ALGORITHM FOR THE AUTOMATIC MEASUREMENT OF CERVICAL LENGTH FROM TRANSVAGINAL ULTRASOUND IMAGES

MARIA FERNANDA VALENZUELA SANCHEZ

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

ALGORITHM FOR THE AUTOMATIC MEASUREMENT OF CERVICAL LENGTH FROM TRANSVAGINAL ULTRASOUND IMAGES

MARIA FERNANDA VALENZUELA SANCHEZ

Degree work presented as a requirement to qualify for the title of Electronic Engineer

Advisor
William Andrés Cancino Rey
Electronic Engineer

Codirector
Said David Pertuz Arroyo
Título académico completo de mayor rango

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

DEDICATION

To my parents, because there are no words enough to thank you for everything you have done for me. To you, who have been my shelter, my strength, and my greatest example in life. To my dad, my silent hero, who always protects me with tenderness and pride. Thank you for making me feel like your little girl, for believing in me even when I doubted myself, and for teaching me to be strong with a heart full of love. To my mom, my confidant, my guide, my unconditional friend. Thank you for always listening to me, for walking by my side through every step, for your wise advice, and for holding me through every fall. Your love has been the support that has lifted me up time and again. This achievement is not mine alone it's yours too. Without your love, support, and faith in me, I wouldn't be here today.

ACKNOWLEDGMENTS

First and foremost, I thank God for giving me life, health, strength, and wisdom to complete this stage. His guidance and presence have been fundamental throughout this journey. To my family, thank you for your unconditional love, constant support, and for believing in me every step of the way. Your encouragement and presence have been the driving force that kept me going, even during the most challenging moments. I am deeply grateful to my advisor, William Cancino, for his continuous support, academic guidance, and dedication to this work. His experience and commitment were key to achieving this result. I also extend my sincere thanks to my co-advisor, Said Pertuz, for his support, technical contributions, and valuable advice, which greatly enriched this research process. To the CPS Research Group, thank you for providing me with a space for growth, learning, and collaboration, and for trusting in my work within the team. To my friends from university, thank you for your companionship, motivation, and for making this journey more enjoyable and meaningful. Your emotional support and friendship meant a great deal to me throughout this process.

TABLE OF CONTENTS

LIST OF FIGURES

page.

LIST OF TABLES

page.

LIST OF APPENDICES

RESUMEN

TÍTULO ALGORITMO PARA LA MEDICIÓN AUTOMÁTICA DE LA LONGITUD CERVICAL A PARTIR DE IMÁGENES DE ULTRASONIDO TRANSVAGINAL. [*]

AUTOR: María Fernanda Valenzuela Sánchez [**]

PALABRAS CLAVE: Parto prematuro, longitud cervical, ecografía transvaginal, redes neuronales convolucionales, detección de calibrador y segmentación de imágenes.

DESCRIPCIÓN: El parto prematuro (PP) es un importante problema de salud pública a nivel mundial, que afecta aproximadamente a 15 millones de embarazos cada año. Uno de los indicadores más fiables del riesgo de PP es la longitud cervical (LC), la cual se mide manualmente en imágenes de ecografía transvaginal (ETV). Sin embargo, las mediciones manuales están sujetas a variabilidad entre operadores, lo que limita su fiabilidad. Para mejorar la consistencia y la automatización, este estudio propone un enfoque basado en aprendizaje profundo para la evaluación de la LC. Se desarrolló un modelo basado en redes neuronales convolucionales (CNN) para detectar las posiciones de los calibradores y estimar la LC. Para lograr esto, primero se segmentó el tercio medio del cuello uterino en imágenes ETV del primer trimestre utilizando una red U-Net. Estas segmentaciones se usaron luego para entrenar el modelo de detección de calibradores, permitiendo la medición automática de la LC. El conjunto de datos incluyó a 311 mujeres embarazadas de dos centros especializados en salud materno-fetal. El modelo de segmentación fue evaluado utilizando el coeficiente de Dice, la distancia media de superficie (MSD) y la tasa de error de área (AER), logrando una puntuación Dice de 0.83 (IC del 95 %: 0.78–0.88), una MSD de 8.89 mm (IC del 95 %: 5.47–12.31), y una AER de 607 mm^2 (IC del 95 %: 402–812). Para la medición de la LC, el modelo basado en CNN obtuvo un error relativo (RE) de 0.19 (IC del 95 %: 0.14, 0.25) en la fase de prueba. Si bien las estimaciones fueron cercanas a la longitud real, se observaron algunas discrepancias debido a errores de segmentación y a la variabilidad en la colocación de los calibradores. Este estudio representa un avance hacia la automatización de la medición de la LC, reduciendo la subjetividad.

---

[*] Tesis de pregrado

[**] Facultad de Ingenieria Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: William Andrés Cancino Rey. Ingeníero Electrónico. Codirector: Said David Pertuz Arroyo. PhD.

ABSTRACT

TITLE: ALGORITHM FOR THE AUTOMATIC MEASUREMENT OF CERVICAL LENGTH FROM TRANSVAGINAL ULTRASOUND IMAGES [*]

AUTHOR: María Fernanda Valenzuela Sánchez [**]

Keywords: Preterm birth, cervical length, transvaginal ultrasound, convolutional neural networks, Caliper detection, image segmentation.

Description: Preterm birth (PTB) is a major global public health concern, affecting approximately 15 million pregnancies annually. One of the most reliable indicators of PTB risk is cervical length (CL), which is manually measured in transvaginal ultrasound (TVUS) images. However, manual measurements are subject to inter-operator variability, limiting their reliability. To improve consistency and automation, this study proposes a deep learning-based approach for CL evaluation. A CNN-based model was developed to detect caliper positions and estimate CL. To achieve this, the mid-third of the cervix in first-trimester TVUS images was first segmented using a U-Net network. These segmentations were then used to train the caliper detection model, enabling the automated measurement of CL. The dataset included 311 pregnant women from two specialized maternal-fetal health centers. The segmentation model was evaluated using Dice coefficient, mean surface distance (MSD), and area error rate (AER), achieving a Dice score of 0.83 (95 % CI: 0.78–0.88), an MSD of 8.89 mm (95 % CI: 5.47–12.31), and an AER of 607 mm² (95 % CI: 402–812). For CL measurement, the CNN-based model obtained a relative error (RE) of 0.19 (95 % CI: 0.14, 0.25) in the test phase. While the estimates were close to the actual length, some discrepancies remained due to segmentation errors and variability in caliper placement. This study represents a step toward automating CL measurement, reducing subjectivity, and improving PTB risk detection in early pregnancy.

---

[*] Undergraduate Thesis

[**] Facultad de Ingenieria Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: William Andrés Cancino Rey. Ingeníero Electrónico. Codirector: Said David Pertuz Arroyo. PhD.

INTRODUCTION

Preterm birth (PTB) refers to the delivery of a baby before the normal gestational period is completed, which is generally between 37 and 42 weeks [1]. This phenomenon can occur for various reasons and can have serious consequences for the health of both the baby and the mother [2]. Preterm birth is a global public health problem that affects millions of families worldwide. According to the World Health Organization (WHO), approximately 15 million babies are born prematurely each year, representing around 11 % of total births worldwide [3]. In Colombia, Santander ranks as one of the departments with the highest rates of preterm births [4]. It is estimated that up to 70 % of preterm births occur without an apparent cause [5].

Premature babies are prone to a variety of short- and long-term health problems, such as respiratory diseases, underdevelopment of vital organs, jaundice, hypothermia, and

---

1 NATIONAL CANCER INSTITUTE. Definition of preterm birth - Cancer terms. In: Cancer.gov. Available at: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/preterm-birth.

2 BAÑOS, Núria; et al. Quantitative analysis of cervical texture by ultrasound in mid pregnancy and association with spontaneous preterm birth. In: Ultrasound In Obstetrics & Gynecology, vol. 51, no. 5, pp. 637–643, 2018. DOI: 10.1002/uog.17525.

3 WORLD HEALTH ORGANIZATION.. Preterm birth. In: World Health Organization. 2023. Available at: https://www.who.int/es/news-room/fact-sheets/detail/preterm-birth.

4 JAIMES, Gustavo. Nacimientos prematuros: una problemática de salud en Santander. In: FCV, 2022. Available at: https://www.fcv.org/co/prensa/noticias/nacimientos-prematuros-una-problematica-de-salud-en-santander.

5 AHUMADA, Juan; BARRERA, Ángela. Incidencia y factores de riesgo asociados a parto prematuro en una muestra de gestantes de Bogotá durante el periodo 2014-2017. Trabajo de grado. In: Universidad El Bosque, 2018. Available at: https://repositorio.unbosque.edu.co/server/api/core/bitstreams/67fe62de-e098-46d8-930e-789b62f80dca/content.

digestive system issues [6]. Additionally, preterm birth (PTB) is the leading cause of death in children under five years old, highlighting the importance of early detection. PTB can also lead to maternal health complications after delivery, including postpartum hemorrhage and uterine infections [7]. Identifying women at high risk of PTB early in pregnancy is essential to adopting intervention strategies and providing appropriate prenatal care. However, the heterogeneous etiology of PTB poses a challenge for the development of detection tools. The manual measurement of cervical length (CL) through transvaginal ultrasound imaging (see figure 1) is routinely used in clinical practice to estimate the risk of preterm birth [8] [9]. A CL of less than 25 mm is usually interpreted as an indicator of a higher probability of preterm birth [10].

Manual measurement of cervical length (CL) is a subjective method that heavily relies on the visual interpretation of images. This subjectivity introduces inter- and intra-operator variability in measurements (see figure 2), which can lead to erroneous clinical decisions. Additionally, the lack of standardization makes comparison and reproducibility of results difficult. This has driven the development of computerized approaches aimed at reducing

---

[6] AMERICAN PREGNANCY ASSOCIATION. Premature birth complications. In: American Pregnancy Association, 2021. Available at: https://americanpregnancy.org/es/healthy-pregnancy/labor-and-birth/premature-birth-complications.

[7] AMERICAN PREGNANCY ASSOCIATION. Premature birth complications. In: American Pregnancy Association, 2021. Available at: https://americanpregnancy.org/es/healthy-pregnancy/labor-and-birth/premature-birth-complications.

[8] BAÑOS, Núria; et al. Quantitative analysis of cervical texture by ultrasound in mid pregnancy and association with spontaneous preterm birth. In: Ultrasound In Obstetrics & Gynecology, vol. 51, no. 5, pp. 637–643, 2018. DOI: 10.1002/uog.17525.

[9] 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. In: Scientific Reports, vol. 11, no. 1, 2021. DOI: 10.1038/s41598-021-86906-8.

[10] LEE, Keun; AHN, Kyo Hwa. Application of artificial intelligence in early diagnosis of spontaneous preterm labor and birth. In: Diagnostics, vol. 10, no. 9, 2020. DOI: 10.3390/diagnostics10090733.

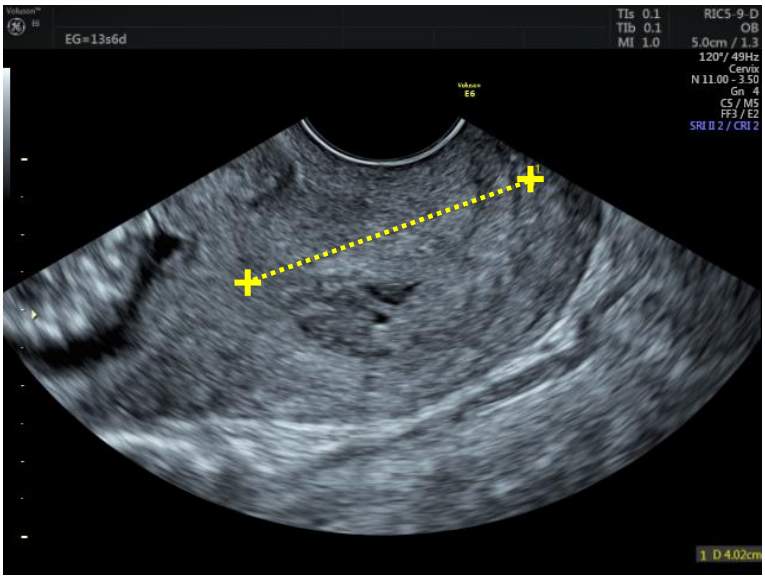


Figure 1. Cervical length measurement. Manual measurement performed by a specialist using transvaginal ultrasound.

variability and providing support to clinical specialists. Given the existing evidence on cervical changes during pregnancy [11] [12], several studies have proposed methods to quantify these changes through automated analysis of transvaginal ultrasound (TVUS) images. However, these images often include irrelevant regions that can hinder the accurate assessment of cervical morphology and CL measurement. Therefore, it is essential to define a region of interest (ROI) that contains only the anatomically relevant portions of the cervix. This not only contributes to the standardization of the analysis but also helps reduce variability in preterm birth risk assessment.

---

[11] HOUSE, M.; SOCRATE, S.; SURESH, S.; KAPLAN, D. Mechanical properties of cervical tissue: insights from animal and human studies. In: *Experimental Mechanics*, vol. 49, pp. 105–115, 2009. DOI: 10.1007/s11340-008-9134-0.

[12] MYERS, K. M.; SOCRATE, S.; TZERANIS, D.; HOUSE, M. Changes in the biochemical constituents and morphologic appearance of the human cervical stroma during pregnancy. In: *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 144, Supplement 1, pp. S82–S89, 2009. DOI: 10.1016/j.ejogrb.2009.02.012.

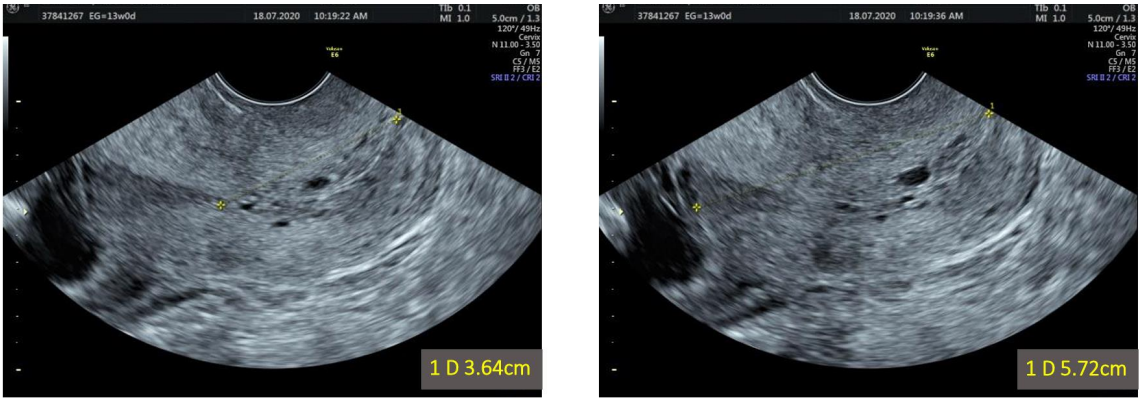


Figure 2. Variation in cervical length measurement. Differences in measurements taken in the same patient.

Various segmentation approaches have been developed for cervical analysis in TVUS images. Some studies have implemented neural networks for automatic CL measurement, improving accuracy and reducing variability compared to manual assessment [13]. Other works have focused on segmenting regions associated with cervical consistency, extracting quantitative information about tissue texture to complement CL measurement [14]. Recently, our group presented a deep learning-based method for cervical structure segmentation in TVUS images [15]. In these approaches, image annotation (ground truth) has been manually performed by obstetrics specialists. In this work, we propose a method for the automatic measurement of CL from transvaginal ultrasound images and their gene-

---

[13] AHUMADA GÓMEZ, J. S.; BARRERA REGALADO, Á. M.; CANOSSA PAREDES, D. Y.; CÁRDENAS MORÓ, L. C. Incidencia y factores de riesgo asociados a parto prematuro en una muestra de gestantes de Bogotá durante el periodo 2014-2017. Tesis de pregrado, Universidad El Bosque, Facultad de Medicina, Bogotá, 2018. Dirigida por: ROMERO, X.; URIEL, M.; IBAÑEZ, E.

[14] NATIONAL CANCER INSTITUTE. Definition of preterm birth - Cancer terms. In: Cancer.gov. Available at: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/preterm-birth.

[15] CANCINO, W.; BECERRA-MOJICA, C. H.; PERTUZ, S. Radiomic analysis of transvaginal ultrasound cervical images for prediction of preterm birth. In: *Lecture Notes in Computer Science*, vol. 14860, pp. 414–424. Springer, Cham, 2024. DOI: 10.1007/978-3-031-66958-3_30.

rated segmentations. To achieve this, we first develop a fully automated process to detect a region of interest containing only the anatomically relevant portions of the cervix in TVUS images obtained during the first trimester of pregnancy. Unlike previous studies that rely on manual annotation by specialists, our approach adopts a fully automated image segmentation process under the supervision of an obstetrics specialist. Moreover, this approach not only segments the cervix but also enables automatic CL measurement, contributing to standardization and reducing variability in its evaluation.

1. OBJECTIVES

1.1. GENERAL OBJECTIVE

To develop an algorithm for the automatic measurement of cervical length from transvaginal ultrasound images.

1.2. SPECIFIC OBJECTIVES

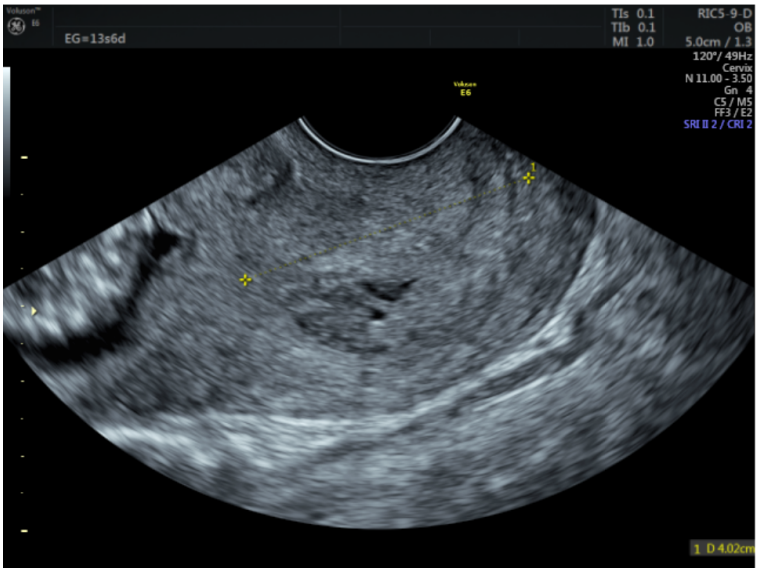
1.2.1. To curate a database of transvaginal ultrasound images in order to remove redundant, incomplete, or low-quality data that could affect the algorithm's performance. To develop and implement an ultrasound image processing algorithm to automatically measure cervical length. To evaluate the algorithm using standard performance metrics.

2. MATERIALS AND METHODS

2.1. IMAGING DATA

This was a retrospective cohort study that included transvaginal ultrasound (TVUS) images from 249 women, collected from two specialized medical centers in Bucaramanga, Colombia: the Centro de Atención Materno-Fetal INUTERO and the Hospital Universitario de Santander. The age of participants ranged from 16 to 43 years, and all images were acquired by gynecologists with experience in performing ultrasounds. From the total number of participants, 298 images were obtained where the cervical length was measured.

Figure 3. Example of Transvaginal Ultrasound Image from the Dataset.

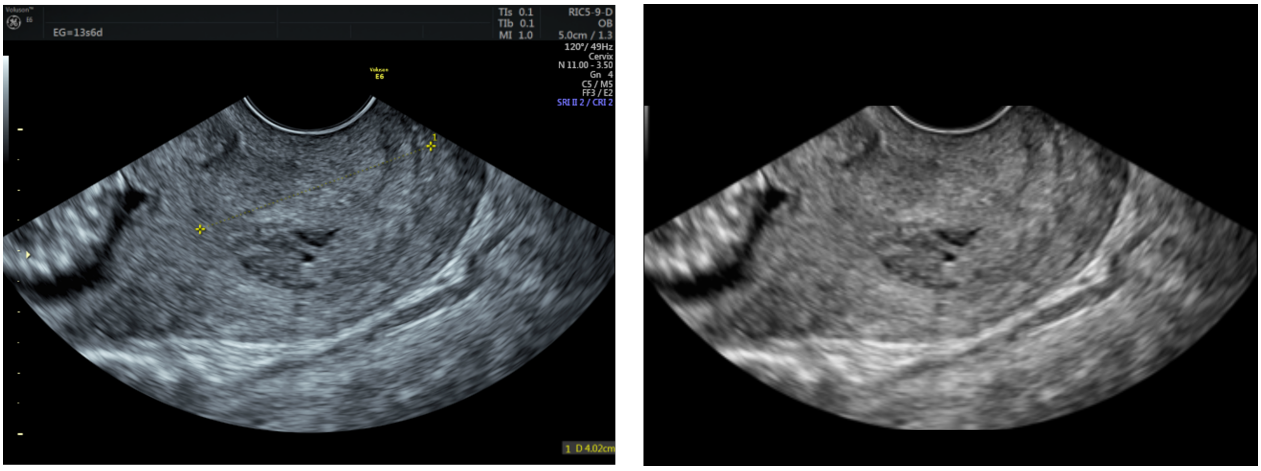


2.2. IMAGE PREPROCCESING

Ultrasound image preprocessing is a key step in improving quality and optimizing the segmentation of relevant structures. First, a portion of the image header is removed to focus

the analysis on the region of interest. Then, the image is converted to grayscale and binarized to identify the largest connected region, which is assumed to be the field of view (FOV). Morphological opening and closing operations are applied to improve the delineation of this region. Subsequently, artifacts present in the image are extracted based on the difference between color channels. To remove the detected artifacts, morphological filters are applied. Additionally, the inpainting method based on Telea's algorithm is used, which fills regions affected by noise or missing information through the interpolation of neighboring pixels [16]. Finally, the processed image is returned for further analysis and segmentation (see figure 4).

Figure 4. Image preprocessing for artifact removal and standardization of transvaginal ultrasound images. Left: original image with artifacts. Right: preprocessed image.



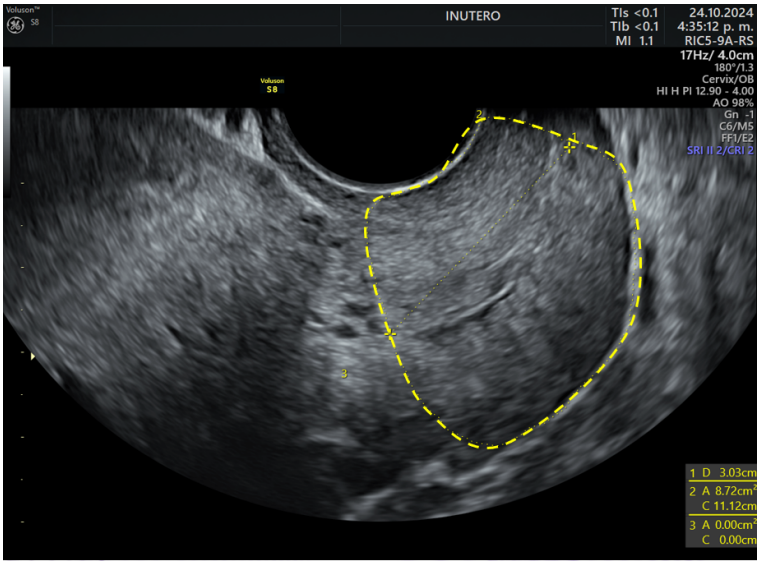
2.3. IMAGE ANOTATION

The annotation process was based on a set of reference images provided by an expert gynecologist. These images were manually delineated to define the anatomical regions

[16] TELEA, A. An image inpainting technique based on the fast marching method. In: *Journal of Graphics Tools*, vol. 9, no. 1, pp. 23–34, 2004. DOI: 10.1080/10867651.2004.10487596.

of interest, establishing a standardized criterion for segmentation. The segmentation was performed on the 224 images of the dataset, which were reviewed and approved by the gynecologist. Specifically, the annotation focused on the middle third of the cervix, as this region contains the most relevant features for clinical analysis and automatic segmentation (see Figure 5). The use of these references allowed for a precise definition of structural boundaries and served as an objective guide for evaluating the quality of the segmentations generated by the model.

Figure 5. Manual selection of the mid-cervical third for segmentation.



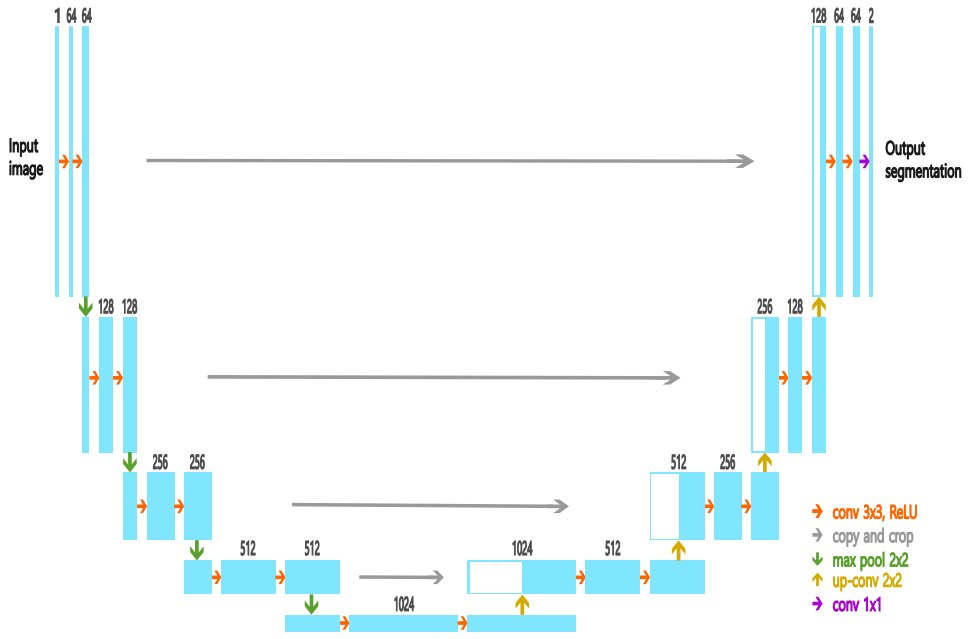
2.4. SEGMENTATION MODEL

The U-Net architecture used in this work was based on the model proposed by T. Wlodarczyk et al. [17], maintaining the characteristic encoder-decoder structure of this type of segmentation network. The fundamental principles of the original design were preserved

[17] WLÓDARCZYK, Tomasz et al. Spontaneous preterm birth prediction using convolutional neural networks. In: arXiv preprint, 2020. Available at: https://arxiv.org/abs/2008.07000.

while adapting the number of filters and input image dimensions to the specific requirements of the problem addressed (see Figure 6). The model operates with images of size 176 × 320 pixels, ensuring a uniform input across all network layers. The encoder phase consists of four convolutional blocks, each composed of two 3×3 convolutional layers with ReLU activation and an increasing number of filters (64, 128, 256, and 512), followed by a 2×2 max pooling operation that reduces spatial resolution. At the network's deepest level (bottleneck), two convolutional layers with 1024 filters are included, allowing the extraction of high-level abstract representations before the decoder phase. The decoder progressively restores the spatial structure of the image using 2×2 transposed convolutions. At each decoder level, feature maps from the encoder are concatenated, preserving detailed edge information and segmented structures. The network output is a 1×1 convolution with a sigmoid activation function, producing a binary segmentation map. For training, the Adam optimizer was used with a learning rate of 1e-4 and binary cross-entropy as the loss function. A batch size of 8 was used, along with a data generator that normalizes images and segmentation masks in real-time. Additionally, an early stopping mechanism was implemented, monitoring validation loss with a patience of 10 epochs, preventing overfitting and ensuring that the best-performing weights were selected during training.

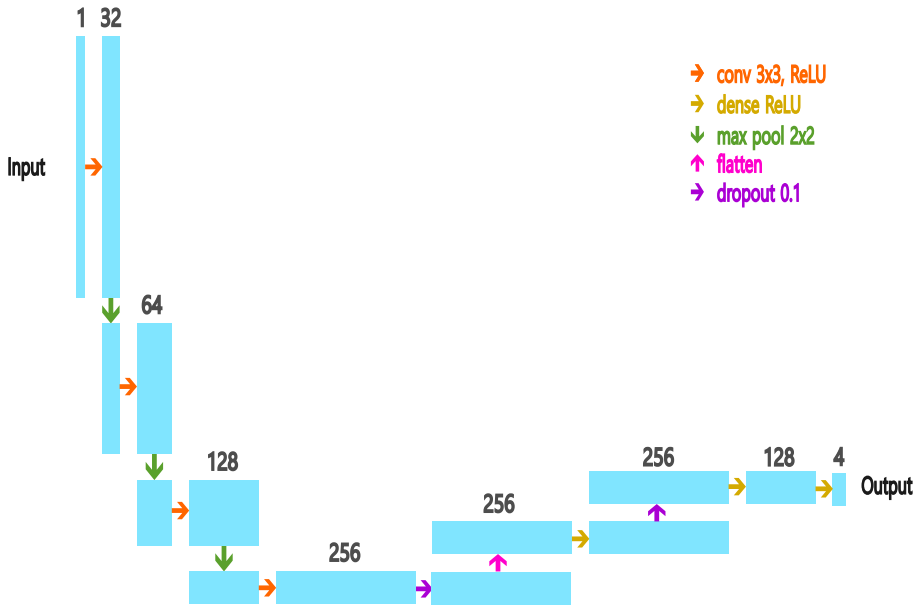
Figure 6. Architecture of the Implemented U-Net Network.



2.5. CALIPER DETECTION MODEL

The convolutional neural network (CNN) was designed to automatically detect the position of calipers in transvaginal ultrasound (TVUS) images. This model takes as input both the original ultrasound image and its corresponding segmentation mask, ensuring that predictions focus on anatomically relevant regions (see Figure 7). The model architecture consists of a series of convolutional layers that progressively extract spatial and texture features from the input images. The network operates with images of size 176 × 320 pixels, maintaining a uniform resolution across all layers.

Figure 7. CNN Architecture for Caliper Detection.



In the feature extraction phase, the CNN uses four convolutional blocks, each consisting of a 3×3 convolutional layer with ReLU activation, followed by a max pooling operation. These layers reduce spatial dimensions while preserving the most relevant features. As information flows through the network, the number of filters progressively increases (32, 64, 128, and 256), facilitating the capture of increasingly complex patterns in the image. After feature extraction, the flattened network output passes through dense layers to process the learned information. A fully connected layer with 256 neurons is included, followed by a dropout layer (0.1) to improve model generalization and prevent overfitting. Subsequently, a dense layer with 128 neurons further refines the representations before the output layer. The model's final layer consists of four neurons with linear activation, corresponding to the predicted coordinates of the calipers in the image. These coordinates allow precise localization of the cervical length measurement points. For model training, the Adam optimizer was used with a learning rate of 5e-4 and mean squared error (MSE)

as the loss function, as it is suitable for coordinate regression tasks. The model was trained with a batch size of 16 images, applying early stopping with a patience of 10 epochs to prevent overfitting.

2.6. EVALUATION METRICS

To evaluate the performance of the segmentation model, three key metrics were implemented: the *Dice coefficient*, the *mean surface distance* (MSD), and the *area error rate* (AER).For the *cervical length* model, the *Mean Relative Error of Distance* (Distance MRE) was used. Additionally, to evaluate the accuracy of *caliper* placement, the *Mean Relative Error* (MRE) was employed.

The *Dice coefficient* measures the overlap between the predicted segmentation (S_p) and the reference segmentation (S_r). It is defined as twice the intersection area divided by the sum of both segmented areas, where a value close to 1 indicates greater similarity. A higher Dice coefficient suggests better segmentation performance, as it signifies a higher proportion of correctly segmented pixels relative to the total segmented regions [18], where $S_p \cap S_r$ represents the intersection of the predicted and reference segmentations, and $|S_p|$ and $|S_r|$ denote the number of pixels in each segmentation. The equation for the Dice coefficient is:

$$\text{Dice} = \frac{2|S_p \cap S_r|}{|S_p| + |S_r|} \tag{1}$$

---

[18] RUSSELL, Richard G. F.; et al. Evaluation of medical image segmentation performance: A review. In: Journal of Medical Imaging, vol. 6, no. 4, 2019, p. 043001. Available at: https://doi.org/10.1117/1.JMI.6.4.043001.

The *mean surface distance* (MSD) quantifies the average difference between the contours of the predicted and reference segmentation by calculating the shortest Euclidean distance from each contour point of one segmentation to the other. Lower values indicate more precise segmentation, as they reflect smaller deviations between the predicted and reference boundaries [19], where ∂S_p and ∂S_r represent the contours of the predicted and reference segmentations, respectively, and the function $d(x, \partial S_r)$ computes the shortest Euclidean distance from a point x in the predicted contour to the reference contour, and vice versa for $d(y, \partial S_p)$. The MSD is calculated as:

$$\text{MSD}(S_p, S_r) = \frac{1}{|\partial S_p| + |\partial S_r|} \sum_{x \in \partial S_p} d(x, \partial S_r) + \sum_{y \in \partial S_r} d(y, \partial S_p) \tag{2}$$

The *area error rate* (AER) quantifies the discrepancy between the predicted segmentation (Sp) and the reference segmentation (Sr) by measuring the total non-overlapping area. It is calculated as the sum of the area of the predicted segmentation that does not intersect with the reference and the area of the reference segmentation that does not intersect with the prediction. A lower AER value indicates higher segmentation accuracy, as it reflects a closer match between the predicted and reference segmentations in terms of total area [20]. The AER is defined as:

---

[19] C.C.C.L.; et al. Evaluation of segmentation algorithms for medical images using the mean surface distance. In: *Medical Image Analysis*, vol. 7, no. 3, pp. 229–240, 2003. DOI: 10.1016/S1361-8415(03)00008-4.

[20] TAHA, A. A.; HANBURY, A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. In: *IEEE Transactions on Medical Imaging*, vol. 35, no. 3, pp. 796–806, 2015. DOI: 10.1109/TMI.2015.2481439.

$$\text{AER} = |S_p \cap \neg S_r| + |\neg S_p \cap S_r| \quad \text{mm}^2 \tag{3}$$

The *Mean Relative Error of Distance* (Distance MRE) evaluates the relative discrepancy between the predicted and actual cervical length by normalizing the error with respect to the actual length. This metric allows for comparing the model's accuracy across different cervical length scales, making it particularly useful when lengths vary between samples [21]. A lower MRE indicates that the model maintains a low error proportion relative to the actual length, ensuring more reliable measurements under different conditions. The equation for distance MRE is:

$$\text{MRE}_{\text{distance}} = \frac{1}{n} \sum_{i=1}^{n} \frac{|d_{p,i} - d_{r,i}|}{d_{r,i} + \epsilon} \tag{4}$$

In addition to evaluating cervical length, it is crucial to analyze the accuracy in caliper localization, which determines the endpoints of the measurement, where c_p and c_r represent the predicted and actual caliper coordinates, respectively. To this end, the *Mean Relative Error* (MRE) of the calipers are used to measure the difference between the predicted and actual coordinates. The equation is:

$$\text{MRE}_{\text{calipers}} = \frac{1}{n} \sum_{i=1}^{n} \frac{|c_{p,i} - c_{r,i}|}{c_{r,i} + \epsilon} \tag{5}$$

---

[21] WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. In: *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005. DOI: 10.3354/cr030079.

To validate the models, a hold-out validation approach was used, splitting the dataset into 70 % for training (156 images), 15 % for validation (34 images), and 15 % for testing (34 images). Additionally, 95 % confidence intervals were calculated for each metric to assess the variability and robustness of the results. The confidence intervals were estimated using the bootstrap resampling method, which involves repeatedly sampling with replacement from the test set and computing the desired metric across multiple resamples. These metrics provide a comprehensive evaluation of the models, considering spatial overlap and contour accuracy in segmentation, as well as the accuracy of caliper detection and the consistency in cervical length measurement.

3. RESULTS AND DISCUSSION

Table 1 shows the results obtained for the evaluation metrics in the segmentation task, including the Dice coefficient, Mean Surface Distance, and Area Error Rate. The average values are reported along with their respective 95 % confidence intervals. The results exhibit a decrease in segmentation performance from the training to the validation and test sets, as reflected in the Dice coefficient values of 0.87, 0.84, and 0.83, respectively. This decline suggests that while the model generalizes well, there is still a performance gap when applied to unseed data. The Mean Surface Distance and Area Error Rate exhibit a similar trend, with higher values in the validation and test sets compared to training, indicating that segmentation performance is reduced in more complex or previously unseen cases.

Table 1. Performance metrics for the segmentation model.

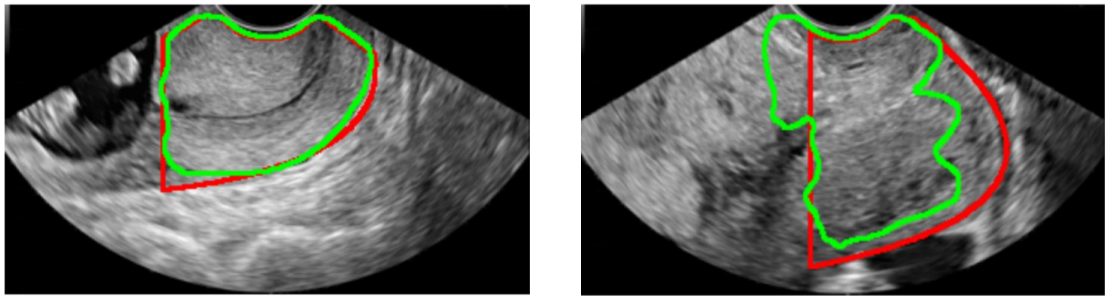
Metric	**Train**	**Validation**	**Test**
Dice (95 % CI)	0.88 (0.86, 0.89)	0.84 (0.79, 0.88)	0.84 (0.78, 0.88)
MSD (mm) (95 % CI)	6.64 (6.09, 7.19)	8.17 (6.54, 9.80)	8.89 (5.47, 12.31)
AER (mm^2) (95 % CI)	473 (439, 507)	588 (496, 681)	607 (402, 812)

A visual inspection of the results further illustrates these observations. Figure 8 presents two examples within the same figure: on the right, a case where the segmentation model successfully delineates the structure of interest, showing strong agreement between the predicted contours (green) and the manual annotations (orange); on the left, a case where the model exhibits segmentation errors. These errors manifest as two common issues: under-segmentation, where the predicted region fails to fully encompass the expected structure, and over-segmentation, where the model extends the predicted region beyond the actual boundaries. These discrepancies may arise due to variations in image

acquisition, differences in transducer orientation, or anatomical variability that the model has not adequately learned.

These challenges indicate that, while the model performs well in most cases, there is room for improvement in segmentation accuracy. Some of these issues could be mitigated through post-processing techniques, such as morphological operations or region-growing algorithms, to refine the final segmentation. Additionally, exploring alternative neural network architectures, such as transformer-based models or hybrid approaches combining U-Net with attention mechanisms, may enhance the model's robustness in more complex cases. Another potential improvement is the application of domain adaptation techniques to enhance generalization across different datasets or imaging conditions.

Figure 8. Comparison of Automatic and Manual Segmentations. Red: Ground truth segmentation, Green: Predicted segmentation. Right: Correct segmentation, Left: Incorrect segmentation.



The performance metrics for the cervical length estimation model and caliper detection were analyzed. Table 2 summarizes the obtained results, reporting the *distance mean relative error* (*distance MRE*) and the *mean relative error* (*MRE*), along with their respective 95 % confidence intervals. The results show stable model performance across different datasets. Specifically, the *distance MRE* in the test set is 0.155, with a confidence interval of

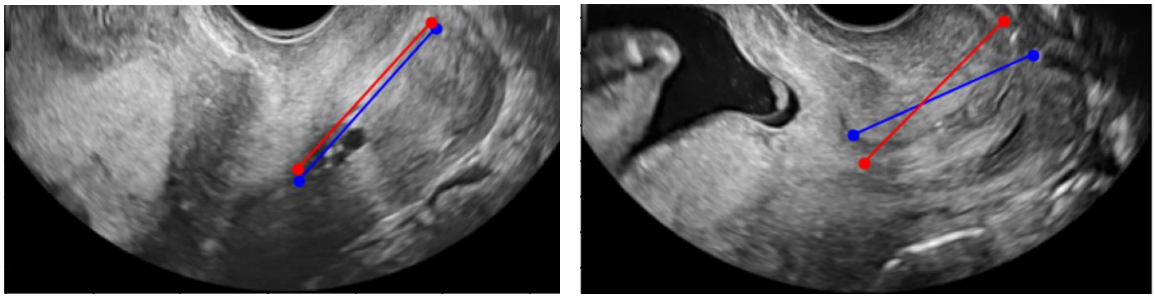
[0.068, 0.310], indicating a low relative error proportion in cervical length estimation. Similarly, the *MRE* in caliper detection presents consistent values, with a test value of 0.19 and a confidence interval of [0.14, 0.25]. These results suggest that the model maintains adequate accuracy in both cervical length measurement and caliper localization.

Table 2. Performance metrics for the cervical length estimation and caliper detection model.

Metric	**Train**	**Validation**	**Test**
Distance MRE (95 % CI)	0.13 (0.08, 0.22)	0.15 (0.06, 0.31)	0.15 (0.06, 0.31)
Caliper MRE (95 % CI)	0.16 (0.12, 0.21)	0.19 (0.14, 0.25)	0.19 (0.14, 0.25)

A visual inspection of the results confirms these findings. Figure 9 presents examples of cervical length measurements, where the right side shows a correctly performed measurement, while the left side illustrates an error in caliper placement, affecting measurement accuracy. These results indicate that, although the model achieves good accuracy in most cases, its performance could still benefit from improvements in handling anatomical variability and image quality. Strategies such as incorporating refinement techniques based on spatial attention or using hybrid neural networks could enhance the model's robustness in more challenging imaging conditions.

Figure 9. Model measurement results for CL. Blue: Ground truth caliper detection and CL measurement, Red: Predicted caliper detection and CL measurement. Right: Measurement performed correctly, Left: Measurement error.



Beyond its quantitative performance, the proposed segmentation model and the cervical length estimation model have significant practical implications. The ability to automatically segment anatomical structures with high accuracy and reliably estimate cervical length could facilitate clinical workflows, reducing the manual effort required for annotation and improving the reproducibility of measurements. This advancement is particularly relevant for applications requiring large-scale image analysis, such as population studies or longitudinal patient monitoring. However, to ensure seamless integration into real-world applications, further validation on diverse datasets and real-time optimization strategies should be explored.

4. CONCLUSIONS

A deep learning-based approach was developed for caliper detection and cervical length estimation from transvaginal ultrasound images. The models were implemented in Python using TensorFlow, and their performance was evaluated using multiple metrics.

The U-Net model implemented segmented the middle third of the cervix in transvaginal ultrasound images from the first trimester of pregnancy, achieving a Dice coefficient of 0.83 in the test phase. This result suggests that the model consistently identifies the region of interest; however, some segmentations showed discrepancies compared to the expected anatomy. During evaluation, certain images displayed segmentations that did not correspond to anatomical structures, indicating possible limitations in the model's generalization ability when faced with variations in image quality, contrast, and the presence of artifacts. For cervical length measurement, a model based on convolutional neural networks (CNNs) was trained, obtaining a Mean Relative Error (MRE) of 0.19 in the test phase. This result indicates that while the model provides estimates close to the actual length, differences still exist compared to the manual measurements obtained using calipers. These discrepancies may be due to variability in caliper placement, as well as potential inaccuracies in cervix segmentation that affect the determination of the measurement endpoints.

The results show that the proposed segmentation model effectively captures anatomical structures with high accuracy; however, certain limitations persist in challenging cases. These limitations could be addressed through improved architectures, post-processing techniques, or additional data augmentation strategies to enhance its performance. Similarly, the accuracy of cervical length measurement could be optimized by refining caliper detection using more robust deep learning approaches or applying adjustment techniques based on anatomical features.

Finally, it is essential to continue research and refine the methods used to improve the model's accuracy and reduce the occurrence of erroneous segmentations and inaccurate measurements. Future work should focus on enhancing the model's generalization ability and exploring its integration into fully automated workflows for clinical or research applications.

BIBLIOGRAPHY

Ahumada, Juan et al.: “Incidencia y factores de riesgo asociados a parto prematuro en una muestra de gestantes de Bogotá durante el periodo 2014-2017”. Supervised by: X. Romero, M. Uriel, E. Ibañez. Tesis de mtría. Universidad El Bosque, 2018.

American Pregnancy Association: *Premature birth complications*. https://americanpregnancy.org/es/healthy-pregnancy/labor-and-birth/premature-birth-complications/. Accessed: 2025-04-02. 2025.

Baños, Núria 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. DOI: 10.1002/uog.17525.

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: *Scientific Reports* 11.1 (2021). DOI: 10.1038/s41598-021-86906-8.

C.C.C.L. et al.: *Evaluation of segmentation algorithms for medical images using the mean surface distance*. En: *Medical Image Analysis* 7.3 (2003), págs. 229-240. DOI: 10.1016/S1361-8415(03)00008-4.

Cahill, Laura S. et al.: *Determination of fetal heart rate short-term variation from umbilical artery Doppler waveforms*. En: *Ultrasound in Obstetrics & Gynecology* 57.4 (2021), págs. 631-639. DOI: 10.1002/uog.23145.

Cancino, Wilson, Cristian H. Becerra-Mojica y Santiago Pertuz: “Radiomic Analysis of Transvaginal Ultrasound Cervical Images for Prediction of Preterm Birth”. En: *Lecture Notes in Computer Science*. Vol. 14860. Springer, Cham, 2024, págs. 414-424. DOI: 10.1007/978-3-031-66958-3_30.

Eunice Kennedy Shriver National Institute of Child Health and Human Development: *Who is at risk for preterm labor and birth?* https://www.nichd.nih.gov/health/topics/preterm/conditioninfo/who_risk. Accessed: 2023-10-25. 2023.

Fischer, Anna M. et al.: *End-to-end learning with interpretation on electrohysterography data to predict preterm birth*. En: *Comput. Biol. Med.* 158 (2023). DOI: 10.1016/j.compbiomed.2023.106846.

Fundación Cardiovascular de Colombia: *Nacimientos prematuros, una problemática de salud en Santander*. https://www.fcv.org/co/en/prensa/noticias/nacimientos-prematuros-una-problematica-de-salud-en-santander. Accessed: 2025-04-02. 2025.

Gao, Cheng et al.: *Deep learning predicts extreme preterm birth from electronic health records*. En: *J. Biomed. Inform.* 100 (2019). DOI: 10.1016/j.jbi.2019.103334.

House, Michael et al.: *Mechanical properties of cervical tissue: insights from animal and human studies*. En: *Experimental Mechanics* 49 (2009), págs. 105-115. DOI: 10.1007/s11340-008-9134-0.

Kalengo, N. H. 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 (2020), e0239037. DOI: 10.1371/journal.pone.0239037.

Lee, Kyung-Ah y Ki Hoon Ahn: *Application of artificial intelligence in early diagnosis of spontaneous preterm labor and birth*. En: *Diagnostics* 10.9 (2020). DOI: 10.3390/diagnostics100907

Lozano-Mosquera, Sergio J. et al.: *Cervicometría menor o igual a 25 mm para identificar parto a los siete días en pacientes con amenaza de parto prematuro*. En: *Revista Colombiana De Obstetricia Y Ginecología* 65.2 (2014), pág. 112. DOI: 10.18597/rcog.59.

Myers, Kristin M. et al.: *Changes in the biochemical constituents and morphologic appearance of the human cervical stroma during pregnancy*. En: *European Journal of Obstetrics & Gynecology and Reproductive Biology* 144 (2009), S82-S89. DOI: 10.1016/j.ejogrb.2009.02.012.

National Cancer Institute: *Preterm birth*. https://www.cancer.gov/publications-/dictionaries/cancer-terms/def/preterm-birth. Accessed: 2025-04-02. 2025.

Russell, R.G.F. et al.: *Evaluation of Medical Image Segmentation Performance: A Review*. En: *Journal of Medical Imaging* 6.4 (2019), pág. 043001. DOI: 10.1117/1.JMI.6.4.043001.

Taha, Ahmed A. y Allan Hanbury: *Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool*. En: *IEEE Transactions on Medical Imaging* 35.3 (2015), págs. 796-806. DOI: 10.1109/TMI.2015.2481439.

Telea, Alexandru: *An image inpainting technique based on the fast marching method*. En: *Journal of Graphics Tools* 9.1 (2004), págs. 23-34. DOI: 10.1080/10867651.2004.10487596.

Willmott, C. J. y K. Matsuura: *Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance*. En: *Climate Research* 30.1 (2005), págs. 79-82. DOI: 10.3354/cr030079.

Włodarczyk, Tomasz et al.: *Spontaneous preterm birth prediction using convolutional neural networks*. En: *arXiv preprint* (2020). DOI: 10.48550/arXiv.2008.07000. eprint: arXiv:2008.07000.

World Health Organization: *Preterm birth*. https://www.who.int/news-room/fact-sheets/detail/preterm-birth. Accessed: 2025-04-02. 2025.

Zhang, Yuan et al.: *Artificial Intelligence-Aided Thermal Model Considering Cross-Coupling Effects*. En: *IEEE Transactions on Power Electronics* 35.10 (2020), págs. 9998-10002. DOI: 10.1109/TPEL.2020.2980240.

*

APPENDICES

Anexo A. GitHub repository

The source codes and evaluation scripts are publicly available at `https://github.com/Mafe-Valenzuela/Algorithm_for_the_automatic_measurement_of_cervical_length_from_transvaginal_ultrasound_images.git` ensuring reproducibility and facilitating future research in this area.