Primary Research Articles

Prediction of spontaneous preterm birth among twin gestations using machine learning and texture analysis of cervical ultrasound images

Sandra Fiset (MHSc)[1,2]; Anne Martel (PhD)[1,3]; Phyllis Glanc (MD, FRCPC, ABR)[1,2]; Jon Barrett (MBBch, FRCOG, MD, FRCSC)[1,2]; Nir Melamed (MSc, MD)[1,2]

[1]Sunnybrook Research Institute, 2075 Bayview Avenue, Toronto, ON, Canada, M4N 3M5.
[2] Faculty of Medicine, University of Toronto, 1 King's College Circle, Medical Sciences Building, Toronto, ON, Canada, M5S 1A8.
[3]Department of Medical Biophysics, University of Toronto, Princess Margaret Cancer Research Tower, MaRS Centre, 101 College Street, Room 15-701, Toronto, ON, Canada, M5G 1L7.

Abstract

Background/Introduction: Preterm birth (PTB) is the main cause of neonatal mortality and morbidity in twin pregnancies. There is a need for new biomarkers to improve the predictive accuracy for PTB. The aim of this study was to use quantitative texture analysis of ultrasound images of the cervix combined with machine learning algorithms to detect patterns that are predictive of PTB in twins.

Methodology: This proof-of-concept study involved a retrospective cohort of women with twin gestation who underwent serial monitoring of cervical length in a single referral center between 2016–2017. Ultrasound images of the cervix at $22^{+0}-26^{+6}$ weeks of gestation were extracted and analyzed. An automated software module was developed to extract texture features from four regions of interest. Texture features were applied to a random forest classifier machine learning algorithm to predict PTB.

Results: A total of 170 images from 98 women with twins (61/98 PTB) were analyzed. The predictive value for PTB, expressed as the area under the receiver operating characteristic curve (AUC), was higher for samples obtained from the external os versus internal os, and was highest for the anterior aspect compared with the posterior aspect of the external os (0.75, 95% CI 0.61–0.78). The sensitivity and specificity for PTB for samples obtained from the anterior external os were 69% and 70%, respectively.

Conclusion: In this preliminary proof-of-concept study we have found a novel biomarker that may improve the prediction of PTB.

Corresponding Author:
Sandra Fiset
Sandra.fiset@mail.utoronto.ca

Introduction

Preterm birth is the leading cause of neonatal mortality and short- and long- term morbidity.[1,2] Identification of women at risk of spontaneous preterm birth (PTB) will allow for closer monitoring and early intervention to decrease associated morbidity and mortality. This is particularly important for twin gestations, approximately 50% of which result in PTB.[1] Many methods of predicting PTB have been proposed. However, most have failed to provide repeatable predictive ability or noteworthy clinical adoption. Among the few methods that are being used in clinical practice, the measurement of maternal cervical length has emerged as an important predictor of PTB, including in twin gestations.[3–13] One of the main limitations of cervical length is its low specificity and positive predictive value, as only approximately 25% of women with a short cervix will deliver preterm,[14] thereby leading to unnecessary treatments, bed rest, and anxiety in the majority of women with a short cervix.[15–17] Thus, there is a need for superior biomarkers to predict PTB.

In the weeks leading to parturition, a process termed cervical ripening occurs. Cervical ripening involves enzymatic processes that alter the cervical stroma tissue hydration status, as well as collagen alignment and organization.[18,19] It is known to precede birth and PTB by up to several weeks.[20] Identifying the onset of cervical ripening is the basis for a new biomarker being proposed for the prediction of PTB. It is hypothesized that the detection of cervical ripening on mid-trimester ultrasound images may provide prognostic information to predict whether a woman will give birth imminently (i.e. preterm). The changes that occur with cervical ripening are micro-structural and, therefore, not visible by gross inspection of imaging. Texture analysis, a means of quantitative imaging, is proposed as a method of detecting micro-structural changes in the cervical stroma and thus identifying women at risk of PTB. Texture analysis is a mathematical and statistical representation of the texture in an image, which can be described as the spatial relationship between pixels of various intensities. It is thought that texture analysis plays an important role in assessing pixel organization of different tissues and organs, thus capturing histological heterogeneity from medical images.[21–23]

In ground-breaking work, texture analysis of the cervix has been shown to provide increased sensitivity and specificity

Primary Research Articles

Prediction of spontaneous preterm birth among twin gestations using machine learning and texture analysis of cervical ultrasound images

in the prediction of PTB when compared with cervical length measurement in singleton pregnancies.[24,25] These pioneering studies have provided the rationale for further research into the applicability of texture analysis as a biomarker for PTB. It is of the utmost importance that biomarkers be tested for generalizability, specifically with respect to high-risk groups such as twin gestations.

This proof-of-concept study aims to demonstrate the use of quantitative texture analysis of ultrasound images of the cervix combined with a machine learning algorithm to detect patterns that are predictive of PTB in twins.

Methods

Study Population

This retrospective single-centre cohort study involved 98 nulliparous women with twin gestations who delivered in Sunnybrook Health Sciences Centre between 2016 and 2017. The study was approved by the institutional research board, with waiver of informed consent. From the 98 women in this study, 170 transvaginal sagittal cervical ultrasound images were acquired between 22^{+0} and 26^{+6} weeks' gestation by experienced sonographers as per routine institutional practice. Image inclusion criteria included: visualization of the entire cervical structure in one frame, the cervix located in the centre of the image, no compression of the anterior lip by the probe, and no cervical funnelling. The GE Voluson E8 ultrasound unit was used to acquire all images. For all analyses, preterm was defined as birth at less than 37 weeks of gestation. A range of cervical lengths was included in this study in order to evaluate the predictive capability of cervical texture independent of cervical length. The independent predictive ability of cervical length in this population was evaluated using a receiver operating characteristic (ROC) curve.

Exclusion criteria similar to those of previous studies evaluating the prediction of preterm birth using cervical length were employed in this study.[8,26] Notably, women with known risk factors for preterm birth such as pre-eclampsia, eclampsia, twin-twin transfusion syndrome, and cervical interventions (e.g. cerclage or vaginal progesterone) were excluded from the study population.

Feature Extraction

Raw DICOM images were collected from the hospital picture archiving and communication system (PACS). Using custom in-house software, four regions of interest (ROI) measuring 40x40 pixels were selected from each image: at the anterior and posterior lip, and internal and external os, as illustrated in Figure 1. ROIs were selected such that they included exclusively cervical stroma, avoiding glandular tissue and surrounding structures.

ROIs were converted into numerical matrices of 8-bit pixel greyscale values ranging from 0 to 255. Texture analysis was computed on the resulting numerical matrices. Descriptive statistics, or first order features, were computed on the histogram of grey level intensities in the ROI. This group of features included mean intensity, variance, skewness, kurtosis, and coefficient of variation. Grey level co-occurrence matrices were calculated using 16 grey levels; angles 0, 45, 90, and 135 degrees; and radii of 1, 3, 5, and 7. Contrast, correlation, energy, homogeneity, and angular second moment were calculated using the resulting grey level co-occurrence matrices which resulted in 96 features. Finally,
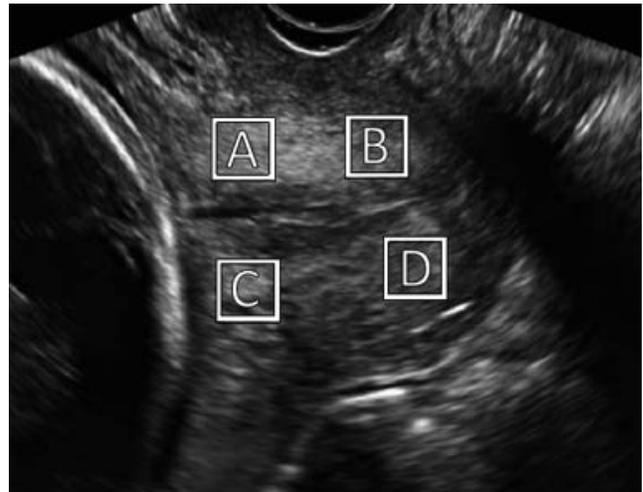
Figure 1. Sagittal view of a transvaginal cervical ultrasound image. The white overlaying boxes demonstrate the four regions of interest (ROIs) that are selected. (A) Anterior lip, internal os. (B) Anterior lip, external os. (C) Posterior lip, internal os. (D) Posterior lip, external os.

a local binary pattern created a binary number by thresholding surrounding pixels according to a centre pixel. A radius of 1 and a resolution of 16 bit were used. A histogram of uniform patterns of binary numbers in the ROI was recorded which resulted in 17 features. A total of 119 features were collected for each ROI.

Machine Learning

Initially, feature selection was performed to mitigate confounding in the machine learning algorithm from features with no discriminatory power between term and preterm images. Two criteria for feature inclusion in the algorithm were selected: (1) a univariate t-test with significance $p<0.05$ between term and preterm images, and (2) a Pearson correlation coefficient r <0.9 with other features.

A 5-fold cross validation technique was utilized to maximize generalizability of the model. A random forest classifier was employed using Scikit-learn package for Python.[27] The maximum number of features to optimize split was determined based on literary standards and the number of trees was set to 50. For the patients with multiple images, all images were included in either training or testing so as to not overestimate the accuracy of the machine learning algorithm.

Statistical Analysis

The outcomes of the learning algorithm prediction were compared to the images' known outcome of "term" or "preterm". Receiver operating characteristic curves and areas under the curve (AUC) with a 95% confidence interval were used to quantify the predictive ability of the learning algorithm. All statistical analyses were performed with Scikit-learn package as well as SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.).

Results

All 98 women in this study were included and all 170 images were deemed suitable for analysis (entire cervix visible and

Primary Research Articles

Prediction of spontaneous preterm birth among twin gestations using machine learning and texture analysis of cervical ultrasound images

centered in image). For the patients with multiple images, the average time interval between image acquisitions was 1.9 weeks (range 1–3 weeks). The patients' cervical lengths, chorionicity, and gestational age (GA) at delivery are outlined in Table 1. Similar to rates reported in the literature, over half (62.2%) of the women in this study delivered prematurely (at less than 37 weeks of gestation) yielding similar class sizes and pretest probabilities for term and pre-term.[1]

Table 1. Demographics of the study population. (no: number, wks: weeks)

|	
Number of participants	98
Number of images	170
Average cervical length – cm (range)	3.3 (0.9, 5.37)
Maternal Demographics	
Maternal age – years (SD)	34.5 (4.3)*
Pre-pregnancy body mass index – kg/m² (SD)	27.4 (8.0)*
Chorionicity – (%)	
Dichorionic diamniotic	63
Monochorionic diamniotic	18
Unknown	17
Average GA at delivery – wks. (range)	35 (26–38)
Preterm – no. (%)	
<37 weeks	61 (62.2%)
<34 weeks	23 (23.5%)
<32 weeks	15 (15.3%)
Caesarian Section	50 (51.0%)
Premature Rupture of Membranes (PPROM)	3 (3.1%)
Intra-Uterine Growth Restriction (IUGR)	3 (3.1%)
Malpresentation	9 (9.2%)

* Maternal age and body-mass index were recorded for 80 patients.

Cervical lengths of the study population ranged from 9 mm to 53 mm. The ROC curve for cervical length was independently evaluated from cervical texture for the study population. The resulting AUC was 0.59 (95% CI: 0.457–0.722).

Table 2. Area under the ROC curve (AUC) of each ROI region resulting from classification into "term" and "preterm" by the machine learning algorithm. (CI: confidence interval)

Patch Location	AUC (95% CI)
Posterior lip, external os	0.70 (0.61–0.78)
Posterior lip, internal os	0.68 (0.59–0.77)
Anterior lip, external os	0.75 (0.67–0.82)
Anterior lip, internal os	0.59 (0.49–0.67)

The prediction algorithm was performed on the ROIs from each of the four quadrants of the cervix. An average of 10 features were used for the random forest classifier following application of feature selection criteria. Figure 2 demonstrates the ROC curve for each of the four quadrants. The diagnostic capability varied significantly for the four quadrants as demonstrated by the AUCs listed in Table 2. Three of the evaluated regions demonstrated statistically significant classification accuracy as their 95% confidence intervals excluded the "luck line." The anterior lip,
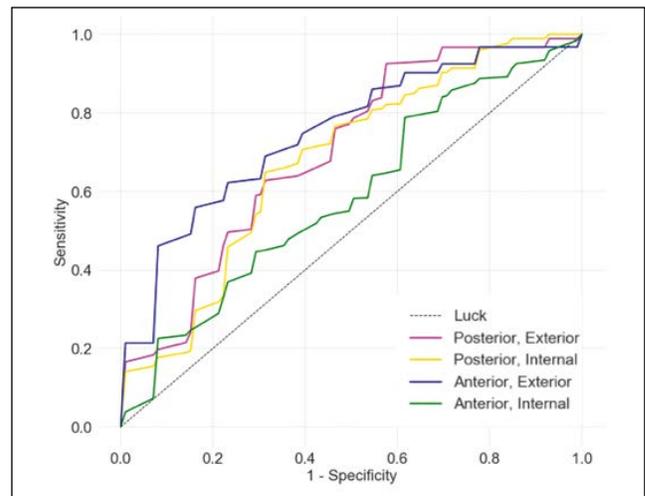
Figure 2. Receiver Operating Characteristic (ROC) curve resulting from the classification of ROIs into "term" and "preterm" by the machine learning algorithm. Each coloured line represents the diagnostic accuracy of the machine learning algorithm when tested with ROIs from different regions of cervical tissue.

external os demonstrated the highest AUC of 0.75 (95% CI: 0.6–0.82). The optimal sensitivity and specificity for this ROI were 69% and 70%, respectively. The quadrants at the external os performed better than the internal os given that the second most accurate quadrant was the posterior lip, external os with an AUC of 0.70 (95% CI: 0.61–0.78). The least predictive region was the anterior lip, internal os, which did not show statistically significant diagnostic capability given that the lower bound of the confidence interval was 0.49. The AUC for the anterior lip, internal os was 0.59 (95% CI: 0.49–0.67).

Discussion

This study proposed a novel biomarker based on quantitative imaging and machine learning for the prediction of PTB in twins. Using such a measurable biomarker, accurate identification of mothers at risk of PTB would allow for close monitoring and early intervention to reduce PTB-associated morbidity and mortality. This is especially important for twins, the focus of this study, given that over 50% are born prematurely.

The texture analysis-based machine learning algorithm utilized in this study achieved a maximum classification accuracy AUC of 0.75 (95% CI: 0.67–0.82) for an ROI located in the anterior lip close to the external os. The results of this study agree with the recent pioneering studies evaluating the use of quantitative imaging for the prediction of PTB in singleton pregnancies.[24,25] Specifically, Baños and colleagues used a machine learning algorithm (discriminant analysis regression) to classify mid-trimester transvaginal cervical ultrasound images using texture analysis features.[25] Their results showed a significantly increased area under the receiver operating characteristic curve when compared with cervical length (0.77 vs. 0.60, p=0.03) in the prediction of spontaneous preterm birth.[25] Accordingly, meta-analyses of the use of cervical length estimate the AUC for mid-trimester cervical length to be 0.6–0.7.[25,28] Indeed, for the population evaluated in this study, the AUC was 0.59. Therefore the AUCs in this study match, if not surpass, the

Primary Research Articles

Prediction of spontaneous preterm birth among twin gestations using machine learning and texture analysis of cervical ultrasound images

predictive values of the current industry standard (cervical length) in the literature.

The cervical regions in proximity to the external os showed improved AUC (posterior lip AUC 0.70, anterior lip AUC 0.75) when compared with the internal os (posterior lip AUC 0.68, anterior lip AUC 0.59). This finding is unexpected given that it is commonly reported that cervical ripening begins at the internal os and proceeds distally.[18] There are two possible explanations for this finding. First, this may contribute to the hypothesis that spontaneous preterm birth is attributed to microbially-induced inflammation ascending from the lower genital tract and therefore beginning at the external os.[29] Second, the external os was most often within the focal zone of the ultrasound image resulting in the best lateral resolution at that location. Further investigation into the differences of the cervical zones should be undertaken in order to maximize the predictive capability of a machine learning biomarker.

Though the AUC for two of the cervical regions in this study were above 0.7, which is commonly considered a fair AUC value,[30] the predictive capability of the machine learning algorithm requires external generalizability and further validation testing. Additionally, several limitations in this study must be addressed. Notably, due to the retrospective nature of this study, the ultrasound images were post-processed (e.g. speckle reduction, smoothing) which may have introduced noise in the texture findings. Additionally, maternal age and BMI were not recorded for all patients in this study, and limited past medical history for the patients was recorded. Images with outlying features were excluded from the study to mitigate inter-image variability. The sample size was relatively small for a machine learning study, however, the number of participants in "preterm" and "term" were kept equal to minimize the effect of class imbalance on the algorithm. Finally, the traditional machine learning algorithm used in this study has been shown to be outperformed by deep convolutional neural networks on medical image classification tasks. Continued research is required for critical analysis of this tool in order to address limitations in this study and improve the accuracy of the tool. Future work should focus on the use of advanced learning algorithms, multivariate analysis with other known predictors of preterm birth, and creation of a comprehensive score with multiple imaging and clinical factors.

Conclusion

In conclusion, this preliminary proof-of-concept study demonstrated that texture analysis and machine learning can be used to discriminate term twin gestations from preterm twin gestations based on mid-trimester cervical ultrasound images with sensitivity and specificity of 69% and 70%, respectively. Further research is warranted to increase the classification accuracy of this biomarker and ultimately offer physicians a tool to provide their patients with optimal care.

Acknowledgements

References

1. Ananth CV, Vintzileos AM. Epidemiology of preterm birth and its clinical subtypes. J Matern Neonatal Med. 2006; 19(12):773–82.

2. Quinn JA, Munoz FM, Gonik B, et al. Preterm birth: Case definition & guidelines for data collection, analysis, and presentation of immunisation safety data. Vaccine. 2016; 34(49):6047–56.

3. Hiersch L, Yogev Y, Domniz N, et al. The role of cervical length in women with threatened preterm labor: is it a valid predictor at any gestational age? Am J Obstet Gynecol. 2014; 211(5):532.e1-532.e9.

4. Melamed N, Hiersch L, Gabbay-Benziv R, et al. Predictive value of cervical length in women with twin pregnancy presenting with threatened preterm labor. Ultrasound Obstet Gynecol. 2015; 46(1):73–81.

5. Rosen H, Hiersch L, Freeman H, et al. The role of serial measurements of cervical length in asymptomatic women with triplet pregnancy. J Matern Neonatal Med. 2018; 31(6):713–9.

6. Hiersch L, Melamed N, Aviram A, et al. Role of cervical length measurement for preterm delivery prediction in women with threatened preterm labor and cervical dilatation. J Ultrasound Med. 2016; 35(12):2631–40.

7. Melamed N, Pittini A, Hiersch L, et al. Do serial measurements of cervical length improve the prediction of preterm birth in asymptomatic women with twin gestations? Am J Forensic Med Pathol. 2016; 215(5):616.e1-616.e14.

8. Melamed N, Pittini A, Hiersch L, et al. Serial cervical length determination in twin pregnancies reveals four distinct patterns with prognostic significance for preterm birth. Am J Obstet Gynecol. 2016; 215(4):476.e1-476.e11.

9. Hiersch L, Rosen H, Okby R, et al. The greater risk of preterm birth in triplets is mirrored by a more rapid cervical shortening along gestation. Am J Obstet Gynecol. 2016; 215(3):357.e1-357.e6.

10. To MS, Skentou CA, Royston P, et al. Prediction of patient-specific risk of early preterm delivery using maternal history and sonographic measurement of cervical length: A population-based prospective study. Ultrasound Obstet Gynecol. 2006; 27(4):362–7.

11. Kim RS, Gupta S, Lam-Rachlin J, et al. Fetal fibronectin, cervical length, and the risk of preterm birth in patients with an ultrasound or physical exam indicated cervical cerclage. J Matern Fetal Neonatal Med. 2016; 29(22):3602–5.

12. Conde-Agudelo A, Romero R. Predictive accuracy of changes in transvaginal sonographic cervical length over time for preterm birth: A systematic review and metaanalysis. Am J Obstet Gynecol. 2015; 213(6):789–801.

13. Leung TN, Pang MW, Leung TY, et al. Cervical length at 18-22 weeks of gestation for prediction of spontaneous preterm delivery in Hong Kong Chinese women. Ultrasound Obstet Gynecol. 2005; 26(7):713–7.

14. Grobman WA, Lai Y, Iams JD, et al. Prediction of spontaneous preterm birth among nulliparous women with a short cervix. J Ultrasound Med. 2016; 35(6):1293–7.

15. Fonseca EB, Celik E, Parra M, et al. Progesterone and the risk of preterm birth among women with a short cervix. N Engl J Med. 2007; 357(5):462–9.

16. Neches WH, Tatum GH. Activity restriction among women with a short cervix. Pediatr Hear Dis A Pract Guid. 2012; 121(6):280–1.

17. Vis JY, Kuin RA, Grobman WA, et al. Additional effects of the cervical length measurement in women with preterm contractions: A systematic review. Arch Gynecol Obstet. 2011; 284(3):521–6.

18. Nott JP, Bonney EA, Pickering JD, et al. The structure and function of the cervix during pregnancy. Transl Res Anat. 2016; 2(1):1–7.

19. Feltovich H, Hall TJ, Berghella V. Beyond cervical length: emerging technologies for assessing the pregnant cervix. Am J Obstet Gynecol. 2012; 207(5):345–54.

20. Vink J, Feltovich H. Cervical etiology of spontaneous preterm birth. Semin Fetal Neonatal Med. 2016; 21(2):106–12.

21. Sardanelli F. Trends in radiology and experimental research. Eur Radiol Exp. 2017; 1(1).

22. Bologna M, Corino VDA, Montin E, et al. Assessment of stability and discrimination capacity of radiomic features on apparent diffusion coefficient images. J Digit Imaging. 2018; 31(6):879–94.

23. Scalco E, Rizzo G. Texture analysis of medical images for radiotherapy applications. Br J Radiol. 2017; 90(1070):20160642.

24. Tekesin I, Hellmeyer L, Heller G, et al. Evaluation of quantitative ultrasound tissue characterization of the cervix and cervical length in the prediction of premature delivery for patients with spontaneous preterm labor. Am J Obstet Gynecol. 2003; 189(2):532–9.

25. Baños N, Perez-Moreno A, Julià C, et al. Quantitative analysis of the cervical texture by ultrasound in the mid-pregnancy is associated with spontaneous preterm birth. Ultrasound Obstet Gynecol. 2018; 51(5):637–43.

26. Melamed N, Hiersch L, Domniz N, et al. Predictive value of cervical length in women with threatened preterm labor. Obstet Gynecol. 2013; 122(6):1279–87.

27. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2012; 12:2825–30.

28. Baños N, Perez-Moreno A, Migliorelli F, et al. Quantitative analysis of the cervical texture by ultrasound and correlation with gestational age. Fetal Diagn Ther. 2017; 41:265–72.

29. Romero R, Dey SK, Fisher SJ. Preterm labor: One syndrome, many causes. Science. 2014; 345(6198):760–5.

30. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. Radiology. 2003; 229(1):3–8.