Sleep and Activity Prediction for Type 2 Diabetes Management Using Continuous Glucose Monitoring

Kimmo Karkkainen[1,2*] **Gregory Lyng**[1] **Brian L. Hill**[1]
Kailas Vodrahalli[1,3*] **Jeffrey Hertzberg**[1] **Eran Halperin**[1]
[1]Optum Labs, MN, USA [2]UCLA, CA, USA [3]Stanford University, CA, USA
{kimmo.karkkainen,gregory.lyng,brian.l.hill}@optum.com
{kailas.vodrahalli,jeffrey.hertzberg,eran.halperin}@optum.com

Abstract

Continuous glucose monitors (CGMs) generate frequent glucose measurements, and numerous studies suggest that these devices may improve diabetes management. These devices give people with diabetes visibility into how lifestyle factors, i.e., meals, physical activity, sleep, stress, and medication adherence, impact their glucose levels. While earlier studies have shown that individual's actions can influence their CGM data, it has not been clear whether CGM data can provide information about these actions. This is the first study to show on a large cohort that CGM can provide information about sleep and physical activities (as aggregated from an activity tracker). We first train a neural network model to determine the sequence of daily activities from CGM signals, and then extend the model to use additional data, such as individual demographics and medical claims history. Using data from 6,981 participants in a Type 2 diabetes (T2D) management program, we show that a model combining an individual's CGM, demographics, and claims data is highly predictive of sleep (AUROC 0.947, AUPRC 0.884), and moderately predictive of physical activity or certain indicators of physical activity (AUROCs/AUPRCs up to 0.817/0.401), as inferred by an activity tracker. These results show that CGM may have wider utility in diabetes management than previously known.

1 Introduction

As of 2021, an estimated 536.6 million people worldwide had diabetes, and the number is expected to increase [1]. Among people with diabetes, as many as 90-95% have T2D [2], which appears to be caused by complex interactions between genetic, environmental, and lifestyle factors. Daily actions, such as physical exercise, healthy eating, and sufficient sleep, are important factors in both preventing or delaying the onset of T2D as well as its management.

In this paper, we demonstrate the strong connection between daily activities and continuous glucose monitor (CGM) output by showing that the sequence of daily activities (sleep, walking, and elevated heart rate, as aggregated from an activity tracker) can be predicted directly from the CGM signal. In addition, this can be used to analyze CGM and activity jointly to better understand their relationship and effects on health outcomes. We then develop the model further to include demographic information as well as medical claims data, and we show that this additional information further improves the prediction results.

There have been several studies using CGM data in predictive models. The most common use of this data is to predict if a person with type 1 diabetes will develop hypoglycemia [3]–[7] or hyperglycemia

*This work was conducted while the authors were doing an internship at Optum Labs.

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

[7]. In addition, CGM data has been used to determine additional insulin dose requirement at mealtimes [8]. A few studies have also used CGM data to determine people's daily actions. For example, CGM data has been used to determine a person's adherence to their daily insulin injection [9], and other studies have detected eating activities [10][11]. To the best of our knowledge, there has been only one study using CGM data to predict physical activity [12]. However, that study was limited to 11 people with type 1 diabetes, and its generalizability is therefore limited.

2 Methods

2.1 Dataset and Preprocessing

Our dataset consisted of de-identified CGM and fitness tracker data collected from participants in a commercial diabetes care program. Participants in this program had type 2 diabetes at the time of program enrollment. All participants were given the same CGM device and fitness tracker. As some participants did not wear the devices consistently, we ensured data quality by only using data for calendar days where the participant had at least 22 hours of overlapping CGM and fitness tracker data. In addition, we required them to have at least 1 hour of sleep during that day to ensure the fitness tracker was collecting data accurately. We used data for all participants who had at least one day of data. The resulting dataset consisted of 6,981 participants with an average of 65.1 days of data (median 28.0, standard deviation (SD) 85.6). The mean age of participants was 54.3 years (SD 9.1), distribution of female/male was 51.3%/48.7%, and the mean body mass index (BMI) was 34.8 (SD 8.6).

Glucose level was measured using a CGM device which provided interstitial glucose measurements every 5 minutes, and physical activity and sleep were tracked using a fitness watch which reported average heart rate and step count once a minute, as well as the intervals when the participant was determined to be asleep based on their heart rate and accelerometer data. Due to the difference in time sampling frequencies, the fitness tracker data was aggregated for each 5-minute time block to match the CGM. Heart rates were categorized using the fitness tracker's personalized heart rate zones into resting, fat burn, and cardio zones. Walking activity was binarized using a threshold of 400 steps per five minutes, which is slightly below a typical moderate-intensity walking cadence [13]. To reduce noise, we filtered out individual positive values from the labels, thereby requiring each activity to last for at least 10 minutes at a time. After these preprocessing steps, the average prevalence of sleep activity was 30.7%, fat burn zone was 21.0%, cardio zone was 0.8%, and walking was 0.6%.

In addition, we accounted for participants' medical history by means of their medical claims data. This data included the information necessary for the reimbursement of medical services including International Classification of Diseases (ICD10) diagnosis codes [14], and service dates. In addition, we grouped similar ICD codes together by only considering the first 3 letters/numbers of each code. On average, participants had 188.2 (SD 178.2) ICD codes in their claims data, with 40.7 (SD 23.8) unique codes.

2.2 Model

Our model was based on the U-Net architecture (see Figure 1) [15], which was originally designed for medical image segmentation. Modified versions of the U-net architecture have been shown to be effective for a wide variety of tasks, such as volumetric medical image segmentation [16], [17], image denoising [18]–[21], audio source separation [22]–[24], speech enhancement [25], [26], and physiological signal imputation [27]–[30].

As our goal was to transform a sequence of glucose measurements to a sequence of daily activities, we modified the original model to use 1-dimensional convolutions instead of 2-dimensional convolutions. The input matrix $\mathbf{X}$ was a sequence of 288 glucose measurements (24 hours * 12 measurements/hour) along with the rate of change computed by the CGM device. These two values were combined as separate channels, i.e., $\mathbf{X} \in \mathbb{R}^{2\times 288}$. Prediction targets included four activities: sleeping, walking, elevated heart rate (Fat burn zone), and high heart rate (Cardio zone). These values were measured at the same time intervals as the glucose values and treated as separate channels, therefore the target matrix $\mathbf{Y} \in \mathbb{R}^{4\times 288}$. The prediction was treated as a multi-label classification task, i.e., multiple target values were allowed to be positive simultaneously (e.g., walking and high heart rate).

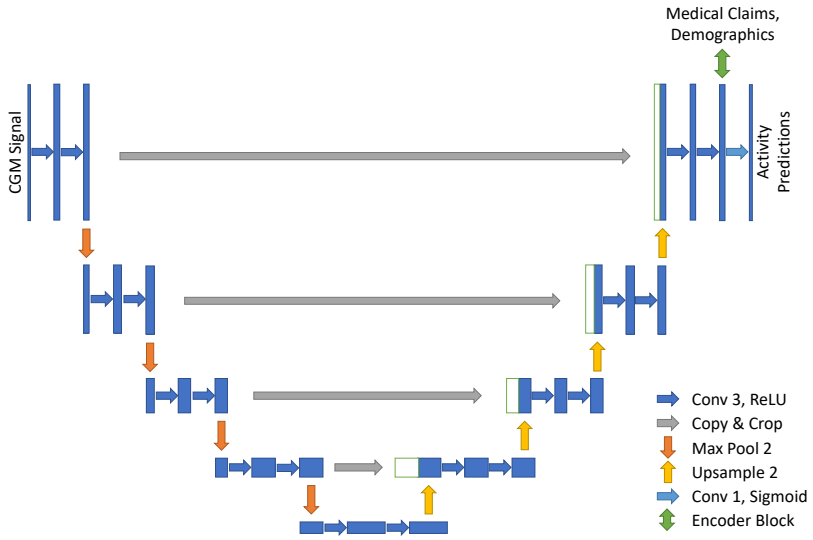

Figure 1: Modified U-Net architecture with an encoder for demographics and medical claims.

Next, to incorporate claims data into the model, we converted the ICD codes that occured prior to the current time window into vectors $\mathbf{x_{icd}} \in \mathbb{R}^{300}$. This conversion was done using pretrained ICD2Vec embeddings [31]. As our hypothesis was that some ICD codes might be more relevant for the prediction task than others, we wanted the model to be able to learn which ICD codes to use. In addition, we wanted the model to be able to distinguish between acute and chronic diseases, as acute diseases might be relevant only for a few days or weeks while chronic ones could have effects across many years. Including ICD codes into the model was implemented using a multi-head attention layer [32], as shown in Figure 2. Time was first converted into larger time blocks to avoid overly sparse time values using conversion $f(t) = \lfloor log(t) \rfloor$, where t is the number of days between the current time window and the claim. This time block was then converted into a vector $\mathbf{t} \in \mathbb{R}^{50}$ using step encoding shown in [32] and concatenated to the ICD vector $\mathbf{x_{icd}}$. The sequence of combined ICD and time vectors was used as both the key $\mathbf{k}$ and value $\mathbf{v}$ for the attention layer while a constant vector was used as the query $\mathbf{q}$.

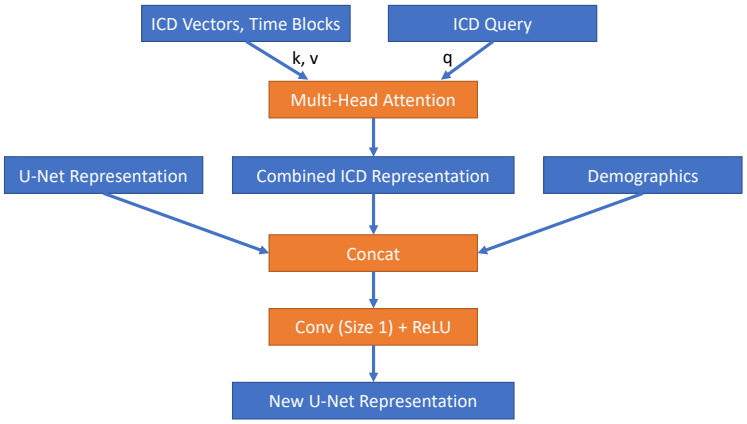

Figure 2: Encoder block which combines ICD vectors using Multi-Head Attention and adds the combined vector along with demographics to U-Net's internal representation.

We evaluated including the combined ICD representations in three locations within the U-Net model: in the input layer, between encoder and decoder, as well as before the last convolutional layer. These vectors were concatenated with the U-Net's internal representation and passed through a convolution with size 1 and a rectified linear unit (ReLU) layer. The resulting vector was then returned to the

same location of the U-Net model where the original vector was. Demographics (age, gender, height, weight, BMI) were standardized and combined with the internal representations the same way.

2.3 Training

Participants were randomly divided 60%-20%-20% to training, validation, and test sets respectively, such that all data for one participant belonged to only one set. Models were trained using AdamW [33] algorithm for 35 epochs, which was enough to reach convergence. In each training epoch, a random day of data was chosen for each participant such that each participant appeared exactly once in each epoch, but different epochs might have used different time windows for participants. In validation and test phases, one day of data was randomly chosen for each participant but this same day was used in every epoch and experiment. This ensured that the results were consistent across different experiments and each participant had an equal impact on the results.

3 Experiments and Results

We evaluated our model with and without using demographics and ICD codes. We also compared the predictive performance when changing the location where demographics and ICD codes were incorporated into the model. Results of these ablation experiments are shown in Table 1. As the results show, including ICD codes and demographics improved heart rate zone and walk predictions. Overall, including demographics and ICD codes later in the model provided better results than including them early or between the encoder and decoder. Sleep predictions were not affected by demographics or ICD codes. Due to the significant class imbalance, we also evaluated area under precision-recall curve (AUPRC) scores, which show a similar pattern. The highest AUPRC score for sleep prediction was 0.884 (2.9 times higher than prevalence), for fat burn zone prediction 0.401 (1.9 times higher than prevalence), for cardio zone prediction 0.026 (3.0 times higher than prevalence), and for walk prediction 0.052 (8.7 times higher than prevalence).

Table 1: Comparison of AUROC scores using demographics and ICD codes at different parts of the model. The first row shows results without demographics or ICD codes, the following rows show results when either ICD codes or demographics are added at different locations (early, middle, late), and the last row shows results when both demographics and ICD codes are added into the model.

Demographics	ICD Codes	Sleep AUROC	HR Fat Burn AUROC	HR Cardio AUROC	Walk AUROC
-	-	0.946 ($\pm$0.001)	0.708 ($\pm$0.001)	0.727 ($\pm$0.004)	0.776 ($\pm$0.003)
-	Early	0.936 ($\pm$0.001)	0.687 ($\pm$0.001)	0.707 ($\pm$0.003)	0.770 ($\pm$0.004)
-	Middle	0.945 ($\pm$0.001)	0.704 ($\pm$0.001)	0.715 ($\pm$0.004)	0.790 ($\pm$0.003)
-	Late	0.945 ($\pm$0.001)	0.719 ($\pm$0.001)	0.744 ($\pm$0.004)	0.804 ($\pm$0.003)
Early	-	0.942 ($\pm$0.001)	0.669 ($\pm$0.001)	0.740 ($\pm$0.004)	0.809 ($\pm$0.002)
Middle	-	0.945 ($\pm$0.001)	0.716 ($\pm$0.002)	0.758 ($\pm$0.006)	0.797 ($\pm$0.003)
Late	-	0.946 ($\pm$0.001)	0.716 ($\pm$0.001)	0.752 ($\pm$0.004)	**0.820** ($\pm$0.003)
Late	Late	**0.947** ($\pm$0.001)	**0.722** ($\pm$0.001)	**0.768** ($\pm$0.005)	0.817 ($\pm$0.003)

Figure 3 shows an example of predictions made for an individual participant. The model was typically very confident in the sleep predictions except in the beginning and end of the sleep period. The lower confidence in both ends can be partially explained by the noise in the ground truth data, as the sleep times were estimated by the fitness tracker based on movement and heart rate patterns. Walking probabilities were skewed toward zero because of the rarity of the event, but despite the miscalibration the model did give higher probabilities when the participant was truly walking. However, the probability sometimes increased slightly at other times also, possibly when the participant was doing other physical activities which had a similar impact on the glucose measurements as walking. Similarly, heart rate predictions often had higher values when the heart rate was elevated, but the predictions were slightly noisy.

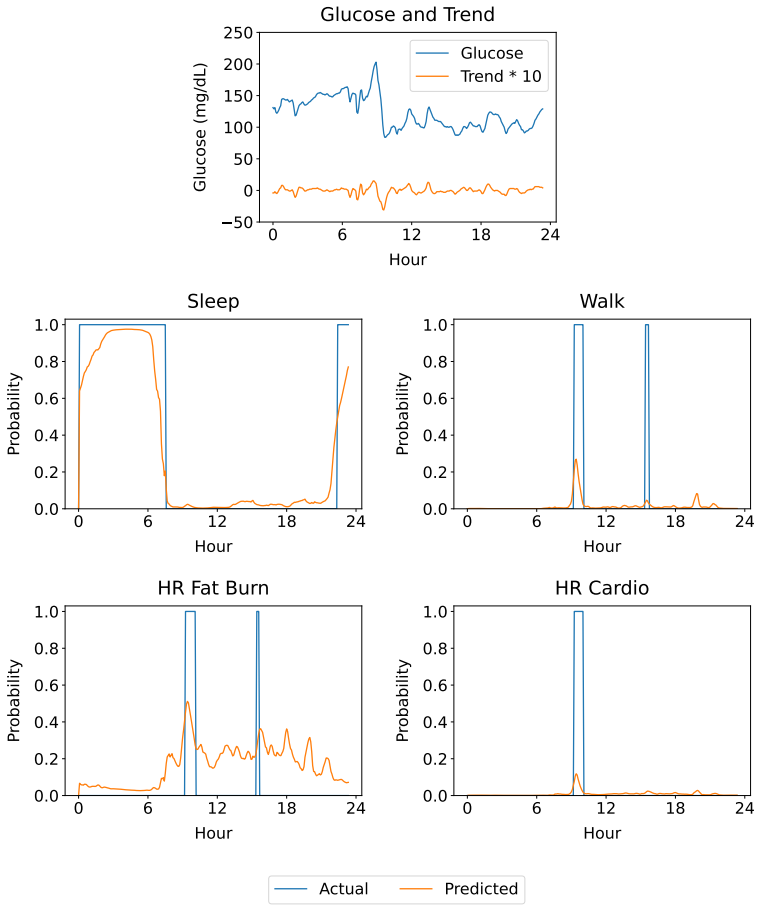

Figure 3: One participant's CGM signal and predictions for one day. X-axis shows hours starting from midnight.

4 Conclusion and Future Work

In this paper, we have shown for the first time using a large cohort of people with T2D that CGM data is highly predictive of sleep and moderately predictive of physical activity or certain indicators of physical activity, as inferred by an activity tracker. To do so, we developed a neural network architecture for predicting daily activities using CGM, demographics, and claims data. We first modified the U-Net architecture so that it can be used with 1-dimensional CGM signals, and then incorporated claims data using an attention layer which allowed the model to learn which claims were important for the task. A limitation of our study is that the data used for ground truth is collected by an activity tracker as opposed to the gold standard measurements, such as polysomnography. Since activity tracker data may sometimes be an inaccurate predictor of sleep, our model may capture different information than sleep, and rather capture what is perceived by the activity tracker as sleep. Thus, comparing CGM to a gold standard sleep activity measurement remains an important direction for future work. Our work should also be taken into account in the context of T2D patients, and future work is needed to understand whether the results are generalizable to the entire population.

In our future work, we will investigate using alternative approaches to determine physical activity, such as combining heart rate and walking into a more informative label, as neither of these labels is a perfect measure of physical activity on its own. For example, walking is not the only type of physical activity for many people and it might have a similar CGM pattern as other physical activities. On the other hand, estimated heart rate zones can be imprecise for some of the participants. We expect these changes to improve our physical activity predictions even further. Our results open new possibilities for the utilization of CGM devices for diabetes management and allow us to gain a better understanding of CGM signal during activities.

References

[1] H. Sun, P. Saeedi, S. Karuranga, *et al.*, "IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes Research and Clinical Practice*, vol. 183, p. 109 119, 2022, ISSN: 0168-8227. DOI: https://doi.org/10.1016/j.diabres.2021.109119.

[2] Centers for Disease Control and Prevention, *Type 2 diabetes*, Dec. 2021. [Online]. Available: https://www.cdc.gov/diabetes/basics/type2.html (visited on 09/20/2022).

[3] Y. Jian, M. Pasquier, A. Sagahyroon, and F. Aloul, "A machine learning approach to predicting diabetes complications," *Healthcare*, vol. 9, no. 12, p. 1712, Dec. 2021, ISSN: 2227-9032. DOI: 10.3390/healthcare9121712.

[4] D. Dave, D. J. DeSalvo, B. Haridas, *et al.*, "Feature-based machine learning model for real-time hypoglycemia prediction," *Journal of Diabetes Science and Technology*, vol. 15, no. 4, pp. 842–855, Jul. 2021, ISSN: 1932-2968, 1932-2968. DOI: 10.1177/1932296820922622.

[5] W. Seo, Y.-B. Lee, S. Lee, S.-M. Jin, and S.-M. Park, "A machine-learning approach to predict postprandial hypoglycemia," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, p. 210, Dec. 2019, ISSN: 1472-6947. DOI: 10.1186/s12911-019-0943-4.

[6] V. B. Berikov, O. A. Kutnenko, J. F. Semenova, and V. V. Klimontov, "Machine learning models for nocturnal hypoglycemia prediction in hospitalized patients with type 1 diabetes," *Journal of Personalized Medicine*, vol. 12, no. 8, 2022, ISSN: 2075-4426. DOI: 10.3390/jpm12081262.

[7] Y. Marcus, R. Eldor, M. Yaron, *et al.*, "Improving blood glucose level predictability using machine learning," *Diabetes/Metabolism Research and Reviews*, vol. 36, no. 8, e3348, 2020. DOI: https://doi.org/10.1002/dmrr.3348.

[8] G. Noaro, G. Cappon, M. Vettoretti, G. Sparacino, S. D. Favero, and A. Facchinetti, "Machine-learning based model to improve insulin bolus calculation in type 1 diabetes therapy," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 247–255, 2021. DOI: 10.1109/TBME.2020.3004031.

[9] D. N. Thyde, A. Mohebbi, H. Bengtsson, M. L. Jensen, and M. Mørup, "Machine learning-based adherence detection of type 2 diabetes patients on once-daily basal insulin injections," *Journal of Diabetes Science and Technology*, vol. 15, no. 1, pp. 98–108, 2021, PMID: 32297804. DOI: 10.1177/1932296820912411.

[10] V. Palacios, D. M.-K. Woodbridge, and J. L. Fry, "Machine learning-based meal detection using continuous glucose monitoring on healthy participants: An objective measure of participant compliance to protocol," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021, pp. 7032–7035. DOI: 10.1109/EMBC46164.2021.9630408.

[11] L. Bertrand, N. Cleyet-Marrel, and Z. Liang, "The role of continuous glucose monitoring in automatic detection of eating activities," in *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, 2021, pp. 313–314. DOI: 10.1109/LifeTech52111.2021.9391849.

[12] M. R. Askari, M. Rashid, X. Sun, *et al.*, "Meal and physical activity detection from free-living data for discovering disturbance patterns of glucose levels in people with diabetes," *BioMedInformatics*, vol. 2, no. 2, pp. 297–317, 2022, ISSN: 2673-7426. DOI: 10.3390/biomedinformatics2020019.

[13] J. Slaght, M. Sénéchal, T. J. Hrubeniuk, A. Mayo, and D. R. Bouchard, "Walking cadence to exercise at moderate intensity for adults: A systematic review," *Journal of Sports Medicine*, vol. 2017, pp. 1–12, 2017, ISSN: 2356-7651, 2314-6176. DOI: 10.1155/2017/4641203.

[14] World Health Organization, *International Classification of Diseases (ICD)*. [Online]. Available: https://www.who.int/standards/classifications/classification-of-diseases (visited on 09/20/2022).

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[16] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, Springer, 2016, pp. 424–432.

[17] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, IEEE, 2016, pp. 565–571.

[18] B. Park, S. Yu, and J. Jeong, "Densely connected hierarchical network for image denoising," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019.

[19] W. Liu, Q. Yan, and Y. Zhao, "Densely self-guided wavelet network for image denoising," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 432–433.

[20] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level wavelet-cnn for image restoration," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 773–782.

[21] F. Jia, W. H. Wong, and T. Zeng, "Ddunet: Dense dense U-net with applications in image denoising," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 354–364.

[22] D. Stoller, S. Ewert, and S. Dixon, "Wave-u-net: A multi-scale neural network for end-to-end audio source separation," *ISMIR Conference*, 2018.

[23] A. Jansson, E. Humphrey, N. Montecchio, R. Bittner, A. Kumar, and T. Weyde, "Singing voice separation with deep U-net convolutional networks," *ISMIR Conference*, 2017.

[24] N. Takahashi, N. Goswami, and Y. Mitsufuji, "MMDenseLSTM: An efficient combination of convolutional and recurrent neural networks for audio source separation," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2018, pp. 106–110. DOI: 10.1109/IWAENC.2018.8521383.

[25] S. Pascual, A. Bonafonte, and J. Serrà, "Segan: Speech enhancement generative adversarial network," *Proc. Interspeech 2017*, pp. 3642–3646, 2017.

[26] H.-S. Choi, J.-H. Kim, J. Huh, A. Kim, J.-W. Ha, and K. Lee, "Phase-aware speech enhancement with deep complex u-net," in *International Conference on Learning Representations*, 2018.

[27] B. L. Hill, N. Rakocz, Á. Rudas, *et al.*, "Imputation of the continuous arterial line blood pressure waveform from non-invasive measurements using deep learning," *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.

[28] M. Chan, V. G. Ganti, and O. T. Inan, "Respiratory rate estimation using U-Net-based cascaded framework from electrocardiogram and seismocardiogram signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2481–2492, 2022. DOI: 10.1109/JBHI.2022.3144990.

[29] G. Cathelain, B. Rivet, S. Achard, J. Bergounioux, and F. Jouen, "U-net neural network for heartbeat detection in ballistocardiography," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2020, pp. 465–468.

[30] F. Bousefsaf, D. Djeldjli, Y. Ouzar, C. Maaoui, and A. Pruski, "iPPG 2 cPPG: Reconstructing contact from imaging photoplethysmographic signals using U-Net architectures," *Computers in Biology and Medicine*, vol. 138, p. 104 860, 2021.

[31] Y. C. Lee, S.-H. Jung, A. Kumar, *et al.*, *ICD2Vec: Mathematical representation of diseases*. Jul. 2021. DOI: 10.21203/rs.3.rs-692012/v1.

[32] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[33] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019.