Automatic calculation of pelvis morphology from CT images

P. Guijarro Martínez[1], D. Sánchez García[2], L. Cubero [1], J. De Leon-Luis[3,4], J. Pascau[1,5],

[1] Departamento de Bioingeniería, Universidad Carlos III de Madrid, Madrid, Spain pguijarr@pa.uc3m.es , jpascau@ing.uc3m.es
[2] Departamento de Radiología, Hospital Universitario Gregorio Marañón, Madrid, Spain
[3] Servicio de Obstetricia y Ginecología, Hospital Gregorio Marañón, Madrid, Spain
[4] Departamento de Salud Pública y Materno-Infantil, Universidad Complutense de Madrid, Madrid, Spain
[5] Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain

Abstract

Pelvimetry is the study of the pelvis morphology in women for labor planning and medical assessment. This can be achieved by manually annotating pelvic CT images for extracting several measures of interest, which can be both time-consuming and subjective. While machine learning has achieved significant success in 2D landmarking applications, results in pelvic CT images are still limited, particularly with small datasets. This paper presents a two-step approach for detecting 3D landmarks in pelvic CT images. First, a simple CNN coarsely estimates landmark locations, serving as a starting point for further refinement. Then, higher resolution 3D patches and independent neural networks are used to obtain the final position for each landmark. Our model has shown promising results, obtaining an average distance error of 6.71 mm across 7 landmarks. These values allowed the calculation of the morphological measurements, demonstrating a strong correlation with the manual values. The proposed model has shown promising results, offering efficient and accurate predictions of the anatomical landmarks in CT examinations.

1. Introduction

As described in the literature [1, 2], pelvimetry is a common procedure for pelvic analysis in women. It allows assessing the shape and dimensions of the pelvic bone, enabling clinical evaluation and labor planning. Understanding pelvic anatomy and morphology requires calculating several morphometric measurements such as the subpubic angle, the anatomical transverse diameter and the obstetric conjugate diameter. These measurements provide valuable information about the overall size of the pelvis and whether they may limit natural birth delivery. According to size and shape, the female pelvis can be classified into four distinct types: anthropoid, gynecoid, android and platypelloid [3]. The above-mentioned measures could also be related to the Symphysis–Ischial Spine Distance (SID), which has been shown to depend on the sex and height of each patient [4]. The SID is of great importance for determining the type of pelvis in women, as well as its size and shape. To replicate these results and understand the relationship between SID and labor complications, radiologists at Hospital Gregorio Marañón are manually annotating a large CT image dataset, identifying several anatomical landmarks. Automating this process would facilitate their research and provide robust and replicable measurements.

In comparison to the clinical approach [5], Computed Tomography (CT) pelvimetry has emerged as a key improvement, providing precise and reliable measurements. Analysing high-resolution 3D images offers radiologists a better understandment of the overall structure of the pelvis. However, this process requires manually annotating each CT scan, which can be time-consuming. Furthermore, it demands the knowledge of skilled specialists, heavily relying on their criteria and expertise.

Hereby, Machine Learning (ML) algorithms have gained importance in this area, providing accurate positioning of anatomical landmarks and therefore relieving the workload of the professionals. These studies have revolutionized multiple fields of image analysis, achieving great success in tasks such as facial landmark detection in 2D images [6] and anatomical landmark detection in medical datasets [7]. These advances have noticeably increased the efficiency and accuracy of landmark identification. However, when it comes to identifying pelvic landmarks in CT scans, the number of published studies is limited. This can be attributed to the high computational cost associated with 3D images and the unavailability of annotated medical datasets. However, Payer et al [6] predicted landmarks in T1-weighted 3D gradient echo hand MR scans using a U-Net where each landmark was represented by a Gaussian. The results on 85 training sets resulted in an average error of 1.18 mm. Bier et al. [8] detected multiple pelvic X-ray landmarks independlently of their viewing direction by utilizing a CNN, which achieved an average error of 5.6 mm. Luo et al. [7] used decision forests followed by morphological image processing techniques on thoracoabdominopelvic CT scans, achieving an average distance error of 13.7mm.

In this work, we have developed a model for the automatic calculation of pelvis morphology based on the prediction of 3D anatomical landmarks. The detected landmarks allowed calculating morphometric measures of interest, offering objective and quantitative data about the overall dimensions of the pelvis.

2. Materials and Methods

2.1 Subjects

The selected dataset for this study was provided by Hospital General Universitario Gregorio Marañón (HGUGM), containing 87 pelvic CT images in DICOM format from women aged from 20 to 44 years. These images were collected from women who undergone a medical examination previous to birth delivery. All images were carefully selected and annotated by three expert

ISBN: 978-84-17853-76-1

radiologists, who then classified them into one of the four types of pelvis mentioned. The images were acquired on several CT devices: Ingenuity (Philips), Brilliance 64 (Philips) and Revomaxima (General Electric). All images were pseudo-anonymized to protect personal data.

Type of pelvis	Number of samples
Anthropoid	48
Gynecoid	31
Android	5
Platypelloid	3

Tabla 1. *Number of samples for each type of pelvis in women*

As shown in Table 1, a significant class imbalance was observed, which is certainly expected given the prevalence of specific pelvic types in women. There were a total of 7 annotated landmarks, corresponding to the subpubic angle (L0, L1, L2), the obstetric conjugate diameter (L3, L4) and the longitudinal diameter (L5, L6), as shown in Figure 1. For instance, the subpubic angle, is formed by the inferior pubic rami and the lower part of the pubic symphysis. The anatomical transverse diameter represents the largest distance of the superior aperture of the pelvis, also known as the pelvic inlet. Finally, the obstetric conjugate diameter, measures the distance from the sacral promontory to the point below the upper inner edge of the symphysis. The importance of these measures relies on understanding the overall size of the pelvis so as to determine its dimensions as they may limit natural birth delivery.

The annotations were initially provided as 2D snapshots on the 3D volume render of each CT. To obtain the corresponding 3D coordinate for each landmark, we manually annotated and saved their positions on 3D Slicer [9], a software platform used for medical image analysis and visualization. In some cases, the provided images did not show the full pelvic bone, as they were cropped at the pubic rami. For these cases, the landmarks of the subpubic angle were strategically positioned at key points due to the absence of bone in those images.

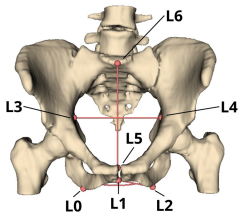

Figure 1. *Landmarks location annotated in 3D Slicer.*

2.2 Data pre-processing

Voxel size and field of view (FOV) differed among patients depending on the original diagnostic purpose of each scan. To standardize image size, manual cropping was performed in the z-axis using 3D Slicer. A reference point was established at the L5 vertebrae, beyond which all images were cropped for homogenization.

Then, the pelvic bone was segmented to eliminate unwanted structures and improve the overall performance of our model. First of all, a gaussian filter was applied with a standar deviation of 2 mm for reducing image noise. Then, we applied intensity-based segmentation with min and max thresholds of 150 and 1050 Hounsfield Units, corresponding to the pixel values where pelvic bone is observed [10]. The resulting mask was then processed with morphological dilation and erosion techniques and finally applied to the original image.

To address variations in image size and resolution, intensity-based registration was applied as part of the preprocessing. Resolution in the X and Y axis ranged from 0.52 to 0.97 mm, with mean 0.76 mm. For the Z axis resolution ranged from 0.5 to 5 mm, with mean 1.11 mm. After analyzing the sizes and resolutions of the dataset, four images were discarded as the overall resolution was significantly worse compared to the rest of the images, so they were removed from the dataset. The image with the largest size and highest resolution ratio was chosen as the target image for the registration process. This reference image had matrix size (582, 444, 280) pixels and voxel spacing (0.625, 0.625, 1.0) mm. All images were normalized in intensity and aligned to the target image with rigid registration with Mean Squared Error as cost function. This process aimed to align all images without introducing scaling or deformation, preserving the relative positions of the landmarks.

The large size of the resulting volumes did not allow training the models at full resolution in the available GPUs due to memory limitations, so all images were resized by a factor of (3, 3, 2) in x,y and z, respectively, increasing the voxel size and therefore reducing the overall resolution of the images. This resizing factor was also chosen to achieve similar voxel size in each dimension.

2.2 Training

Our proposed landmark prediction model involves a two-step approach based on Convolutional Neural Networks. All tests were run in a workstation equipped with AMD Ryzen 7 5800X 8-Core Processor and 64 GB RAM, along with one Quadro RTX 8000 GPU with 48 GB memory. For training our model the following libraries were installed: keras v2.8.0, numpy v1.22.3, pandas v2.0.0, SimpleITK v2.1.1.2, matplotlib v3.5.2 and TensorFlow v2.8.0.

Our proposed approach consists of two steps. First, a simple CNN, named `CNN_LR`, takes as input the resized images of size (194, 148, 105) pixels with (1.875, 1.875, 2) mm spacing. In this step, low-resolution images will be used for an initial prediction of the landmarks. The final step involves training a CNN for each landmark, named `CNN_HR`, using the same architecture on patches of size 100x100x40 pixels around each landmark obtained from the original high-resolution images. This second model focuses on the local volume surrounding each landmark and preserves the original resolution, but fits in the GPU memory due to the drecreased image size.

During training of the `CNN_HR`, the generated patches were centered around the ground truth labels of the orig-

ISBN: 978-84-17853-76-1

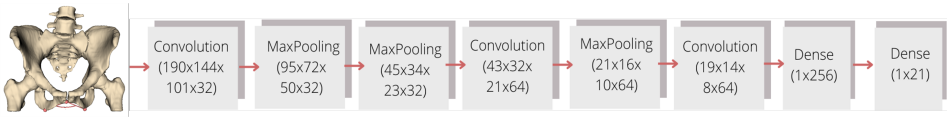

Figura 2. *Model architecture.*

inal image. To better simulate the inference, where the predicted landmark from the `CNN_LR` will have some error, we applied a random translation between -10 and 10 pixels in all axis. This translation led to variations in position within the patch, so the training landmarks were not always at the center of the patch, and the network could learn the spatial relationships and true positions of the landmarks. For the test subset, the generated patches were obtained from the predictions of the `CNN_LR` rather than the real annotations from the original images. To determine the corresponding position of these predictions in the original image, we applied a reverse scaling operation by multiplying the predicted positions by the scaling factors (3, 3, 2). For this subset, random translation was not necessary as the real predictions were not in the center of the generated patch.

In both phases of the pipeline, we used the same model architecture, as shown in Figure 2. This architecture consisted of conventional 3D convolutions and max pooling layers to capture relevant features in the input data. Additionally, we included several dropout layers to regularize the network and reduce overfitting. The loss function was the Mean Squared Error which penalized large errors due to the squaring of the differences between the predicted and true values. For evaluating the results obtained in each stage, the Euclidean distance between the prediction and the real coordinate was calculated for each landmark.

We divided the dataset into train, test and validation sets using a 60-20-20 relation, maintaining proportional class distribution across sets, ensuring representative sampling and enhancing model evaluation reliability. This supports robust CNN training and accurate performance assessment on unseen test data.

Furthermore, data augmentation was applied to deal with limited data and interpatient variability, increasing the training data by a factor of 10. Given the importance of preserving the structural and compositional characteristics of the original data, we implemented random scaling and rotation. First, images were randomly scaled with a probability of 0.5 and a scaling factor from 0.9 to 1.15, which was centered at the image midpoint. In addition, random rotation from -15 to 15 degrees was applied around X and Y axes. The rotation was also centered in the middle of the image. Data augmentation noticeably increased the training data by making minor changes in the orientation and spatial positioning of the pelvic structures.

For the evaluation of the models performance we used the correlation coefficient metric. This measures the strength of the linear relationship between the morphometric measures of interest and their corresponding predictions. It ranges from -1 to 1, where negative values correspond to inverse correlation and values near zero indicate that there is no linear relationship between values.

3. Results and discussion

In the first stage, the mean distance error obtained between the predicted points and the ground truth labels was 9.32 mm with a standar deviation of 4.18 mm and a median of 9.17 mm. The distance errors for each landmark are illustrated in Figure 3. For the final step, the average prediction error for the corresponding landmarks was significantly smaller, achieving a mean value of 6.71 mm, with a standar deviation of 3.28 mm. The median obtained in this stage was 6.58 mm. As observed, the second CNN improves the results decreasing the error for all landmarks. L0 and L2 obtained the largest errors, possibly due to their location within the pubic rami. For cropped images, these coordinates were no longer anatomical points but virtual ones, potentially hindering the model's ability to learn their accurate positioning.

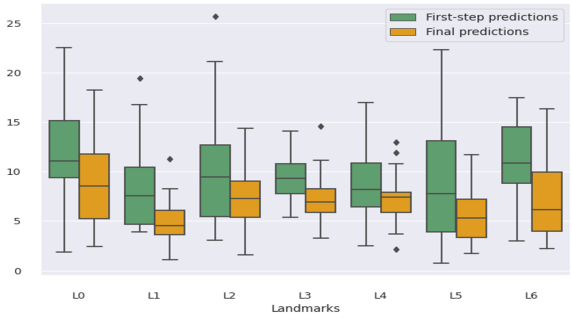

Figure 3. *Landmarks location annotated in 3D Slicer.*

The main goal of this study was the prediction of anatomical landmarks from pelvic CT images, but also calculate the corresponding measures, thus enabling pelvimetry studies on large datasets without manual labelling. To better analyze the similarity between the predicted and the ground truth measurements, scatter plots are provided to analyze the results achieved in each stage quantitatively. The scatter plots show that the ground truth labels and the predictions are more correlated after the refinement of the landmarks. This is especially noticeable in the subpubic angle, displayed in Figure 4, for which the coarse predictions were highly uncorrelated, with correlation coefficient 0.01, while the linear relationship significantly increased after the refinement, with a final r value of 0.57. Nevertheless, the

ISBN: 978-84-17853-76-1

calculation of the subpubic angle is highly dependent on the exact positions of the predicted landmarks, as small changes in the coordinates could lead to significant differences in the angle. Due to the fact that L0 and L2 had large errors as observed in the boxplot representation, we also expected higher errors in the subpubic angle calculation.

In the case of the transverse diameter, Figure 5, the correlation coefficient improved from 0.37 to 0.85. This had the most correlated predictions with respect to the real annotations, which matches the observed results in the boxplot. Finally, the correlation coefficient obtained for the obstetric conjugate diameter was improved from 0.35 to 0.58 (Figure 6). In this case, the low correlation could be associated with the large errors obtained for L6 predictions due to the position of this landmark in the vertebrae.

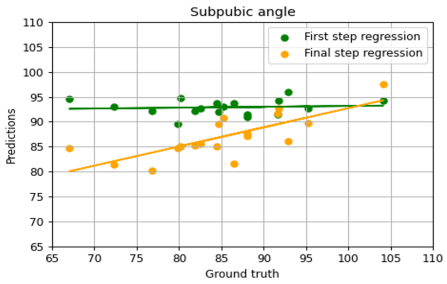

Figure 4. *Scatter plot and linear regression for the subpubic angle measurements.*

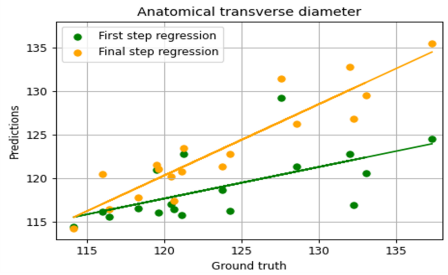

Figure 5. *Scatter plot and linear regression for the transverse diameter.*

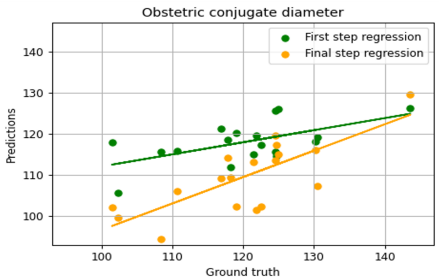

Figure 6. *Scatter plot and linear regression for the obstetric conjugate diameter..*

4. Conclusions

This study highlighted the potential of machine learning algorithms for improving pelvic landmark detection. The lack of available literature in this area made it difficult to develop an efficient method for predicting 3D landmarks on our limited dataset. Additionally, cropped images difficulted the manual annotation in landmarks L0 and L2, limiting the learning possibilities of the neural network. However, our model showed promising results, providing efficient and accurate predictions of the anatomical landmarks in CT scans. We have successfully developed a two-step approach that will reduce the workload of the radiologists, serving as a valuable tool for pelvimetry. Future investigations will explore testing in larger datasets and classifying the different types of pelvis based on our predictions.

Acknowledgments

Research supported by projects PI122/00601 and AC20/00102 (Ministerio de Ciencia, Innovación y Universidades, Instituto de Salud Carlos III, Asociación Española Contra el Cáncer and European Regional Development Fund “Una manera de hacer Europa”), project PerPlanRT (under the frame of ERA PerMed), TED2021-129392B-I00 and TED2021-132200B-I00 (MCIN/AEI/10.13039/501100011033 and European Union “NextGenerationEU”/PRTR).

References

[1] P. Frémondière *et ál.* Which foetal-pelvic variables are useful for predicting caesarean section and instrumental assistance. *Medical Principles and Practice*, 26(4):359–367, 2017.

[2] Miriam S Lenhard *et ál.* Pelvimetry revisited: analyzing cephalopelvic disproportion. *European journal of radiology*, 74(3):e107–e111, 2010.

[3] S. Kilpatrick *et ál.* Normal labor and delivery. *Obstetrics: Normal and problem pregnancies*, 5:303–321, 2007.

[4] D. Sánchez *et ál.* Effects of sex, age and height on symphysis–ischial spine distance measured on a pelvic ct. *Journal of Clinical Medicine*, 11(9):2395, 2022.

[5] I. Salk *et ál.* Pelvimetry by three-dimensional computed tomography in non-pregnant multiparous women who delivered vaginally. *Polish journal of radiology*, 81:219, 2016.

[6] C. Payer *et ál.* Regressing heatmaps for multiple landmark localization using cnns. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 230–238. Springer, 2016.

[7] A. Jimenez-Pastor *et ál.* Automated vertebrae localization and identification by decision forests and image-based refinement on real-world ct data. *La radiologia medica*, 125:48–56, 2020.

[8] B. Bier *et ál.* Learning to detect anatomical landmarks of the pelvis in x-rays from arbitrary views. *International journal of computer assisted radiology and surgery*, 14:1463–1473, 2019.

[9] A. Fedorov *et ál.* 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9):1323–1341, Nov 2012.

[10] Tami D DenOtter *et ál.* Hounsfield unit. 2019.