Employing Support Vector Machine Regression to Estimate the Fetal Gestational Age

Maisam Wahbah[1], Raghad Al Sakaji[2], Kiyoe Funamoto[3], Anita Krishnan[4], Yoshiyuki Kasahara[5,6], Yoshitaka Kimura[5], Ahsan H. Khandoker[3]

[1] College of Engineering & Information Technology, University of Dubai, Dubai 14143, UAE
[2] Department of Engineering Systems and Management, Khalifa University, Abu Dhabi 127788, UAE
[3] Health Engineering Innovation Center (HEIC), Khalifa University, Abu Dhabi 127788, UAE
[4] Division of Cardiology, Children's National Hospital, Washington, DC, United States
[5] Department of Maternal and Fetal Therapeutics, Tohoku University, Sendai, Japan
[6] Department of Maternal and Child Health Care Medical Science, Tohoku University, Sendai, Japan

Abstract

The accurate estimation of the Gestational Age (GA) in fetal development studies has the potential to detect health issues at early stages of pregnancy. In this article, we adopt the Support Vector Machine (SVM) tool to investigate whether gold standard GA can be reliably estimated by using maternal as well as fetal Heart Rate Variability (HRV) features. The study considered Electrocardiogram (ECG) signals from 60 pregnant women. Maternal and fetal HRV parameters were calculated, and SVM regression with the linear kernel function was utilized to produce a robust estimate of fetal age. By evaluating the cross-validation performances, we found that maternal electrophysiological parameters contribute to the correct estimation of the GA. Results showed that the linear kernel maintains better performance over the radial basis function kernel in the SVM-based regression models. Compared with gold standard GA identified by CRL, the proposed model resulted in an error of 5.11 weeks, Bland–Altman estimated bias of -0.31 weeks and limits of agreement of 8.97 and -9.59 weeks, and Pearson correlation coefficient of 0.63. It can be speculated that the fetal GA can be more reliably estimated when incorporating maternal along with fetal HRV parameters using 1 min of ECG signals.

1. Introduction

The estimation error when using the Crown-Rump Length (CRL) to measure the Gestation Age (GA) of the fetus can reach up to 7 days [1]. Some of the challenges associated with CRL can include human errors and the requirement to have good clinical practice [2]. Such requirements, however, might not be feasible in some settings, and it is thus required to have an approach that is more robust when estimating the GA while mitigating the challenges.

It is reported that fetal growth can be estimated using the Fetal Heart Rate (FHR) and its variability [3]. One advantage of this method is that it can be applied without the need for heavy training nor expensive equipment. This is essential for countries that have limited resources [4]. In early pregnancies, the estimated GA from FHR has been compared with that of the CRL method in an early study, which showed insignificant differences [5]. However, this study did not take into account maternal physiological factors, such as the Heart Rate Variability (HRV).

Our previous studies [6, 7] showed that fetal and maternal Heart Rate (HR) coupling strengths as well as fetal and maternal HRV features are important when estimating the GA. Generalized linear regression has been used as the adopted methodology in that study. However, it is interesting to apply new technologies when assessing fetal development to improve fetal GA estimation accuracy.

The Machine Learning (ML) framework [8] was employed in [9] to predict fetal GA based on ultrasound brain image appearance. Of the various types of ML models, Support Vector Machine (SVM) has been a popular choice in small samples setting [10]. For example, [11] used SVM to develop a large-for-gestational-age classification system. There have been no studies, however, that utilize SVM together with fetal and maternal HRV features for GA estimation in fetal development studies, which is a fundamental aspect in fetal neurological screening and an essential information for reducing fetal deaths.

In this article, the SVM is utilized in a novel regression approach for the estimation of the GA by using maternal and fetal HRV features computed from ECG abdominal signals of 60 pregnant women with a recording length of 1 min, which is potentially easy to obtain in limited

resources clinical settings. To determine a final model, the model utilizes the SVM tool in conjunction with the nonparametric linear kernel. An essential point that this study reveals is the important contribution of fetal along with maternal HRV parameters to estimate the fetal development in a correct manner. We refer to the proposed model as the Support Vector Machine estimator with the Linear kernel based on fetal and maternal HRV parameters (i.e. the SVML-MF estimator).

The proposed SVML-MF estimator is compared with that of other models that also use the SVM tool with the linear kernel, but are based on either maternal or fetal HRV features. For completion, comparisons are carried out with three other SVM models that are based on the same categories of HRV features (i.e. maternal, fetal, and maternal-along-with-fetal), but use the Radial Basis Function (RBF) kernel function (instead of the linear kernel).

2. Methods

2.1. Processing of Participants ECG Dataset

The dataset consists of abdominal ECGs from 60 healthy pregnant women in Japan and the USA. The Institutional Review Boards (IRB: 2015-2-80-1) have approved the study protocols with appropriate institutional agreements. Abdominal signals with 12 channels were recorded for 10 min [12], and sampled using 16 bit resolution for 1 ms. The fetal ECG had been separated from the abdominal composite signal using maternal ECG cancellation in combination with blind source separation with a reference as reported in [13]. A MATLAB routine program has been customized to detect the fetal and maternal QRS peak locations.

2.2. Heart Rate Variability

Time-domain HRV parameters include the Standard Deviation of NN intervals in Maternal or Fetal HR (MSDNNHR or FSDNNHR), Root Mean Square of Successive Differences between normal Maternal or Fetal heartbeats (MRMSSDHR or FRMSSDHR), and Mean value of Maternal or Fetal HR (MMHR or FMHR). These metrics were evaluated from RR intervals of 1 min length of the recorded ECG signals, which may be sufficient to correctly measure such variables for healthy individuals as long as artifacts are carefully removed [14]. In addition, it is potentially easy and more practical to record 1 min duration of ECG signals in limited resources clinical settings. Due to the article length limitation, scatterplots of the mean values of maternal and fetal HR are shown here only (Fig. 1).

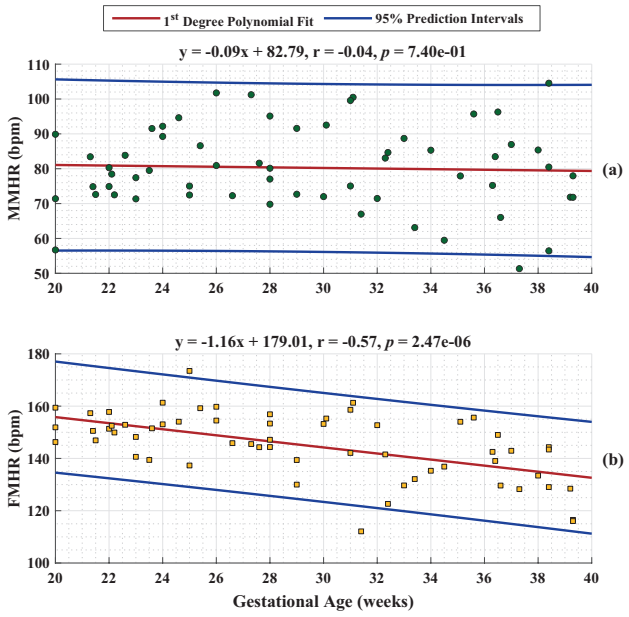

Figure 1. Scatterplots of the mean values of heart rate for the: (a) mother (MMHR), and (b) fetus (FMHR).

2.3. SVM Regression Models and Statistics

The proposed SVM-based model (i.e. SVML-MF) uses the linear kernel and combines maternal and fetal HRV parameters. Five other SVM-based models are developed based on different combinations of fetal and maternal HRV features, and use the linear kernel (SVML-MF, SVML-M and SVML-F) or the RBF kernel (SVMRBF-MF, SVMRBF-M and SVMRBF-F). All the models were generated using MATLAB's `fitrsvm`.

Consider the training dataset that includes predictor variables (x) of N observations and observed response values (y), that is

$$T = \Big\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\Big\},$$

where x_i and $y_i \in R^n$, and $i \in \{1, 2, \ldots, N\}$ [15]. The goal of the SVM regression algorithm is to produce a function $f(x)$ that deviates from y by a value no greater than ε (defined within as half the width of the ε-insensitive band) for each of the training points in x, and is as flat as possible. This requires selecting an appropriate kernel function ($k(x_i, x_j)$) and a penalty parameter of the error term (C) to construct and find the solution to the problem

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j k(x_i, x_j) - \sum_{i=1}^{N} \alpha_i \qquad (1)$$

such that $\sum_{i=1}^{N} \alpha_i y_i = 0$, where $0 \le \alpha_i \le C$. Finding the

optimal solution: $\alpha^* = \{\alpha_1^*, \alpha_2^*, \ldots, \alpha_N^*\}$. The next step is to select a component $0 < \alpha^* < C$ and calculate $b^* = y_j - \sum_{i=1}^{N} \alpha^* y_i\, k(x_i, x_j)$. The decision function can now be constructed as $f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i\, k(x, x_i) + b\right)$. In this study, two kernel functions are considered: linear and RBF.

The dataset (60 subjects) was divided randomly into two halves to obtain the considered models. The two parts of the dataset (i.e. Subjects#1–30 and Subjects#31–60) consider the time segments of the fetal and maternal ECG signals to be different when preparing the testing and training data. This is an essential step to overcome potential data dependency and increase reproducibility. Fig. 2 shows a flowchart of the SVM-based models.

When preparing the training data, the fetal and maternal ECG signals were partitioned into ten divisions. The first time segments of the ECGs were considered for half of the dataset, and the last time segments were considered for the other half. To prepare the testing data, all segments were taken into consideration except the segments used in training. The results obtained using the testing dataset were averaged to find one final result. This methodology has been implemented to establish a balanced representation of the dataset and avoid any systematic bias.

3. Results

The proposed models for estimating the GA against gold standard age identified by CRL were validated using the cross-validation scheme, and the estimation error was measured by computing the mean Root Mean Square Error (mRMSE). Results are shown for the best performing model only due to space limitations.

Table 1 lists the cross-validation results evaluated by mRMSE, Pearson correlation coefficient (r), and Bland–Altman results (bias and Limits of Agreement (LoA) ($\pm 1.96 \times$SD)) for the six introduced models. The model that produced the lowest mRMSE value is the SVML-MF estimator (5.11 weeks), and is hence the best performing model. Fig. 3 illustrates the correlation ($p < 0.05$) between the gold standard GA and estimated values by SVML-MF estimator with an r value of 0.63. Additionally, the figure shows the Bland–Altman plot which validates that the GA values estimated by the best performing model are within the LoA (8.97 and -9.59 weeks), and that the bias (i.e. estimated mean differences) is -0.31 weeks.

Table 2 lists the correlation results (r) between the different combinations of maternal and fetal HRV features. Although MMHR has no relationship ($p < 0.05$) with the GA individually (see Fig. 2), there exists a relationship ($p < 0.05$) between MMHR and FMHR.

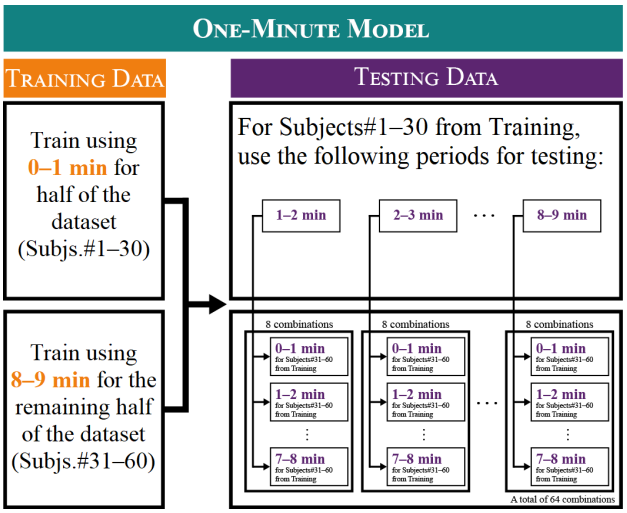

Figure 2. Flowchart for the proposed SVM-based models.

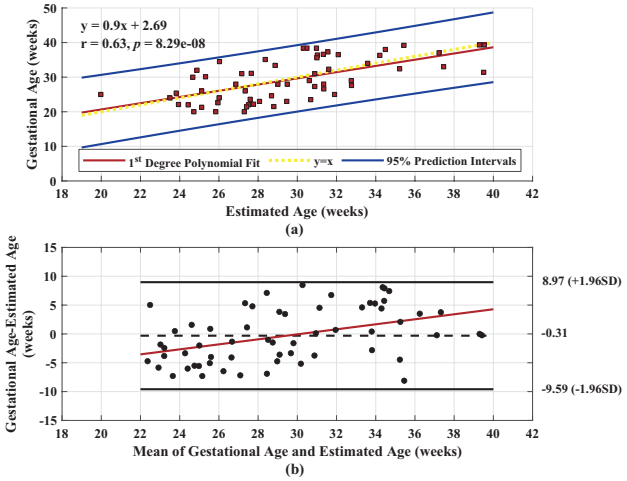

Figure 3. (a) Pearson correlation, and (b) Bland–Altman.

4. Discussion

This study has successfully demonstrated that the SVM model with linear kernel function based on both of maternal and fetal HRV features computed from recorded ECG abdominal signals for 1 min could estimate the GA more reliably than that of similar models with either maternal or fetal HRV features. The model combined fetal in conjunction with maternal HRV features rather than fetal features only, which highlights the importance and significance of maternal cardiac factors on the development of the fetal.

The proposed SVML-MF model produced higher values of r. This can be speculated due to HRV features being linear. Additionally, the overall cross-validation error is less than that compared to the other SVM-based models. The value of ε for all six models is equal to 0.82. It is interesting to implement an algorithm for selecting the best

Table 1. Cross-validation performance evaluated by mRMSE Pearson correlation coefficient (r), and Bland–Altman results

Kernel Type	Maternal-based		Fetal-based		Maternal and Fetal	
	(SVML-M)	(SVMRBF-M)	(SVML-F)	(SVMRBF-F)	(SVML-MF)	(SVMRBF-MF)
mRMSE (weeks)	5.97	5.98	5.17	5.57	**5.11**	5.94
r	0.21	0.31	0.58	0.55	**0.63**	0.35
Bias (weeks)	0.02	-0.10	-0.26	-0.15	-0.31	-0.01
ULoA (weeks)	11.62	11.23	9.40	10.27	8.97	11.72
LLoA (weeks)	-11.58	-11.43	-9.91	-10.57	-9.59	-11.73

feature subset for every model and observe the effect on the values of ε and C. Moreover, the study requires additional validation on a bigger sample size, and various lengths of the recorded signals.

Table 2. Correlation between different combinations of maternal and fetal HRV features ($p < 0.05$).

	MMHR	MSDNNHR	MRMSSDHR
FMHR	0.39*	0.01	-0.10
FSDNNHR	-0.01	0.05	0.07
FRMSSDHR	-0.24	0	0.09

5. Conclusions

This article presented an approach for accurately estimating the fetal GA by adopting the SVM algorithm with the linear kernel function based on maternal and fetal physiological parameters computed from recorded ECG signals for 1 min. Interestingly enough, combining maternal along with fetal HRV features could result in a more reliable estimation of the GA than that of similar models with either maternal or fetal features. The study successfully showed that using the linear kernel instead of the radial basis function kernel produces a proper estimate of GA, which is likely due to HRV features being linear. Further research work could consider the effect of abnormal cases of fetuses on the estimation of GA for the various scenarios of heart arrhythmias and anomalies.

References

[1] Tunón K, et al. Gestational age in pregnancies conceived after *in vitro* fertilization: A comparison between age assessed from oocyte retrieval, crown-rump length and biparietal diameter. Ultrasound Obstetrics Gynecol Jan. 2000; 15(1):41–46.

[2] Barnova K, et al. Hybrid methods based on empirical mode decomposition for non-invasive fetal heart rate monitoring. IEEE Access 2020;8:51200–51218.

[3] Hoyer D, et al. Fetal functional brain age assessed from universal developmental indices obtained from neuro-vegetative activity patterns. PLoS ONE Sep. 2013;8(9). Art. no. e74431.

[4] Hofmeyr G, et al. Obstetric care in low-resource settings: What, who, and how to overcome challenges to scale up? Int J Gynaecol Obstetrics Oct. 2009;107:S21–S45.

[5] Tezuka N, et al. Comparison of the accuracy of gestational age estimation from fetal heart rate and crown-rump length. Primary Care Update OBGYNS Jul. 1998;5(4):193.

[6] Wahbah M, et al. Estimating gestational age from maternal-fetal heart rate coupling parameters. IEEE Access 2021; 9:65369–65379. ISSN 2169-3536.

[7] Khandoker AH, et al. Estimating fetal age by fetal maternal heart rate coupling parameters. In 2020 Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), volume 42. Montreal, QC, Canada. ISSN 1558-4615, Jul., 20–24, 2020; 604–607.

[8] Garcia-Canadilla P, et al. Machine learning in fetal cardiology: What to expect. Fetal Diagn Ther May 2020; 47(5):363–372.

[9] Namburete AIL, et al. Learning-based prediction of gestational age from ultrasound images of the fetal brain. Med Image Anal Apr. 2015;21(1):72–86.

[10] Yang L, et al. Predictive models of hypertensive disorders in pregnancy based on support vector machine algorithm. Technol Health Care Jun. 2020;28(S1):181–186.

[11] Akhtar F, et al. Diagnosis and prediction of large-for-gestational-age fetus using the stacked generalization method. Appl Sci Oct. 2019;9(20):4317–4334.

[12] Doshi A, et al. Feasibility of non-invasive fetal electrocardiographic interval measurement in the outpatient clinical setting. Pediatric Cardiol Aug. 2019;40(6):1175–1182.

[13] Sato M, et al. A novel extraction method of fetal electrocardiogram from the composite abdominal signal. IEEE Trans Biomed Eng Jan. 2007;54(1):49–58.

[14] Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Frontiers Public Health Sep. 2017;5:1–17.

[15] Lunghi F, et al. Detection of fetal distress though a support vector machine based on fetal heart rate parameters. In 2005 CinC. Lyon, France, Sep. 2005; 247–250.

Address for correspondence:

Maisam Wahbah (maisam.wahbah@icloud.com)