RISK ESTIMATION OF PRETERM BIRTH FROM COMPUTERIZED ANALYSIS OF TRANSVAGINAL ULTRASOUND CERVICAL IMAGES

WILLIAM ANDRÉS CANCINO REY

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
BUCARAMANGA
2025

RISK ESTIMATION OF PRETERM BIRTH FROM COMPUTERIZED ANALYSIS OF TRANSVAGINAL ULTRASOUND CERVICAL IMAGES

WILLIAM ANDRÉS CANCINO REY

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Electronic Engineering

Advisor
Said Pertuz, PhD

Co-advisor
Carlos Hernán Becerra Mojica, MD

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
BUCARAMANGA
2025

DEDICATION

A mis padres, Mercedes y Andelfo, por su amor y apoyo incondicional.

A mis hermanos, Andel y Camilito, que me motivan e inspiran para alcanzar cada uno de mis objetivos.

ACKNOWLEDGMENTS

Quiero agradecer especialmente a mi director, Said Pertuz, por su constante apoyo en cada idea y proyecto que surgió a lo largo de este proceso. Sus valiosos consejos académicos, profesionales y personales fueron fundamentales en esta etapa.

Agradezco también a mi co-director, Dr. Carlos Becerra, por su guía y paciencia durante este camino. Sus conocimientos y experiencias compartidas me permitieron explorar y profundizar en nuevos aspectos del ámbito profesional.

Finalmente, quiero expresar mi agradecimiento a los miembros del Grupo de Investigación en Conectividad y Procesamiento de Señales (CPS), en especial a aquellos con los que compartí momentos de risas, charlas y discusiones enriquecedoras, así como a quienes me brindaron su apoyo y palabras de aliento.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ANNEXES

RESUMEN

TÍTULO: ESTIMACIÓN DE RIESGO DE PARTO PREMATURO A PARTIR DEL ANÁLISIS COMPUTARIZADO DE IMÁGENES DE ULTRASONIDO TRANSVAGINAL CERVICALES [*]

AUTOR: WILLIAM ANDRÉS CANCINO REY [**]

PALABRAS CLAVE: PARTO PREMATURO ESPONTÁNEO, APRENDIZAJE AUTOMÁTICO, ULTRASONIDO TRANSVAGINAL, ANÁLISIS RADIÓMICO, CUELLO UTERINO, PRIMER TRIMESTRE

DESCRIPCIÓN: El parto prematuro (PP) sigue siendo un problema de salud global. Los esfuerzos de los investigadores se han centrado en desarrollar métodos para detectar a mujeres con alto riesgo de PP en etapas tempranas del embarazo. Aunque existen biomarcadores que intentan detectar PP en el segundo trimestre, la detección en el primer trimestre es lo deseable para implementar estrategias de intervención más tempranas. Este estudio tiene como objetivo desarrollar un algoritmo para estimar el riesgo de parto prematuro espontáneo (PPE) a partir de imágenes de ultrasonido transvaginal (TVUS) del primer trimestre. Se llevó a cabo un estudio retrospectivo de cohorte con mujeres que asistieron a un examen de tamizaje por ultrasonido entre las semanas 11+0 y 13+6. Se recopilaron variables clínicas, que incluyeron información demográfica, longitud cervical (LC), índice de consistencia cervical (ICC) e historial obstétrico, junto con imágenes de TVUS. Estas imágenes fueron sometidas a un análisis radiómico computarizado para extraer características que fueron usadas en el entrenamiento de cinco modelos de aprendizaje automático. El análisis radiómico permite la extracción automática de información clínicamente relevante de las imágenes médicas, transformándolas en variables cuantitativas basadas en medidas de intensidad, morfometría y textura. El área bajo la curva ROC (AUC) y los intervalos de confianza (IC) del 95 % evaluaron el rendimiento de los modelos para predecir PPE. Se utilizó una regresión logística para calcular los puntajes de riesgo basados en las medidas cervicales (LC e ICC) individuales y en la combinación del análisis radiómico con las variables clínicas. De 253 mujeres incluidas, 28 experimentaron un PPE. El modelo de Regresión Logística entrenado con las características radiómicas (RadF) mostró el mejor desempeño, con un AUC de 0.67 (IC 95 %, 0.57 - 0.77). Al combinar el análisis radiómico con LC e ICC, el AUC aumentó a 0.68 (IC 95 %, 0.56 - 0.79) y 0.74 (IC 95 %, 0.64 - 0.83), respectivamente. Al integrar RadF con las variables clínicas, el AUC alcanzó 0.79 (IC 95 %, 0.69 - 0.88). Estos hallazgos resaltan el potencial del análisis radiómico en la predicción de PPE en el primer trimestre y la importancia de integrar múltiples biomarcadores.

---

[*] Tesis

[**] Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Said Pertuz, PhD. Codirector: Carlos Hernán Becerra Mojica, MD.

ABSTRACT

TITLE: RISK ESTIMATION OF PRETERM BIRTH FROM COMPUTERIZED ANALYSIS OF TRANSVAGINAL ULTRASOUND CERVICAL IMAGES [*]

AUTHOR: WILLIAM ANDRÉS CANCINO REY [**]

Keywords: SPONTANEOUS PRETERM BIRTH, MACHINE LEARNING, TRANSVAGINAL ULTRASOUND, RADIOMIC ANALYSIS, CERVIX, FIRST TRIMESTER

Description: Preterm birth (PTB) remains a global health problem. Researchers' efforts have focused on developing methods to detect women at high risk of PTB early in pregnancy. Although existing biomarkers attempt to detect PTB in the second trimester, detection in the first trimester is desirable in order to implement earlier intervention strategies. This study aims to develop an algorithm to estimate the risk of spontaneous preterm birth (sPTB) from first trimester transvaginal ultrasound (TVUS) images. A retrospective cohort study was conducted with women attending a screening ultrasound examination between 11+0 and 13+6 weeks. Clinical variables were collected, including demographic information, cervical length (CL), cervical consistency index (CCI) and obstetric history, along with TVUS images. These images were subjected to computerized radiomic analysis to extract features that were used in training five machine learning models. Radiomic analysis allows the automatic extraction of clinically relevant information from medical images, transforming them into quantitative variables based on measures of intensity, morphometry and texture. The area under the ROC curve (AUC) and 95 % confidence intervals (CI) evaluated the performance of the models in predicting sPTB. Logistic regression was used to calculate risk scores based on individual cervical measurements (CL and CCI) and the combination of radiomic analysis with clinical variables. Out of 253 women included, 28 experienced sPTB. The Logistic Regression model trained with radiomic features (RadF) showed the best performance, with an AUC of 0.67 (95 % CI, 0.57 - 0.77). When combining radiomic analysis with CL and CCI, the AUC increased to 0.68 (95 % CI, 0.56 - 0.79) and 0.74 (95 % CI, 0.64 - 0.83), respectively. When integrating RadF with clinical variables, the AUC reached 0.79 (95 % CI, 0.69 - 0.88). These findings highlight the potential of radiomic analysis in predicting sPTB in the first trimester and the relevance of integrating multiple biomarkers.

---

[*] Thesis

[**] Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Advisor: Said Pertuz, PhD. Co-advisor: Carlos Hernán Becerra Mojica, MD.

INTRODUCTION

Preterm birth (PTB), defined as birth before 37 weeks gestation, is a major global health problem and the leading cause of mortality in children under five years of age. According to the World Health Organization, more than 13 million cases of PTB occur annually [1]. In 2023, the rate of preterm births in Colombia was 11.1 % [2]. Santander was one of the regions with the highest incidence, with 113 preterm births per 1000 live births. Babies born prematurely face an increased risk of short- and long-term health complications, including respiratory [3], gastrointestinal [4] and cognitive problems [5]. In addition, the risk of mortality and morbidity increases with degree of prematurity [6]; the earlier the birth, the greater the risk. Preterm birth can be spontaneous (sPTB) or induced. sPTB occurs when labor begins on its own before 37 weeks gestation and is the most common type of PTB [7].

---

1 World Health Organization et al. *Born too soon: decade of action on preterm birth*. World Health Organization, 2023, ix, 108 p.

2 Departamento Administrativo Nacional de Estadística (DANE). *Estadísticas Vitales (EEVV) - Nacimientos en Colombia*. [Online]. Available: `https://www.dane.gov.co/files/operaciones/EEVV/bol-EEVV-Nacimientos-IVtrim2023.pdf`. Accessed: Sept. 04, 2024. Mar. de 2023.

3 John Lowe, Sarah J. Kotecha y Sailesh Kotecha. "Long Term Effects Following Extreme Prematurity: Respiratory Problems". En: *Emerging Topics and Controversies in Neonatology*. Springer International Publishing, 2020, 351–366.

4 Kazue Nakamura et al. "Exclusively Breastfeeding Modifies the Adverse Association of Late Preterm Birth and Gastrointestinal Infection: A Nationwide Birth Cohort Study". En: *Breastfeed Med* 15.8 (2020). PMID: 32543213, págs. 509-515.

5 Mohammed Y Al-Hindi et al. "Screening for Neurodevelopmental Delay for Preterm Very Low Birth Weight Infants at Tertiary Care Center in Saudi Arabia". En: *Cureus* (dic. de 2021).

6 Eric O Ohuma et al. "National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis". En: *The Lancet* 402.10409 (oct. de 2023), 1261–1271.

7 Robert L Goldenberg et al. "Epidemiology and causes of preterm birth". En: *The Lancet* 371.9606 (ene. de 2008), 75–84.

The heterogeneous etiology of preterm birth makes it difficult to develop methods to identify women at high risk early in pregnancy. A previous PTB is a significant risk factor [8], but one study found that only 30 % of women who experienced one had a subsequent PTB [9]. Furthermore, this risk factor does not apply to nulliparous women. Cervical length less than 25 mm, measured by transvaginal ultrasound (TVUS) imaging in the second trimester, is considered the gold standard for risk assessment of PTB [10]. However, previous studies have reported its low sensitivity [11], which highlights the need to explore new methods for PTB prediction. Some studies have focused their efforts on identifying biomarkers that quantify the changes that the cervix undergoes due to pregnancy. The cervix plays a crucial role as a biomechanical barrier, keeping the uterus closed and protecting the developing fetus. For delivery, the cervix has to soften, dilate and reconstitute itself, which requires the activation of microstructural events in a precise manner [12]. Scientific evidence has shown that microstructural changes occur in the cervix composition during

---

[8] T Tingleff et al. “Risk of preterm birth in relation to history of preterm birth: a population-based registry study of 213 335 women in Norway”. En: *BJOG* 129.6 (nov. de 2021), págs. 900-907.

[9] Nathaniel Halide Kalengo et al. “Recurrence rate of preterm birth and associated factors among women who delivered at Kilimanjaro Christian Medical Centre in Northern Tanzania: A registry based cohort study”. En: *PLoS One* 15.9 (sep. de 2020). Ed. por Joel Msafiri Francis, e0239037.

[10] Lee Reicher, Yuval Fouks y Yariv Yogev. “Cervical Assessment for Predicting Preterm Birth-Cervical Length and Beyond”. En: *J Clin Med* 10.4 (feb. de 2021).

[11] Jeanine van der Ven et al. “The capacity of mid-pregnancy cervical length to predict preterm birth in low-risk women: a national cohort study”. En: *Acta Obstet Gynecol Scand* 94.11 (sep. de 2015), págs. 1223-1234.

[12] Joy Vink y Helen Feltovich. “Cervical etiology of spontaneous preterm birth”. En: *Seminars in Fetal and Neonatal Medicine* 21.2 (2016). Prediction and prevention of preterm birth and its sequelae, págs. 106-112.

pregnancy at term birth and sPTB [13], [14], which has prompted research on computerized analysis of this structure. One study implemented a quantitative analysis of cervix texture by extracting features from an anatomically relevant region of TVUS images using local binary patterns [15]. The machine learning model trained with these features obtained an area under the curve (AUC) of 0.77, outperforming cervical length (AUC = 0.60). Recently, a work employed convolutional neural networks to predict sPTB from first- and second-trimester TVUS images [16], achieving an AUC of 0.72. Although the results are promising, these studies use second-trimester images, when cervical changes related to PTB are more pronounced. Identifying women at risk of preterm birth earlier in pregnancy is crucial to the effectiveness of prevention strategies.

Radiomic analysis, the extraction of high-throughput computational features from medical images, has been used to extract clinically relevant information from radiological images. Its application has been extended to various imaging modalities, such as magnetic reso-

---

[13] Helen Feltovich, Kibo Nam y Timothy J. Hall. “Quantitative Ultrasound Assessment of Cervical Microstructure”. En: *Ultrasonic Imaging* 32.3 (jul. de 2010), 131–142.

[14] Helen Feltovich, Timothy J. Hall y Vincenzo Berghella. “Beyond cervical length: emerging technologies for assessing the pregnant cervix”. En: *American Journal of Obstetrics and Gynecology* 207.5 (nov. de 2012), 345–354.

[15] N. Baños et al. “Quantitative analysis of cervical texture by ultrasound in mid-pregnancy and association with spontaneous preterm birth”. En: *Ultrasound in Obstetrics & Gynecology* 51.5 (2018), págs. 637-643.

[16] Tomasz Włodarczyk et al. “Spontaneous Preterm Birth Prediction Using Convolutional Neural Networks”. En: *Lecture Notes in Computer Science*. Springer International Publishing, 2020, 274–283.

nance imaging [17], computed tomography [18] and ultrasound [19]. Radiomic analysis allows the transformation of an image into quantitative variables based on measures of intensity, morphometry and texture. In this way, it can reveal subtle relationships between image features and disease status that may be missed during visual inspection by a specialist [20]. Motivated by the success of radiomic analysis in a variety of medical applications, we studied its use in TVUS imaging to predict sPTB. Features extracted from these images were used to train five traditional machine learning models. Unlike other studies, our work used images from the first trimester of pregnancy, when changes associated with sPTB are less pronounced, making their detection more difficult. In addition, we also explored the potential of combining radiomic analysis with clinical variables to predict sPTB. Clinical variables include demographic information, cervical measurements (cervical length and cervical consistency index), and obstetric history. Our results demonstrate that radiomic analysis shows promise in this prediction. Furthermore, the construction of multivariate models integrating radiomic features with existing biomarkers improves the identification of women at high risk of sPTB.

---

[17] Ahmad Chaddad, Christian Desrosiers y Matthew Toews. “Multi-scale radiomic analysis of sub-cortical regions in MRI related to autism, gender and age”. En: *Sci Rep* 7.1 (mar. de 2017).

[18] Chia-Hung Chen et al. “Radiomic features analysis in computed tomography images of lung nodule classification”. En: *PLoS One* 13.2 (feb. de 2018), págs. 1-13.

[19] Jionghui Gu y Tian’an Jiang. “Ultrasound radiomics in personalized breast management: Current status and future prospects”. En: *Front Oncol* 12 (ago. de 2022).

[20] Michal R. Tomaszewski y Robert J. Gillies. “The Biological Meaning of Radiomic Features”. En: *Radiology* 298.3 (2021). PMID: 33399513, págs. 505-516.

1. OBJECTIVES

General objective

- To develop an algorithm for estimating the risk of spontaneous preterm birth from first-trimester transvaginal ultrasound images.

Specific objectives

1. To develop and implement an algorithm for the extraction of computerized features from transvaginal ultrasound images.
2. To build a machine learning-based model that uses the extracted features for estimating the risk of spontaneous preterm birth.
3. To evaluate the performance of the machine learning-based model and compare it with cervical length and cervical consistency index.

2. METHODS

2.1. Study design

This was a retrospective cohort study that included pregnant women who attended their screening ultrasound in the first trimester (11+0 to 13+6 weeks). Inclusion criteria were women with a singleton pregnancy within that same range of weeks and who planned to stay in a nearby area during gestation and delivery. Exclusion criteria were participants with a medical history of cancer, human immunodeficiency virus infection, or history of cervical surgery, as well as women whose gestation ended before 37 weeks due to medical indication based on maternal or fetal pathologies. TVUS images were obtained for cervical length (CL) measurement and cervical consistency index (CCI) assessment. In addition, demographic information and obstetric history of pregnant women were included. Data were collected between November 2019 and June 2021 in two maternal-fetal medicine units of two centers in Bucaramanga, Colombia: the Centro de Atención Materno-Fetal INUTERO and the Hospital Universitario de Santander. The imaging and clinical data used in this work were collected in two separate primary studies. The collection, data usage and informed consents, and authorization for usage in secondary research studies were reviewed and approved by the Institutional Review Board of Universidad Industrial de Santander (CEINCI) for those studies. In compliance with the requirements established in the primary studies, the usage of imaging and clinical data in this study was reviewed and approved by the Institutional Review Board of Universidad Industrial de Santander. Because data was collected and authorized in previous research studies, no informed consent was required for this study.

2.2. Cervical measurements

Transvaginal ultrasound images were obtained by maternal-fetal medicine specialists. During the scans, cervical length (CL) measurement and cervical consistency index (CCI) assessment were performed following established technical recommendations [21], [22]. All patients had an empty bladder and were in the lithotomic position. The vaginal transducer was then directed through the anterior vaginal fornix. To obtain an adequate image, the endocervical canal was identified in a sagittal view, and the image was magnified until the cervix occupied 75 % of the screen. CL measurement was performed by drawing a line between the internal and external os as reference points, avoiding the isthmus and ensuring proper orientation, as shown in Figure 1. The CCI assessment was performed by two measurements, making sure to position the callipers on the bright line of the anterior and posterior edge of the cervix. The first measurement corresponded to the basal anterior-posterior diameter (AP1) of the cervix, performed without applying pressure with the transducer. In the second measurement, the operator applied soft and gradual pressure with the transducer to the cervix until no further compression was observed. At this point, the anteroposterior diameter of the cervix at its maximum compression (AP2) was measured, as illustrated in Figure 2 (right side). CCI was calculated by dividing AP2 by AP1, obtaining a value between 0 and 1. The aim of CCI is to assess cervix softening. The images were acquired and stored in General Electric Voluson E6, General Electric Voluson E8 and General Electric Voluson S8 ultrasound equipment.

---

[21] J. Sonek y C. Shellhaas. “Cervical sonography: a review”. En: *Ultrasound Obstet Gynecol* 11.1 (1998), págs. 71-78.

[22] Carlos Hernan Becerra-Mojica et al. “Cohort profile: Colombian Cohort for the Early Prediction of Preterm Birth (COLPRET): early prediction of preterm birth based on personal medical history, clinical characteristics, vaginal microbiome, biophysical characteristics of the cervix and maternal serum biochemical markers”. En: *BMJ Open* 12.5 (2022).

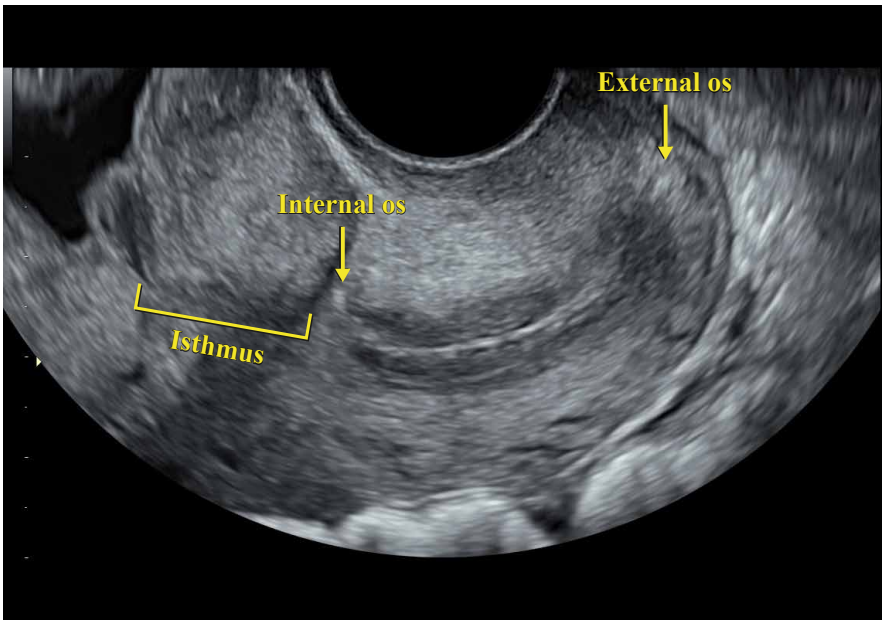

Figure 1. Characteristics of TVUS image for measurement of cervical length.

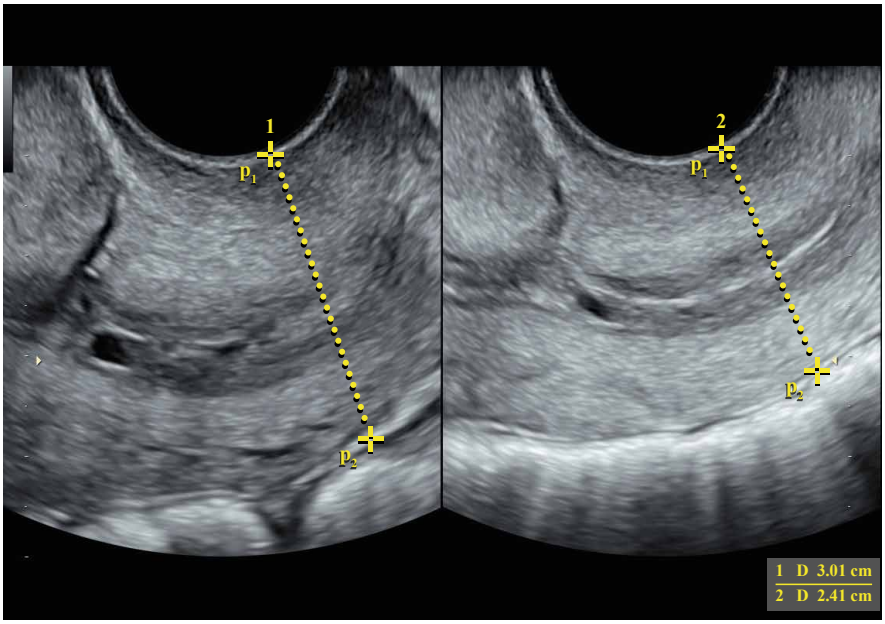

Figure 2. Anteroposterior diameter of the cervix without compression (left) and anteroposterior diameter at maximum compression (right) to obtain the cervical consistency index.

2.3. Preprocessing of TVUS images

For the subsequent computerized analysis, we use only the images acquired for the CCI assessment. Prior to the analysis, preprocessing is essential to improve the quality of the images and minimize its negative impact on the results. This preprocessing includes two fundamental steps: pixel size standardization and artifact removal.

Pixel size standardization. Because the images were acquired with different equipment and configurations, they have variations in resolution. These differences can affect the comparison and analysis of the data. To address this, we standardized all images to a pixel size of 0.21 mm, ensuring that the higher resolution images do not lose information.

Artifact removal. The first step involved automatic detection of the field of view (FOV), which contains the visible area of the tissue captured by the transducer. This detection was performed by morphological operations and made it possible to remove elements outside the FOV and peripheral areas such as the header and footer. Even within the FOV, unnecessary elements such as callipers, lines and text remain. Since their removal implies loss of tissue information, an exemplar-based image inpainting algorithm was employed [23]. It generates a coherent representation of the tissue using the surrounding pixels and ensures the visual integrity of the affected areas. The preprocessed image is shown in Figure 3.

2.4. Radiomic analysis

The radiomic analysis pipeline consists of two steps: the identification of the region of interest (ROI) and the extraction of radiomic features. Initially, the ROI delineation was

---

[23] Anupam, Pulkit Goyal y Sapan Diwakar. “Fast and Enhanced Algorithm for Exemplar Based Image Inpainting”. En: *2010 Fourth Pacific-Rim Symposium on Image and Video Technology*. 2010, págs. 325-330.

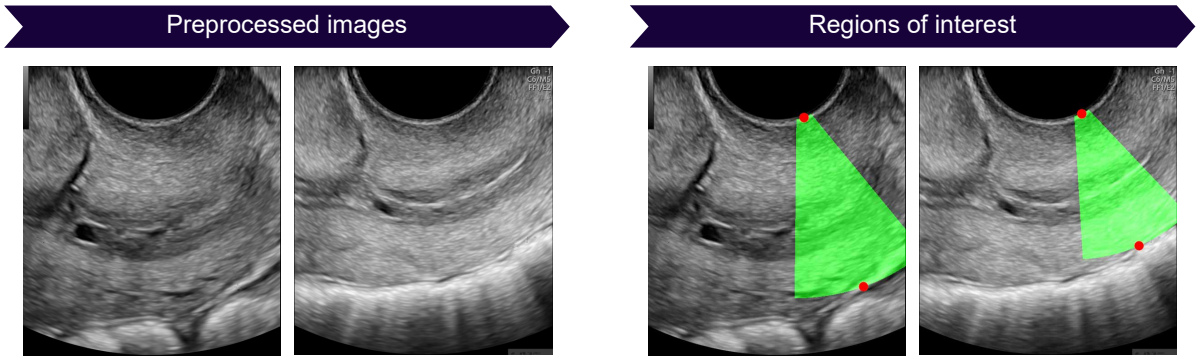

Figure 3. Images resulting from the preprocessing (left) and ROI identification (right) stages. Image preprocessing involved pixel size standardization and artifact removal. The ROIs (in green) were defined from the automatically detected callipers (in red).

based on the detection of callipers in the preprocessing stage. For this, a point p_0 was identified and located above the first calliper (p_1), ensuring a fixed distance (r = 50 pixels) between both points. Then, an angular scan of 20° to the left and 20° to the right from point p_0 was performed, defining the lower boundary of the ROI at point p_2. Figure 3 illustrates the resulting region. The definition of this ROI was supervised by an obstetrician-gynecologist for two purposes: to ensure a reproducible and fully automatic process, and to ensure that the ROI is anatomically relevant for cervix analysis. Subsequently, 33 radiomic features were extracted from that ROI using the OpenBreast package [24]. The extracted radiomic features are classified into five fundamental groups: statistical features, gray-level co-occurrence features, gray-level run-length features, gradient-based features, and spatial-frequency analysis. The statistical features aim to describe the properties of the gray-scale histogram of pixel intensities. The gray-level co-occurrence features seek to represent the statistical distribution of image intensities that co-occur at certain distances and pixel orientations. Gray-level run-length features describe the length and distribution of consecutive gray-level values in the image. Gradient-based features evaluate the change

[24] Said Pertuz et al. “Open Framework for Mammography-based Breast Cancer Risk Assessment”. En: *IEEE EMBS International Conference on Biomedical & Health Informatics*. 2019, págs. 1-4.

in pixel intensities in the horizontal and vertical directions of the image. Spatial-frequency analysis focuses on calculating descriptors that are based on texture properties, either in the spatial domain, the frequency domain, or both.

2.5. Risk assessment

The risk assessment was based on two key components: feature selection and risk scoring. We implemented five traditional machine learning models for sPTB prediction: Decision Trees (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machines (SVM), and Random Forests (RF). These models are typically used for detection, diagnosis or risk assessment of various medical conditions [25], [26]. The previously extracted features were standardized by z-score. Due to the low incidence of sPTB compared to term births, class imbalance is present, which may affect the generalization capability and performance of the model (see Annex A). To mitigate this, we applied SMOTE, an oversampling technique that balances the minority class [27].

Feature selection. One of the main challenges in machine learning applications is the curse of dimensionality [28], referred to as working with high dimensional data (many features) and a limited number of samples, which can lead to model overfitting. To mitigate this problem, feature selection is commonly used. A set of features can be classified in-

---

[25] Shumaila Batool y Saima Zainab. “A comparative performance assessment of artificial intelligence based classifiers and optimized feature reduction technique for breast cancer diagnosis”. En: *Computers in Biology and Medicine* 183 (2024), pág. 109215.

[26] Rasha M. Abd El-Aziz y Alanazi Rayan. “Early detection of sepsis using machine learning algorithms”. En: *Alexandria Engineering Journal* 111 (2025), págs. 47-56.

[27] N. V. Chawla et al. “SMOTE: Synthetic Minority Over-sampling Technique”. En: *Journal of Artificial Intelligence Research* 16 (jun. de 2002), 321–357.

[28] Naomi Altman y Martin Krzywinski. “The curse(s) of dimensionality”. En: *Nature Methods* 15.6 (mayo de 2018), 399–400.

to four groups: irrelevant features, redundant features, weakly relevant but not redundant features, and strongly relevant features. Optimal feature selection algorithm should identify those features that are strongly relevant and non-redundant, reducing overfitting and improving both the generalization and interpretability of the model. For this, we used sequential forward feature selection [29], a method that starts with an empty set and iteratively adds the feature that most improves the model's performance. At each iteration, the model performance is evaluated with each of the available unselected features, choosing the single feature that generates the largest improvement. This process stops until the performance improvement is less than a predefined threshold, leaving a subset of features. In this case, we use a threshold of 0.0001 in terms of AUC as a stopping criterion.

Risk scoring. Each model was trained with the feature subset selected in the previous step to generate a risk score, where values close to one indicate an increased risk of sPTB. The proposed workflow until the risk scores are obtained is shown in Figure 4. We used logistic regression to obtain a risk score for the cervical measurements (CL and CCI) individually. In the multivariate analysis, which combines risk scores from radiomic analysis with clinical variables (demographic information, cervical measurements and obstetric history), logistic regression was also used to generate a new risk score.

2.6. Statistical analysis

Maternal demographic characteristics, obstetric history and birth outcomes were compared between the sPTB and term birth (TB) groups. Normality of data was assessed by inspection of histograms and quantile-quantile plots. Quantitative data were presented as median and interquartile range, while categorical data were expressed as numbers and percentages. Quantitative variables were compared using the Mann-Whitney U-test (non-

---

[29] Girish Chandrashekar y Ferat Sahin. "A survey on feature selection methods". En: *Computers & Electrical Engineering* 40.1 (2014). 40th-year commemorative issue, págs. 16-28.

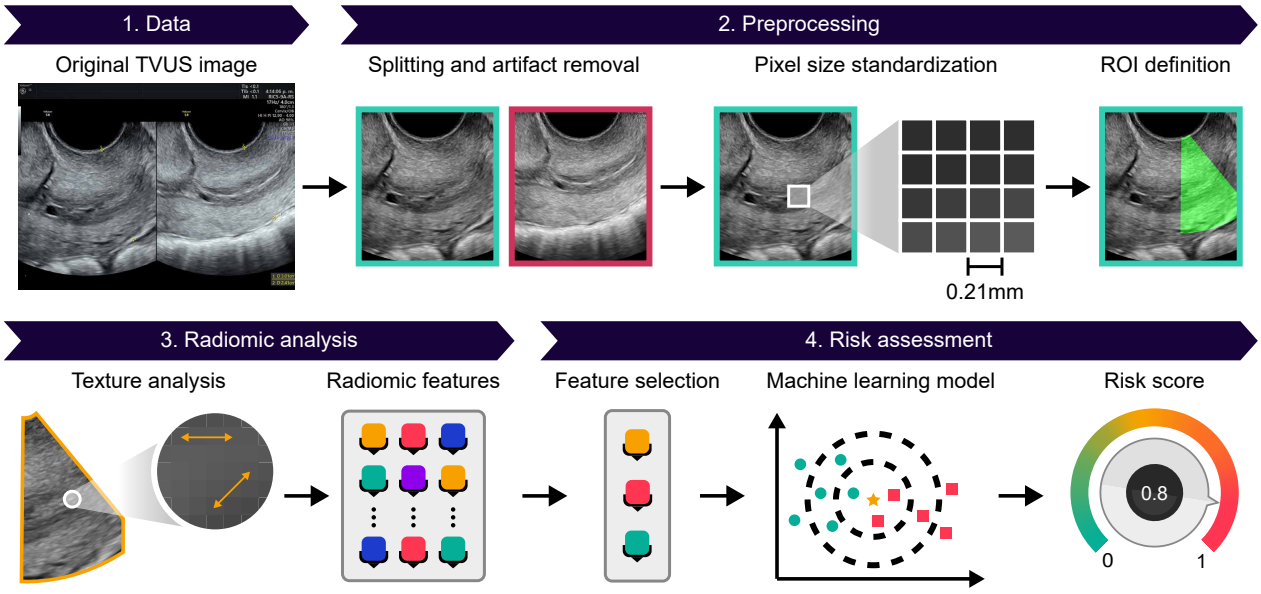

Figure 4. Proposed workflow for risk estimation of spontaneous preterm birth.

normal distributions) and categorical variables were compared using the chi-square test. The association between biomarkers and birth outcome was investigated by odds ratio (OR), calculated together with their 95 % confidence intervals using the normal distribution. A p-value less than 0.05 was considered statistically significant. The AUC was used to evaluate the performance of machine learning (ML) models and the predictive effectiveness of biomarkers based on radiomic features, CL, CCI and clinical variables. The AUC and its 95 % confidence intervals were calculated using the bootstrap resampling method. To compare the performance in terms of AUC between ML models, the DeLong test was applied. For the evaluation, we employed 5-fold cross-validation. In this evaluation method, the dataset is divided into five subsets of equal size. The model is trained on four of the subsets and tested on the remaining one. This process is repeated five times, each time using a different subset for testing, ensuring that all data points are used for both training and testing. An hyperparameter tuning of the five models was performed by Bayesian Search and 5-fold cross-validation, in order to improve their predictive ability while avoiding overfitting.

3. RESULTS

In this retrospective cohort study, 253 women were included: 225 (89 %) with term birth (TB) and 28 (11 %) with sPTB. Table 1 shows the demographic information, obstetric history and cervical measurements of both groups. Statistically significant differences were found only between the medians of CCI. Table 2 shows the performance of five machine learning models in terms of AUC for sPTB prediction. Each model was trained in two scenarios: using radiomic features from the image of the cervix without compression (UNC) and at its maximum compression (COM). The model that achieved the best performance was LR, using uncompressed cervix features. When comparing the performance of the models trained with these features, statistically significant differences were observed between DT and LR, as well as between LR and SVM.

Table 1. Demographic data, obstetric history and cervical measurements of the outcome groups. *Comparison between TB and sPTB. Data are given as median (interquartile range) or n (%). The statistical significance of differences in continuous data was calculated by Mann-Whitney U-test (non-normal distribution) and in categorical data by chi-square test. BMI: body mass index; CL: cervical length; CCI: cervical consistency index; GA: gestational age; PTB: preterm birth; TB: term birth.

Characteristic	**Total (n=253)**	**TB (n=225)**	**sPTB (n=28)**	**p-value***
Maternal age (years)	28.0 (24.0 - 33.0)	28.0 (23.0 - 33.0)	28.0 (25.0 - 32.3)	0.730
Height (m)	1.60 (1.56 - 1.64)	1.60 (1.56 - 1.64)	1.58 (1.55 - 1.63)	0.218
Weight (kg)	65.0 (57.0 - 75.0)	65.0 (57.0 - 75.0)	69.0 (59.8 - 79.0)	0.225
BMI (kg/m^2)	25.5 (22.8 - 28.3)	25.3 (22.7 - 28.2)	27.1 (24.1 - 30.8)	0.080
Nulliparous	80 (31.6)	74 (32.8)	6 (21.4)	0.310
PTB history	24 (9.5)	18 (8.0)	6 (21.4)	0.052
GA at screening (weeks)	13+1 (12+5 - 13+4)	13+1 (12+5 - 13+4)	13+4 (12+6 - 13+5)	0.088
GA at birth (weeks)	38+5 (38+0 - 39+4)	38+6 (38+1 - 39+4)	36+1 (34+4 - 36+4)	-
CL (mm)	37.0 (35.0 - 40.0)	37.0 (35.0 - 40.0)	35.5 (32.7 - 39.0)	0.067
CCI (%)	82.8 (78.5 - 87.0)	83.3 (79.3 - 87.0)	78.5 (73.5 - 82.1)	$<$0.01

The LR model trained with the radiomic features (RadF) of the cervix without compression was compared and combined with other biomarkers, and it will now be referred to as the RadF biomarker. Table 3 presents the performance of six biomarkers in sPTB prediction.

Table 2. Model performance and association assessment using radiomic features of the cervix without compression (UNC) and at maximum compression (COM). *Statistical significance (p-value < 0.05). DT: decision trees; KNN: k-nearest neighbors; LR: logistic regression; SVM: support vector machines; RF: random forests.

Status	Model	AUC (95 % CI)	OR (95 % CI)
UNC	DT	0.49 (0.39-0.59)	0.93 (0.61-1.40)
	KNN	0.60 (0.48-0.70)	1.37 (0.93-2.03)
	LR	**0.67 (0.57-0.77)**	1.75 (1.17-2.61)*
	RF	0.58 (0.47-0.70)	1.36 (0.95-1.95)
	SVM	0.55 (0.45-0.65)	0.86 (0.52-1.44)
COM	DT	0.57 (0.47-0.68)	1.31 (0.92-1.86)
	KNN	**0.64 (0.52-0.75)**	1.63 (1.10-2.41)*
	LR	0.44 (0.33-0.55)	0.86 (0.59-1.26)
	RF	0.60 (0.48-0.71)	1.32 (0.92-1.88)
	SVM	0.45 (0.34-0.57)	0.87 (0.58-1.28)

In the first two scenarios, the performance of cervical measurements (CL and CCI) was evaluated individually. In the third scenario, radiomic features were used exclusively. The next two scenarios combined RadF with cervical measurements. The last biomarker explored the complementary value of radiomic features together with clinical variables (CV). The clinical variables considered were age, height, weight, body mass index, gestational age at cervical measurement, history of preterm birth, number of previous preterm births, cervical length, cervical consistency index, AP1 diameter and AP2 diameter. The RadF+CV combination showed the best performance, with an AUC of 0.79. In addition, the association of each biomarker with delivery outcome was analyzed by odds ratio, finding that all biomarkers except CL had a statistically significant association. Table 4 summarizes the predictive performance in terms of sensitivity, specificity, positive likelihood ratio (LR+) and negative likelihood ratio (LR-) for false positive rates (FPR) of 10 % and 20 %. The RadF + CV combination showed a sensitivity of 46.4 %, a specificity of 87.1 %, an LR+ of 3.60 and an LR- of 0.62 for an FPR of 10 %.

Table 3. Biomarker performance in sPTB prediction and association assessment. *Statistical significance (p-value < 0.05).

Biomarker	AUC (95 % CI)	OR (95 % CI)
CL	0.61 (0.48-0.73)	0.71 (0.46 - 1.10)
CCI	0.68 (0.55-0.79)	0.57 (0.39 - 0.83)*
RadF	0.67 (0.57-0.77)	1.75 (1.17-2.61)*
RadF + CL	0.68 (0.56-0.79)	1.87 (1.34 - 2.62)*
RadF + CCI	0.74 (0.64-0.83)	1.70 (1.23 - 2.34)*
RadF + CV	**0.79 (0.69-0.88)**	2.24 (1.58 - 3.18)*

Table 4. Predictive performance of biomarkers at different false positive rates.

(a) False positive rate at 10 %

Biomarker	Sensitivity (%)	Specificity (%)	LR+ (95 % CI)	LR- (95 % CI)
CL	25.0	89.8	2.45 (0.90 - 4.88)	0.84 (0.64 - 1.01)
CCI	32.1	88.9	2.89 (1.29 - 5.52)	0.76 (0.56 - 0.96)
RadF	14.3	93.3	2.14 (0.50 - 5.78)	0.92 (0.75 - 1.05)
RadF + CL	25.0	87.6	2.01 (0.78 - 3.82)	0.86 (0.67 - 1.04)
RadF + CCI	25.0	89.8	2.45 (0.83 - 5.06)	0.84 (0.64 - 1.02)
RadF + CV	46.4	87.1	3.60 (1.95 - 6.03)	0.62 (0.40 - 0.84)

(b) False positive rate at 20 %

Biomarker	Sensitivity (%)	Specificity (%)	LR+ (95 % CI)	LR- (95 % CI)
CL	42.9	80.0	2.14 (1.22 - 3.40)	0.71 (0.48 - 0.94)
CCI	46.4	79.6	2.27 (1.33 - 3.55)	0.67 (0.43 - 0.91)
RadF	35.7	80.4	1.83 (0.91 - 3.07)	0.80 (0.57 - 1.02)
RadF + CL	46.4	78.7	2.18 (1.24 - 3.45)	0.68 (0.43 - 0.93)
RadF + CCI	53.6	77.3	2.36 (1.48 - 3.64)	0.60 (0.34 - 0.85)
RadF + CV	53.6	80.4	2.74 (1.68 - 4.12)	0.58 (0.35 - 0.82)

Figure 5 shows the risk scores for three biomarkers. The medians are higher in the sPTB group than in the TB group, indicating a higher associated risk. The largest difference between the medians of the two groups was observed in RadF + CV. In all three biomarkers analyzed, the differences between the medians were statistically significant (p-value < 0.01).

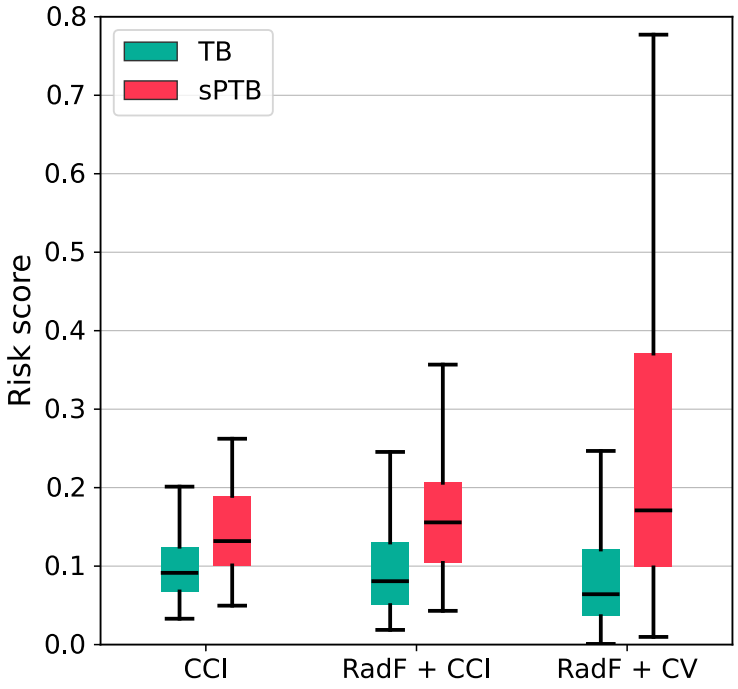

Figure 5. Risk scores of the TB and sPTB groups using the biomarkers CCI, RadF + CCI, and RadF + CV.

4. DISCUSSION

4.1. Principal findings

Our study shows the potential of radiomic analysis for the prediction of sPTB, with an AUC of 0.67. Combining RadF with cervical measurements improved performance, and interestingly, integrating RadF with clinical variables yielded the best discriminatory ability (AUC = 0.79) between sPTB and TB. These results suggest the need to explore the complementary value of multiple biomarkers rather than relying on a single one for prediction. Radiomic analysis showed a statistically significant association with delivery outcome, highlighting the importance of using imaging data. These data can reveal patterns and abnormal changes in cervical tissue related to sPTB that might be missed by a human observer, but can be identified by the computerized analysis.

4.2. Comparison with other works

Previous studies have used computerized techniques to analyze cervical texture and predict PTB. One study used the Quantus Prematurity tool to automatically analyze cervical texture from second trimester TVUS images, achieving an AUC of 0.63 [30]. In comparison, the exclusive use of CL for prediction showed an AUC of 0.55, being inferior to texture analysis. However, by combining both approaches, the AUC increased to 0.67, which is in line with our findings. Another study used convolutional neural networks to analyze first- and second-trimester TVUS images, obtaining an AUC of 0.72 [16]. In contrast to those studies, we exclusively analyzed first-trimester TVUS images, obtaining promising

[30] Xavier P. Burgos-Artizzu et al. “Mid-trimester prediction of spontaneous preterm birth with automated cervical quantitative ultrasound texture analysis and cervical length: a prospective study”. En: *Sci Rep* 11.1 (abr. de 2021).

results despite the early stage of pregnancy, when abnormal changes in cervical tissue are less evident. Our work paves the way for future research on the use of TVUS imaging in early pregnancy to predict sPTB. A recent study explored the prediction of PTB in the first trimester by combining maternal factors, obstetric history, and biomarkers of placental function [31], achieving an AUC of 0.63. In contrast, maternal factors alone only achieved an AUC of 0.55. These results support our hypothesis that the integration of multiple biomarkers improves discrimination between groups, with the simplicity that our approach incorporates imaging data that can be obtained in routine clinical practice.

4.3. Clinical implications

Our findings suggest that radiomic analysis of TVUS images in the first trimester may be a useful tool for predicting sPTB, providing valuable information that could facilitate the implementation of timely interventions by clinicians. Combining clinical variables with imaging data improves the ability to discriminate between TB and sPTB. This approach has the potential to improve early risk stratification, which could lead to more personalized care for women at high risk. In addition, integration of this approach into clinical workflows could be easier, since it uses data and images routinely acquired in clinical practice. Our results also show that CCI has a better discrimination ability between groups during the first trimester, compared to CL, where no statistically significant differences were found. These results are in agreement with those of Nguyen et al. [32] who suggest that cervical softening precedes shortening. Further research is required to confirm this hypothesis and to better understand the cervical changes associated with sPTB. Nevertheless, these

---

[31] C. P. H. Chiu et al. “Prediction of spontaneous preterm birth and preterm prelabor rupture of membranes using maternal factors, obstetric history and biomarkers of placental function at 11–13 weeks”. En: *Ultrasound Obstet Gynecol* 60.2 (2022), págs. 192-199.

[32] L. Nguyen-Hoang et al. “Longitudinal evaluation of cervical length and shear wave elastography in women with spontaneous preterm birth”. En: *Ultrasound Obstet Gynecol* 63.6 (2024), págs. 789-797.

findings provide a baseline for the adoption of intervention strategies aimed at reducing accelerated cervical softening and ultimately preventing sPTB.

4.4. Strengths and limitations

The main strength of this work is the use of first-trimester TVUS images, which is a promising strategy for early identification of women at high risk for sPTB. Computerized analysis and machine learning-based methods can detect abnormal cervical features related to sPTB that might be missed by a human observer. In addition, radiomic analysis follows clear and well-established guidelines [33], facilitating the reproducibility and comparability of results. Another point in favor is that the gestational age at delivery was estimated from the crown-rump length in the first trimester. Although our algorithm was used offline, i.e., not in real time during the acquisition of the ultrasound images, its integration into the ultrasound equipment software would be straightforward, avoiding the need for complex infrastructure and reducing the associated costs. The main limitation of this study is the sample size, which may not be sufficient to make robust conclusions. In addition, class imbalance (more term births than preterm births) is a common challenge in machine learning algorithms and may affect their generalization ability. As the CL and CCI measurements are based on manual annotations, there is inter- and intra-operator variability that could influence the comparability of results. Data collection at only two centers may introduce selection bias and underrepresent certain population characteristics. We are aware of the importance of validating the algorithm in a large-scale prospective cohort study to ensure its general applicability. Another limitation of this study is the variability in the results when using different machine learning models to predict sPTB. The sample size does not allow us to understand why some models showed lower performance. It is unclear whether the

---

[33] Alex Zwanenburg et al. “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping”. En: *Radiology* 295.2 (2020). PMID: 32154773, págs. 328-338.

better performing models were able to distinguish more effectively between groups due to their structure or whether this variability is due to the methodology based on radiomic analysis. The sample size needs to be increased to investigate this variability.

4.5. Other explorations

Deep learning generally requires large amounts of data to obtain good results. However, on small datasets, these methods can show shortcomings in their performance. For this reason, deep learning was not our first choice for this study. Instead, radiomic analysis is a widely studied method in medical image analysis, less susceptible to overfitting bias in small datasets [33]. Nevertheless, we adopted a deep learning-based approach to quantify cervix deformations. Specifically, we used the image evaluating the CCI (Figure 2) to estimate the optical flow, i.e., the displacement of pixels from the uncompressed cervix image to the image of the cervix at its maximum compression. We then performed a radiomic analysis of this optical flow image and trained five traditional machine learning models to predict the sPTB, whose results are presented in Annex B. On the other hand, an exploration was conducted in Annex C following the same workflow described in Figure 4, but using the image measuring cervical length (CL) as input. Although these explorations did not outperform our proposed main method, the results warrant further research, particularly with more complex methods such as deep learning.

5. CONCLUSION

We proposed a methodology based on radiomic analysis and machine learning techniques for sPTB prediction in the first trimester. The best performing model was logistic regression trained with the radiomic features extracted from the cervix without compression. The radiomic analysis obtained an AUC of 0.67, while its combination with CL, CCI and clinical variables increased the AUC to 0.68, 0.74 and 0.79, respectively. These findings highlight the potential of radiomic analysis to detect subtle patterns associated with sPTB in early pregnancy, as well as the added value of integrating multiple sources of information. This work paves the way for future research exploring the incorporation of multiple biomarkers into sPTB risk estimation models.

CONTRIBUTIONS

Journal Articles

- Cancino, W., Becerra-Mojica, C.H., & Pertuz, S. (2024). The added value of radiomic analysis for predicting spontaneous preterm birth in the first trimester. Ultrasound in Medicine and Biology.
  Status: Under review

Conference Proceedings

- Cancino, W., Becerra-Mojica, C.H., & Pertuz, S. (2024). Radiomic Analysis of Transvaginal Ultrasound Cervical Images for Prediction of Preterm Birth. In Medical Image Understanding and Analysis (MIUA 2024). Lecture Notes in Computer Science, vol 14860. Springer, Cham. DOI: 10.1007/978-3-031-66958-3_30
  Status: Published
- Cancino, W., Becerra-Mojica, C.H., & Pertuz, S. (2024). Análisis computarizado de imágenes de ultrasonido transvaginal cervicales para la estimación de riesgo de parto prematuro espontáneo. In IV Encuentro Internacional de Ciencias de la Salud: La Ciencia Transforma y el IV Congreso Internacional del Proceso de Enfermería y del Lenguaje Estandarizado. Revista Salud UIS, vol 56. Available: `https://revistas.uis.edu.co/index.php/revistasaluduis/article/view/15959/13764`
  Status: Published

Collaborations

- Pertuz, S., Ortega, D., Suarez, É., Cancino, W., Africano, G., Rinta-Kiikka, I., Arponen, O., Paris, S., & Lozano, A. (2023). Saliency of breast lesions in breast cancer detection using artificial intelligence. Scientific Reports, vol. 13, no. 1. Springer Science and Business Media LLC, p. 20545. DOI: 10.1038/s41598-023-46921-3
  Status: Published

REFERENCES

Abd El-Aziz, Rasha M. y Alanazi Rayan. “Early detection of sepsis using machine learning algorithms”. En: *Alexandria Engineering Journal* 111 (2025), págs. 47-56 (vid. pág. 21).

Al-Hindi, Mohammed Y et al. “Screening for Neurodevelopmental Delay for Preterm Very Low Birth Weight Infants at Tertiary Care Center in Saudi Arabia”. En: *Cureus* (dic. de 2021) (vid. pág. 11).

Altman, Naomi y Martin Krzywinski. “The curse(s) of dimensionality”. En: *Nature Methods* 15.6 (mayo de 2018), 399–400 (vid. pág. 21).

Anupam, Pulkit Goyal y Sapan Diwakar. “Fast and Enhanced Algorithm for Exemplar Based Image Inpainting”. En: *2010 Fourth Pacific-Rim Symposium on Image and Video Technology*. 2010, págs. 325-330 (vid. pág. 19).

Batool, Shumaila y Saima Zainab. “A comparative performance assessment of artificial intelligence based classifiers and optimized feature reduction technique for breast cancer diagnosis”. En: *Computers in Biology and Medicine* 183 (2024), pág. 109215 (vid. pág. 21).

Baños, N. et al. “Quantitative analysis of cervical texture by ultrasound in mid-pregnancy and association with spontaneous preterm birth”. En: *Ultrasound in Obstetrics & Gynecology* 51.5 (2018), págs. 637-643 (vid. pág. 13).

Becerra-Mojica, Carlos Hernan et al. “Cohort profile: Colombian Cohort for the Early Prediction of Preterm Birth (COLPRET): early prediction of preterm birth based on personal medical history, clinical characteristics, vaginal microbiome, biophysical characteristics of the cervix and maternal serum biochemical markers”. En: *BMJ Open* 12.5 (2022) (vid. pág. 17).

Burgos-Artizzu, Xavier P. et al. “Mid-trimester prediction of spontaneous preterm birth with automated cervical quantitative ultrasound texture analysis and cervical length: a prospective study”. En: *Sci Rep* 11.1 (abr. de 2021) (vid. pág. 28).

Chaddad, Ahmad, Christian Desrosiers y Matthew Toews. “Multi-scale radiomic analysis of sub-cortical regions in MRI related to autism, gender and age”. En: *Sci Rep* 7.1 (mar. de 2017) (vid. pág. 14).

Chandrashekar, Girish y Ferat Sahin. “A survey on feature selection methods”. En: *Computers & Electrical Engineering* 40.1 (2014). 40th-year commemorative issue, págs. 16-28 (vid. pág. 22).

Chawla, N. V. et al. “SMOTE: Synthetic Minority Over-sampling Technique”. En: *Journal of Artificial Intelligence Research* 16 (jun. de 2002), 321–357 (vid. pág. 21).

Chen, Chia-Hung et al. “Radiomic features analysis in computed tomography images of lung nodule classification”. En: *PLoS One* 13.2 (feb. de 2018), págs. 1-13 (vid. pág. 14).

Chiu, C. P. H. et al. “Prediction of spontaneous preterm birth and preterm prelabor rupture of membranes using maternal factors, obstetric history and biomarkers of placental function at 11–13 weeks”. En: *Ultrasound Obstet Gynecol* 60.2 (2022), págs. 192-199 (vid. pág. 29).

Departamento Administrativo Nacional de Estadística (DANE). *Estadísticas Vitales (EEVV) - Nacimientos en Colombia*. [Online]. Available: `https://www.dane.gov.co/files/operaciones/EEVV/bol-EEVV-Nacimientos-IVtrim2023.pdf`. Accessed: Sept. 04, 2024. Mar. de 2023 (vid. pág. 11).

Esplin, M. Sean et al. “Predictive Accuracy of Serial Transvaginal Cervical Lengths and Quantitative Vaginal Fetal Fibronectin Levels for Spontaneous Preterm Birth Among Nulliparous Women”. En: *JAMA* 317.10 (mar. de 2017), pág. 1047.

Feltovich, Helen, Timothy J. Hall y Vincenzo Berghella. “Beyond cervical length: emerging technologies for assessing the pregnant cervix”. En: *American Journal of Obstetrics and Gynecology* 207.5 (nov. de 2012), 345–354 (vid. pág. 13).

Feltovich, Helen, Kibo Nam y Timothy J. Hall. “Quantitative Ultrasound Assessment of Cervical Microstructure”. En: *Ultrasonic Imaging* 32.3 (jul. de 2010), 131–142 (vid. pág. 13).

Goldenberg, Robert L et al. “Epidemiology and causes of preterm birth”. En: *The Lancet* 371.9606 (ene. de 2008), 75–84 (vid. pág. 11).

Gu, Jionghui y Tian’an Jiang. “Ultrasound radiomics in personalized breast management: Current status and future prospects”. En: *Front Oncol* 12 (ago. de 2022) (vid. pág. 14).

Kalengo, Nathaniel Halide et al. “Recurrence rate of preterm birth and associated factors among women who delivered at Kilimanjaro Christian Medical Centre in Northern Tanzania: A registry based cohort study”. En: *PLoS One* 15.9 (sep. de 2020). Ed. por Joel Msafiri Francis, e0239037 (vid. pág. 12).

Lowe, John, Sarah J. Kotecha y Sailesh Kotecha. “Long Term Effects Following Extreme Prematurity: Respiratory Problems”. En: *Emerging Topics and Controversies in Neonatology*. Springer International Publishing, 2020, 351–366 (vid. pág. 11).

Nakamura, Kazue et al. “Exclusively Breastfeeding Modifies the Adverse Association of Late Preterm Birth and Gastrointestinal Infection: A Nationwide Birth Cohort Study”. En: *Breastfeed Med* 15.8 (2020). PMID: 32543213, págs. 509-515 (vid. pág. 11).

Nguyen-Hoang, L. et al. “Longitudinal evaluation of cervical length and shear wave elastography in women with spontaneous preterm birth”. En: *Ultrasound Obstet Gynecol* 63.6 (2024), págs. 789-797 (vid. pág. 29).

Ohuma, Eric O et al. “National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis”. En: *The Lancet* 402.10409 (oct. de 2023), 1261–1271 (vid. pág. 11).

Organization, World Health et al. *Born too soon: decade of action on preterm birth*. World Health Organization, 2023, ix, 108 p. (Vid. pág. 11).

Pertuz, Said et al. “Open Framework for Mammography-based Breast Cancer Risk Assessment”. En: *IEEE EMBS International Conference on Biomedical & Health Informatics*. 2019, págs. 1-4 (vid. pág. 20).

Reicher, Lee, Yuval Fouks y Yariv Yogev. “Cervical Assessment for Predicting Preterm Birth-Cervical Length and Beyond”. En: *J Clin Med* 10.4 (feb. de 2021) (vid. pág. 12).

Sonek, J. y C. Shellhaas. “Cervical sonography: a review”. En: *Ultrasound Obstet Gynecol* 11.1 (1998), págs. 71-78 (vid. pág. 17).

Teed, Zachary y Jia Deng. “RAFT: Recurrent All-Pairs Field Transforms for Optical Flow”. En: *Computer Vision – ECCV 2020*. Ed. por Andrea Vedaldi et al. Cham: Springer International Publishing, 2020, págs. 402-419 (vid. pág. 40).

Tingleff, T et al. “Risk of preterm birth in relation to history of preterm birth: a population-based registry study of 213 335 women in Norway”. En: *BJOG* 129.6 (nov. de 2021), págs. 900-907 (vid. pág. 12).

Tomaszewski, Michal R. y Robert J. Gillies. “The Biological Meaning of Radiomic Features”. En: *Radiology* 298.3 (2021). PMID: 33399513, págs. 505-516 (vid. pág. 14).

UIS, Facultad de Salud. “Memorias IV Encuentro internacional de ciencias de la salud: la ciencia transforma IV Congreso internacional del proceso de enfermería y del lenguaje estandarizado”. En: *Salud UIS* 56 (2024).

Ven, Jeanine van der et al. “The capacity of mid-pregnancy cervical length to predict preterm birth in low-risk women: a national cohort study”. En: *Acta Obstet Gynecol Scand* 94.11 (sep. de 2015), págs. 1223-1234 (vid. pág. 12).

Vink, Joy y Helen Feltovich. “Cervical etiology of spontaneous preterm birth”. En: *Seminars in Fetal and Neonatal Medicine* 21.2 (2016). Prediction and prevention of preterm birth and its sequelae, págs. 106-112 (vid. pág. 12).

World Health Organization (WHO). *15 Million babies born too soon*. [Online]. Available: `https://www.who.int/news/item/02-05-2012-15-million-babies-born-too-soon`. Accessed: Oct. 20, 2023. Mayo de 2012.

Włodarczyk, Tomasz et al. “Spontaneous Preterm Birth Prediction Using Convolutional Neural Networks”. En: *Lecture Notes in Computer Science*. Springer International Publishing, 2020, 274–283 (vid. págs. 13, 28).

Zwanenburg, Alex et al. “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping”. En: *Radiology* 295.2 (2020). PMID: 32154773, págs. 328-338 (vid. págs. 30, 31).

ANNEXES

Annex A. Model performance when class balancing is not applied

Model performance was evaluated without using class balancing techniques. Based on Table 5, most models obtained a lower AUC compared to when SMOTE was used for class balancing. Furthermore, statistically significant association was not observed in any case, suggesting that class imbalance may negatively affect the performance of the models.

Table 5. Model performance and association assessment without class balancing, using radiomic features of the cervix without compression (UNC) and at maximum compression (COM). DT: decision trees; KNN: k-nearest neighbors; LR: logistic regression; SVM: support vector machines; RF: random forests.

Status	Model	AUC (95 % CI)	OR (95 % CI)
UNC	DT	0.56 (0.45-0.67)	1.08 (0.76-1.54)
	KNN	0.54 (0.42-0.65)	1.18 (0.81-1.71)
	LR	0.57 (0.46-0.68)	1.11 (0.78-1.58)
	RF	0.57 (0.46-0.68)	1.17 (0.81-1.69)
	SVM	0.42 (0.32-0.53)	0.91 (0.60-1.39)
COM	DT	0.50 (0.40-0.61)	0.97 (0.65-1.46)
	KNN	0.61 (0.50-0.72)	1.41 (0.97-2.03)
	LR	0.47 (0.35-0.58)	0.89 (0.60-1.32)
	RF	0.54 (0.43-0.64)	1.08 (0.74-1.57)
	SVM	0.53 (0.41-0.65)	1.38 (0.95-1.99)

Annex B. Analysis of cervix deformation

Based on the results obtained by the CCI (AUC = 0.68) in the first trimester, we decided to explore computerized methods to study cervix deformation. We analyzed cervix images both without compression and at maximum compression to estimate optical flow, which calculates pixel displacement between consecutive images. For this purpose, we

employed the Recurrent All-Pairs Field Transforms (RAFT) deep network architecture [34], specifically, a version pre-trained on the KITTI dataset. RAFT extracts per-pixel features, constructs 4D multiscale correlation volumes for all pixel pairs, and iteratively updates the flow field through a recurrent unit that performs searches on the correlation volumes. Before using RAFT, translational and rotational corrections were performed on the cervix image with compression to ensure coincidence of the anatomical regions of the cervix. This alignment was necessary due to the movements introduced by the operator's manipulation of the transducer, which generate apparent displacements unrelated to the pressure applied with the transducer. The content of both images was extracted based on the region of interest defined for the cervix image with compression, and the optical flow in these regions was estimated. The processing pipeline until obtaining the optical flow is shown in Figure 6. Radiomic features were extracted from the optical flow images using PyRadiomics and subsequently used to train five traditional machine learning models. As shown in Table 6, the best model was LR. The results obtained were poor, probably because displacements introduced by transducer manipulation still persist. These unwanted movements prevent the analyzed regions from being coincident between images. In future research, it will be necessary to employ advanced alignment techniques to ensure correspondence of anatomical regions and allow objective deformation analysis.

---

[34] Zachary Teed y Jia Deng. "RAFT: Recurrent All-Pairs Field Transforms for Optical Flow". En: *Computer Vision – ECCV 2020*. Ed. por Andrea Vedaldi et al. Cham: Springer International Publishing, 2020, págs. 402-419.

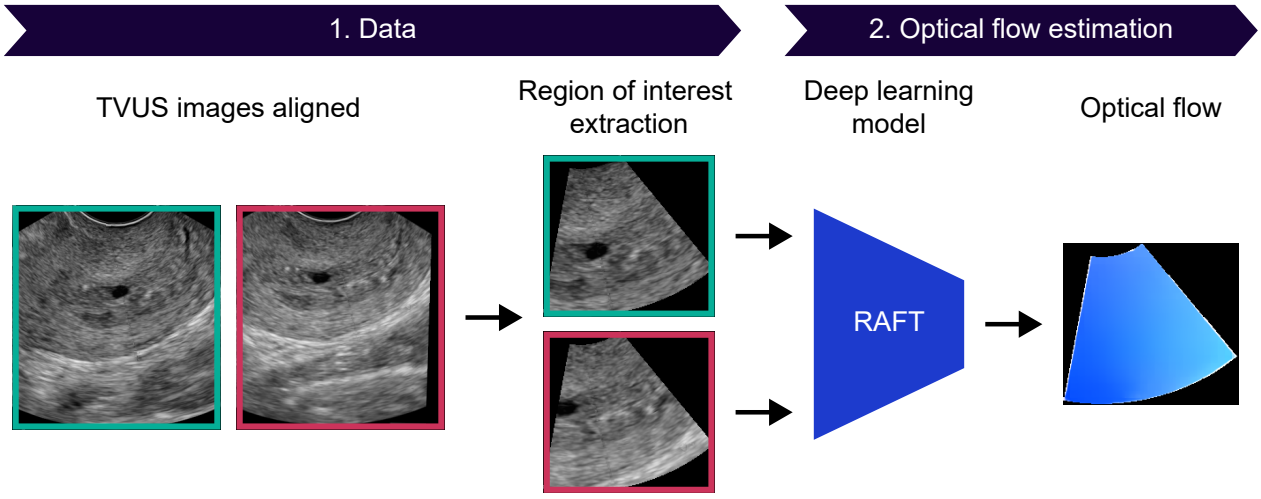

Figure 6. Proposed workflow for optical flow estimation.

Table 6. Model performance and association assessment using deformation analysis.

Model	AUC (95 % CI)	OR (95 % CI)
DT	0.44 (0.35-0.55)	0.80 (0.52-1.24)
KNN	0.48 (0.36-0.60)	0.98 (0.66-1.45)
LR	**0.50 (0.38-0.61)**	1.12 (0.76-1.67)
RF	0.45 (0.34-0.57)	0.88 (0.59-1.33)
SVM	0.49 (0.37-0.62)	1.10 (0.79-1.54)

Annex C. Prediction of sPTB using cervical length images

We also used the image measuring cervical length (Figure 1) to predict sPTB, which provides greater coverage of the cervix compared to the images used to assess CCI. From a region of interest defined in Figure 7, 33 features were extracted using OpenBreast. These features were used to train five machine learning models, evaluated using the 5-fold cross-validation scheme. According to the results shown in Table 7, KNN was the

only model that outperformed CL (AUC = 0.62). One aspect to improve in the future is the definition of the region of interest, which may not be capturing all anatomically relevant portions of the cervix.

Figure 7. Region of interest for images where cervical length is measured.

Table 7. Model performance and association assessment using cervical length images. *Statistical significance (p-value < 0.05).

Model	**AUC (95 % CI)**	**OR (95 % CI)**
DT	0.50 (0.41-0.60)	0.95 (0.64-1.43)
KNN	**0.63 (0.53-0.72)**	1.52 (1.01-2.28)*
LR	0.50 (0.38-0.62)	0.99 (0.67-1.47)
RF	0.51 (0.41-0.60)	0.90 (0.59-1.35)
SVM	0.53 (0.40-0.65)	1.04 (0.72-1.51)