UC Irvine
UC Irvine Previously Published Works

Title

Preterm birth risk stratification through longitudinal heart rate and HRV monitoring in daily life.

Permalink

https://escholarship.org/uc/item/61h666w9

Journal

Scientific Reports, 14(1)

Authors

Feli, Mohammad

Azimi, Iman

Sarhaddi, Fatemeh

et al.

Publication Date

2024-08-27

DOI

10.1038/s41598-024-70773-0

Peer reviewed

scientific reports

OPEN

Preterm birth risk stratification through longitudinal heart rate and HRV monitoring in daily life

Mohammad Feli[1✉], Iman Azimi[2], Fatemeh Sarhaddi[1], Zahra Sharifi-Heris[3], Hannakaisa Niela-Vilen[4], Pasi Liljeberg[1], Anna Axelin[4] & Amir M. Rahmani[2,5]

Preterm birth (PTB) remains a global health concern, impacting neonatal mortality and lifelong health consequences. Traditional methods for estimating PTB rely on electronic health records or biomedical signals, limited to short-term assessments in clinical settings. Recent studies have leveraged wearable technologies for in-home maternal health monitoring, offering continuous assessment of maternal autonomic nervous system (ANS) activity and facilitating the exploration of PTB risk. In this paper, we conduct a longitudinal study to assess the risk of PTB by examining maternal ANS activity through heart rate (HR) and heart rate variability (HRV). To achieve this, we collect long-term raw photoplethysmogram (PPG) signals from 58 pregnant women (including seven preterm cases) from gestational weeks 12–15 to three months post-delivery using smartwatches in daily life settings. We employ a PPG processing pipeline to accurately extract HR and HRV, and an autoencoder machine learning model with SHAP analysis to generate explainable abnormality scores indicative of PTB risk. Our results reveal distinctive patterns in PTB abnormality scores during the second pregnancy trimester, indicating the potential for early PTB risk estimation. Moreover, we find that HR, average of interbeat intervals (AVNN), SD1SD2 ratio, and standard deviation of interbeat intervals (SDNN) emerge as significant PTB indicators.

Preterm birth (PTB) is a major cause of newborn complications and death, affecting up to 16% of pregnancies globally[1]. By definition, PTB is a birth that occurs before the completion of 37 weeks of gestation (less than 259 days) from the first day of conception[2]. In 2020 alone, the World Health Organization (WHO) reported 13.4 million instances of PTB worldwide[1], making it a leading cause of neonatal mortality. Alarmingly, PTB is responsible for over one-third of neonatal deaths and stands as the primary cause of death for children under the age of five. Despite technological and healthcare advancements, recent data from WHO indicates a concerning lack of progress in reducing PTB rates, with a minimal change from 9.8% in 2010 to 9.9% in 2020[3]. Predicting PTB remains challenging due to the various factors associated with it and the complexity of understanding all the contributing factors.

Studies indicate a negative correlation between gestational age and maternal-neonatal morbidity and mortality[4]. The lower the gestational weeks are, the higher the risks for the preterm infant. PTB survivors often require prolonged hospitalization in neonatal intensive care units, imposing additional psychological strain and financial burdens on families while also placing significant pressure on healthcare systems[5,6]. According to the literature, parents of preterm infants often experience heightened stress, anxiety, depression, and post-traumatic stress symptoms, especially when the baby is born before 32 weeks. This situation raises immediate concerns about the baby's health and introduces uncertainties about long-term well-being. Additionally, children born very preterm indicated an increased risk of mental health difficulties, including Attention Deficit Hyperactivity Disorder (ADHD), Autism Spectrum, and anxiety disorders[7]. Premature birth can also disrupt brain development, leading to neurodevelopmental disorders in premature children. According to the evidence, the period between 22 and 36 weeks of gestation plays a crucial role in the physiological maturation, structural, and functional development of various nervous systems, including motor, neurosensory, cognitive, and neurovegetative

[1]Department of Computing, University of Turku, Turku, Finland. [2]Department of Computer Science, University of California, Irvine, USA. [3]School of Nursing, University of California, Los Angeles, USA. [4]Department of Nursing Science, University of Turku, Turku, Finland. [5]School of Nursing, University of California, Irvine, USA. ✉email: mohammad.feli@utu.fi

systems. Prematurity adversely affects neurobehavioral development and potentiates poor pregnancy outcomes during this critical period[8].

PTB is a complex phenomenon influenced by various factors, making it challenging to pinpoint a singular cause. PTB often is either a spontaneous response to the existing dysfunction or an intentional considered strategy by the care providers in response to the existing life-threatening medical conditions[9]. Spontaneous PTB is associated with risk factors like infections, maternal health conditions, uterine or cervical structural dysfunctions, short interpregnancy intervals, extremes in maternal age, lifestyle factors, poor nutrition, stress, psychosocial factors, and a history of previous preterm birth[10,11]. Non-spontaneous PTB may result from complications such as preeclampsia, multiple pregnancies, intrauterine growth restriction, and cervical infections. Notably, PTB can also occur without an identifiable cause[12,13]. The pathophysiological pathways of spontaneous PTB involve complex interactions, incorporating structural inefficiencies and contractility influenced by inflammation, hormonal imbalances, uterine overdistension, vascular changes, and idiopathic factors. Ultimately, compromised blood flow, premature rupture of the chorioamniotic membrane, and structural insufficiency collectively contribute to triggering premature labor[9].

This complexity explains why the existing methodologies for estimating and predicting PTB mostly rely on conventional risk factors derived from electronic health records (EHR)[14–17]. However, leveraging patient data from EHRs faces challenges, including incomplete records for pregnant women due to socio-economic barriers[18]. Some studies have adopted biomedical signal acquisition methods to predict PTB, such as electrohysterogram (EHG) signals, which indicate uterine electrical activity from the pregnant woman's abdominal surface and have been utilized in various investigations[19–21].

Discovering the pathophysiological pathway and mechanism of action of PTB may serve as a complementary approach to comprehending this complex phenomenon. The Autonomic Nervous System (ANS) serves as a vital regulatory mechanism in the body, regulating involuntary physiological functions such as those of the cardiovascular, digestive, and respiratory systems in response to internal and external stimuli. Comprising the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS), these branches collaborate during pregnancy to maintain a harmonious balance in bodily functions amid the unique demands of gestation[22]. Heart rate variability (HRV) emerges as a crucial measure, reflecting the dynamic interplay between the ANS branches. HRV refers to the variation in time intervals between successive heartbeats, reflecting the adaptability and resilience of the ANS in diverse situations, ensuring healthy adaptation. A relatively higher HRV often indicates a more adaptable and resilient ANS, capable of responding flexibly to different situations assuring the healthy adaptation[23]. While HRV is conventionally linked to the cardiovascular system, its capacity to mirror the intricate interactions within the ANS branches extends beyond the heart, making it a relevant and versatile biomarker in assessing overall well-being[24–26].

Electrocardiogram (ECG) signals have been conventionally employed to explore the ANS function in women with a history of recurrent pregnancy loss[27]. The HRV when assessed by ECG is linked to behavioral and preexisting medical conditions known to affect PTB, including stress[28,29], physical activity[30], sleep[31], mental health[32], smoking[33], BMI[34], drug abuse[35], gestational hypertension and preeclampsia[36], and fetal growth[37]. However, biosignal methodologies for PTB prediction are limited to short-term and intermittent data collected mainly in predefined clinical settings, limiting comprehension of dynamic real-life scenarios linked to PTB.

Recent studies have leveraged wearable technologies to investigate pregnancy complications, covering the gaps related to the existing snapshot assessments[38]. Wearable devices facilitate the ongoing and instantaneous gathering of data on maternal health, including physiological parameters such as heart rate (HR) and HRV[39]. Instead of ECG, photoplethysmography (PPG) is applied as the prominent method for acquiring HRV parameters through wearables[40]. The continuous monitoring of pregnant women using wearable devices allows for the exploration of various pregnancy issues, such as stress[28,41], hypertension[42], mental problems[43], and obesity[44]. Moreover, it enables the assessments of maternal sleep[45], physical activity[46,47], and fetal parameters, including heart activity[48–50] and movement[51,52].

Previous studies have used wearable devices for monitoring maternal ANS activity[53]. However, to the best of our knowledge, they have not utilized these devices for the long-term, continuous monitoring of maternal ANS activity, particularly in relation to dysregulations associated with PTB risk during pregnancy. The primary challenge stems from the absence of user-friendly wearable devices capable of providing long-term access to raw PPG signals. Current wearables, like the OURA Ring, Apple Watch, and Fitbit, do not offer access to raw PPG waveforms. To overcome this limitation, we have developed a custom smartwatch application designed for Tizen OS-powered watches, including the Samsung Galaxy series. This application is engineered to stream PPG signals to our private servers with millisecond precision. Existing short-term studies, primarily based on ECG and PPG, have demonstrated a link between ANS activity and various preexisting medical conditions[28,29]. Our objective is to explore the risk of PTB by conducting a long-term observational study of ANS, as manifested in HR and HRV parameters, using wearable technology throughout pregnancy.

In this paper, we conduct a longitudinal study to assess the risk of PTB by examining maternal ANS activity. This is achieved through the continuous monitoring of ANS activity via smartwatches during pregnancy. To this end, we collected PPG signals from pregnant women using smartwatches throughout their pregnancies. These signals are then processed to derive accurate HR and HRV measurements. Following this, we develop an autoencoder machine learning (ML) model with SHAP analysis as an explainable AI method that generates an abnormality score, which acts as a predictive indicator for PTB, based on the longitudinal HR and HRV data. Additionally, our research delves into evaluating the impact of HR and HRV on the likelihood of PTB occurrence.

Related works

Previous studies have utilized EHRs and questionnaire data in conjunction with ML techniques to predict PTB and identify associated risk factors[14–17,54,55]. For instance, Raja et al.[15] employed text-based features extracted from EHRs and surveys, applying various ML techniques for PTB prediction. In a study by Arabi et al.[54], ML models were developed using EHR data alongside socio-demographic information, revealing associations between PTB and pregnancy-related complications like diabetes. Another study[55] utilized structured interviews, incorporating parental background, occupation, and environmental parameters, to identify PTB risk factors. However, these approaches may encounter practical challenges arising from patient-related barriers and health system limitations, including incomplete health records for pregnant women due to factors such as lack of insurance and restricted access[18].

Various studies have employed biomedical signal acquisition techniques to gain deeper insights into the pathophysiological pathways linked to PTB. Among these methods, the EHG is a frequently employed technique for recording the electrical activity of the uterine muscle[56]. Some studies have harnessed frequency and statistical features derived from EHG signals, employing various ML models to detect PTB[19–21]. However, the utility of the EHG signal is limited in high-risk individuals with comorbid conditions such as obesity and factors like fetal and placental position[57,58], recognized contributors to PTB. In another approach, Wlodarczyk et al.[59] proposed a PTB prediction model utilizing ultrasound images. Additionally, Kataoka et al.[27] gathered ECG signals to investigate the activity of the ANS in non-pregnant women with a history of recurrent pregnancy loss. This study suggests decreased overall HRV and diminished parasympathetic activity, indicating potential autonomic dysregulation in PTB. However, these methods are often constrained to short-term and intermittent assessments, primarily conducted in predetermined clinical settings and typically shortly before the occurrence of PTB. The temporal and contextual gaps in these approaches restrict their preventive capabilities and raise concerns about their sufficiency for dynamic real-life scenarios associated with PTB, particularly in addressing the needs of high-risk individuals requiring heightened attention.

Recent research has utilized continuous monitoring of pregnant women using wearable technologies to explore pregnancy-related issues[38]. For instance, Galea et al.[46] investigated maternal sleep quality and physical activity by collecting data from twenty pregnant women using a wristwatch-like device. In another study by Sarhaddi et al.[39], trends in maternal HR and HRV collected using smartwatches were assessed during pregnancy and three months postpartum. Lu et al.[60] explored the impact of early nursing intervention on adverse pregnancy outcomes in women with gestational diabetes mellitus utilizing wearable devices. A recent study, by Sharifiheris et al.[61], suggested maternal HRV differences between healthy and complicated pregnant individuals. Assessed continuously using the Oura ring[62] during the second and third pregnancy trimesters, the result of this study supports the potential of tracking HRV for PTB detection. Additionally, Jasinski et al.[63] have explored the relationship between maternal HRV and gestational weeks and time until birth, enrolling 241 pregnant women wearing WHOOP straps[64]. While these approaches highlight the significance of maternal HRV monitoring, they have not investigated the potential of PTB estimation during pregnancy. It is noteworthy that these studies[61,63] have been limited to one HRV parameter, i.e., the root mean square of successive NN interval difference (RMSSD). Therefore, they overlooked the analysis of HRV from different perspectives, which might potentially improve PTB estimation.

Method
Study design
We conducted a longitudinal health monitoring study on pregnant women, where participants wore smartwatches to measure their HR and HRV remotely during daily routines in free-living conditions. The data collection spanned from gestational weeks 12-15 to three months post-delivery.

Participants and recruitment
The recruitment for the study was conducted between January 2019 and March 2020 through advertisements on social media and in maternity clinics in the region of Southwest Finland. The participants were divided into two groups: (1) the high-risk group, including 32 pregnant women with a history of PTB (i.e., birth between gestational weeks 22–36) or late miscarriage (between gestational weeks 12–21), and (2) the low-risk group, comprising 30 pregnant women with a history of full-term births (i.e., birth after gestational week 37) and no prior pregnancy losses. Among all participants, seven experienced PTB, with six from the high-risk group and one from the low-risk group. All participants met specific inclusion criteria, such as singleton pregnancies, recruitment at 12–15 gestational weeks, age over 18 years, proficiency in the Finnish language, and having access to an Android or iOS smartphone.

The study's objective was conveyed to the eligible candidates through direct, in-person meetings. Each participant received a smartwatch, provided written informed consent, and received detailed instructions for the study. Participants were instructed to wear a smartwatch continuously throughout their pregnancy. Out of the initial 62 pregnant women recruited, four opted to withdraw from the study. Background information of the remaining 58 participants, categorized by high-risk versus low-risk and preterm versus term, with their statistical test p-values, is presented in Table 1.

Research ethics
This study received ethical approval from the Ethics Committee of the Hospital District of Southwest Finland, approval number: Dnro: 1/1801/2018. All study methods were carried out in accordance with relevant guidelines and regulations, and the ethics committee approved experimental protocols. Before their involvement, written

Characteristic	High-risk n=28	Low-risk n=30	p-value	Preterm n=7	Term n=51	p-value
Age (years), mean (range)	32.8 (24-43)	31.2 (23-39)	0.32	36.4 (30-41)	31.4 (23-43)	0.009
BMI (kg/m^2), mean (range)	26.9 (18-66.2)	26.6 (19.9-45.9)	0.85	26.6 (18.9-35.7)	26.8 (18-66.2)	0.95
Married/cohabiting, n (%)	27 (96)	30 (100)	1	7 (100)	50 (98)	1
Employed/student, n (%)	23 (82)	28 (93)	0.95	7 (100)	44 (86)	0.43
University education, n (%)	10 (36)	8 (27)	0.04	1 (14)	17(33)	0.16
Pregnancy planned, n (%)	27 (96)	27 (90)	0.23	7 (100)	47 (92)	1
GA at recruitment (week+day), mean (std)	13.8+3.3 (1.5+1.5)	14+2.8 (0.8+1.3)	0.56	14.4+2.9 (1.7+1.7)	13.8+3.1(1.1+1.3)	0.21
GA at delivery (week+day), mean (std)	37.9+2.7 (2.3+1.9)	39.8+2.7 (1.3+1.9)	0.0002	34.7+3.4(1.7+1.4)	39.4+2.6(1.4+2)	0

Table 1. Background information of the 58 participants, categorized by high-risk vs. low-risk and preterm vs. term, with associated statistical test p-values. *BMI* body mass index, *GA* gestational age. P-values for continuous variables were calculated by independent t-test and for categorical variables by chi-square test.

informed consent was asked and received from all the participants signifying their agreement for the publication of results with a commitment to data anonymization to protect their privacy.

Data collection

Data were collected using Samsung Gear Sport smartwatches[65]. The smartwatches are water-resistant and light-weight with the dimensions of $44.6 \times 42.9 \times 11.6$ mm and weigh 67 grams, making them convenient for continuous long-term data collection in free-living conditions. They are equipped with PPG and inertial measurement unit (IMU) sensors. Operating on the Tizen open-source operating system, they facilitate customized data collection and transmission. The integrated sensors in the smartwatches capture PPG and gyroscope/accelerometer data at a sampling frequency of 20 Hz.

During the data collection, participants were instructed to continuously wear the smartwatches on their non-dominant hands. We programmed the watches to record 12 minutes of PPG signals every 2 hours, enabling consistent data collection throughout the day. The setup allowed sufficient battery life, which spanned approximately 2 to 3 days, ensuring uninterrupted monitoring[66]. The collected data were stored within the internal storage of the smartwatch. Additionally, we developed a Tizen app that allows participants to upload the collected data to a remote server using a Wi-Fi connection. We requested participants to upload their data regularly, at least once a week. If participants failed to upload data for two consecutive weeks, they were notified via text messages or emails. The dataset acquired is extensive, including 1,576,080 minutes of PPG signals from pregnant women throughout their pregnancy and three months postpartum.

Data analysis

Overview

Our analysis includes harnessing the collected PPG data to develop a PTB indicator. Illustrated in Fig. 1, our data analysis workflow consists of two key steps. Initially, we employ a reliable PPG processing pipeline[67] to extract accurate HR and HRV data from raw PPG signals. Following this, we leverage an autoencoder model to distinguish HR and HRV data patterns between mothers with term deliveries and those who experienced PTB, resulting in a PTB indicator.
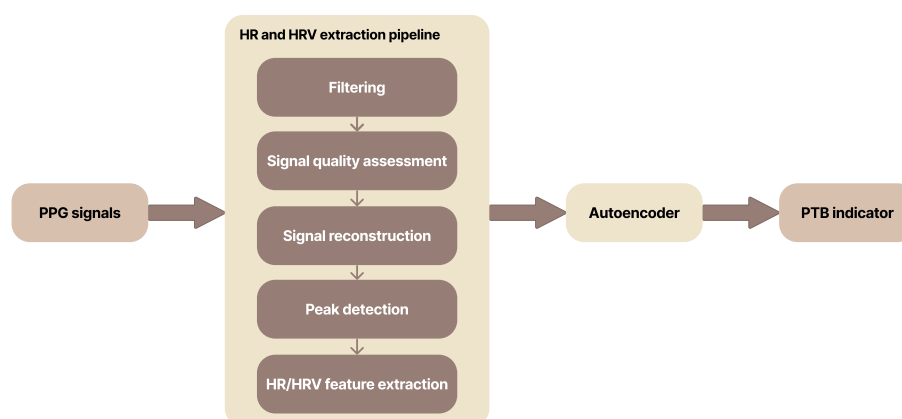

Fig. 1. Data analysis workflow.

HR and HRV data extraction

We extract HR and HRV features from PPG signals collected during the monitoring period. PPG signals are particularly susceptible to corruption when recorded in free-living conditions. The signal quality can be compromised due to both motion artifacts from the subject's physical activities and environmental noises, which may lead to inaccurate HR and HRV measurements. To ensure reliable extraction of HR and HRV data from PPG signals, we leverage a PPG processing pipeline. For the evaluation of the pipeline, refer to[67]. This pipeline comprises several stages, including filtering, signal quality assessment, PPG waveform reconstruction, and peak detection. These stages are briefly described below.

- *Filtering*: We apply a second-order Butterworth highpass filter with a cut-off frequency of 0.5 Hz to eliminate unwanted frequencies from the signal. Raw PPG signals, especially when collected in uncontrolled free-living settings, include various interference and noise elements. Accordingly, we remove signal components that are not in the frequency of interest (i.e., HR and HRV frequencies).
- *Signal quality assessment*: We utilize a signal quality assessment approach, which we developed in our previous work[68], to differentiate clean and noisy PPG signals. Motion artifacts can significantly distort the morphology of the signal, making it impossible to extract reliable HR and HRV. Therefore, assessing the PPG signal quality is essential to identify reliable parts of the signal for analysis. To this end, we employ our previously established PPG signal quality assessment method[68], which can distinguish reliable and unreliable parts of the PPG signal with an accuracy of 0.97.
- *Signal reconstruction*: In this step, minor noisy parts of the PPG signals are reconstructed. PPG signals inherently exhibit quasi-periodic characteristics, representing the rhythmic activity of the cardiovascular system. When subjected to slight noise interference, the corrupted part can be reconstructed by leveraging the information contained in the preceding clean parts. To accomplish this, we employ a PPG reconstruction approach[69] based on a deep convolutional generative adversarial network to reconstruct noisy parts for durations of up to 15 seconds.
- *Peak detection*: In our analysis, we utilize a PPG peak detection technique to identify systolic peaks within PPG signals. Systolic peaks are crucial elements for deriving vital signs from PPG signals, as they represent essential characteristics associated with the cardiovascular system. We employ a deep-learning-based method that we have developed and previously published[70], specifically designed for the extraction of systolic peaks from PPG signals.
- *HR and HRV features extraction*: We first derive interbeat intervals (IBIs) based on the systolic peaks detected in the previous step. Then, HR and several HRV features are computed based on the derived IBI values. A window length of 5 minutes is employed, ensuring the reliable extraction of the features. A list of HRV features obtained in this study is provided in Table 2.

Autoencoder

We utilize longitudinal HR and HRV data collected during pregnancy to build an autoencoder model for estimating the risk of PTB. Autoencoders are artificial neural networks commonly used in anomaly detection tasks because they can effectively learn normal data representations and identify instances that deviate from these learned representations[71–73].

In our approach, we consider PTB as an anomalous event, framing our analysis as an anomaly detection task, which yields a PTB abnormality score. During the training phase, the autoencoder is trained exclusively on HR and HRV data associated with term deliveries. In the testing phase, the autoencoder generates an abnormality score for each sample by calculating the reconstruction error-the difference between the input data and the autoencoder's reconstruction of that data.

Category	Feature	Units	Description
Time-domain	AVNN	ms	Average of NN intervals
	SDNN	ms	Standard deviation of NN intervals
	RMSSD	ms	Root mean square of successive NN interval differences
	pNN50	%	Percentage of successive NN intervals that differ by more than 50 ms
Frequency-domain	LF	s^2	Spectral power of low frequencies (0.04-0.15 Hz)
	HF	s^2	Spectral power of high frequencies (0.15-0.4 Hz)
	LFHF	%	LF to HF ratio
Non-linear	SD1	ms	Standard deviation of Poincaré plot perpendicular to the line-of-identity
	SD2	ms	Standard deviation of the Poincaré plot along the line-of-identity
	SD1SD2	%	SD1 to SD2 ratio

Table 2. List of HRV features. NN (normal-to-normal) intervals refer to the time intervals between consecutive R peaks in an ECG signal, equivalent to interbeat intervals (IBIs) measured from a PPG signal.

This methodology is based on the principle that the autoencoder better reconstructs term (normal) instances due to their close alignment with learned representations from the training data. Conversely, preterm (anomalous) instances show significant deviations and higher reconstruction errors, primarily because the model has not encountered such anomalies before. As a result, the model can detect preterm cases by identifying these anomalies in the HR and HRV data, providing a PTB indicator.

The architecture of our autoencoder is illustrated in Fig. 2, consisting of two main components: an encoder and a decoder. The encoder receives the HR and HRV data, including features outlined in Table 2 through an input layer with a dimension of 11. Given the input vector X representing the HR and HRV features, the encoder component maps X to the latent representation Z through nonlinear transformations as follows:

$$Z = f(w_3(f(w_2(f(w_1 X + b_1) + b_2) + b_3)) \tag{1}$$

where f is the activation function, and w_i and b_i are the weight and bias matrices in layer i. The encoder involves three dense layers with Rectified Linear Units (ReLU) activation functions. The three hidden layers include 9, 7, and 5 units, respectively.

Conversely, the decoder component receives the encoded representation and reconstructs the original HR and HRV data by mapping it back to the original input space. Given the latent representation Z, the original data X is reconstructed to $\hat{X}$ as follows:

$$\hat{X} = f(w_6(f(w_5(f(w_4 Z + b_4) + b_5) + b_6)) \tag{2}$$

The decoder consists of two hidden layers with a Sigmoid activation function (7 and 9 units) followed by an output layer with 11 units, matching the dimension of the input layer.

During the training process, the goal is to minimize the reconstruction error (L). To this end, the mean square error (MSE) is utilized as the loss function to measure the difference between the input data and its reconstructed representation.

$$L(X, \hat{X}) = argmin_{W,B} \left(\frac{1}{n} \right) \sum_{i=1}^{n} \| X_i - \hat{X}_i \|^2 \tag{3}$$

where W and B are the weight and bias matrices of the model, n represents the number of input samples, and X_i and $\hat{X}_i$ represent the original input and corresponding reconstructed output of the model. We utilize the Adam optimizer in the training process for optimizing model performance and convergence.

In our analysis, we conduct an inter-participant experiment, where training and testing participants do not overlap. For training, we use HR and HRV data from the beginning of the monitoring to gestational week 32 from the participants who underwent term delivery. This specific time window, up to week 32, is chosen to prevent potential bias in the model, given that all participants, regardless of term or preterm status, had delivery dates after week 32. For testing, we utilize data from the beginning of the monitoring to the delivery date (up to week 37) from the participants who experienced term and preterm deliveries.

Abnormality scoring and evaluation metrics
The trained autoencoder generates an abnormality score for each test sample. The score reflects the extent to which the sample differs from the learned normal representation. The abnormality score is calculated using the Root Mean Square Error (RMSE), which measures the difference between two vectors of values, as demonstrated in the following equation:
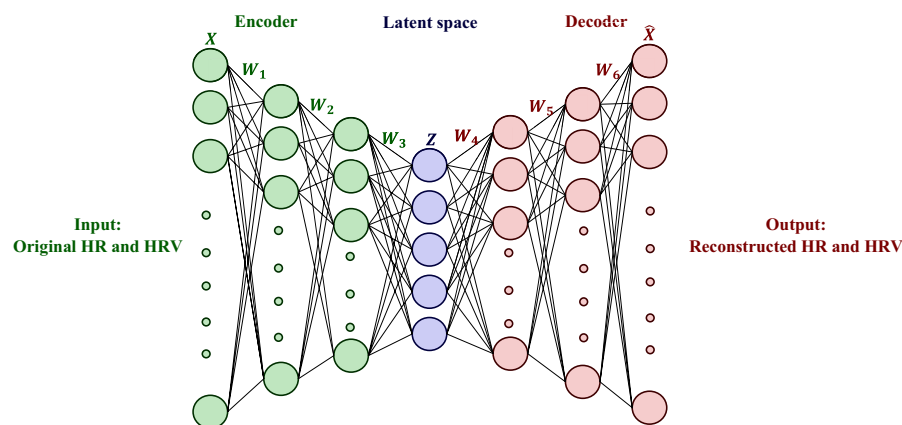

Fig. 2. Autoencoder architecture.

$$AS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} \tag{4}$$

where x and $\hat{x}$ represent the original and reconstructed HR and HRV samples.

To evaluate the performance of the autoencoder, we employ several evaluation metrics, including Receiver Operating Characteristic (ROC) curves, Area Under the Curve (AUC), Accuracy, Precision, Recall, and F1-score[74]. ROC curve represents the trade-off between the true positive rate (TPR) and the false positive rate (FPR) across different threshold values. AUC also quantifies the overall performance of the model, with higher values indicating better discrimination. Accuracy measures the proportion of correctly classified samples among all samples. Precision quantifies the proportion of true positives among all predicted positives, while Recall measures the proportion of actual positives correctly identified by the model. The F1-score is the harmonic mean of Precision and Recall, which provides a balanced measure of the performance of the model. For more detailed information about these metrics, refer to the reference[74].

Results
Overview
In this section, we present our findings from the PTB risk estimation enabled by longitudinal HR and HRV data collected during pregnancy. We first compare PTB abnormality scores between preterm and term individuals. Subsequently, we indicate the significance and contribution of the HR and HRV features in estimating the risk of PTB. Finally, we evaluate the performance of our model using a cross-validation approach.

Comparing PTB abnormality scores between preterm and term individuals
We utilize the autoencoder model—trained on the data from forty-four mothers with term pregnancy—to obtain the PTB abnormality scores for seven mothers who experienced PTB and seven with term deliveries. The preterm deliveries occurred at gestational weeks 32, 33, 34 (three individuals), and 36 (four individuals), while the term deliveries occurred at gestational weeks 39 (two individuals), 40 (three individuals), and 41 (two individuals). The aim is to differentiate between these two user categories by exploring the PTB abnormality score across the two groups.

Figure 3 illustrates the weekly average PTB abnormality scores for preterm users (red lines) and term users (blue lines) during the second trimester (gestational week 13 to week 27) and the third trimester (gestational week 27 to week 37). The breaks in the lines are due to missing data. The abnormality scores of five mothers with PTB are notably higher than mothers with term deliveries. This difference in the abnormality scores was observed from the early weeks of the second trimester. The scores of the five mothers remained elevated until their labors. For the remaining two mothers with PTB, the scores were relatively low in the second trimester but increased considerably throughout the pregnancy.

The average of PTB abnormality scores of the two groups is indicated in Fig. 4. The averages are calculated from gestational weeks 13 to 37. As indicated, the average abnormality scores of preterm users is considerably higher than the scores of term users. For preterm users, starting around 0.06 in week 13, with a sharp decline in week 17, the average PTB abnormality score remains relatively consistent until the end of the second trimester (week 27). This stability persists into the early stages of the third trimester and then increases in week 34. In contrast, for the term users, the initial average PTB abnormality score is around 0.03 in week 13, dropping below 0.03 in week 15, and maintains this level throughout the entire second trimester, reaching 0.03 again in the last week (week 37). Overall, the average abnormality score for term users is around 0.03, whereas for preterm users, it is approximately 0.055.

To indicate the differences between the two groups, we also present the scores' statistical properties and distribution via kernel density estimation (KDE)[75] plot in Table 3 and Fig. 5, respectively. The Preterm class
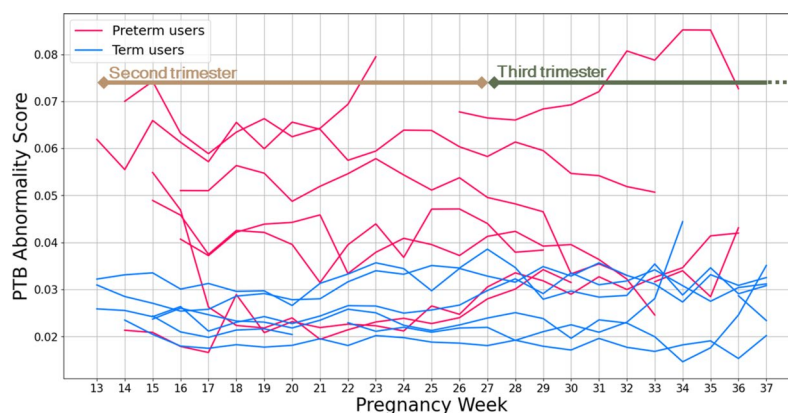

Fig. 3. Weekly average abnormality scores for 7 pregnant women with PTB and 7 pregnant women with term delivery during the second and third trimesters.
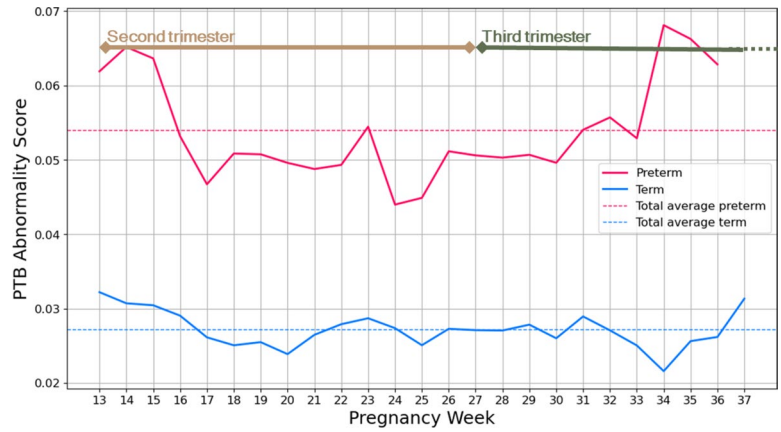
Fig. 4. Average of weekly average PTB abnormality scores of preterm and term users.

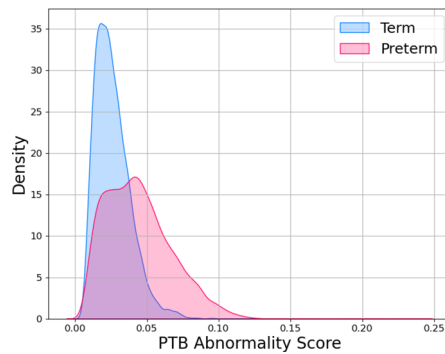| Class	Mean	Median	Std deviation	Q1	Q3
Term	0.026	0.024	0.013	0.017	0.033
Preterm	0.044	0.042	0.023	0.026	0.058

Table 3. Summary Statistics for Abnormality Scores. T-test P-value < 0.0001.

Fig. 5. Distribution of PTB abnormality score for term and preterm classes.

exhibits higher scores compared to the term class. The KDE plot for term users reveals a concentration of scores around the mean, with a narrower spread. In contrast, the KDE plot for preterm users shows a broader, right-shifted, and positively skewed distribution, indicating a wider range of PTB abnormality scores. Additionally, we conducted a two-independent sample t-test to directly compare the PTB abnormality scores between these two groups, resulting in a P-value below 0.0001, which underlines the significant difference in PTB abnormality scores between these two groups. This distinction highlights the potential of our PTB abnormality score – obtained from smartwatches HR and HRV – to discriminate between the term and preterm groups.

Feature analysis: importance and SHAP explanibility
Feature importance analysis
We assess the contribution of each HR and HRV feature in estimating the PTB abnormality scores. To this end, we compute the abnormality score for each HR and HRV feature separately. As outlined in Eq. 4, the abnormality score for a sample is calculated using the RMSE between the original and reconstructed HR and HRV features of that sample. To calculate the abnormality score for each specific feature, we measure the difference between the original and reconstructed values of the feature. A higher abnormality score for a feature indicates greater importance in distinguishing between the two classes.

Figure 6 demonstrates the PTB abnormality score for each HR and HRV feature for preterm (in red) and term (in blue) samples. The findings indicate that for certain features such as HR, AVNN, and SD1SD2, the abnormality scores for preterm samples are significantly higher than those of term samples. Specifically, for HR values below 70, AVNN values exceeding 800, and SD1SD2 values higher than around 1, notable differences are observed in the PTB abnormality score of the two classes, highlighting the substantial importance of these
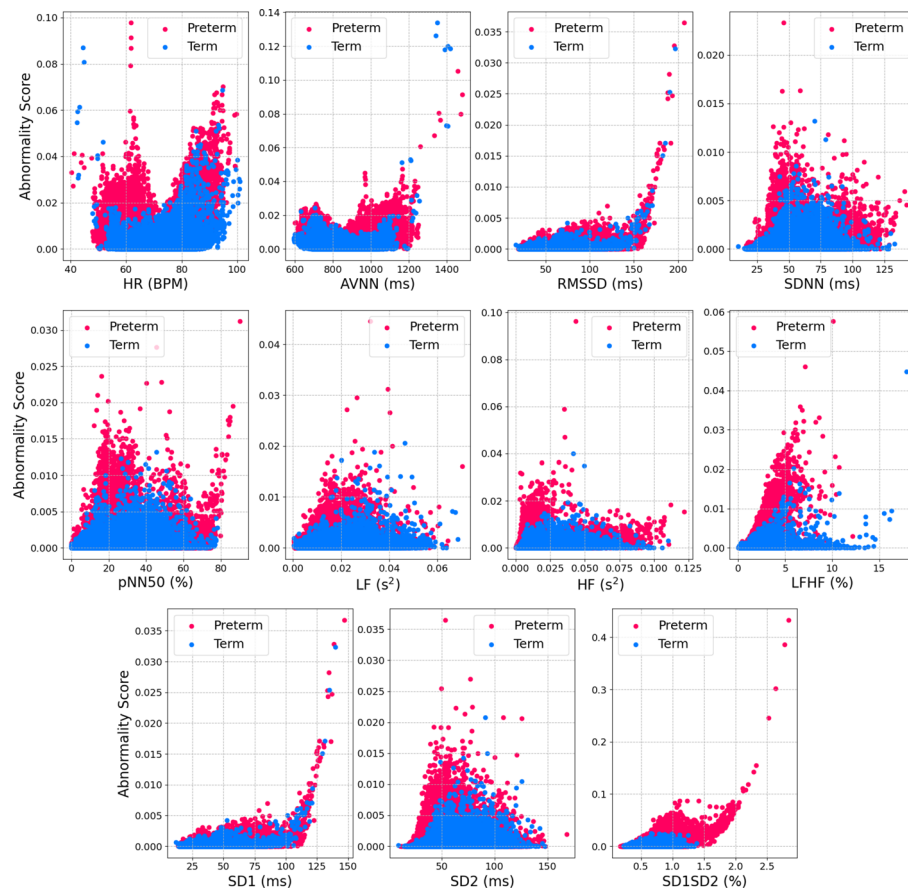
Fig. 6. PTB abnormality score of HR and HRV features for term and preterm samples.

features in distinguishing preterm samples from term samples. In contrast, the difference in abnormality score between the two groups is relatively less considerable for RMSSD and SD1, indicating the comparatively lower significance of these two features in our approach. It should be noted that these two features are nearly identical since they are derived from the time intervals between adjacent peak-to-peak intervals in the signal[76].

Moreover, we calculate the importance of each HR and HRV feature separately. To this end, we repeat the experiments by intentionally removing a feature and calculating the abnormality scores. Therefore, we can assess the feature's influence on the PTB abnormality score. Features that lead to a significant decrease in the abnormality score upon removal are considered more important. To ensure a meaningful comparison, the feature importance values are normalized between 0 and 1, where an importance value of 1 corresponds to the most significant feature, and an importance of 0 represents the lowest significance.

Figure 7 illustrates the importance of each HR and HRV feature. HR is the feature with the highest degree of influence, showing an importance value of 1. Furthermore, AVNN and SD1SD2 exhibit considerable importance, with values of 0.59 and 0.5, respectively. In contrast, SD1 and RMSSD features demonstrate relatively lower importance, indicating their limited impact on the PTB abnormality score.

SHAP explainability for PTB samples
We employ SHapley Additive exPlanations (SHAP) framework[77], as an explainability method, to evaluate the influence of feature interactions and combinations on the model's output[78]. Ranking features based on the individual impact on the model's output (i.e., abnormality score in our case) might overlook the complex interaction and combinations between features.

SHAP is an approach in explainable Artificial intelligence (AI) that breaks down a model's prediction by measuring the individual contributions of each feature, providing transparent explanations for predictions. It assigns a Shapley value to each feature, indicating its contribution to a specific prediction. The method we adopt introduced by Antwarg et al.[79], which is specifically tailored to explain anomalies detected by the autoencoder. It focuses on the connection between the features with high reconstruction error and those wielding the most effect on the reconstruction error of the high-error features. Using this method, we identify features that contribute the most to preterm samples.

Figure 8 represents the mean SHAP values corresponding to each feature for preterm samples. A higher SHAP value attributed to a feature indicates that the given feature exhibits a more significant impact in pushing the sample toward the Preterm class. As illustrated, HR has the highest SHAP value. SD1SD2, SDNN, and SD2 also are the following features with notable SHAP values, showing their substantial contributions. In contrast,
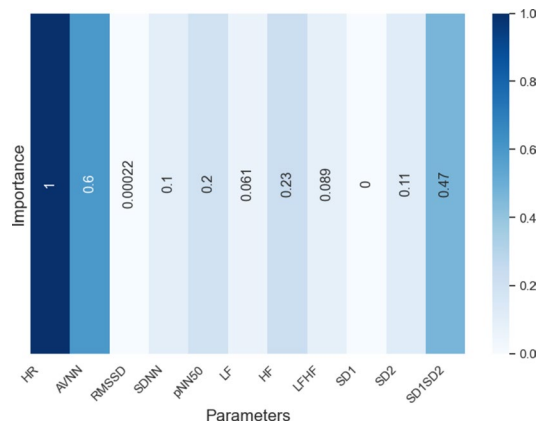
Fig. 7. The importance of each HR and HRV feature in PTB abnormality score.
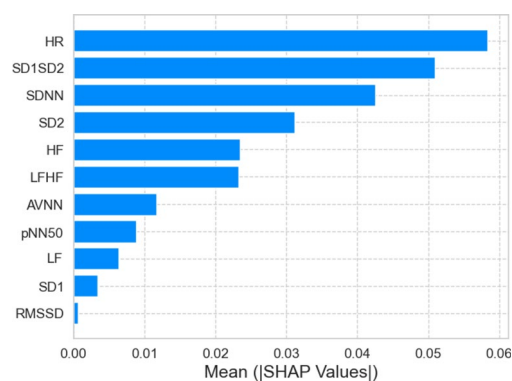

Fig. 8. Mean SHAP values of each HR and HRV feature for preterm samples.

RMSSD and SD1 demonstrate the least influence in pushing samples toward Preterm class, as indicated by their minimal SHAP values.

Moreover, we demonstrate how high-impact features influence the model's output while considering their interactions with other features. Figure 9 shows the SHAP dependence plots for the four features with the highest SHAP values: HR, SD1SD2, SDNN, and SD2. On these plots, the X-axis represents feature values, while the Y-axis represents SHAP values. The color bar demonstrates the values of a secondary feature, which has the most significant interaction with the feature of interest. Our findings reveal that this secondary feature is LFHF, which emerges as the feature with the most influential interaction with the high-SHAP features.

As shown in the dependence plots, when HR values increase and are combined with elevated LFHF values, it results in high SHAP values. In the case of SD1SD2, interacting moderate values of this feature, around 0.6, with high values of LFHF, yields significant SHAP values. Furthermore, for SDNN, both low and high LFHF values interacting with SDNN lead to SHAP values for SDNN averaging less than 0.05. Similarly, when both low and high LFHF values are paired with SD2, this combination results in SHAP values for SD2 of around 0.03.

Model performance evaluation

To evaluate the robustness of our model, we perform a 5-fold cross-validation with 51 term users. This approach partitions the term data into five folds. In each iteration, four folds are used to train the model, and the remaining fold is used to validate. This process is repeated five times, with each fold used once as the validation set. It is important to note that the preterm data, consisting of 7 preterm users, remain unchanged and are preserved for validation in all iterations.

Figure 10 presents the ROC curves and the AUC values for each fold, with the mean of the cross-validation. Ideally, a perfect ROC curve would have a TPR of 1 (top of the plot) and an FPR of zero (left of the plot). The curves demonstrate that the model consistently performs well across all folds, distinguishing between the term and preterm classes. The AUC values obtained for each fold are 0.79, 0.83, 0.78, 0.75, and 0.81, respectively, with a mean AUC of 0.79. The consistency of the AUC values across the folds indicates that the model performs reliably across different subsets of the term data for training, demonstrating its robustness.

In addition, we evaluate the classification metrics, including accuracy, precision, recall, and F1-score for each fold. The results are summarized in Table 4. The mean accuracy across the folds is 0.75, with precision, recall, and F1-score values of 0.74, 0.74, and 0.73, respectively. These metrics indicate consistent performance across different subsets of the term data for training and validation, demonstrating the model's robustness.
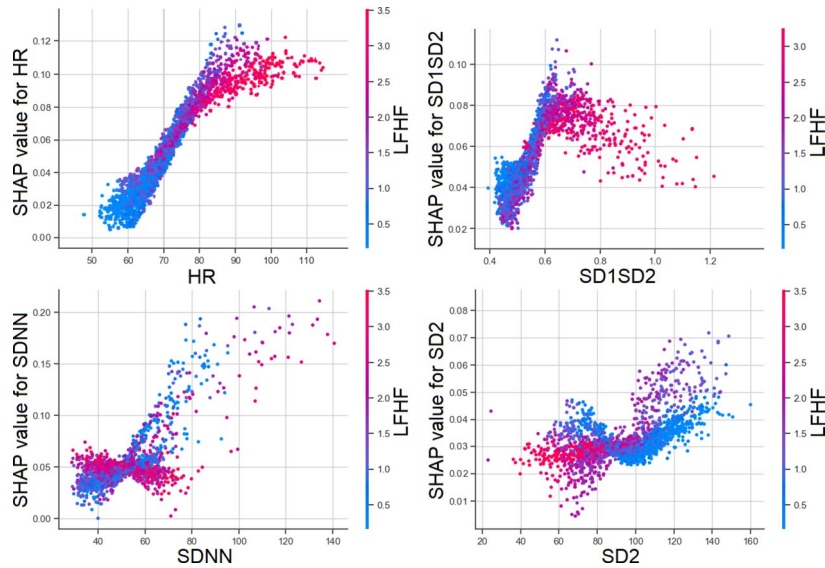
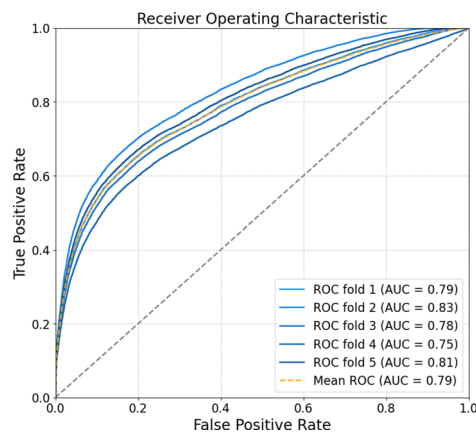**Fig. 9.** Dependence plot of four HR and HRV features with the highest SHAP value.

Fig. 10. Receiver operating characteristic curves of 5-fold cross-validation.

Fold	Accuracy	Precision	Recall	F1-Score
Fold 1	0.74	0.74	0.73	0.73
Fold 2	0.77	0.77	0.75	0.75
Fold 3	0.74	0.73	0.74	0.72
Fold 4	0.73	0.71	0.72	0.70
Fold 5	0.75	0.75	0.74	0.74
Mean	0.75	0.74	0.74	0.73

Table 4. Classification metrics for each fold and mean of cross-validation.

Discussion
Principal findings
To the best of our knowledge, our study is the first effort to explore the risk of PTB by continuously monitoring maternal ANS function through HR and HRV features in everyday activities during pregnancy. Our findings revealed different patterns in the PTB abnormality scores during the second and third trimesters of pregnancy. Significantly, the early second trimester exhibited a substantial difference in abnormality scores between preterm and term users. During this period, five preterm users consistently demonstrated higher abnormality

scores compared to their term counterparts. This early differentiation highlights the potential of our approach to estimating PTB risk in the relatively early stages of pregnancy, offering a valuable timeframe for possible intervention and support.

According to the existing literature, various HRV metrics provide valuable insights into the functioning of the ANS, differentiating between sympathetic and parasympathetic influences. Specifically, LF and SD2 are typically associated with sympathetic activity, reflecting stress responses and the balance between sympathetic and parasympathetic modulation. HF, RMSSD, AVNN, PNN50, and SD1 are indicative of parasympathetic (vagal) activity relating to relaxation, recovery processes, and the body's ability to manage stress through vagal tone. The SDNN, LF/HF, and SD1/SD2 ratios encompass overall HRV, representing contributions from both the sympathetic and parasympathetic branches of the ANS[76,80].

In our study, illustrated in Fig. 6, it is evident that contributing to abnormality score, all HRV features for both sympathetic and parasympathetic nervous systems are higher in preterm births compared to term births. This may suggest an increase in both sympathetic and parasympathetic activity in preterm pregnancies. Elevated sympathetic activity in preterm births may reflect a heightened stress response, whereas increased parasympathetic activity suggests enhanced vagal tone, possibly as a compensatory mechanism. The overall higher HRV in preterm births suggests a complex interaction between these two branches of the ANS, highlighting the importance of HRV analysis in understanding autonomic regulation and its implications in healthy pregnancy outcomes.

Analyzing the importance of individual HR and HRV features improved understanding of the specific physiological factors influencing PTB. HR appeared to be the most influential feature, indicating its critical role in discriminating between term and PTB samples during continuous and long-term monitoring. Additionally, features such as AVNN and SD1SD2 exhibited vital importance, highlighting the significance of the average of interbeat intervals and the unpredictability of the interbeat intervals in the occurrence of PTB, respectively. Regarding the interaction between the features, HR, SD1SD2, SDNN, and SD2 emerged as high-impact features when combined with other features, pushing a sample toward the PTB class. Furthermore, a significant interaction was observed between these high-impact features and LFHF, the ratio between the sympathetic nervous system and the parasympathetic nervous system.

In understanding the sociodemographic impact, in our results, preterm pregnancies were associated with a significantly higher average maternal age (36.4 years, range 30–41) compared to term pregnancies (31.4 years, range 23–43). The lower HR in older ages, identified as a top contributor to the abnormality score, may offer a potential explanation for the association between age and preterm birth risk. Evidence suggests that as maternal age increases, there is a decrease in cardiac output and an increase in peripheral vascular resistance, resulting in a decreased heart rate. This association is due to physiological changes such as decreased elasticity of blood vessels and altered autonomic nervous system regulation, which are more pronounced in older pregnant individuals[81,82].

Regarding BMI, higher BMI is linked to increased sympathetic nervous system activity, greater cardiovascular workload, and chronic low-grade inflammation, all of which can elevate resting HR. Individuals with higher BMI often have increased metabolic demands and vascular resistance, necessitating the heart to pump more frequently leading to elevated resting HR[83].

In our study, we examined the relationship between maternal age, BMI, and HR using ordinary least squares (OLS) regression and mixed linear models. The OLS regression analysis indicated a statistically significant but very weak negative association between maternal age and HR, with an R-squared value of 0.008. This suggests a slight decrease in HR as maternal age increases. However, when using the mixed linear model, which accounts for repeated measurements within individuals, the relationship was statistically insignificant, indicating that maternal age is not a strong predictor of HR in our dataset. Similarly, the OLS regression for BMI showed a statistically significant but weak positive association with HR, with an R-squared value of 0.009, suggesting a slight increase in HR with higher BMI. This relationship also proved to be statistically insignificant in the mixed linear model. These results imply that while there are minor trends of decreasing HR with advancing maternal age and increasing BMI, neither maternal age nor BMI are strong predictors of HR during pregnancy in our study.

Comparison with previous studies

Several studies explored maternal HR and HRV to investigate pregnancy complications and establish associations. For instance, Kataoka et al.[27] explored HRV data in postpartum women with a history of recurrent pregnancy loss (RPL), revealing significantly lower levels of HRV features such as HF, SDNN, and total power in women with RPL compared to control subjects. In our study, HF emerged as the fourth important feature, and SDNN ranked third with a notable SHAP value in estimating PTB among the 11 features considered. While Kataoka et al. observed no significant difference in the LFHF ratio between women with RPL and controls, our findings highlighted LFHF as the feature with the most influential interaction with high-SHAP features in estimating PTB. In addition, we explored trends in nocturnal maternal HR and HRV of this longitudinal health monitoring study in another work[39]. In Ref.[39], no significant differences were found in HR and HRV of high-risk and low-risk pregnant women. However, in this paper, we aimed to estimate PTB risk by exploring the association between the ANS and PTB. We showed that long-term maternal HR and HRV exhibited a substantial difference in abnormality scores between preterm and term users.

Moreover, exploratory findings from the study by Sharifiheris et al.[61] demonstrated maternal HRV differences between healthy pregnant women and one participant who experienced a placental rupture in the early third trimester (34 weeks). They discovered significant differences in RMSSD between this case and healthy participants, starting from the study's onset at 23 weeks of gestational age until the complication occurrence at 34 weeks. However, in our study, RMSSD showed the least influence in estimating PTB. Furthermore, Jasinski et al.[63] explored maternal HRV trends and their association with gestational age and time until birth. While this study showed a strong fit for weeks until birth in predicting maternal HRV changes in both term and preterm

pregnancies, our study extends this understanding by discovering distinct patterns in PTB abnormality scores of term and preterm pregnancies based on multiple maternal HRV parameters.

Limitations

A limitation of our study is the small sample size, comprising only 28 high-risk pregnant women, seven of whom experienced PTB. This limited number of preterm cases in our dataset may impact the generalizability of our findings and the robustness of our PTB risk estimation approach. Moreover, it may restrict the exploration of patterns and variations within the long-term HR and HRV data from the preterm cases. Our future work will consider recruiting a more diverse cohort of individuals with high-risk pregnancies.

Another limitation lies in the susceptibility of the PPG technique integrated into smartwatches to motion artifacts and noise. Although the PPG signals are convenient for continuous monitoring, PPG signals – and subsequently the derived HR and HRV – may be distorted due to artifacts caused by the user's physical activity or environmental noises. In this study, we addressed this problem by leveraging a deep learning-based PPG processing pipeline including filtering, signal denoising, signal quality assessment, and systolic peak detection methods[67]. The future work in this direction should focus on developing techniques to further mitigate the impact of noise on PPG signals and enhance the accuracy of HR and HRV parameters.

Conclusion

In this study, we conducted a PTB risk estimation by exploring long-term continuous monitoring of maternal physiological parameters, i.e., HR and several HRV features. We extracted these features from longitudinal PPG data collected from 58 pregnant women, including seven preterm cases using wearable devices in free-living conditions. Our findings showed distinctive patterns in PTB abnormality scores obtained from our trained autoencoder model during the second trimester. These findings indicate a critical window for timely interventions and support to mitigate PTB. Moreover, our results highlighted the significance of HR, AVNN, SD1SD2, and SDNN as crucial PTB indicators. Consequently, this study underscored the potential of longitudinal continuous monitoring of ANS activity reflected in HR and HRV for early PTB risk estimation.

Data availability

The data utilized in this study include sensitive health information, and the informed consent signed by the participants strictly prohibits the public release of the data due to ethical constraints. According to the current approval by the Ethics Committee of the Hospital District of Southwest Finland, the participants explicitly granted permission to use the collected data solely for the purposes outlined in the consent agreement. Requests for data access are subject to individual consent and require ethics committee approval. Researchers interested in accessing the data should reach out to the Ethics Committee of the Hospital District of Southwest Finland, with contact details as follows: Tyks U-hospital, Kiinamyllynkatu 4-8, UB3, PO Box 52, FI-20521 TURKU, Finland; email: eettinen.toimikunta@tyks.fi. Additionally, we recommend initiating contact with the Principal Investigator of the research project, Associate Professor Anna Axelin, at the University of Turku, Department of Nursing Science, 20014 University of Turku, Finland; email: anmaax@utu.fi.

References

1. World Health Organization. Preterm birth. https://www.who.int/news-room/fact-sheets/detail/preterm-birth (2018). (Accessed January 2024).
2. World Health Organization. Born too soon: the global action report on preterm birth. https://iris.who.int/bitstream/handle/10665/44864/9789244503430_rus.pdf (2012). (Accessed January 2024).
3. World Health Organization. *Born too Soon: Decade of Action on Preterm Birth* (World Health Organization, 2023).
4. Manuck, T. A. et al. Preterm neonatal morbidity and mortality by gestational age: A contemporary cohort. *Am. J. Obstet. Gynecol.* **215**, 103.e1 (2016).
5. Lean, R. E., Rogers, C. E., Paul, R. A. & Gerstein, E. D. Nicu hospitalization: Long-term implications on parenting and child behaviors. *Curr. Treat. Options Pediatrics* **4**, 49–69 (2018).
6. Cheah, I. G. S. Economic assessment of neonatal intensive care. *Trans. Pediatrics* **8**, 246 (2019).
7. Treyvaud, K. & Brown, S. J. Mental health of children and parents after very preterm birth. *World Psychiatry* **21**, 148 (2022).
8. Wallois, F., Routier, L. & Bourel-Ponchel, E. Impact of prematurity on neurodevelopment. In *Handbook of Clinical Neurology* Vol. 173 (ed. Wallois, F.) 341–375 (Elsevier, 2020).
9. Goldenberg, R. L., Culhane, J. F., Iams, J. D. & Romero, R. Epidemiology and causes of preterm birth. *Lancet* **371**, 75–84 (2008).
10. Di Renzo, G. C. et al. Guidelines for the management of spontaneous preterm labor: Identification of spontaneous preterm labor, diagnosis of preterm premature rupture of membranes, and preventive tools for preterm birth. *J. Matern.-Fetal Neonatal Med.* **24**, 659–667 (2011).
11. Menon, R. Spontaneous preterm birth, a clinical dilemma: Etiologic, pathophysiologic and genetic heterogeneities and racial disparity. *Acta Obstet. Gynecol. Scand.* **87**, 590–600 (2008).
12. Gyamfi-Bannerman, C., Fuchs, K. M., Young, O. M. & Hoffman, M. K. Nonspontaneous late preterm birth: Etiology and outcomes. *Am. J. Obstet. Gynecol.* **205**, 456.e1 (2011).
13. Harrison, M. S. et al. Pathways to preterm birth: Case definition and guidelines for data collection, analysis, and presentation of immunization safety data. *Vaccine* **34**, 6093 (2016).
14. Sharifi-Heris, Z. et al. Machine learning approach for preterm birth prediction using health records: Systematic review. *JMIR Med. Inform.* **10**, e33875 (2022).
15. Raja, R., Mukherjee, I. & Sarkar, B. K. A machine learning-based prediction model for preterm birth in rural India. *J. Healthc. Eng.* **2021**, 6665573 (2021).
16. Sun, Q. et al. Machine learning-based prediction model of preterm birth using electronic health record. *J. Healthc. Eng.* **2022**, 9635526 (2022).

17. Zhang, Y., Lu, S., Wu, Y., Hu, W. & Yuan, Z. The prediction of preterm birth using time-series technology-based machine learning: Retrospective cohort study. *JMIR Med. Inform.* **10**, e33835 (2022).

18. Devers, K. et al. *The Feasibility of Using Electronic Health Data for Research on Small Populations* (Assistant Secretary for Planning and Evaluation (ASPE), 2013).

19. Selvaraju, V., Karthick, P. & Swaminathan, R. Analysis of frequency bands of uterine electromyography signals for the detection of preterm birth. In MIE, 283–287 (2021).

20. Fergus, P., Idowu, I., Hussain, A. & Dobbins, C. Advanced artificial neural network classification for detecting preterm births using ehg records. *Neurocomputing* **188**, 42–49 (2016).

21. Chen, L., Hao, Y. & Hu, X. Detection of preterm birth in electrohysterogram signals based on wavelet transform and stacked sparse autoencoder. *PLoS ONE* **14**, e0214712 (2019).

22. Waxenbaum, J. A., Reddy, V. & Varacallo, M. *Anatomy, Autonomic Nervous System* (StatPearls Publishing, 2019).

23. Shaffer, F., McCraty, R. & Zerr, C. L. A healthy heart is not a metronome: An integrative review of the heart's anatomy and heart rate variability. *Front. Psychol.* **5**, 1040 (2014).

24. Mulcahy, J. S., Larsson, D. E., Garfinkel, S. N. & Critchley, H. D. Heart rate variability as a biomarker in health and affective disorders: A perspective on neuroimaging studies. *Neuroimage* **202**, 116072 (2019).

25. Arakaki, X. et al. The connection between heart rate variability (hrv), neurological health, and cognition: A literature review. *Front. Neurosci.* **17**, 1055445 (2023).

26. Lu, W.-C. et al. Correlation between health-related quality of life in the physical domain and heart rate variability in asymptomatic adults. *Health Qual. Life Outcomes* **14**, 1–10 (2016).

27. Kataoka, K. et al. Altered autonomic nervous system activity in women with unexplained recurrent pregnancy loss. *J. Obstet. Gynaecol. Res.* **41**, 912–918 (2015).

28. Niela-Vilen, H. et al. Comparing prenatal and postpartum stress among women with previous adverse pregnancy outcomes and normal obstetric histories: A longitudinal cohort study. *Sex. Reprod. Healthc.* **35**, 100820 (2023).

29. Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H. & Koo, B.-H. Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investig.* **15**, 235 (2018).

30. Henje Blom, E., Olsson, E. M., Serlachius, E., Ericson, M. & Ingvar, M. Heart rate variability is related to self-reported physical activity in a healthy adolescent population. *Eur. J. Appl. Physiol.* **106**, 877–883 (2009).

31. Khatibi, E. et al. Impact of covid-19 pandemic on sleep including hrv and physical activity as mediators: A causal ml approach. medRxiv 2023–06 (2023).

32. Jafarlou, S. et al. Objective monitoring of loneliness levels using smart devices: A multi-device approach for mental health applications. medRxiv 2023–06 (2023).

33. Murgia, F. et al. Effects of smoking status, history and intensity on heart rate variability in the general population: The chris study. *PLoS ONE* **14**, e0215053 (2019).

34. Triggiani, A. I. et al. Heart rate variability reduction is related to a high amount of visceral adiposity in healthy young women. *PLoS ONE* **14**, e0223058 (2019).

35. Qiu, H. et al. Increased vulnerability to atrial and ventricular arrhythmias caused by different types of inhaled tobacco or marijuana products. *Heart Rhythm* **20**, 76–86 (2023).

36. Yousif, D. et al. Autonomic dysfunction in preeclampsia: A systematic review. *Front. Neurol.* **10**, 816 (2019).

37. Sharifi-Heris, Z. et al. Heart rate variability and pregnancy complications: Systematic review. *Interact. J. Med. Res.* **12**, e44430 (2023).

38. Alim, A. & Imtiaz, M. H. Wearable sensors for the monitoring of maternal health—a systematic review. *Sensors* **23**, 2411 (2023).

39. Sarhaddi, F. et al. 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* **10**, e33458 (2022).

40. Castaneda, D., Esparza, A., Ghamari, M., Soltanpur, C. & Nazeran, H. A review on wearable photoplethysmography sensors and their potential future applications in health care. *Int. J. Biosensors Bioelectron.* **4**, 195 (2018).

41. Oti, O. et al. Iot-based healthcare system for real-time maternal stress monitoring. In: Proc. 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, 57–62 (2018).

42. Lopez, B. D. B., Aguirre, J. A. A., Coronado, D. A. R. & Gonzalez, P. A. Wearable technology model to control and monitor hypertension during pregnancy. In *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)* (ed. Lopez, B. D. B.) 1–6 (IEEE, 2018).

43. Sarhaddi, F. et al. Maternal social loneliness detection using passive sensing through continuous monitoring in everyday settings: Longitudinal study. *JMIR Format. Res.* **7**, e47950 (2023).

44. Saarikko, J. et al. Supporting lifestyle change in obese pregnant mothers through the wearable internet-of-things (slim)-intervention for overweight pregnant women: Study protocol for a quasi-experimental trial. *PLoS ONE* **18**, e0279696 (2023).

45. Azimi, I. et al. Personalized maternal sleep quality assessment: An objective iot-based longitudinal study. *IEEE Access* **7**, 93433–93447 (2019).

46. Galea, J. T. et al. The use of wearable technology to objectively measure sleep quality and physical activity among pregnant women in urban Lima, Peru: A pilot feasibility study. *Matern. Child Health J.* **24**, 823–828 (2020).

47. Ullah, F. et al. A framework for maternal physical activities and health monitoring using wearable sensors. *Sensors* **21**, 4949 (2021).

48. Yang, C., Antoine, C., Young, B. K. & Tavassolian, N. A pilot study on fetal heart rate extraction from wearable abdominal inertial sensors. *IEEE Sens. J.* **19**, 10773–10781 (2019).

49. Skrivanos, A. G. et al. Fetus heart rate monitoring: A preliminary research study with remote sensing. *IEEE Consumer Electron. Magaz.* **11**, 32–44 (2021).

50. Zhang, Y. et al. Wearable fetal ecg monitoring system from abdominal electrocardiography recording. *Biosensors* **12**, 475 (2022).

51. Delay, U. et al. Non invasive wearable device for fetal movement detection. In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)* (ed. Delay, U.) 285–290 (IEEE, 2020).

52. Du, Y.-C., Yen, L. B., Kuo, P.-L. & Tsai, P.-Y. A wearable device for evaluation of relative position, force, and duration of fetal movement for pregnant woman care. *IEEE Sens. J.* **21**, 19341–19350 (2021).

53. Kuncoro, C. B. D., Efendi, A. & Sakanti, M. M. Wearable sensor for psychological stress monitoring of pregnant woman-state of the art. Measurement 113556 (2023).

54. Arabi Belaghi, R., Beyene, J. & McDonald, S. D. Prediction of preterm birth in nulliparous women using logistic regression and machine learning. *PLoS ONE* **16**, e0252025 (2021).

55. Chen, H.-Y., Chuang, C.-H., Yang, Y.-J. & Wu, T.-P. Exploring the risk factors of preterm birth using data mining. *Expert Syst. Appl.* **38**, 5384–5387 (2011).

56. Alexandersson, A., Steingrimsdottir, T., Terrien, J., Marque, C. & Karlsson, B. The icelandic 16-electrode electrohysterogram database. *Sci. Data* **2**, 1–9 (2015).

57. Euliano, T. Y. et al. Monitoring uterine activity during labor: A comparison of 3 methods. *Am. J. Obstet. Gynecol.* **208**, 66.e1 (2013).

58. Almeida, M. et al. Electrohysterography extracted features dependency on anthropometric and pregnancy factors. *Biomed. Signal Process. Control* **75**, 103556 (2022).

59. Włodarczyk, T. et al. Spontaneous preterm birth prediction using convolutional neural networks. In *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis: First International Workshop, ASMUS 2020, and 5th International Workshop,*

PIPPI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 1 (ed. Włodarczyk, T.) 274–283 (Springer, 2020).

60. Lu, L. & Huang, T. Effects of early nursing monitoring on pregnancy outcomes of pregnant women with gestational diabetes mellitus under internet of things. *Computat. Math. Methods Med.* **2022**, 8535714 (2022).
61. Sharifiheris, Z. et al. Phenotyping autonomic nervous system in pregnancy using remote sensors: Potential for complication prediction. *Front. Physiol.* **14**, 1293946 (2023).
62. Oura ring. https://ouraring.com/. (Accessed January 2024).
63. Jasinski, S. R., Rowan, S., Presby, D. M., Claydon, E. & Capodilupo, E. R. Wearable-derived maternal heart rate variability as a novel digital biomarker of preterm birth. medRxiv 2022–11 (2022).
64. WHOOP. WHOOP Strap. https://www.whoop.com/. (Accessed January 2024).
65. Gear sport 42mm smartwatch (bluetooth). https://www.samsung.com/us/mobile/wearables/smartwatches/gear-sport-blue-sm-r600nzbaxar/. (Accessed January 2024).
66. Sarhaddi, F. et al. Long-term iot-based maternal monitoring: System design and evaluation. *Sensors* **21**, 2281 (2021).
67. Feli, M. et al. End-to-end ppg processing pipeline for wearables: From quality assessment and motion artifacts removal to hr/hrv feature extraction. In IEEE International Conference on Bioinformatics and Biomedicine (BIBM'23) (2023).
68. Feli, M., Azimi, I., Anzanpour, A., Rahmani, A. M. & Liljeberg, P. An energy-efficient semi-supervised approach for on-device photoplethysmogram signal quality assessment. *Smart Health* **28**, 100390 (2023).
69. Wang, Y., Azimi, I., Kazemi, K., Rahmani, A. M. & Liljeberg, P. Ppg signal reconstruction using deep convolutional generative adversarial network. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (ed. Wang, Y.) 3387–3391 (IEEE, 2022).
70. Kazemi, K., Laitala, J., Azimi, I., Liljeberg, P. & Rahmani, A. M. Robust ppg peak detection using dilated convolutional neural networks. *Sensors* **22**, 6054 (2022).
71. Khan, S. S. & Taati, B. Detecting unseen falls from wearable devices using channel-wise ensemble of autoencoders. *Expert Syst. Appl.* **87**, 280–290 (2017).
72. Amarbayasgalan, T., Lee, J. Y., Kim, K. R. & Ryu, K. H. Deep autoencoder based neural networks for coronary heart disease risk prediction. In *VLDB Workshop on Data Management and Analytics for Medicine and Healthcare* (ed. Amarbayasgalan, T.) 237–248 (Springer, 2019).
73. Chandola, V., Banerjee, A. & Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. (CSUR)* **41**, 1–58 (2009).
74. Rainio, O., Teuho, J. & Klén, R. Evaluation metrics and statistical tests for machine learning. *Sci. Rep.* **14**, 6086 (2024).
75. Stanisław, W. Kernel density estimation and its application. *ITM Web Conf.* **23**, 00037 (2018).
76. Shaffer, F. & Ginsberg, J. P. An overview of heart rate variability metrics and norms. *Front. Public Health* **5**, 258 (2017).
77. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. Advances in neural information processing systems **30** (2017).
78. Horel, E. *Towards Explainable AI: Feature Significance and Importance for Machine Learning Models* (Stanford University, 2020).
79. Antwarg, L., Miller, R. M., Shapira, B. & Rokach, L. Explaining anomalies detected by autoencoders using shapley additive explanations. *Expert Syst. Appl.* **186**, 115736 (2021).
80. Pham, T., Lau, Z. J., Chen, S. A. & Makowski, D. Heart rate variability in psychology: A review of hrv indices and an analysis tutorial. *Sensors* **21**, 3998 (2021).
81. Ling, H., Garcia Jara, P., Nicolaides, K. & Kametas, N. Effect of maternal age on cardiac adaptation in pregnancy. *Ultrasound Obst. Gynecol.* **58**, 285–292 (2021).
82. Ogliari, G. et al. Resting heart rate, heart rate variability and functional decline in old age. *CMAJ* **187**, E442–E449 (2015).
83. Plaza-Florido, A., Ruiz, J. R. & Alcantara, J. M. Resting heart rate but not heart rate variability is associated with the normal-weight obesity phenotype. *Am. J. Hum. Biol.* **36**, e24043 (2024).

Acknowledgements

Author contributions

MF performed data analysis, conducted experiments, and drafted the manuscript; IA contributed to study design, data collection, and data analysis and played a major role in manuscript revision; FS participated in study design, data collection, and assisted in manuscript drafting; ZSH contributed to both drafting and revising the manuscript; HNV and AA were major contributors to study design, participants' recruitment, and manuscript revision; and PL and AR contributed to study design, data collection, and assisted in manuscript revision.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.