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 related to the fetus at early stages of pregnancy, which could possibly reduce obstetric interventions, morbidity and cost. In this article, we adopt the Support Vector Machine (SVM) tool to investigate whether fetal 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 healthy pregnant women with no records of fetal abnormalities. 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. Cross-validation performances were evaluated by the mean square root of the average of squared errors (mRMSE) between age values estimated by the proposed models and gold standard GA identified by Crown-Rump Length (CRL). 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 mRMSE 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. In conclusion, 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 is ±5–7 days [1]. Some challenges exist when evaluating the GA using methods such as CRL or conventional ultrasonography, which include human errors and the requirement to have good clinical practice [2]. Such requirements, however, might not be feasible in some settings, such countries with low- and middle-income population. Due to this, it is required to have an approach that is more robust when estimating the GA while mitigating the current 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 and income [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).

In our previous studies [6, 7], the results showed that fetal and maternal Heart Rate (HR) coupling strengths as well as fetal and maternal HRV features are important when estimating the GA. 16 pregnant women (with healthy conditions) were used as patients for collecting ECG abdominal signals that were recorded for 10 min. It revealed the importance of considering maternal-fetal HR coupling parameters when estimating fetal 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.

Machine Learning (ML) models are built through a learning approach to perform a specific task [8]. The ML framework 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 no records of fetal abnormalities 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 pregnant women with healthy medical conditions and was obtained from the Tohoku University Hospital (60%) and Kanagawa Children's Medical Center (15%) in Japan, as well as the Children's National Hospital in the US (25%). The Tohoku University Institutional Review Board (IRB: 2015-2-80-1) and Children's National Hospital IRB have approved the study protocols with appropriate institutional agreements.

Abdominal signals with 12 channels (i.e. probes) were recorded from the electrodes that were adjusted on the maternal abdomen. The dataset has been sampled using 16 bit resolution for 1 ms. Abdominal ECG signals were recorded for 10 min while the participant was in the supine position. 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 [12]. A detailed description of the experimental set up has been included in our previous study [13]. A MATLAB routine program has been customized to detect the fetal and maternal QRS peak locations.

2.2. Heart Rate Variability of the Mother and the Fetal

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). In this study, 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).

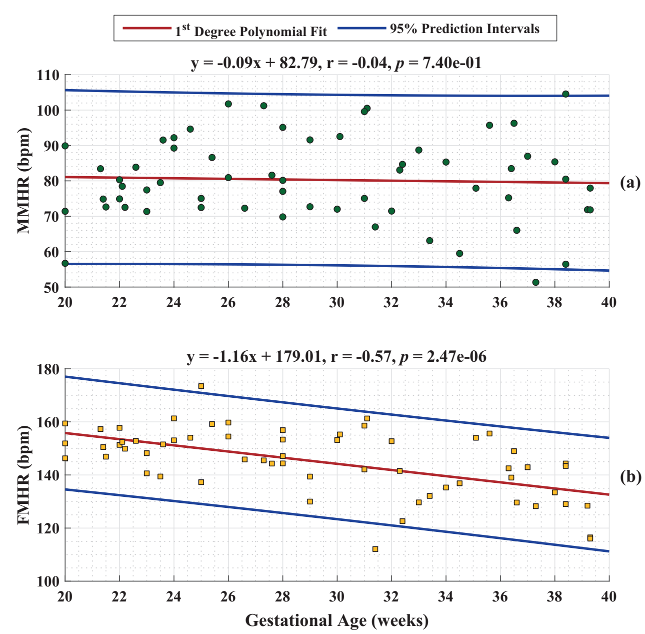

Figure 1. Scatterplots of the mean values of heart rate for the: (a) mother (MMHR), and (b) fetus (FMHR). bpm, beats per minute.

2.3. Support Vector Machine Regression Models and Statistics

To produce a reliable estimation of the fetal age, the proposed SVM-based model (i.e. SVML-MF) uses the linear kernel and combines maternal along with fetal HRV parameters. In this study, 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 of the considered models were generated using MATLAB's `fitrsvm`. Evaluations are carried out using the cross-validation scheme.

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 (also known as the L_1 loss and defined here 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 $0 \leq \alpha_i \leq C$. 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 $k(x_i, x_j)$ functions are considered: the linear and RBF kernels.

The overall dataset that consists of 60 pregnant women was divided randomly into two halves to obtain the considered regression 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 dataset. 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 abdominal 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 the training dataset. In this study, 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.

Results are presented in the next section, and are only shown for the best performing model due to space limitations. Validation has been implemented by using cross-validation for multiple times. The mean Root Mean Square Error (mRMSE) has been used to measure the estimation error due to the gold standard GA identified by CRL and the introduced models.

3. Results

The proposed models for estimating the GA against gold standard age identified by CRL were validated using the cross-validation scheme that had been implemented for multiple times. 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 identified by CRL and estimated values by the best performing model (i.e. 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 introduced SVML-MF estimator (i.e. 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 the mean value of maternal HR (MMHR) has no relationship ($p < 0.05$) with the GA individually (see Fig. 2), there exists a relationship ($p < 0.05$) between the mean values of maternal and fetal HR (MMHR and FMHR).

4. Discussion

This study has demonstrated successfully 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

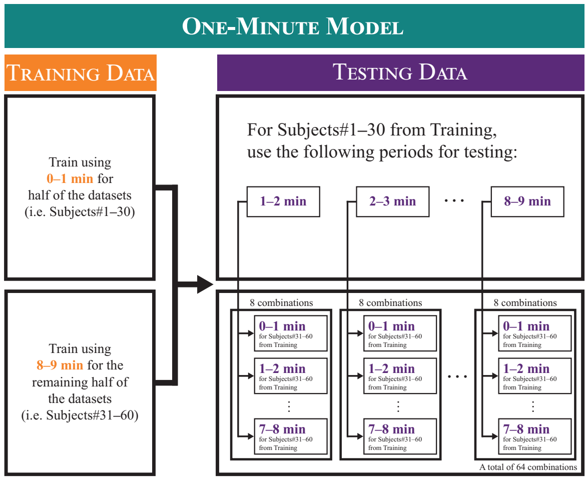

Figure 2. Flowchart for the proposed SVM-based regression models.

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 factors on the development of the fetal. An essential point that this study reveals is the important contribution of maternal HRV features to correctly estimate the fetal development. In conclusion, it has been clearly showed by the results that combining fetal along with maternal cardiac features leads to a robust approach that reliably estimate the fetal age.

The proposed SVM-based model with the linear kernel produced higher values of r. This can be speculated due to HRV features being linear. Additionally, the overall cross-validation error produced by the proposed model is less than that compared to the other SVM-based models considered in this study. The value of ε for all six models is equal to 0.82. It is interesting to implement an algorithm for selecting the best feature subset for every model and observe the effect on the values of ε and C.

The approach proposed in this article is fully automated, and is thus less likely to be affected by human errors in comparison with last menstrual period and sonography approaches. The study, however, requires additional validation on a bigger sample size, and various lengths of the recorded signals. It would also be interesting to apply another methodology for regression, such as a broader family of ML methods like deep learning; particularly, the long short-term memory architecture.

5. Conclusions

This article presented a novel approach for accurately estimating the fetal GA by adopting the SVM algorithm with the linear kernel function based on maternal and fetal

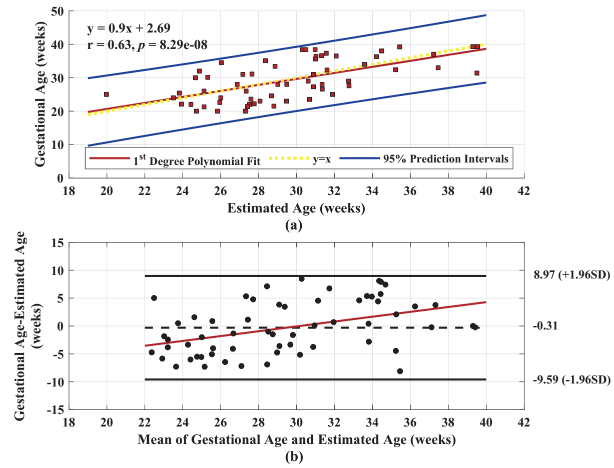

Figure 3. The panels in this figure are: (a) Pearson correlation plot between gold standard Gestational Age (GA) identified by Crown-Rump Length (CRL) and computed age by the proposed Support Vector Machine estimator with the Linear kernel based on Maternal and Fetal heart rate variability parameters (i.e. SVML-MF) for 60 fetuses with no abnormal health conditions. The identity line is represented as dotted yellow line. The linear polynomial fit line in addition to the 95% prediction intervals are shown as solid red and blue lines, respectively. (b) Bland–Altman plot for the estimated and CRL-based GA by the proposed SVML-MF (sample size = 60 dataset). Regression fit of the differences on the means is shown as solid red line, bias is represented as dashed black line, and limits of agreement ($\pm 1.96 \times$SD) are represented as solid black lines,.

physiological parameters computed from recorded ECG signals for 1 min. Interestingly enough, combining maternal along with fetal HRV features could result in a more reliably 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.

Acknowledgments

The study was supported in part by an internal grant awarded to Ahsan H. Khandoker, PhD (CIRA 2019-023 grant Project 8474000174). The authors also acknowledge the contributions from the group of Obstetricians and Research nurses at Kanagawa Children's Medical Center in Japan, as well as the contributions from Ms. Catherine Ingbar and the Research Team at Children's National Hospital in Washington, DC, USA for providing the ECG signals

Table 1. Cross-validation performance evaluated by the mean square root of the average of squared errors (mRMSE) between gestational age values estimated by the proposed models and gold standard age identified by crown-rump length, Pearson correlation coefficient (r), and Bland–Altman results (bias and limits of agreement ($\pm 1.96 \times$SD)). ULoA, Upper Limit of Agreement; LLoA, Lower Limit of Agreement

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

Table 2. Correlation results (r) between the different combinations of maternal and fetal HRV features measured in beats per minute. $p < 0.05$ was considered significant.

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

* $p < 0.05$

for this study.

References

[1] Tunón K, Eik-Nes SH, Grøttum P, Von Düring V, Kahn JA. 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, Martinek R, Jaros R, Kahankova R. 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, Saito H, Hiroi M. 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, Al Sakaji R, Funamoto K, Krishnan A, Kimura Y, Khandoker AH. Estimating gestational age from maternal-fetal heart rate coupling parameters. IEEE Access 2021;9:65369–65379. ISSN 2169-3536.

[7] Khandoker AH, Wahbah M, Al Sakaji R, Funamoto K, Krishnan A, Kimura Y. 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, Sanchez-Martinez S, Crispi F, Bijnens B. Machine learning in fetal cardiology: What to expect. Fetal Diagn Ther May 2020;47(5):363–372.

[9] Namburete AIL, Stebbing RV, Kemp B, Yaqub M, Papageorghiou AT, Alison Noble J. 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] 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.

[13] 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.

[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, Magenes G, Pedrinazzi L, Signorini MG. 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
College of Engineering & Information Technology, University of Dubai, Dubai 14143, United Arab Emirates
maisam.wahbah@icloud.com