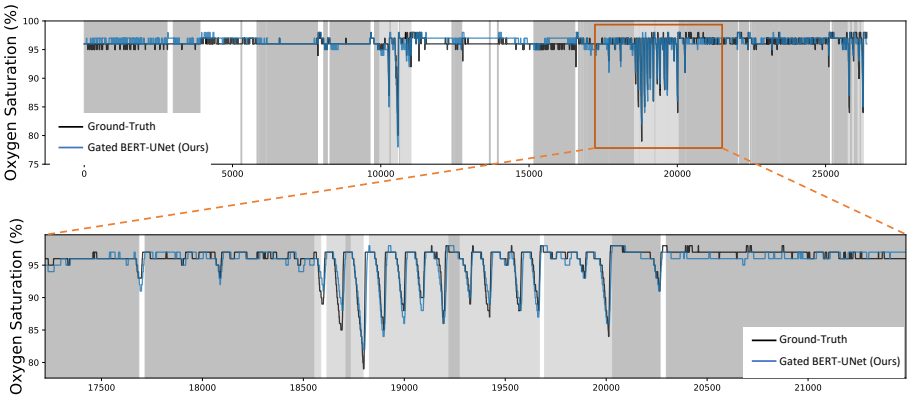


Figure 10: *MESA* Example 2.

MESA. The example in Figure 9 shows our model's ability to capture the fluctuations of oxygen saturation. At the same time, the example in Figure 10 shows that our model accurately detects the region of low oxygen saturation, which highlights its usefulness in monitoring patients.

MrOS. From the zoomed-in regions (a) and (b) in Figure 11, we see that the model exhibits a larger error when the ground truth SpO2 reading is very low. It is mainly caused by the imbalanced labels in the training set since subjects usually experience much less time of having low oxygen level (e.g., below 90%) than having a normal oxygen level between 94% to 100%. Figure 12 shows another example of the dynamics. As shown in the zoomed range, oxygen fluctuations correlate with one's sleep stage: in 'REM' (colored by light gray), the oxygen fluctuates drastically while in 'Non-REM sleep' (colored by dark gray), the oxygen is much more stable and fluctuates in a small range. Our model makes accurate predictions since it leverages the sleep stage information.

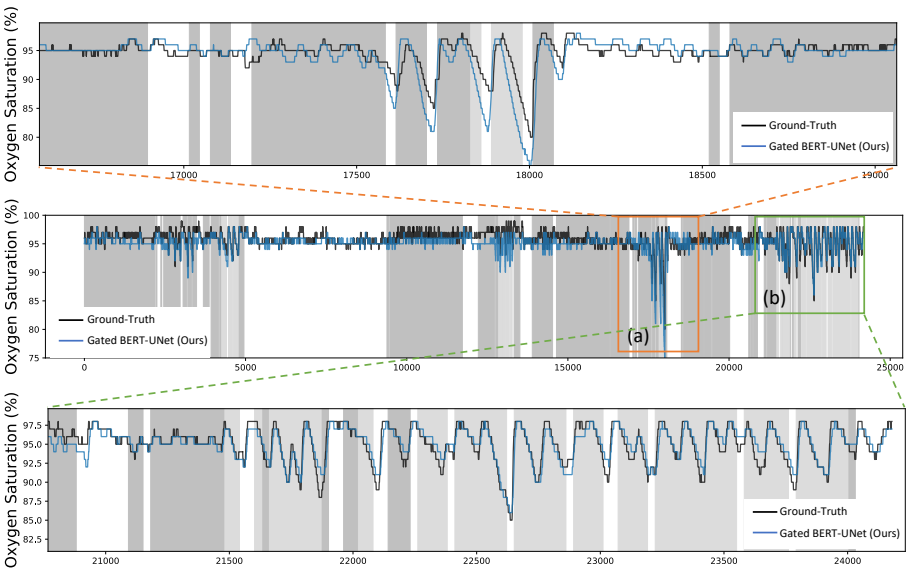


Figure 11: *MrOS* Example 1.

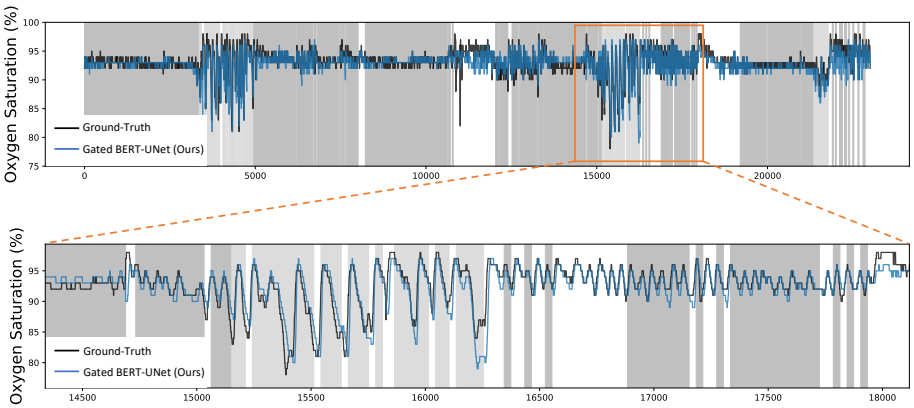


Figure 12: *MrOS* Example 2.

I.3 Results for Relevant Diseases

As we mentioned earlier, using breathing as the input to the neural network model allows us to both train and test on large respiration dataset from past sleep studies. Since these datasets contain diverse people with a variety of diseases, it makes it possible to check how the model generalizes to unhealthy individuals. Particularly we are interested in diseases that interact with oxygen saturation including pulmonary diseases such as chronic obstructive pulmonary disease (COPD), chronic bronchitis, asthma, emphysema, and others like diabetes and coronary heart disease.

Diabetes is a disease in which the patient's blood sugar levels are too high. Research has shown that diabetes is a risk factor for severe nocturnal Hypoxemia in obese patients [22]. Further diabetic patients tend to have 3% to 10% lower lung volumes than adults without the disease. Figure 13 shows an example of a diabetic patient who has an oxygen level that keeps oscillating between 85% and 95%. From the zoomed-in region, we can see our model captures the oscillating oxygen dynamics and accurately predicts the oxygen values.

Chronic obstructive pulmonary disease (COPD) refers to a chronic inflammatory lung disease that obstructs airflow from the lungs. Severe COPD can cause hypoxia, an extremely low oxygen level. Figure 14 shows an example of a COPD patient who experiences several oxygen droppings during the REM period (indicated by the orange box). As we can see, our model successfully predicts the events of oxygen dropping.

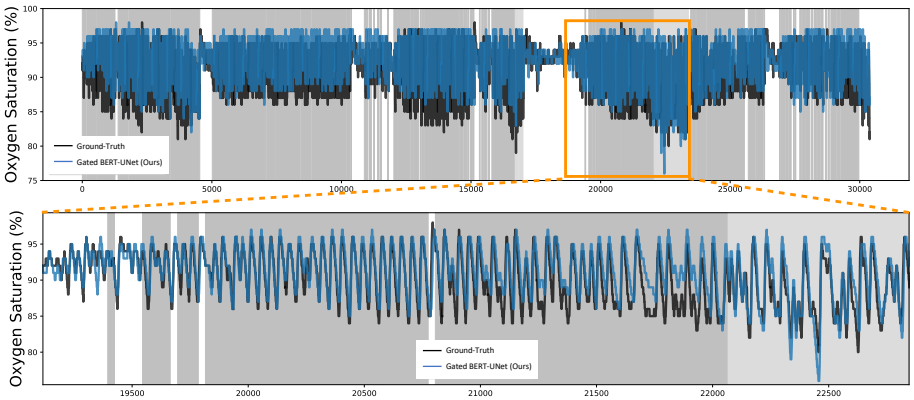


Figure 13: Diabetes Patient Example. The first row shows the full-night oxygen saturation, while the second row zooms into the first row's orange box region. The black and blue curves are the ground-truth oxygen and our prediction. The background color indicates the subject's sleep stages. The dark grey, light grey and white corresponds to Sleep, REM and Awake respectively.

Chronic bronchitis refers to long-term inflammation of the bronchi. Chronic bronchitis patients can have shortness of breath which affects oxygen levels. Figure 15 is an example of a chronic bronchitis patient whose oxygen level keeps osculating between normal and low during the night. Our model captures the trend well.

Asthma is a condition in which a person's airways narrow, swell, and produce extra mucus. An asthma patient's oxygen levels can be irregular due to the breathing difficulty caused by the disease. As shown in the example in Figure 16, the patient has a normal oxygen level for most of the time, but the oxygen occasionally drops to low levels. Our model works well on detecting those abnormal oxygen levels from the person's respiration.

Emphysema refers to a lung condition in which the air sacs in the person's lung are damaged. Patients with emphysema usually have breathing issues that affect oxygen saturation. Figure 17 shows an example of an emphysema patient. The patient experiences a long period of low oxygen during sleep (as highlighted by the orange box). Our model successfully predicts such unusual oxygen dynamics.

Coronary heart disease develops when the arteries of the heart are too narrow to deliver enough oxygen-rich blood to the heart. A deficiency in providing oxygen-rich blood can lead one's oxygen saturation to deviate from normal level. Figure 18 is an example of a person with coronary heart disease who experiences several severe oxygen drops during sleep. Our model accurately tracks their oxygen level and detects the oxygen reduction events.

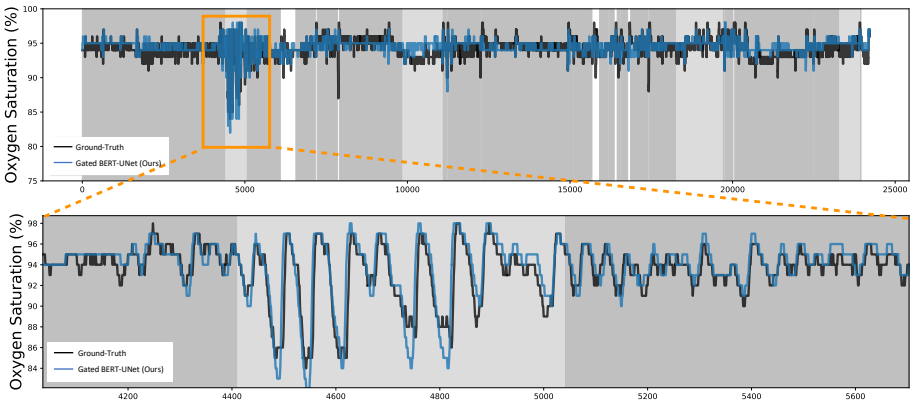


Figure 14: Chronic Obstructive Pulmonary Disease (COPD) Patient Example. The first row shows the full-night oxygen level, while the second row zooms into the first row's orange box region. The black and blue curves are the ground-truth oxygen and our prediction. The background color indicates the subject's sleep stages. The dark grey, light grey and white means Sleep, REM and Awake.

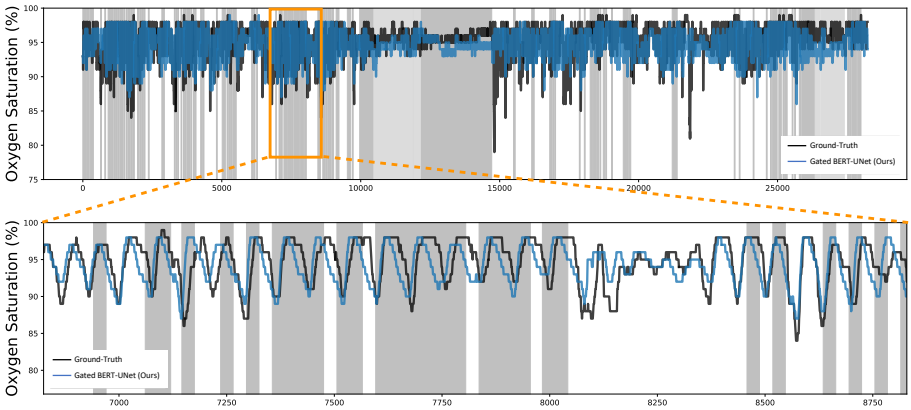


Figure 15: Chronic Bronchitis Patient Example. The first row shows the full-night oxygen saturation, while the second row is a zoomed-in visualization of the orange box region in the first row. The black and blue curves are the ground-truth oxygen and our prediction. The background color indicates the subject's sleep stages. The dark grey, light grey and white means Sleep, REM and Awake.

J Concluding Remarks

This paper introduces the new task of inferring oxygen saturation from radio signals. It develops a new gated transformer architecture to deliver this application and adapt deep models to auxiliary categorical variables. We note the work have some limitations. First, the paper focuses on special use cases (e.g., at-home oxygen monitoring in very old adults or during sleep), but is not suitable for other use cases (e.g., measuring oxygen to optimize performance during exercise). Second, the results in the paper provide an initial proof of concept that the shape and dynamics of the breathing signal include sufficient clues to infer a useful estimate of oxygen saturation. However, before this system can be used in clinical care, one needs clinical studies to quantify the performance for different disease conditions.

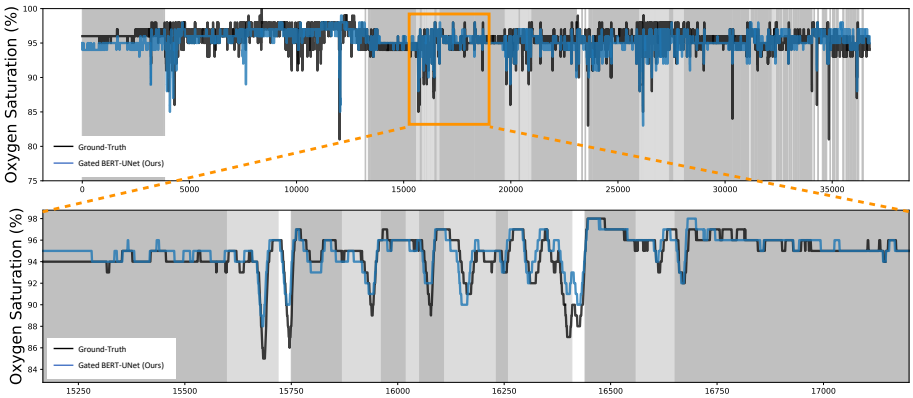


Figure 16: Asthma Patient Example. The first row shows the full-night oxygen saturation, while the second row is a zoomed-in visualization of the orange box region in the first row. The black and blue curves are the ground-truth oxygen and our prediction. The background color indicates the subject's sleep stages. The dark grey, light grey and white means Sleep, REM and Awake.

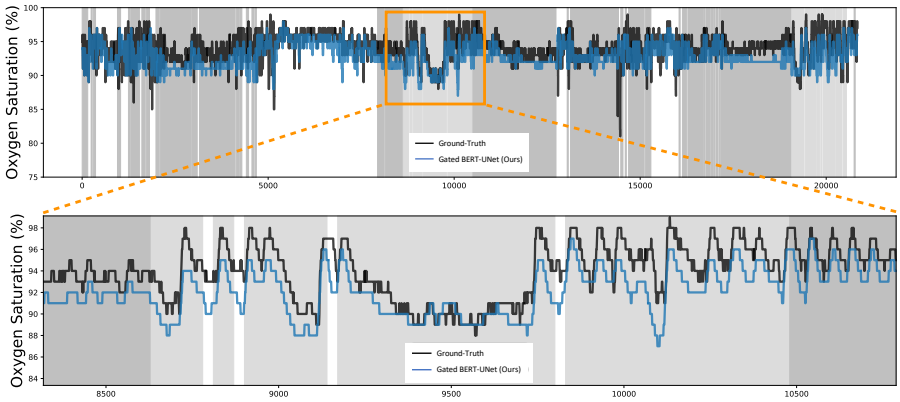


Figure 17: Emphysema Patient Example. The first row shows the full-night oxygen saturation, while the second row is a zoomed-in visualization of the orange box region in the first row. The black and blue curves are the ground-truth oxygen and our prediction. The background color indicates the subject's sleep stages. The dark grey, light grey and white means Sleep, REM and Awake.

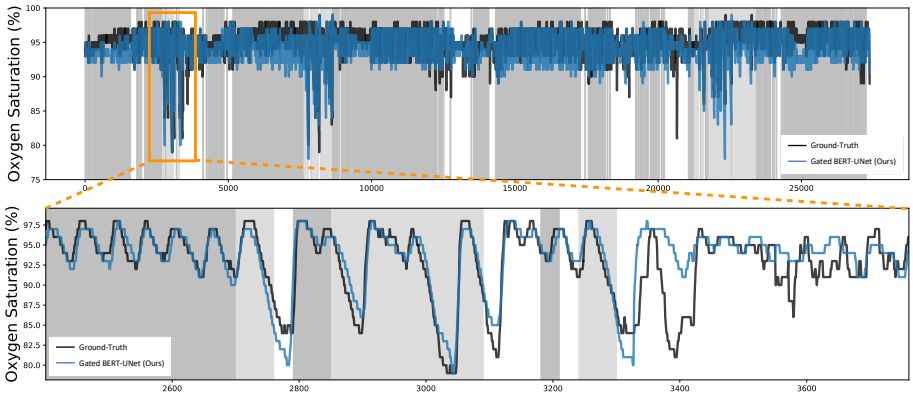


Figure 18: Coronary Heart Disease Patient Example. The first row is the full-night oxygen saturation prediction results, where the black curve indicates the ground-truth oxygen saturation and the blue curve indicates our predicted oxygen saturation. The second row is the zoom-in version of the orange box region in the first row. The background color indicates sleep stages. The dark grey, light grey and white corresponds to non-REM Sleep, REM and Awake, respectively.