Personalized Graph Attention Network for Multivariate Time-series Change Analysis: A Case Study on Long-term Maternal Monitoring

Yuning Wang
University of Turku
Turku, Finland
yuning.y.wang@utu.fi

Iman Azimi
University of California, Irvine
California, USA

Mohammad Feli
University of Turku
Turku, Finland

Amir M. Rahmani
University of California, Irvine
California, USA

Pasi Liljeberg
University of Turku
Turku, Finland

ABSTRACT

Internet-of-Things-based systems have recently emerged, enabling long-term health monitoring systems for the daily activities of individuals. The data collected from such systems are multivariate and longitudinal, which call for tailored analysis techniques to extract the trends and abnormalities in the monitoring. Different methods in the literature have been proposed to identify trends in data. However, they do not include the time dependency and cannot distinguish changes in long-term health data. Moreover, their evaluations are limited to lab settings or short-term analysis. Long-term health monitoring applications require a modeling technique to merge the multisensory data into a meaningful indicator. In this paper, we propose a personalized neural network method to track changes and abnormalities in multivariate health data. Our proposed method leverages convolutional and graph attention layers to produce personalized scores indicating the abnormality level (i.e., deviations from the baseline) of users' data throughout the monitoring. We implement and evaluate the proposed method via a case study on long-term maternal health monitoring. Sleep and stress of pregnant women are remotely monitored using a smartwatch and a mobile application during pregnancy and 3-months postpartum. Our analysis includes 46 women. We build personalized sleep and stress models for each individual using the data from the beginning of the monitoring. Then, we compare the two groups by measuring the data variations. The abnormality scores produced by the proposed method are compared with the findings from the self-report questionnaire data collected in the monitoring and abnormality scores generated by an autoencoder method. The proposed method outperforms the baseline methods in exploring the changes between high-risk and low-risk pregnancy groups. The proposed method's scores also show correlations with the self-report data. Consequently, the results indicate that the proposed method effectively detects the abnormality in multivariate long-term health monitoring.

CCS CONCEPTS

• **Applied computing** → **Health informatics**;

KEYWORDS

Multivariate Time-series, Graph Attention Network, Change Analysis, Maternal Health, Long-term Monitoring

ACM Reference Format:
Yuning Wang, Iman Azimi, Mohammad Feli, Amir M. Rahmani, and Pasi Liljeberg. 2023. Personalized Graph Attention Network for Multivariate Time-series Change Analysis: A Case Study on Long-term Maternal Monitoring. In *Proceedings of ACM SAC Conference (SAC'23).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3555776.3577675

1 INTRODUCTION

Internet-of-Things (IoT) technology is fundamentally transforming traditional and hospital-based healthcare into ubiquitous health monitoring, by which individuals are remotely monitored while they engage in their daily activities [9]. Such monitoring systems exploit a distinct set of paradigms – such as smart devices (including smartwatches and smartphones), communication infrastructure, and computing resources – to track users' health and physical activity over months or years. Recent studies show a growing demand for such systems in home-based healthcare applications, for example, to track sleep [18] and different physiological signs [8].

Ubiquitous health monitoring systems produce a large volume of data, which accumulates over time. The collected data in such systems are often time-series with temporal dependencies and correlations. The longitudinal data enable the investigation and detection of certain diseases [25]. Moreover, IoT-based systems include multi-sensory and -modal data collection, where various health parameters are captured simultaneously. Such multimodal data might represent different aspects of an event or situation. For example, various heart rate variability (HRV) parameters in time and frequency domains can be exploited to evaluate the individual's stress level [10].

Multivariate and longitudinal data call for novel analysis techniques to extract the *trends* and *abnormalities (e.g., deviations or*

SAC'23, March 27 –March 31, 2023, Tallinn, Estonia

ACM ISBN 978-1-4503-9517-5/23/03…$15.00
https://doi.org/10.1145/3555776.3577675

changes from a baseline trend) over the course of monitoring. Data models are required to receive the multimodal data and merge them into a univariate variable, which can be an *explicit representation* of the entity. The univariate parameter allows us to visualize and evaluate the variations and trends throughout the monitoring. It also enables the comparison between different entities (e.g., different groups of patients). For example, various sleep parameters - such as total sleep time, sleep efficiency, and sleep fragmentation - can be collected for a sleep quality assessment application [21]. Each parameter represents partial information of sleep quality. Therefore, the application needs a method to merge the parameters into one parameter or score to comprehensively evaluate the sleep quality changes over time [2].

Traditionally, rule-based methods have been employed to extract a meaningful parameter from multimodal health data. For example, Early Warning Score (EWS) methods are broadly used in hospitals to detect patients' health deterioration [1, 16]. Various vital signs – e.g., heart rate (HR), respiration rate, and body temperature – are recorded periodically. The EWS methods generate a score (ranging from 0 to 3) for each vital sign using several thresholds. Finally, the sum of the scores shows the health deterioration level. Although the rule-based methods are easy to implement (e.g., can be done manually), the selection of the thresholds highly affects accuracy.

Regression methods are also used to estimate trends [3]. The data are fitted to, for example, a line by which the changes in time are illustrated. Moreover, linear mixed methods create models to fit the data, including both fixed and random effects in an analysis [12]. Hierarchical linear modeling has been exploited in different studies for nested data to investigate data variations [17, 23]. These methods can be tailored to compare the trends of data in different groups. However, they might fail to fit personalized models to multivariate data that might highly vary throughout long-term health monitoring.

Recently, machine learning-based methods have been proposed to model the data and identify trends in the data. For instance, Su et al. [24] introduced stochastic recurrent neural networks to extract the normal pattern of multivariate data. Similarly, an autoencoder was developed to identify abnormalities in sleep quality data collected in a health monitoring application [2]. In [13], a clustering-based approach has been proposed to detect abnormalities in the shape and amplitude of the multivariate health time series data. A fuzzy clustering technique was employed to discover structures in a generated multivariate ECG set. These methods extract the collective behavior of data and represent the difference between observed data and a system's normal behavior. However, these studies are restricted to short-term data evaluation or do not consider time dependency in data throughout the monitoring [2].

Longitudinal multi-modal time-series data require a modeling method to learn the trends in different aspects of the parameter and fuse them into a meaningful high-level parameter. Such a parameter should be an explicit indicator of the data, showing the abnormalities and trends throughout the monitoring. The required method needs to build models for each user, leveraging the spatial and temporal dependencies among multiple attributes.

In this paper, we develop a deep learning-based method to create personalized models for multivariate data collected in health monitoring. The proposed method exploits convolutional and graph attention layers to produce personalized scores showing the abnormality level of users' data throughout the monitoring. We implement and evaluate the proposed method via a case study of maternal health monitoring conducted by our team including high-risk and low-risk pregnancy groups. We construct personalized sleep and stress models for each individual using the data from the beginning of the pregnancy. Then, we compare the two groups by measuring the variations and abnormalities of the monitoring data. Finally, we evaluate the proposed method in comparison with the findings from 1) self-report questionnaire data and 2) abnormality scores generated by an autoencoder method.

2 PROPOSED METHOD

In this study, we develop a Gated Recurrent Units (GRU) network together with convolutional and graph attention layers as feature extractors to track the changes in long-term ubiquitous monitoring. The GRU has shown its superior performance in anomaly detection for sequential data [7]. In this work, we leverage GRU layers to capture long-term dependencies in data. Our method identifies the individuals' daily changes, using temporal relationships between the features over time.

The proposed method builds a model that is an explicit representation of the multivariate data collected in the monitoring. For personalization, we create a model for each individual using their own data showing user's normal data pattern. The model is trained to reproduce the input data and their succeeding points, and the model's training goal is to minimize the output loss by optimizing the network's weights (see Fig 1). In the testing phase, the variations in the monitoring are investigated by feeding the user's (test) data to the model and obtaining the loss value. The model provides a low loss value if the test data is similar to a training sample, which the model has already seen in the training phase. In contrast, the loss value is high if the sample is new to the model. We consider this loss value as an abnormality/change score showing the distance between the test data and the personalized model.

Our approach is inspired by the neural network architecture known as MTAD-GAT, introduced by Zhao et al. [29]. This architecture is primarily proposed for anomaly detection and used in detecting anomalies in cluster resources and services (e.g., CPU, network bandwidth, etc.). In contrast, we use a modified architecture to build models to learn users' normal patterns and analyze changes in long-term trends. The authors, in [29], proposed to use a spectral residual head to remove anomalies in the training dataset. However, our model is trained with all the data at the beginning of monitoring, since we consider them as the "normal" representation of users' behaviour. Moreover, the method in [29] detects anomalies by calculating an inference score for each variable, but our method is developed to provide one overall score for the multivariate input data. The modified architecture, illustrated in Fig 1, consists of 1) a feature extractor, 2) a forecasting model, and 3) a reconstruction model.

The feature extractor includes a 1D convolutional layer, by which the features are extracted from a fixed window of the sequential input. The kernel size is 7, and the shape of the output window feature is the same as the input data size. Then, the extracted convolutional features are fed separately to two graph attention layers:

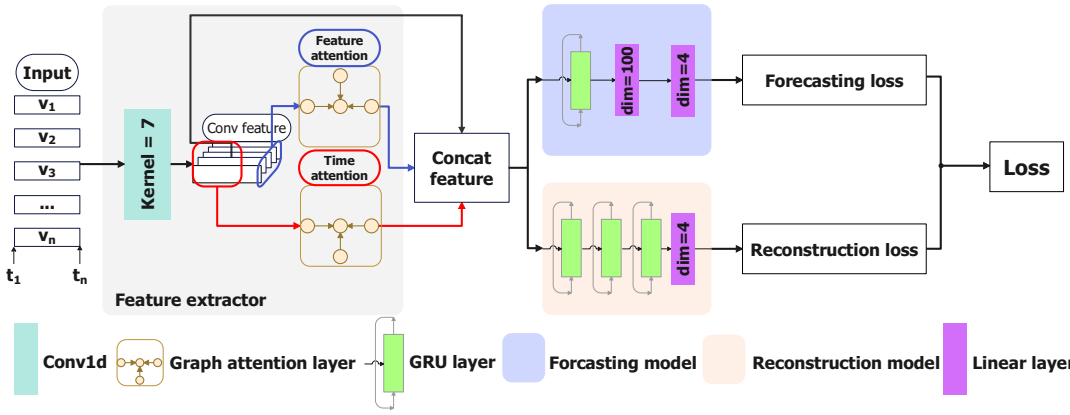

Figure 1: The architecture of the proposed method

feature attention and time attention layers (see Fig 1). The graph attention network, as a data modeling method, has recently been used in models to avoid feature contamination and to capture the temporal correlation between data points [5, 28]. In our model, each node in the graph represents one feature, and each edge represents the attention-based weight of each adjacent node to the current node. In the feature attention layer, the nodes are the features from multivariate data. In the time attention layer, the nodes are data points at a certain time in the window. Finally, the obtained convolutional, feature-attention, and time-attention features are fed to the forecasting and reconstruction models.

The forecasting model stacks one GRU layer and two linear layers to predict the succeeding data. The root mean square error (RMSE) of the predicted output and succeeding data is calculated as the loss of the forecasting model. The reconstruction model stacks three GRU layers with the hidden dimension of 200, followed by a linear layer to flatten the output. The model is trained to reconstruct the input data. The RMSE of the reconstructed output and the input data is the loss of the reconstruction model. The models are optimized by minimizing the sum of the reconstruction loss and the forecasting loss during the training phase. All the activation functions in our model are Leaky Rectified Linear Units (LeakyReLU), and the dropout rate is 0.2.

3 CASE STUDY: MATERNAL HEALTH MONITORING

3.1 Participants and Data Collection

We evaluate the proposed method using a case study on maternal health monitoring performed from January 2019 to March 2020 [22]. The study includes 62 pregnant women who belonged to high-risk and low-risk pregnancy groups. 32 individuals had prior preterm births (i.e., delivery date between gestational week 22 and 36) or late miscarriages (i.e., loss of pregnancy between gestational week 12 and 21) experience. The other 30 women had full-term births (i.e., delivery date between gestational weeks 37 and 42) experience with no pregnancy losses. The pregnant women were recruited when they were at the gestational week 12 to 15. They were asked to wear a smartwatch and to use a cross-platform mobile application during pregnancy and for 3-months postpartum.

The smartwatch used in the data collection was a Samsung Gear Sport watch [4]. During the monitoring, the participants were asked to wear the watch on the non-dominant hand continuously. We programmed the watch –run by Tizen OS– to track various health parameters. Our data collection app captured 12-minute photoplethysmogram (PPG) signals at a frequency of 20 Hz every two hours. The signals were analyzed to extract HR and HRV parameters [23]. The physical activity and sleep data are also provided by the watch. [15].

In addition, we developed a cross-platform mobile application to collect self-report questionnaires during the monitoring. The participants were asked to rate on a scale of 0 to 100 their stress level and sleep quality every week.

3.2 Data Analysis

We build personalized models for the pregnant women to investigate their health parameters' variations during and after pregnancy. We train the models leveraging the data before gestational week 21. We assume that the data represent the *normal pattern* of the women. A sliding window with a size of 10 days is used to select the input and moves at a step of 1 during the training phase.

In our analysis, a sleep model and a stress model are trained for each pregnant woman. The sleep model is trained using four sleep attributes: i.e., total sleep time, wake after sleep onset, sleep fragmentation, and average hand's movement captured by the smartwatch [21]. We extract the parameters from the sleep data provided by the watch.

The stress model is trained using HR and HRV parameters, which are controlled by the autonomic nervous system and correlated with user's stress [11]. The HRV parameters in our analysis are average interval between normal heart beats (AVNN), root mean square of successive differences of NN intervals (RMSSD), standard deviation of NN intervals (SDNN), power in low-frequency range (LF), power in high-frequency range (HF), and LF to HF ratio (LF/HF) [23]. We extract the parameters from the raw PPG signals collected by the watch. In this regard, we first use a bandpass Butterworth filter with the cut-off frequencies of 0.7 and 3.5 Hz to remove noises that are not in the cardiac signal ranges. Second, unreliable PPG segments are detected and dropped using a support vector machine classifier fed by different features: i.e., skewness, kurtosis, approximate entropy,

Shannon entropy, and spectral entropy [14]. Then, we employ a peak detection method [26] enabled by an adaptive threshold to find the systolic peaks. Finally, the HR and HRV parameters are obtained using the detected peaks [23]. Note that the HRV parameters are normalized by the average of HR to exclude the effect of HR changes. Moreover, all the sleep and stress features are standardized via Z-score before feeding them into the models.

The trained models, as representatives of the training data, are used to reflect the changing trend of the collected data (i.e., test set) from each patient during pregnancy (i.e., after the gestational week 21) and the postpartum. The test data (per day) are fed to the models, and the obtained loss is considered as the abnormality score. Eventually, we obtain the (sleep and stress) scores per day during the monitoring.

An example of the daily abnormality scores of a pregnant woman (selected randomly) is illustrated in Fig 2. The sleep and stress models are in green and blue colors, respectively. As shown in the figure, the abnormality scores rise after the week 36. To put it simply, the sleep and stress of the mother considerably changed after week 36 in comparison to her normal patterns (i.e., data before week 21).

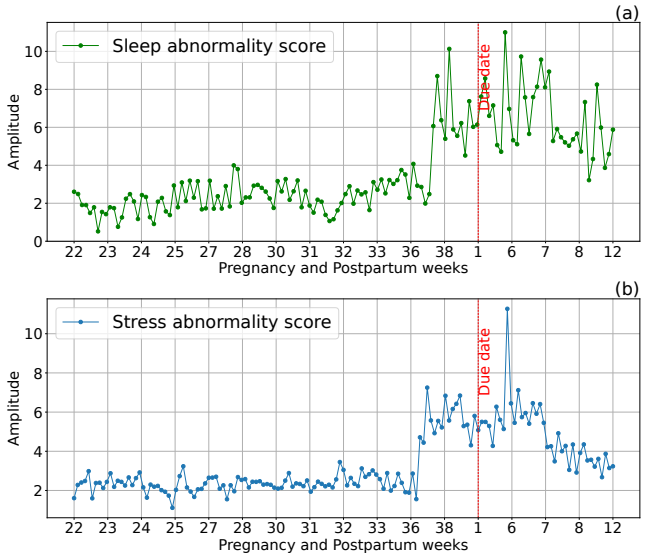

Figure 2: Daily abnormality scores of a participant obtained by the proposed method. (a) Sleep abnormality score (b) Stress abnormality score.

Moreover, we extract the sleep and stress changes from the self-report questionnaire data. The relative changes during the test phase (i.e., after the gestational week 21) are obtained compared to the training data collected before week 21. The variations of the self-report sleep data are calculated as:

$$X_{variation} = \left| \frac{X_{reported} - C_{mean}}{C_{std}} \right| \quad (1)$$

where $X_{reported}$ is a score reported by the user (after week 21), and C_{mean} and C_{std} are the mean and standard deviation of the (training) self-reported data. Similarly, we also calculate the variations of the self-report stress data.

4 RESULTS AND EVALUATION

We assess the proposed method in comparison with the self-report questionnaire data and an autoencoder method. The abnormality scores of the proposed method and the autoencoder method, along with the variations in the self-report data are evaluated for the low-risk and high-risk pregnancy groups. We also investigate the correlations between the abnormality scores and the self-report data. Our evaluation includes 46 (i.e., 26 low-risk and 20 high-risk) pregnant women. We exclude 16 participants, as their self-report data after week 21 have more than 50% missing values.

The proposed method was implemented on a Linux machine with AMD Ryzen Threadripper 2920X 12-Core processor, NVIDIA TITAN RTX GPU (24 GB memory), and 126 GB RAM. We used HeartPy, SciPy, and Scikit-learn for PPG analysis and PyTorch for the model training and testing [6, 19, 27]. The training process runs 1000 epochs per model.

4.1 Autoencoder Method

In our evaluation, we compare the proposed method with an autoencoder (AE), as a baseline method proposed in the literature for data modeling and abnormality extraction [2].

The AE model includes an encoder and a decoder. The encoder compresses the input to its representative features in latent space, and the decoder reconstructs the attributes from the bottleneck features. The encoder and decoder each stack two linear layers. Every unit in the input linear layer represents one of the variates in multivariate data. The AE model is optimized by minimizing the difference between the reconstructed and original input. Similar to the proposed method, we train sleep and stress AE models for each pregnant woman. The loss values obtained by the participants' data in the testing phase are used as the abnormality scores of the AE method. LeakyReLU is the activation function of the AE model.

4.2 Sleep and Stress Abnormality Scores

We first investigate how the proposed method, AE method, and self-report questionnaire can differentiate the high-risk and low-risk pregnancy groups during and after pregnancy. To this end, we exploit a Z-test method to determine if the values of the two pregnancy groups are different. The analysis includes both sleep and stress data.

The obtained results are shown in Table 1. The abnormality scores indicate the changing trend of sleep levels during pregnancy and postpartum. Using the proposed method, the sleep abnormality scores of the low-risk group are significantly lower than the scores of the high-risk group during pregnancy (p = 0.015) and postpartum (p = 0.004). The sleep abnormally scores of the AE method only show that the postpartum sleep data of the two groups are statistically different (p = 0.001). However, the self-report sleep data indicates that the groups are not statistically different. Consequently, only the proposed method was able to show that the high-risk pregnant women had greater changes in sleep quality during pregnancy and postpartum, compared to the low-risk group.

In contrast to the sleep analysis, the methods show that the stress scores/values of the two groups are not significantly different during pregnancy but significantly different during postpartum. Therefore, the trends of stress levels in low-risk and high-risk individuals

Table 1: Differences between high-risk and low-risk group using Z-test

Data type	State	Method	High-risk group Mean (SD)	Low-risk group Mean (SD)	Test statistic	P Value
Sleep	Pregnancy	Abnormality score (proposed)	**2.401 (0.755)**	**2.260 (0.656)**	**2.429**	**0.015***
		Abnormality score (AE)	0.315 (0.345)	0.242 (0.253)	1.922	0.055
		Self-report sleep value	1.582 (1.925)	1.554 (1.474)	0.209	0.834
	Postpartum	Abnormality score (proposed)	**3.220 (1.510)**	**2.605 (0.774)**	**4.562**	**0.004****
		Abnormality score (AE)	**0.539 (0.428)**	**0.347 (0.353)**	**4.529**	**0.001****
		Self-report sleep value	2.464 (3.661)	2.150 (2.128)	1.099	0.272
Stress	Pregnancy	Abnormality score (proposed)	3.036 (1.146)	2.924 (0.989)	1.263	0.207
		Abnormality score (AE)	1.035 (0.560)	1.056 (0.604)	1.185	0.236
		Self-report stress value	1.583 (2.360)	1.541 (1.474)	0.278	0.781
	Postpartum	Abnormality score (proposed)	**3.556 (1.245)**	**3.946 (1.664)**	**-2.144**	**0.032***
		Abnormality score (AE)	**1.548 (0.657)**	**1.921 (1.734)**	**-2.326**	**0.020***
		Self-report stress value	**2.495 (4.655)**	**1.682 (1.910)**	**2.449**	**0.014***

* Statistically significant (< 0.05) ** Highly Statistically significant (< 0.01)

were different after the delivery, compared to their own "normal" conditions.

In addition to the Z-test method, we employ the Spearman correlation coefficient test to investigate the correlations between the variations in self-report data and the abnormality scores of the proposed and AE methods. As shown in Table 2, during the pregnancy, the sleep abnormality scores of the proposed method have a low correlation with the self-report sleep data ($p = 0.002$). Moreover, there is a low correlation between the sleep abnormality scores and the self-report stress data ($p = 0.044$) during the postpartum. However, the sleep abnormality scores of the AE method show no correlations with the self-report data. Moreover, the stress abnormality scores of the proposed method have low (significant) correlations with the stress and sleep self-report data during pregnancy and with the sleep self-report data during postpartum. In contrast to the proposed method, the AE only indicates a low (significant) correlation with the self-report sleep data.

Consequently, our proposed model is more successful compared to the AE method, to find the relationships in the data using the time dependencies between the multiple attributes. It can effectively detect the significant changes in both multivariate sleep and stress data. It outperforms the AE method and subjective questionnaires to find the differences between the low-risk and high-risk pregnancy groups. Moreover, although the correlations between the proposed abnormality scores and the self-report data are low, the proposed method's scores show more significant correlations with the pregnancy and postpartum self-report data than the AE method.

5 LIMITATION AND FUTURE DIRECTION

One of the limitations of the study was the small sample size. Our analysis was limited to 46 participants. Other studies in the literature included a higher number of participants, although they were limited to subjective data collection methods. For example, Priya et al. investigated depression, anxiety, and stress among 165 pregnant women in East Delhi by self-report questionnaires [20]. The future directions can be to conduct wearable-based monitoring systems on a larger population and further validate the proposed method's performance. In addition, our study was focused on preterm birth and miscarriage. Future work should consider other complications during pregnancy, such as diabetes and preeclampsia.

Another limitation is the frequency of self-report data collection. The wearable devices provided continuous monitoring data in the monitoring. However, the self-reports on sleep and stress in our study were acquired weekly. To compare with the self-report values, we had to downscale the continuous attributes to one value per week, which caused a loss of data. For future directions, the self-report questionnaires should be scheduled multiple times per day to prevent such data loss.

6 CONCLUSIONS

In this paper, we developed a personalized GRU network with a graph-attention based feature extractor to detect the changing trend in multivariate data in long-term health monitoring. We implemented the method in a maternal health case study, in which pregnant women were monitored for 6-months pregnancy and 3-months postpartum. The proposed method built sleep and stress models for each woman using the data at the beginning of the monitoring. Then, the models were used to identify abnormalities during the rest of the pregnancy and postpartum. We evaluated the proposed method compared to the self-report questionnaire data and an autoencoder method. The proposed method outperformed the autoencoder method and subjective questionnaires to find relationships in the data and differences between the low-risk and high-risk pregnancy groups. The proposed method's scores also showed more significant correlations with the pregnancy and postpartum self-report data, compared to the autoencoder method.

Consequently, the results indicated that our proposed method could identify the changes and abnormality trends in multivariate long-term health monitoring.

Table 2: Spearman correlation between abnormality scores and self-report values

Data type	State	Method	Self-report sleep		Self-report stress	
			corr.	P value	corr.	P value
Sleep	Pregnancy	Abnormality score (proposed)	**0.125**	**0.002****	0.040	0.323
		Abnormality score (AE)	0.041	0.308	-0.020	0.631
	Postpartum	Abnormality score (proposed)	-0.065	0.270	**0.118**	**0.044***
		Abnormality score (AE)	-0.088	0.134	0.069	0.239
Stress	Pregnancy	Abnormality score (proposed)	**0.108**	**0.008****	**0.084**	**0.041***
		Abnormality score (AE)	**0.082**	**0.044***	0.024	0.560
	Postpartum	Abnormality score (proposed)	**0.238**	**<0.001****	0.083	0.153
		Abnormality score (AE)	0.098	0.071	0.094	0.085

* Statistically significant (< 0.05) ** Highly Statistically significant (< 0.01)

ACKNOWLEDGMENTS

This research was funded in part by the Academy of Finland through the SLIM Project under grant numbers 316810 and 316811 and in part by the U.S. National Science Foundation (NSF) through the UNITE Project under grant number SCC CNS-1831918 and the D-CCC Project under grant number FW-HTF CNS-2026614.

REFERENCES

[1] A. Anzanpour et al. 2017. Self-awareness in remote health monitoring systems using wearable electronics. In *DATE*. 1056–61.
[2] I. Azimi et al. 2019. Personalized maternal sleep quality assessment: An objective iot-based longitudinal study. *IEEE Access* 7 (2019), 93433–93447.
[3] A. P. Betrán et al. 2005. National estimates for maternal mortality: an analysis based on the WHO systematic review of maternal mortality and morbidity. *BMC Public Health* 5, 1 (2005), 1–12.
[4] Samsung Electronics. [n. d.]. Samsung Gear Sport. https://www.samsung.com/us/mobile/wearables/smartwatches/gear-sport-black-sm-r600nzkaxar/. Accessed: October 2022.
[5] M. Fang et al. 2021. FTPG: A Fine-Grained Traffic Prediction Method With Graph Attention Network Using Big Trace Data. *T-ITS* (2021), 1–13. https://doi.org/10.1109/TITS.2021.3049264
[6] P. Van Gent et al. 2019. HeartPy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation research part F: traffic psychology and behaviour* 66 (2019), 368–378.
[7] Y. Guo et al. 2018. Multidimensional time series anomaly detection: A gru-based gaussian mixture variational autoencoder approach. In *ACML*. PMLR, 97–112.
[8] S. Huhn et al. 2022. The Impact of Wearable Technologies in Health Research: Scoping Review. *JMIR mHealth UHealth.* 10 (2022), e34384.
[9] M. H. Kashani et al. 2021. A systematic review of IoT in healthcare: Applications, techniques, and trends. *Journal of Network and Computer Applications* 192 (2021), 103164.
[10] H.-G. Kim et al. 2018. Stress and heart rate variability: a meta-analysis and review of the literature. *Psyc. Investig.* 15, 3 (2018), 235.
[11] A. V. Klinkenberg et al. 2009. Heart rate variability changes in pregnant and non-pregnant women during standardized psychosocial stress. *Acta Obstet Gynecol Scand.* 88, 1 (2009), 77–82.
[12] A. Kuznetsova et al. 2017. lmerTest package: tests in linear mixed effects models. *Journal of statistical software* 82 (2017), 1–26.
[13] J. Li et al. 2021. Clustering-based anomaly detection in multivariate time series data. *Applied Soft Computing* 100 (2021), 106919.
[14] A. Mahmoudzadeh et al. 2021. Lightweight photoplethysmography quality assessment for real-time IoT-based health monitoring using unsupervised anomaly detection. *Proc. Comp. Sci.* 184 (2021), 140–7.
[15] M. Mehrabadi et al. 2020. Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. *JMIR mHealth UHealth.* 8, 11 (2020), e20465.
[16] R. Morgan et al. 1997. An early warning scoring system for detecting developing critical illness. *Clin Inten. Care* 8, 2 (1997), 100.
[17] H. Niela-Vilén et al. 2021. Pregnant women's daily patterns of well-being before and during the COVID-19 pandemic in Finland: Longitudinal monitoring through smartwatch technology. *PloS one* 16 (2021).
[18] Q. Pan et al. 2020. Current Status and Future Challenges of Sleep Monitoring Systems: Systematic Review. *JMIR Biomedical Engineering* 5, 1 (2020), e20921.
[19] F. Pedregosa et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830.
[20] Adity Priya, Sanjay Chaturvedi, Sanjiv Kumar Bhasin, Manjeet Singh Bhatia, and Gita Radhakrishnan. 2018. Depression, anxiety and stress among pregnant women: A community-based study. *Indian journal of psychiatry* 60, 1 (2018), 151.
[21] Avi Sadeh. 2015. III. Sleep assessment methods. *Monographs of the Society for Research in Child Development* 80, 1 (2015), 33–48.
[22] F. Sarhaddi et al. 2021. Long-Term IoT-Based Maternal Monitoring: System Design and Evaluation. *Sensors* 21, 7 (2021), 2281.
[23] F Sarhaddi et al. 2022. Trends in Heart Rate and Heart Rate Variability During Pregnancy and the 3-Month Postpartum Period: Continuous Monitoring in a Free-living Context. *JMIR mHealth UHealth.* (2022).
[24] Y. Su et al. 2019. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *KDD '19*.
[25] L. Tikotzky et al. 2019. Infant nocturnal wakefulness: a longitudinal study comparing three sleep assessment methods. *Sleep* 42 (2019), 191.
[26] P. van Gent et al. 2018. Heart rate analysis for human factors: Development and validation of an open source toolkit for noisy naturalistic heart rate data. In *Proceedings of the 6th HUMANIST Conf.* 173–8.
[27] P. Virtanen et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods* 17, 3 (2020), 261–72.
[28] L. Wang et al. 2019. Graph Attention Convolution for Point Cloud Semantic Segmentation. In *CVPR*.
[29] H. Zhao et al. 2020. Multivariate time-series anomaly detection via graph attention network. In *ICDM*. IEEE, 841–850.