NOFOR
NOVEL FORENSIC RESEARCH

DOI: 10.5455/NOFOR.2026.08.012

NOFOR. 2026;5(1):8-26

Original Article

A comprehensive computational pipeline for foramen magnum shape analysis integrating elliptic fourier transform, polygon approximation, and procrustes heatmaps from CT images: Application to forensic sex estimation

Yasin Etli[1], Erhan Kartal[1], Mahmut Asirdizer[2], Yavuz Hekimoglu[3], Siddik Keskin[4], Ugur Demir[5], Alparslan Yavuz[6]

[1]*Van Yüzüncü Yıl University, Faculty of Medicine, Department of Forensic Medicine, Van, Türkiye*
[2]*Bahçeşehir University, Faculty of Medicine, Department of Forensic Medicine, İstanbul, Türkiye*
[3]*Health Sciences University, Ankara City Hospital, Department of Forensic Medicine, Ankara, Türkiye*
[4]*Van Yüzüncü Yıl University, Faculty of Medicine, Department of Biostatistics, Van, Türkiye*
[5]*Harran University, Faculty of Medicine, Department of Forensic Medicine, Şanlıurfa, Türkiye*
[6]*Health Sciences University, Antalya Training and Research Hospital, Department of Radiology, Antalya, Türkiye*

Received February 01, 2026; Accepted April 03, 2026; Available online April 15, 2026

Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Abstract

Aim: The foramen magnum (FM) shows size-based sexual dimorphism, but whether its shape differs independently between sexes is unresolved, as previous studies relied on subjective classification. This study applied computational geometry to computed tomography (CT)-derived FM contours to objectively quantify shape variation and evaluate its value for sex estimation.
Materials and Methods: FM boundaries were automatically extracted from axial CT images of 473 adults (234 males, 239 females) from an Eastern Turkish population and screened by an automated topological quality check. Seven geometric analyses, including normalized elliptic Fourier transform and Procrustes-based heatmaps, were applied; six feature-based analyses yielded 209 descriptors per specimen, classified a priori as size measures (n = 30), scale-dependent contour descriptors (n = 63), or scale-invariant shape descriptors (n = 116). Sex differences were assessed with Mann–Whitney U tests (Bonferroni-corrected) and Cohen's d. Classification used linear discriminant analysis (LDA), support vector machines, and random forest with stratified 10-fold cross-validation and nested forward feature selection.
Results: All 30 size descriptors differed significantly between sexes (Cohen's $d = 0.71–1.33$), whereas only 1 of 116 scale-invariant shape descriptors did (boundary roughness, d = 0.62). Size alone yielded an LDA accuracy of 77.1% (95% confidence interval 73.2 – 80.7%). Scale-invariant shape descriptors reached 64.0% (58.3% after nested forward selection) and Procrustes heatmaps 59.9–60.9%; adding forward-selected shape descriptors to size did not improve classification (74.6%; McNemar p = .08). A six-sided polygon was optimal in both sexes, and the Bayesian information criterion favoured a single mixture component, indicating no discrete within-sex subtypes.
Conclusion: FM sexual dimorphism is almost entirely size-mediated. Once scale is removed, the FM outline carries only weak sex information that does not add to size-based estimation. The automated, openly available pipeline replaces subjective shape categorization with objective, continuous measurements and may be transferable to other skeletal structures.

Keywords: Foramen magnum; sex determination by skeleton; forensic anthropology; fourier analysis; tomography, X-ray computed

INTRODUCTION

The foramen magnum (FM), the largest opening in the skull base formed by the occipital bone, serves as the primary conduit between the cranial cavity and the vertebral canal, transmitting the medulla oblongata, vertebral arteries, spinal accessory nerves, and associated meningeal membranes [1]. Beyond its well-established clinical significance in cranio-cervical surgery and neuropathology, the FM has attracted considerable

CITATION
Etli Y, Kartal E, Asirdizer M, Hekimoglu Y, Keskin S, Demir U, et al. A comprehensive computational pipeline for foramen magnum shape analysis integrating elliptic fourier transform, polygon approximation, and procrustes heatmaps from CT images: Application to forensic sex estimation. NOFOR. 2026;5(1):8-26. DOI: 10.5455/NOFOR.2026.08.012

OPEN ACCESS

Corresponding Author: Yasin Etli, Van Yüzüncü Yıl University, Faculty of Medicine, Department of Forensic Medicine, Van, Türkiye
Email: yasinetli@yyu.edu.tr

DOI: 10.5455/NOFOR.2026.08.012

research interest due to its reported morphological variation across individuals, populations, and between sexes [2-4]. A central question that has persisted in the anatomical and forensic literature is whether sex-related differences in the FM are limited to overall size, with males exhibiting larger dimensions, or whether genuine shape differences also exist between the sexes [5,6].

The dimensional sexual dimorphism of the FM has been well documented across diverse populations. Numerous studies using dry skulls, radiographs, computed tomography (CT), and cone-beam CT have consistently demonstrated that the anteroposterior (AP) diameter, transverse diameter, circumference, and area of the FM are significantly greater in males than in females [3,7,8]. A recent systematic review and meta-analysis confirmed these dimensional differences across multiple populations, with sex estimation accuracies ranging from approximately 62% to 90% depending on the population and analytical method employed [9]. In the Turkish population specifically, Meral et al. reported a maximum multivariate accuracy of 75% [10], while Kartal et al. achieved up to 86.7% with discriminant analysis and 88.2% with artificial neural networks using dimensional measurements from 720 individuals [11]. These findings collectively establish that the FM exhibits moderate but consistent size-based sexual dimorphism that is primarily driven by the overall larger cranial dimensions in males.

In contrast, the question of whether the FM also exhibits sex-related differences in shape, independent of size, remains substantially less resolved. Morphological descriptions in the forensic and anatomical literature have long suggested that the male FM tends to display a “rougher” or more irregular contour, while the female FM is described as “more oval” and “more regular” [2,12]. However, these descriptions are based on qualitative, observer-dependent assessments and have not been quantitatively validated. The most common approach to characterizing FM shape has been subjective visual classification into categorical types such as oval, round, egg-shaped, tetragonal, pentagonal, hexagonal, and irregular forms [12,13]. This approach suffers from several well-documented limitations. First, the shape categories themselves are not standardized across studies, leading to marked inconsistencies in the reported prevalence of each type; for instance, the FM has been reported as oval in 87.3% of crania in one study but in only 7.39% in another that utilized the same visual categorization approach [14]. Second, studies that have investigated whether these categorical shape types differ between sexes have generally concluded that FM shape is not a reliable sex discriminator [5,6], with the FM index (AP/transverse diameter ratio) showing no significant sex difference in several populations [4,5]. Zdilla et al. critically observed that previous shape studies utilized “ambiguous and subjective descriptors” that are “difficult to reproduce” and called for more reproducible methods of characterizing FM morphology [14].

The reliance on subjective categorical classification or simple ratios to characterize a complex, irregularly bounded anatomical contour fundamentally limits the ability to detect subtle but potentially meaningful shape differences. A two-dimensional closed contour such as the FM boundary contains considerably more morphological information than can be captured by two perpendicular diameters or a single aspect ratio [15,16]. Modern computational imaging and geometric analysis methods offer the potential to extract this richer shape information objectively. Elliptic Fourier analysis (EFA), introduced by Kuhl and Giardina [15], enables the mathematical decomposition of any closed contour into a series of harmonic ellipses, capturing both global shape properties and fine local boundary details in a compact numerical representation. This approach has been applied to various anatomical contours, including the FM of the tropical raccoon, where Samuel et al. showed that elliptical Fourier descriptors captured outline details not described by conventional measurements [17]. Hayashizaki et al. applied Fourier analysis to pelvic contours from postmortem CT images and showed that frequency-domain shape descriptors could discriminate sex with greater objectivity than traditional morphometric methods [18]. Similarly, Gonzalez et al. utilized semi-landmark-based geometric morphometrics on the greater sciatic notch and ischiopubic complex to achieve over 90% sex estimation accuracy from shape variables alone [19]. In our own previous work, we applied contour-based geometric analysis methods, including ellipse fitting, to pelvic CT images and demonstrated that computational shape features could objectively capture morphological variations that had previously been described only qualitatively [20]. These studies collectively suggest that computational geometric approaches, when applied to anatomical contour data extracted from CT images, can reveal shape information that is inaccessible to conventional linear measurements and subjective visual classification.

Despite the clear potential of these computational shape analysis methods, no study to date has applied a comprehensive suite of such techniques to the FM contour for the purpose of objectively quantifying and validating the morphological descriptions that have persisted in the anatomical literature. Specifically, it remains unknown whether computationally derived metrics, such as circularity, boundary roughness, contour regularity, left–right and AP asymmetry, and harmonic shape descriptors, differ significantly between male and female FM contours when size effects are controlled for. Furthermore, whether any such shape features, once objectively defined and separated from scale, can contribute to or improve upon size-based sex estimation has not been systematically evaluated. Resolving these questions is important not only for advancing evidence-based anatomical knowledge, but also for practical applications in forensic medicine and anthropology, where the ability to estimate sex from fragmentary cranial remains, particularly the skull base, which is resistant to taphonomic destruction, can be critical in identification cases [21,22].

The present study aims to address these gaps through two integrated objectives. The first objective is to apply a comprehensive array of computational geometric methods to FM boundary coordinates extracted from CT images of a large

DOI: 10.5455/NOFOR.2026.08.012

contemporary Turkish sample (n = 473). These methods comprise least-squares, maximum inscribed, and minimum enclosing ellipse fitting; systematic polygon approximation from 3-sided to 13-sided polygons; normalized elliptic Fourier transform (EFT) with 20 harmonics; radial profile analysis; and point density heatmap construction. The aim is to objectively characterize and quantify morphological features that correspond to classical qualitative descriptions such as "oval," "round," "irregular," and "rough." The second objective is to evaluate the discriminative capacity of these objectively defined shape features for sex estimation, both independently and in combination with conventional size-based measurements. Linear discriminant analysis (LDA), support vector machines (SVM), and random forest (RF) were applied under cross-validation in which all data-driven model choices were confined to training data. Finally, the study investigates whether unsupervised clustering reveals discrete sex-specific morphological subtypes.

MATERIAL AND METHODS

Study Design and Ethical Approval

This retrospective cross-sectional study utilized cranial CT images originally acquired for a prior investigation of FM sexual dimorphism in an Eastern Turkish population [11]. The original study protocol received approval from the Non-Interventional Clinical Research Ethics Committee of Van Yüzüncü Yil University Faculty of Medicine (decision number: 02, dated: 19.02.2017) and was conducted in accordance with the Declaration of Helsinki and the principles of Good Clinical Practice. Since the present study represents a secondary analysis of the same anonymized imaging dataset using computational geometric methods, no additional ethical approval was required. Informed consent was waived due to the retrospective design and full anonymization of all data.

Study Sample

The source dataset comprised 720 cranial CT examinations (360 males and 360 females) from adult patients who presented to Yüzüncü Yıl University Dursun Odabaş Medical Center between January and December 2016 for various clinical indications unrelated to craniocervical pathology [11]. The original sample was constructed using stratified sampling across six adult age groups (21–30, 31–40, 41–50, 51–60, 61–70, and ≥ 71 years) with equal sex distribution to control for potential age-related confounding effects. Exclusion criteria in the original study included skull trauma causing cranial deformities, congenital or acquired cranial bone diseases, any malignancy affecting the skull, and cases with missing or defective FM boundaries [11].

For the present study, all 720 CT examinations were re-evaluated to determine their suitability for automated contour extraction. Cases were included if the axial CT slice displaying the widest FM cross-section yielded a boundary amenable to automated edge detection, specifically, sufficient contrast at the bone–soft tissue interface and a complete, uninterrupted FM margin. Cases with excessive image noise, partial volume artifacts obscuring the FM border, or incomplete boundary delineation after edge detection were excluded. After this additional selection, 473 individuals (234 males and 239 females) met the criteria and constituted the final study sample. The 247 excluded examinations comprised 126 males and 121 females, corresponding to 35.0% of the male and 33.6% of the female source sample (χ^2 test with continuity correction, p = .75). The screening outcome was recorded as a single include/exclude decision per examination; the frequency of each individual exclusion reason could therefore not be reconstructed.

CT Image Acquisition

All CT examinations had been performed on a 16-detector multislice CT system (Somatom Emotion 16-slice, CT2012E; Siemens AG, Berlin and Munich, Germany). Imaging parameters included a tube voltage of 120 kV with automatic tube current modulation ranging between 80 and 120 mA, a section thickness of 1.2 mm, a reconstruction interval of 1.0 mm, a pitch factor of 0.8, a field of view of 220 mm, and a matrix size of 256 × 256. Data acquisition was performed in the caudocranial direction without gantry tilt. In the source study, three-dimensional volume-rendered reformations were generated using Syngo VE.52A software (Siemens Healthcare, Erlangen, Germany) for manual measurement [11]; the present analysis used the axial images. The spatial calibration factor (mm per pixel) was individually extracted from the Digital Imaging and Communications in Medicine (DICOM) metadata of each examination and ranged from 0.271 to 0.513 mm/pixel (mean: 0.371 mm/pixel), ensuring accurate conversion of all pixel-based coordinate measurements to absolute millimeter values.

Foramen Magnum Boundary Extraction and Quality Control

For each specimen, the axial CT slice displaying the widest cross-section of the FM was identified. The FM boundary was extracted using automated edge detection applied to the bone-windowed CT image at this level. Each extracted boundary comprised between 36 and 187 points (mean: 85.5) and was archived as a sequence of 500 equally spaced points obtained by periodic cubic spline interpolation along the cumulative arc length. These archived coordinate files constitute the deposited dataset and the input of the analysis pipeline.

Contour quality was controlled at two levels. First, every extracted contour was visually inspected by one of the authors, overlaid on its source CT slice, to confirm that it followed the bony FM margin. Second, because narrow tracing errors can escape visual inspection, an automated topological check was applied to every contour: a valid FM outline must be a simple closed curve, so all pairs of non-adjacent contour segments were tested for crossings. Twenty-one contours (15 males, 6 females; 4.4%) contained at least one self-intersection, typically at the point where the tracing closed. These crossings were corrected automatically: at each crossing, the shorter of the two arcs (the loop) was removed and replaced by the crossing point, and the procedure was repeated until the contour was simple. A median of 23 of the 500 points (range 5–154) was removed. In 10 contours (8 males, 2 females) the removed loop exceeded 5% of the contour points;

DOI: 10.5455/NOFOR.2026.08.012

in the remaining 11 it comprised 5–23 points (at most 4.6%). All corrections are listed and illustrated in the supplementary material (Table S1; Figure S1), and all principal analyses were repeated without correction as a sensitivity analysis (Table S1).

After quality control, each contour was resampled once to 200 equally spaced points by cubic interpolation along the cumulative arc length, and all subsequent computations were performed on these 200-point contours. All contours were then given the same traversal direction and a homologous starting point (the anterior-most contour point, i.e., the point with the minimum AP image coordinate, as all images were in standard radiological orientation). Because the contours were derived computationally rather than traced manually, contour placement on a given slice and all subsequent computations are deterministic: processing the same CT slice repeatedly yields identical coordinates and features. This removes inter- and intra-observer variability in contour placement, a documented source of error in conventional FM morphometry [3,6]; the reproducibility of the identification of the slice with the widest cross-section was not assessed. The accuracy of the automated boundaries against an independent reference, such as manual expert tracing, was not assessed (see Limitations).

Computational Geometric Analyses

All computational analyses were implemented in Python 3 (NumPy, SciPy, scikit-learn, pandas, and Matplotlib). The complete analysis pipeline, from the archived boundary coordinates onward, is publicly available as a single, self-contained script (FM_Comprehensive_Analysis_v4.py) in the Zenodo repository accompanying this paper [23], together with all anonymized boundary coordinate files and the calibration metadata. The script reproduces all analyses, tables, and figures reported here; the edge-detection step that produced the coordinate files is not part of it. Seven complementary analyses were performed: classical morphometry; ellipse and circle fitting; polygon approximation; polygon elbow analysis; EFT; radial profile analysis; and Procrustes-based point density heatmaps. The first six yielded 209 descriptors per specimen, whereas the heatmap analysis operates directly on the aligned coordinates. A complete dictionary of the descriptors, with their definitions, units, and scale category (see *Feature categories* below), is provided in Table S2.

Classical Morphometric Measurements

Eleven standard morphometric parameters were computed from each contour. The AP diameter was defined as the extent of the contour along the AP (vertical) image axis, and the transverse diameter as its extent along the transverse (horizontal) image axis. The maximum diameter was the largest distance between any two points of the convex hull (maximum Feret diameter), and the minimum diameter was the smaller of the AP and transverse diameters. Perimeter was the sum of consecutive point-to-point distances along the closed contour, and area was computed using the shoelace formula. Five ratios were derived from these quantities: FM index (AP/transverse diameter), circularity ($4\pi A/P^2$, where A is area and P is perimeter), eccentricity (maximum/minimum diameter), area-to-perimeter ratio (A/P), and compactness ($P/\sqrt{A}$). These parameters correspond to those commonly reported in the FM literature [3,4,6,10,11], thereby facilitating direct comparison with previously published dimensional data. Note that the area-to-perimeter ratio has the dimension of length (mm) and therefore scales with FM size.

Ellipse and Circle Fitting

Three ellipse models and one circle model were fitted to each contour. (1) The least-squares ellipse was obtained by minimizing the sum of squared algebraic residuals of the canonical ellipse equation, $[(x'/a)^2 + (y'/b)^2 - 1]^2$, with respect to the centre, semi-axes (a, b), and orientation, using the Nelder–Mead simplex algorithm initialized from the contour centroid, extents, and principal axis; this is an iterative alternative to the direct least-squares solution of Fitzgibbon et al. [24]. Fits whose semi-axes fell outside 0.1–2 times the contour span were to be rejected as implausible; no fit was rejected. (2) The maximum inscribed ellipse was defined as the largest-area ellipse with every contour point on or outside its boundary. It was found by sequential least-squares quadratic programming with a deterministic multistart over four initial orientations. (3) The minimum enclosing ellipse was defined as the smallest-area ellipse containing every contour point and was computed with Khachiyan's algorithm [25]. The inscribed and enclosing ellipses bracket the contour, so the gap between them reflects the departure of the FM outline from an ideal ellipse (Figure S2). For each ellipse model, eight features were recorded: semi-major and semi-minor axes (mm), ellipse area (mm^2), the mean, standard deviation (SD), maximum, and root mean square error (RMSE) of the distances from each contour point to the nearest point of the ellipse (mm), and the ellipse-to-FM area ratio. (4) A circle was fitted by algebraic least squares, and seven features were recorded: radius, circle area, the four distance metrics, and the circle-to-FM area ratio. In total, ellipse and circle fitting yielded 31 features.

Polygon Approximation

To capture FM boundary irregularity at multiple geometric resolution levels, systematic polygon approximation was performed for polygons with 3 to 13 sides (11 polygon orders). For each polygon order n, the n-sided polygon was determined by angular sectoring from the FM centroid, with the vertex of each sector placed at the contour point farthest from the centroid within that sector. The sector grid was rotated through one sector angle in 20 equal steps (19 distinct orientations, as the first and last coincide), and the configuration yielding the minimum RMSE was selected. The polygons are therefore an algorithmic approximation of the outline, not a delineation of anatomical landmarks.

For each polygon order, six features were extracted: the RMSE, mean, SD, and maximum of the point-to-segment distances between the contour and the polygon edges, the polygon area, and the polygon-to-FM area ratio. This yielded 66 polygon features in total.

DOI: 10.5455/NOFOR.2026.08.012

The rationale for this multi-scale approach is that low-sided polygons (3–5 sides) capture gross shape asymmetry and angularity, while high-sided polygons (9–13 sides) approximate the contour more closely and their residual errors primarily reflect fine boundary roughness and local irregularity. The progression of RMSE across polygon orders thus provides an objective, multi-resolution characterization of contour complexity.

Polygon Elbow (Optimal Complexity) Analysis

To identify the polygon order at which additional sides yield diminishing returns in contour approximation accuracy, an elbow-point analysis was performed on each specimen's RMSE curve using the Kneedle algorithm [26]. For each specimen, the RMSE values for polygon orders 3 through 13 were normalized to the unit interval on both axes. A reference line connecting the first and last normalized points was constructed, and the vertical distance of the normalized RMSE curve below this line was computed at each polygon order. The elbow point was defined as the polygon order at which this distance was maximal, representing the point of maximum curvature, that is, the transition from rapid to marginal improvement in contour fit with each additional polygon side [26].

For each specimen, five derived metrics were computed at the identified elbow point: (1) the optimal polygon order (number of sides), (2) the RMSE at the optimal polygon, (3) the RMSE reduction from the 3-sided to the optimal polygon as a percentage of the 3-sided RMSE, (4) the remaining RMSE reduction from the optimal polygon to the 13-sided polygon, also as a percentage of the 3-sided RMSE, and (5) an efficiency ratio defined as metric 3 divided by metric 4.

The elbow analysis was performed at both the individual specimen level and the group level. At the group level, RMSE values were averaged across all specimens within each sex and across the entire sample, and the Kneedle algorithm was applied to these mean curves. The distribution of individually determined optimal polygon orders was compared between males and females using the Mann–Whitney U test and Cohen's *d*. The step-by-step marginal improvement, that is, the percentage reduction in mean RMSE at each polygon-order increment, was computed for males, females, and the pooled sample.

Elliptic Fourier Transform

The EFT was applied to each FM contour following the method of Kuhl and Giardina [15]. The EFT decomposes a closed two-dimensional contour into a series of harmonically related ellipses, where the first harmonic captures the gross elliptical shape and higher harmonics capture progressively finer morphological details. Twenty harmonics were computed, yielding 80 coefficients (four per harmonic: a_n, b_n, c_n, d_n). The coefficients were normalized for size, orientation, and starting point using the first-harmonic ellipse (standard Kuhl–Giardina normalization [15,16]. After this normalization, three first-harmonic coefficients (a_1, b_1, c_1) are constant and were discarded, leaving 77 informative normalized coefficients that reflect shape independent of scale, orientation, and starting location.

In addition, 13 summary descriptors were derived from the harmonic powers ($P_n = a_n^2 + b_n^2 + c_n^2 + d_n^2$) of the unnormalized coefficients: the power ratios of harmonics 1–10 (P_n divided by the total power), a higher-harmonic energy ratio (the proportion of total power in harmonics 4–20, reflecting fine morphological detail relative to gross shape), a symmetry index (the ratio of odd-harmonic to even-harmonic power), and the total spectral power expressed in mm². The power ratios and the symmetry index are dimensionless and invariant to scale, whereas the total power scales with FM size. This yielded 90 EFT features per specimen.

Radial Distance Profile Analysis

To provide intuitive, physically interpretable shape descriptors that correspond directly to classical qualitative morphological descriptions, a radial distance profile was computed for each FM contour. Distances from the centroid to each contour point were calculated and ordered by angular position. Six descriptors were extracted: radial mean and SD (mm), reflecting overall size and radial variability; the radial coefficient of variation (CV), reflecting boundary regularity independent of size; roughness, the mean absolute second difference of the unit-mean radial profile, quantifying high-frequency boundary undulation; and two mirror-asymmetry indices computed from the unit-mean radial profile $r(\theta)$ resampled at 1° intervals. Left–right asymmetry was the mean absolute difference between $r(\theta)$ and its mirror image about the AP (sagittal) axis, $r(\pi - \theta)$. AP asymmetry was the mean absolute difference between $r(\theta)$ and its mirror image about the transverse axis, $r(-\theta)$. Roughness and both asymmetry indices were computed from size-normalized profiles.

Generalized Procrustes Alignment and Point Density Heatmaps

To enable shape comparison independent of size, position, and orientation, all quality-controlled contours were aligned by generalized Procrustes analysis (GPA) [27,28]. Because all contours had a common starting point (the anterior-most point), a common traversal direction, and 200 equally spaced points, corresponding points were treated as homologous semilandmarks. Each configuration was centred on its centroid and scaled to unit root-mean-square radius. Each was then rotated by least squares, without reflection, onto an iteratively updated consensus configuration until convergence. For display, the aligned configurations were rotated so that the long axis of the consensus was vertical with the anterior end uppermost. GPA uses no sex information.

Sex-specific point density heatmaps were generated by pooling the aligned contour points of each sex onto a 200 × 200 grid and applying Gaussian smoothing ($\sigma = 2$ grid units) (Figure 1). These pooled heatmaps were used for visualization only; heatmap-based classification used the cross-validated procedure described below.

DOI: 10.5455/NOFOR.2026.08.012

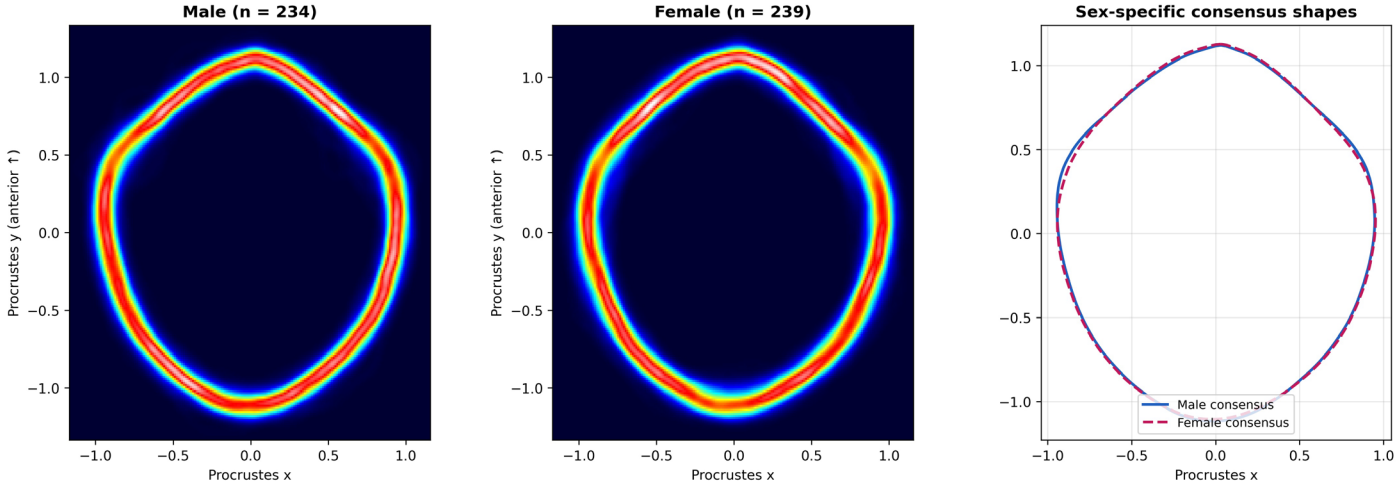

Figure 1. Procrustes-aligned foramen magnum point density heatmaps for males (left, n = 234) and females (middle, n = 239), and the sex-specific consensus outlines (right). Contours were aligned by generalized Procrustes analysis (translation, scaling to unit root-mean-square radius, and rotation without reflection) and are displayed with the anterior end uppermost. Gaussian smoothing (σ = 2 grid units) was applied to the pooled contour points on a 200 × 200 grid; warmer colors indicate higher point density. The heatmaps are descriptive; heatmap-based classification used templates built from training folds only.

Feature Categories

Because several descriptors that are not diameters nonetheless carry information on FM size, every descriptor was assigned a priori to one of three categories according to its physical dimension (Table S2).

Size descriptors (n = 30) are absolute dimensions: the six classical size measurements (four diameters, perimeter, and area), the radial mean, the semi-axes and areas of the fitted ellipses and circle, the polygon areas, and the total EFT power.

Scale-dependent contour descriptors (n = 63) quantify the irregularity of the outline but are expressed in millimetres and therefore scale with FM size. They comprise the area-to-perimeter ratio, the radial SD, all ellipse, circle, and polygon distance metrics, and the RMSE at the optimal polygon.

Scale-invariant shape descriptors (n = 116) are dimensionless and unchanged by isometric rescaling of the contour. They comprise the FM index, circularity, eccentricity, compactness, radial CV, roughness, the two asymmetry indices, the ellipse, circle, and polygon area ratios, the four dimensionless polygon elbow metrics, the 77 normalized EFT coefficients, the 10 harmonic power ratios, the higher-harmonic energy ratio, and the EFT symmetry index. Throughout this paper, "shape" refers exclusively to the scale-invariant category.

Statistical Analysis

Sex Comparison of Individual Features

All 209 descriptors were compared between males and females using the Mann–Whitney U test, chosen for its robustness to non-normality. Effect sizes were quantified using Cohen's *d*, computed as the difference in group means divided by the pooled SD (positive values indicate larger values in males). The Bonferroni correction was applied across all 209 comparisons, and features were considered statistically significant if the Bonferroni-corrected p value was less than .05. All 473 specimens contributed to every analysis; no descriptor had missing values.

Sex Classification

The discriminative capacity of the descriptors for sex estimation was evaluated using four classifiers: LDA, an SVM with a linear kernel (SVM-linear), an SVM with a radial basis function (RBF) kernel (SVM-RBF), and RF (100 trees). Hyperparameters were fixed a priori at the library defaults (SVM: C = 1; RBF kernel: γ = "scale"), and no hyperparameter tuning was performed. All classifiers were evaluated by stratified 10-fold cross-validation with a fixed random seed (42). Within each fold, features were standardized using the mean and SD of the training set only. Performance was summarized as accuracy (mean ± SD across the 10 test folds; the primary estimate reported in the text), the pooled out-of-fold accuracy with its Wilson 95% confidence interval (CI), sensitivity (proportion of males correctly classified), and specificity (proportion of females correctly classified). In the present data, fold-mean and pooled accuracies differed by at most 0.1 percentage points.

Feature Sets

Five fixed feature sets were evaluated: S1 (size), the six classical size measurements (AP, transverse, maximum, and minimum diameters, perimeter, and area); S2 (shape), all 116 scale-invariant shape descriptors; S3 (size + shape), the union of S1 and S2 (122 features); S4 (scale-dependent), the 63 scale-dependent contour descriptors; and S5 (all), all 209 descriptors. In addition, three forward-selected (FS) models were derived from the S2, S3, and S5 pools (FS-Shape, FS-Size+Shape, and FS-All), and two heatmap-based classifiers were evaluated (see below).

DOI: 10.5455/NOFOR.2026.08.012

Nested Forward Feature Selection

Sequential forward selection (SFS) was used to identify small feature subsets. Starting from an empty set, at each step the feature producing the greatest improvement in 10-fold stratified LDA cross-validation accuracy was added, and the process terminated when no remaining feature improved this criterion (maximum 15 features). To obtain unbiased performance estimates, SFS was nested within the outer cross-validation [29,30]. In each of the 10 outer folds, SFS was run using only the nine training folds, with an inner 10-fold split of those data. The selected subset was then fitted on the outer training folds and evaluated once on the untouched outer test fold. The reported accuracies of FS-Shape, FS-Size+Shape, and FS-All are these outer-fold estimates. The mean inner (selection) accuracy is reported alongside them to show the optimism that results when the same data are used both to select features and to estimate performance. Because a different subset may be selected in each outer fold, the frequency with which each feature was selected across the 10 outer folds is reported.

Comparison of Classifiers and Stability

Paired comparisons with the size-only LDA model (S1) were made with the exact McNemar test on the out-of-fold predictions of the same specimens [31]; these p values are reported without adjustment for multiple comparisons. To assess the stability of the estimates with respect to fold assignment, the LDA analyses of S1–S5 were repeated with 10 different random fold partitions, and the nested forward-selection analyses with 5 different partitions.

Heatmap-Based Sex Classification

A heatmap-based classification approach was developed as an alternative to conventional feature-based methods. This method operates directly on the spatial distribution of Procrustes-aligned FM boundary points, comparing each test specimen's contour against sex-specific population density templates. Heatmaps were constructed exclusively from the training folds of the same 10-fold partition, so test specimens were never included in the templates against which they were scored.

Method A – Point sampling: For each fold, sex-specific density heatmaps were constructed from training specimens by accumulating their aligned contour points onto a shared 200 × 200 grid and applying Gaussian smoothing ($\sigma = 2$ grid units). Each test contour was overlaid on both heatmaps, and the mean density at its boundary point positions was computed. The specimen was assigned to the sex whose heatmap yielded the higher score.

Method B – Kernel density estimation (KDE) with overlap-integral scoring: Point sampling evaluates only the exact grid cells occupied by the test contour and is therefore sensitive to minor positional shifts. In the KDE approach, the reference heatmaps and the test contour are each rendered as continuous density fields with the same Gaussian kernel, so that nearby but not exactly coincident boundary points still contribute to the similarity measure. The similarity between the test field and each sex-specific reference field was the overlap integral (the element-wise product of the two fields summed over the grid), a quantity related to the Bhattacharyya coefficient. The specimen was assigned to the sex with the higher overlap. The kernel bandwidth (full width at half maximum: 3, 5, 7, 9, or 11 grid pixels) was selected within each outer training set by an inner 10-fold cross-validation, and the selected value was then applied to the outer test fold.

Within-Sex Shape Structure

To investigate whether the FM exhibits discrete morphological subtypes within each sex, unsupervised clustering was performed separately for each sex. The GPA-aligned coordinates (200 points × 2 = 400 variables) were reduced by principal component analysis to the first 20 components, which retained 96.1% (males) and 96.9% (females) of the shape variance. Gaussian mixture models (GMM) with full covariance matrices were fitted for $k = 1, 2, 3$, and 4 components (10 random initializations each). The number of components was selected by the Bayesian information criterion (BIC; lower values indicate better support) [32]. Including $k = 1$ allows a single, unstructured distribution to be retained when no clustering is supported. As supporting criteria, the silhouette coefficient was computed for $k \geq 2$, and partition stability was assessed with 200 bootstrap resamples: a GMM was refitted to each resample, and its labelling of the full sample was compared with the original labelling by the adjusted Rand index (ARI) [33]. Clusters were to be interpreted only if supported by BIC and reproducible under resampling. Mean values of the shape descriptors are reported for descriptive purposes.

Because a claim of greater within-sex heterogeneity does not require discrete clusters, dispersion was also compared directly between sexes. For outline shape as a whole, the Procrustes distance of each specimen from its own sex-specific consensus was computed. The male–female difference in mean distance was tested by 10,000 label permutations, a test analogous to the permutational analysis of multivariate dispersions (PERMDISP) [34]. For individual indices, equality of variances for eight scale-invariant indices (FM index, circularity, eccentricity, compactness, radial CV, roughness, and the two asymmetry indices) was tested with the Brown–Forsythe test [35], with Bonferroni correction for eight comparisons.

Software and Reproducibility

All analyses were implemented in Python 3 using NumPy, SciPy, scikit-learn, pandas, and Matplotlib. The complete analysis code (FM_Comprehensive_Analysis_v4.py) and all 473 anonymized FM boundary coordinate files are publicly available in a Zenodo repository [23]. The coordinate files contain the 500-point archived contours, and a metadata file gives the sex and mm-per-pixel calibration of each specimen. The code performs the quality control, 200-point resampling, feature extraction, statistical analyses, and figures reported here. The repository is currently under embargo and will be made publicly accessible upon publication; reviewer access can be arranged upon request to the corresponding author. Per-specimen measurements are provided in Table S3.

DOI: 10.5455/NOFOR.2026.08.012

RESULTS

Study Sample and Contour Quality

A total of 473 adult CT scans met the inclusion criteria: 234 males (49.5%) and 239 females (50.5%). The automated topological check identified self-intersections in 21 contours (4.4%; 15 males, 6 females), which were corrected as described in the Methods (Figure S1; Table S1). After correction, all 473 contours were simple closed curves, all ellipse and circle fits converged to plausible solutions, and all 209 descriptors were available for every specimen. Every analysis was therefore performed on the complete sample of 473 specimens (234 males, 239 females).

Descriptive Morphometry

Classical linear and area measurements revealed pronounced sexual dimorphism in all size-related parameters (Table 1). Male FM perimeters averaged 106.14 ± 9.89 mm compared with 94.80 ± 7.02 mm in females (Cohen's d = 1.33, $p < .001$). FM area showed a comparable effect (males: 832.1 ± 160.7 mm², females: 665.8 ± 99.1 mm²; d = 1.25). AP and transverse diameters were both significantly larger in males (AP: 36.52 ± 3.47 vs. 32.68 ± 2.77 mm, d = 1.22; transverse: 31.38 ± 3.19 vs. 28.24 ± 2.31 mm, d = 1.13). On average, male perimeters were 12.0% longer and male areas 25.0% larger than those of females.

Table 1. Descriptive morphometry and sex comparison of key foramen magnum parameters (n = 473; 234 males, 239 females)

Parameter		Male mean ± SD	Female mean ± SD	Cohen's *d*	p (Bonferroni)
Size measurements	**AP diameter (mm)**	36.52 ± 3.47	32.68 ± 2.77	**1.22**	**< .001**
	Transverse diameter (mm)	31.38 ± 3.19	28.24 ± 2.31	**1.13**	**< .001**
	Maximum diameter (mm)	36.85 ± 3.39	33.04 ± 2.62	**1.26**	**< .001**
	Minimum diameter (mm)	31.38 ± 3.18	28.20 ± 2.29	**1.15**	**< .001**
	Perimeter (mm)	106.14 ± 9.89	94.80 ± 7.02	**1.33**	**< .001**
	Area (mm²)	832.1 ± 160.7	665.8 ± 99.1	**1.25**	**< .001**
Scale-dependent contour descriptors	**Area/perimeter ratio (mm)**	7.772 ± 0.762	6.985 ± 0.545	**1.19**	**< .001**
	LS ellipse RMSE (mm)	0.709 ± 0.206	0.657 ± 0.159	0.28	> .99
Scale-invariant shape indices	**FM index**	1.167 ± 0.080	1.160 ± 0.091	0.08	> .99
	Circularity	0.920 ± 0.030	0.926 ± 0.022	-0.22	> .99
	Eccentricity	1.178 ± 0.077	1.175 ± 0.083	0.04	> .99
	Compactness	3.697 ± 0.061	3.685 ± 0.045	0.23	> .99
Radial profile descriptors (scale-invariant)	**Radial CV**	0.073 ± 0.022	0.076 ± 0.021	-0.14	> .99
	Roughness	0.0035 ± 0.0014	0.0029 ± 0.0007	**0.62**	**< .001**
	L–R asymmetry	0.051 ± 0.025	0.056 ± 0.026	-0.19	> .99
	AP asymmetry	0.060 ± 0.023	0.064 ± 0.025	-0.16	> .99
Ellipse and polygon descriptors (scale-invariant)	**LS ellipse/FM area ratio**	1.013 ± 0.007	1.013 ± 0.007	-0.09	> .99
	Inscribed ellipse/FM area ratio	0.881 ± 0.032	0.878 ± 0.028	0.10	> .99
	Enclosing ellipse/FM area ratio	1.148 ± 0.048	1.147 ± 0.039	0.03	> .99
	Optimal polygon order (sides)	5.66 ± 0.92	5.58 ± 0.98	0.08	> .99

Values are mean ± SD. p values are Bonferroni-corrected for 209 comparisons. Bold values indicate statistically significant sex differences ($p < .05$, Bonferroni-corrected). Cohen's d: positive values indicate larger values in males. Scale categories are defined in the Methods; all 209 descriptors are listed in Table S2. Abbreviations: AP, anteroposterior; CV, coefficient of variation; FM, foramen magnum; L–R, left–right; LS, least-squares; RMSE, root mean square error; SD, standard deviation.

In contrast, the classical shape indices showed minimal dimorphism. The FM index did not differ between sexes (males: 1.167 ± 0.080, females: 1.160 ± 0.091; d = 0.08, Bonferroni-corrected $p > .99$), nor did circularity (0.920 ± 0.030 vs. 0.926 ± 0.022; d = −0.22), eccentricity (d = 0.04), or compactness (d = 0.23; all $p > .99$). The left–right and anteroposterior mirror-asymmetry indices likewise showed no sex differences (d = −0.19 and −0.16, respectively; both $p > .99$).

Feature Significance

Of the 209 descriptors, 74 (35.4%) differed significantly between the sexes after Bonferroni correction (Table 1; Table S2). The distribution of significant results followed the scale category of the descriptors closely. All 30 size descriptors were significant (d = 0.71–1.33), led by perimeter (d = 1.33), the least-squares circle radius (d = 1.28), the radial mean (d = 1.28), and the polygon areas (up to d = 1.26). Of the 63 scale-dependent contour

DOI: 10.5455/NOFOR.2026.08.012

descriptors, 43 (68.3%) were significant, including the area-to-perimeter ratio (d = 1.19) and 41 of the 44 polygon distance metrics. Of the 116 scale-invariant shape descriptors, only one was significant: boundary roughness, which was higher in males (0.0035 ± 0.0014 vs. 0.0029 ± 0.0007; d = 0.62, p < .001). None of the 89 scale-invariant EFT descriptors, none of the 15 ellipse, circle, and polygon area ratios, and none of the polygon elbow metrics differed between the sexes. The only significant EFT feature was the total spectral power (d = 1.25), a size descriptor.

Polygon Approximation Analysis

The Kneedle-based elbow detection identified the 6-sided polygon as the optimal approximation in 238 specimens (50.3%), followed by 5-sided (n = 88, 18.6%), 4-sided (n = 82, 17.3%), 7-sided (n = 59, 12.5%), and 8-sided polygons (n = 6, 1.3%) (mean 5.62 ± 0.95 sides; median and mode 6) (Figure 2; Table S4). The group-mean RMSE curves of males, females, and the pooled sample all had their elbow at six sides.

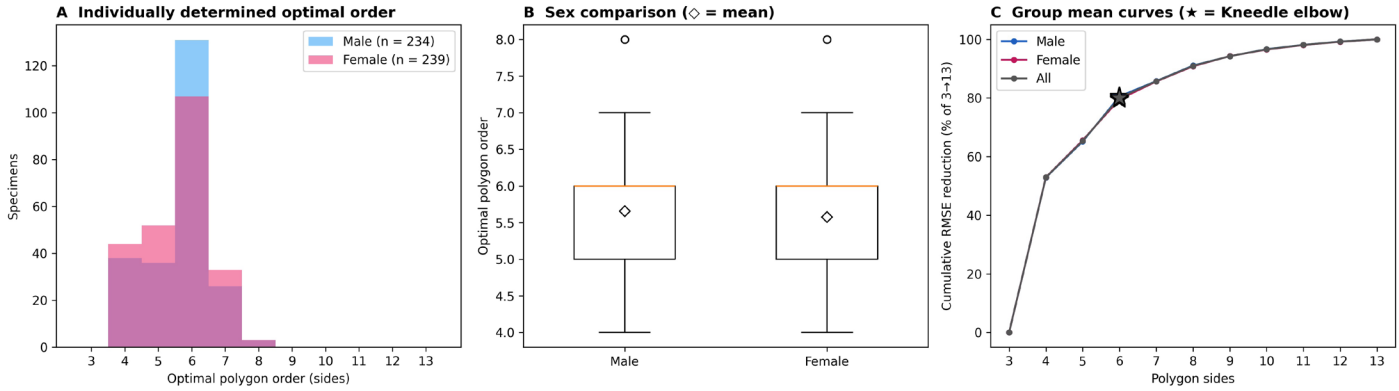

Figure 2. Distribution of individually determined optimal polygon orders (Kneedle algorithm; n = 473). (A) Histogram of optimal polygon order by sex (mode = 6 in both sexes). (B) Box plots by sex (◇ = mean). (C) Cumulative RMSE reduction of the group mean curves as a function of the number of polygon sides, expressed as a percentage of the total reduction from 3 to 13 sides; stars indicate the elbow points.

The transition from 3 to 4 sides provided the largest marginal RMSE improvement (50.3% for males, 50.5% for females), followed by the transition from 5 to 6 sides (38.5% and 35.5%, respectively) (Table 2; Figure 3). Beyond six sides, each additional vertex reduced the RMSE by at most 27.4%, with a declining trend (12.5% and 14.0% for the 12→13 transition). Up to eight sides, transitions to an odd-numbered polygon (4→5 and 6→7) yielded smaller gains than the subsequent transitions to an even-numbered polygon (5→6 and 7→8). The mean RMSE at the optimal polygon was 1.78 ± 0.66 mm, representing a mean reduction of 74.4% ± 8.7% relative to the 3-sided approximation. Males and females did not differ in optimal polygon order (male mean 5.66 ± 0.92, female mean 5.58 ± 0.98; d = 0.08, Mann–Whitney p = .31, uncorrected).

Table 2. Polygon approximation: mean RMSE and marginal improvement for successive polygon orders by sex (n = 473; 234 males, 239 females).

Transition	Male RMSE from (mm)	Male RMSE to (mm)	Male improvement (%)	Female RMSE from (mm)	Female RMSE to (mm)	Female improvement (%)
3→4	7.459	3.707	50.3	6.557	3.246	50.5
4→5	3.707	2.836	23.5	3.246	2.458	24.3
5→6	2.836	1.744	38.5	2.458	1.585	35.5
6→7	1.744	1.372	21.3	1.585	1.206	24.0
7→8	1.372	0.996	27.4	1.206	0.882	26.9
8→9	0.996	0.771	22.6	0.882	0.664	24.7
9→10	0.771	0.600	22.2	0.664	0.525	21.0
10→11	0.600	0.495	17.5	0.525	0.428	18.5
11→12	0.495	0.416	15.9	0.428	0.358	16.3
12→13	0.416	0.364	12.5	0.358	0.308	14.0

Each row shows the reduction in mean RMSE gained by adding one side, expressed as a percentage of the mean RMSE before the step. The largest improvement occurs at the 3→4 transition (≈50%) in both sexes, with a secondary peak at 5→6 (35.5–38.5%). Beyond six sides, marginal gains do not exceed 27.4%. Abbreviations: RMSE, root mean square error.

DOI: 10.5455/NOFOR.2026.08.012

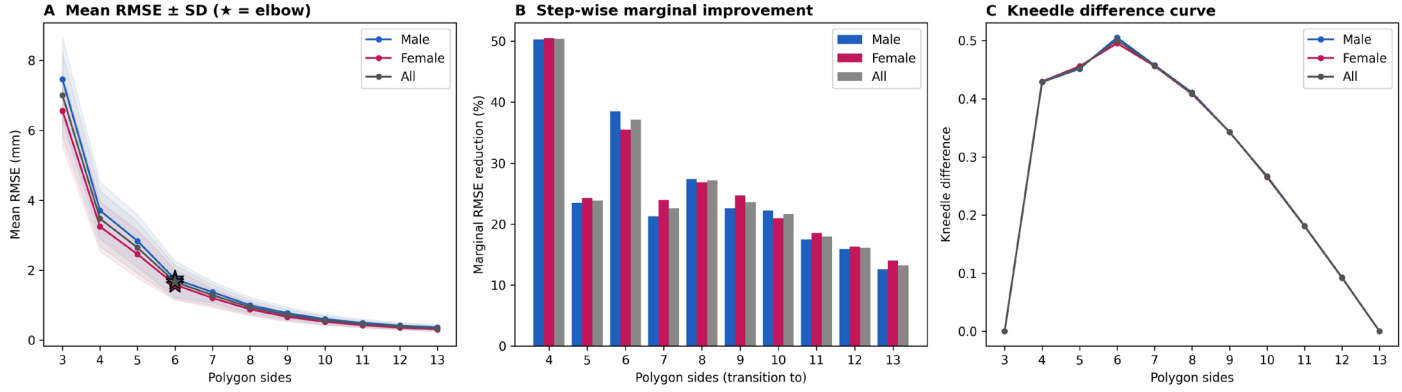

Figure 3. Polygon approximation elbow analysis. (A) Mean RMSE (shaded band, ± SD) versus number of polygon sides for males, females, and the pooled sample, with stars indicating the elbow of each group mean curve. (B) Step-wise marginal RMSE reduction for each additional side. (C) Kneedle difference curves (distance of the normalized RMSE curve below the chord joining its end points), peaking at six sides.

Sex Classification

Classification results for all feature sets and classifiers are given in Table 3 and Table S5 and summarized for LDA in Figure 4.

Size alone (S1) was the most accurate feature set: 77.1% ± 4.9% with LDA (pooled 95% CI 73.2–80.7%; sensitivity 75.2%, specificity 79.1%) and 77.4% ± 5.1% with SVM-RBF. The 116 scale-invariant shape descriptors (S2) achieved 64.0% ± 7.0% with LDA (95% CI 59.6–68.3%) and 58.6–62.4% with the other classifiers. The 63 scale-dependent contour descriptors (S4) reached 71.7–74.4%, approaching size alone, which is consistent with their dependence on scale; with LDA, S4 remained less accurate than S1 (McNemar $p = .016$). Neither combining size with all shape descriptors (S3: 70.1–77.0%) nor using all 209 descriptors (S5: 66.8–75.7%) exceeded the best size-only results (77.1% with LDA, 77.4% with SVM-RBF). With LDA, both were less accurate than S1 (McNemar $p = .015$ and $p < .001$, respectively; unadjusted), consistent with overfitting of high-dimensional feature sets.

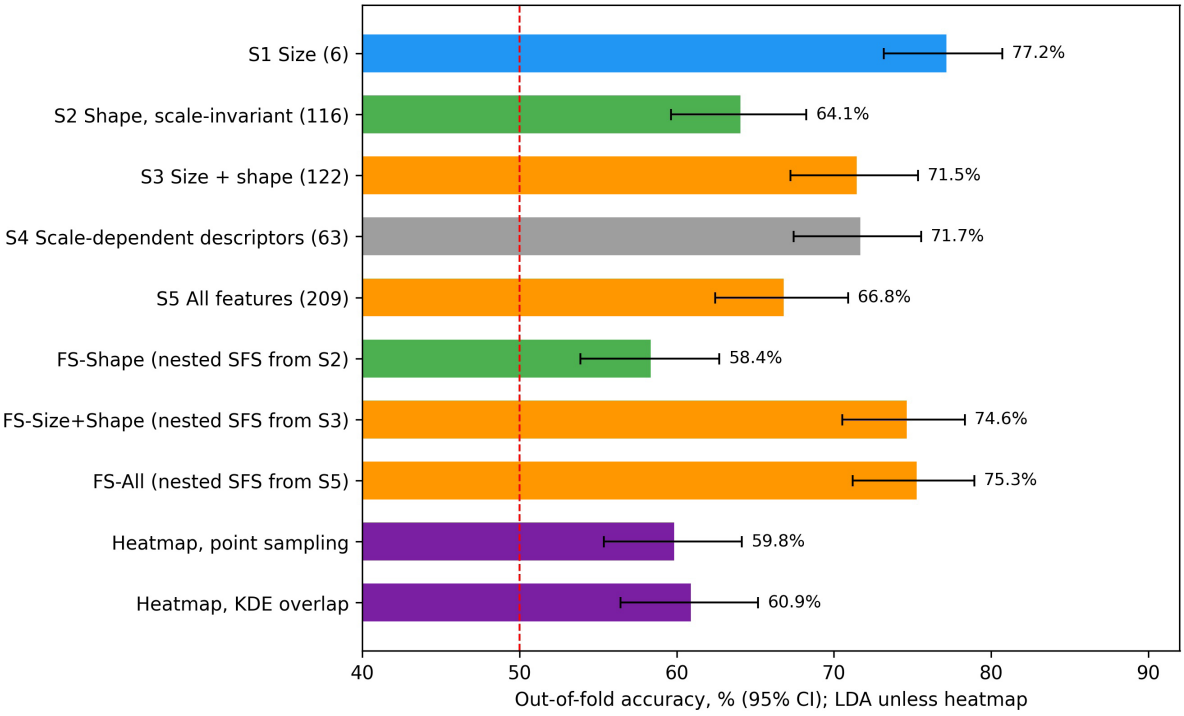

Figure 4. Sex classification performance across feature sets. Bars show the pooled out-of-fold accuracy under stratified 10-fold cross-validation with Wilson 95% CIs, using LDA for the feature sets and the respective scoring rule for the heatmap methods. FS models were evaluated with feature selection nested within the cross-validation. The dashed red line indicates chance level (50%). Blue, size; green, scale-invariant shape; grey, scale-dependent contour descriptors; orange, combined sets; purple, heatmap methods.

DOI: 10.5455/NOFOR.2026.08.012

Table 3. Sex classification performance across feature sets and classifiers under stratified 10-fold cross-validation (n = 473).

Feature set	Features (n)	Classifier	Accuracy, mean ± SD (%)	Pooled accuracy (95% CI) (%)	Sensitivity (%)	Specificity (%)
S1: Size	**6**	**LDA**	**77.1 ± 4.9**	**77.2 (73.2–80.7)**	**75.2**	**79.1**
		SVM-linear	75.9 ± 4.5	75.9 (71.8–79.5)	73.5	78.2
		SVM-RBF	77.4 ± 5.1	77.4 (73.4–80.9)	71.8	82.8
		RF	73.6 ± 5.2	73.6 (69.4–77.3)	70.9	76.2
S2: Shape (scale-invariant)	**116**	**LDA**	64.0 ± 7.0	64.1 (59.6–68.3)	60.7	67.4
		SVM-linear	59.4 ± 9.5	59.4 (54.9–63.7)	60.3	58.6
		SVM-RBF	58.6 ± 4.9	58.6 (54.1–62.9)	59.0	58.2
		RF	62.4 ± 5.6	62.4 (57.9–66.6)	66.7	58.2
S3: Size + shape	**122**	**LDA**	71.4 ± 7.4	71.5 (67.2–75.3)	67.5	75.3
		SVM-linear	70.1 ± 9.3	70.2 (65.9–74.1)	67.9	72.4
		SVM-RBF	75.2 ± 5.3	75.3 (71.2–78.9)	75.6	74.9
		RF	77.0 ± 4.2	77.0 (73.0–80.5)	73.9	79.9
S4: Scale-dependent descriptors	**63**	**LDA**	71.7 ± 7.9	71.7 (67.4–75.5)	69.2	74.1
		SVM-linear	74.0 ± 7.7	74.0 (69.9–77.7)	73.9	74.1
		SVM-RBF	74.4 ± 6.5	74.4 (70.3–78.1)	72.2	76.6
		RF	73.3 ± 4.9	73.4 (69.2–77.1)	72.6	74.1
S5: All descriptors	**209**	**LDA**	66.8 ± 8.2	66.8 (62.4–70.9)	66.2	67.4
		SVM-linear	69.1 ± 9.1	69.1 (64.8–73.1)	67.9	70.3
		SVM-RBF	74.6 ± 5.7	74.6 (70.5–78.3)	73.1	76.2
		RF	75.7 ± 4.0	75.7 (71.6–79.3)	73.1	78.2
FS-Shape (nested)	**2–9**	**LDA**	58.3 ± 8.4	58.4 (53.9–62.7)	48.7	67.8
		SVM-linear	59.6 ± 6.6	59.6 (55.1–63.9)	38.5	80.3
		SVM-RBF	58.5 ± 7.0	58.6 (54.1–62.9)	43.2	73.6
		RF	59.8 ± 7.1	59.8 (55.4–64.2)	55.6	64.0
FS-Size+Shape (nested)	**3–10**	**LDA**	74.6 ± 4.0	74.6 (70.5–78.3)	71.4	77.8
		SVM-linear	74.4 ± 4.5	74.4 (70.3–78.1)	70.9	77.8
		SVM-RBF	74.0 ± 3.9	74.0 (69.9–77.7)	72.2	75.7
		RF	74.4 ± 5.9	74.4 (70.3–78.1)	74.8	74.1
FS-All (nested)	**1–9**	**LDA**	75.3 ± 5.5	75.3 (71.2–78.9)	69.7	80.8
		SVM-linear	76.5 ± 4.4	76.5 (72.5–80.1)	71.4	81.6
		SVM-RBF	74.2 ± 3.5	74.2 (70.1–77.9)	69.2	79.1
		RF	74.6 ± 8.3	74.6 (70.5–78.3)	73.9	75.3
Heatmap: point sampling	**—**	**Density**	59.9 ± 6.7	59.8 (55.4–64.2)	62.8	56.9
Heatmap: KDE overlap	**—**	**Overlap**	60.9 ± 8.2	60.9 (56.4–65.2)	47.9	73.6

Accuracy is the mean ± SD across the 10 test folds; pooled accuracy is the proportion of all 473 out-of-fold predictions that were correct, with the Wilson 95% CI. Sensitivity, males correctly classified; specificity, females correctly classified. For FS models, SFS (LDA criterion) was repeated within each outer training set, and the number of features gives the range across the 10 outer folds. For KDE overlap, the kernel bandwidth was selected by inner cross-validation within each outer training set. All classifiers used fixed default hyperparameters. The bold row (size-only LDA) served as the reference for McNemar tests (Table S5). CI, confidence interval; FS, forward-selected; KDE, kernel density estimation; LDA, linear discriminant analysis; RBF, radial basis function; RF, random forest; SFS, sequential forward selection; SVM, support vector machine.

DOI: 10.5455/NOFOR.2026.08.012

Under nested cross-validation, FS-Shape achieved 58.3% ± 8.4% with LDA (95% CI 53.9–62.7%) and 58.5–59.8% with the other classifiers, significantly below size alone (McNemar $p < .001$). FS-Size+Shape achieved 74.6% ± 4.0% and FS-All 75.3% ± 5.5%. Neither differed significantly from S1 (McNemar $p = .08$ and $p = .21$), and both point estimates were lower. During selection, the inner cross-validation accuracy of the selected subsets averaged 66.6% (FS-Shape), 80.1% (FS-Size+Shape), and 80.1% (FS-All), exceeding the corresponding nested estimates by 8.3, 5.5, and 4.9 percentage points, respectively (Table 4). This optimism arises whenever features are selected and evaluated on the same data. In FS-Shape and FS-Size+Shape, one feature was selected first in every outer fold: roughness and perimeter, respectively. In FS-All, the first feature was always a size measure (perimeter in 5 folds and a polygon area in 5). All other features were selected in at most 3 of the 10 folds, indicating that beyond these, the selected features were unstable and largely fold-specific (Table 4; Table S6).

The estimates were stable across repeated fold partitions (LDA, mean ± SD across partitions): S1 76.6% ± 0.4%, S2 63.5% ± 1.0%, S3 71.5% ± 1.0%, S4 72.8% ± 1.9%, S5 67.3% ± 1.1%, FS-Shape 60.5% ± 1.1%, FS-Size+Shape 75.9% ± 0.5%, and FS-All 75.0% ± 1.1% (Table S5).

Table 4. Sequential forward feature selection nested within 10-fold cross-validation: selection-stage (inner) versus unbiased (outer) accuracy, and features selected consistently across outer folds (n = 473).

Model	Candidate features (n)	Features selected per outer fold, median (range)	Inner CV accuracy during selection, mean (range) (%)	Nested outer CV accuracy, LDA, mean ± SD (%)	Features selected in ≥3 of 10 outer folds (n of folds)
FS-Shape (pool: S2)	116	5 (2–9)	66.6 (64.5–70.0)	58.3 ± 8.4	Roughness (10); EFT harmonic 16, d (3)
FS-Size+Shape (pool: S3)	122	5.5 (3–10)	80.1 (78.8–81.5)	74.6 ± 4.0	Perimeter (10); EFT harmonic 19, a (3); 9-sided polygon area ratio (3); EFT harmonic 3, c (3)
FS-All (pool: S5)	209	4 (1–9)	80.1 (78.8–82.4)	75.3 ± 5.5	Perimeter (5); 11-sided polygon area (3)

In each outer fold, forward selection (LDA, 10-fold inner cross-validation, stopping when no feature improved inner accuracy) was performed on the nine training folds only, and the selected subset was evaluated on the held-out fold. The inner accuracy is a selection-stage estimate of the kind obtained without nesting; its difference from the outer accuracy approximates the optimism of non-nested selection (the non-nested full-sample path is shown in Table S6). Selected features for every outer fold, selection frequencies, and a descriptive full-sample selection path are given in Table S6. Abbreviations: CV, cross-validation; EFT, elliptic Fourier transform; LDA, linear discriminant analysis.

Heatmap-Based Density Classification

With Procrustes-aligned contours and training-fold-only templates, point-sampling heatmap classification achieved 59.9% ± 6.7% (pooled 95% CI 55.4–64.2%; sensitivity 62.8%, specificity 56.9%), and KDE overlap scoring achieved 60.9% ± 8.2% (95% CI 56.4–65.2%; sensitivity 47.9%, specificity 73.6%). The inner cross-validation selected bandwidths between 3 and 9 grid pixels. Both methods performed better than chance but significantly worse than size alone (McNemar $p < .001$ for both), and at a level similar to that of the scale-invariant shape descriptors. This is expected, because Procrustes alignment removes size, and the aligned male and female outlines were nearly superimposable (Figure 1).

Within-Sex Shape Structure

For both sexes, BIC was lowest for a single mixture component (males: −6437 for $k = 1$ vs. −5852, −5105, and −5043 for $k = 2$–4; females: −6413 vs. −5680, −5021, and −4200) (Table 5; Table S7; Figure 5). The alternative criteria agreed. Silhouette coefficients for $k = 2$–4 ranged from 0.04 to 0.08, indicating no separation between the fitted components. Partitions were not reproducible under bootstrap resampling (mean ARI 0.09–0.17; 1 indicates identical partitions and 0 chance agreement). The first two principal components of shape showed a single continuous cloud in each sex (Figure S3). Thus, the data provide no evidence of discrete within-sex FM subtypes, and no subtypes were characterized further.

Dispersion analyses gave a nuanced picture (Table 5). Overall outline shape was not more variable in males: the mean Procrustes distance from the sex-specific consensus was slightly smaller in males than in females (0.0516 vs. 0.0554; permutation $p = .038$) (Figure 6). However, indices of boundary regularity varied more in males: circularity (SD 0.030 vs. 0.022), compactness (SD 0.061 vs. 0.045), and roughness (SD 0.0014 vs. 0.0007) (Brown–Forsythe, Bonferroni-corrected $p < .001$ for all three). Because compactness is a monotonic transform of circularity, these represent two independent findings. The remaining five indices showed no difference in variance.

DOI: 10.5455/NOFOR.2026.08.012

Table 5. Within-sex shape structure: Gaussian mixture model selection and comparison of shape dispersion between sexes (n = 473; 234 males, 239 females).

A. Gaussian mixture model selection (20 principal components of Procrustes coordinates)

Sex	k	BIC	Silhouette	Component sizes	Bootstrap ARI, mean ± SD
Males	**1**	**-6437**	—	234	—
	2	-5852	0.081	159/75	0.15 ± 0.17
	3	-5105	0.072	120/67/47	0.16 ± 0.10
	4	-5043	0.083	134/48/42/10	0.12 ± 0.08
Females	**1**	**-6413**	—	239	—
	2	-5680	0.057	165/74	0.09 ± 0.12
	3	-5021	0.082	139/58/42	0.17 ± 0.09
	4	-4200	0.044	122/49/38/30	0.12 ± 0.05

B. Dispersion comparison between sexes

Measure	Male	Female	Test	p
Procrustes distance to sex consensus, mean	0.0516	0.0554	Permutation (10,000)	.038
FM index, SD	0.080	0.091	Brown–Forsythe	.66
Circularity, SD	0.030	0.022	Brown–Forsythe	**<.001**
Eccentricity, SD	0.077	0.083	Brown–Forsythe	>.99
Compactness, SD	0.061	0.045	Brown–Forsythe	**<.001**
Radial CV, SD	0.022	0.021	Brown–Forsythe	>.99
Roughness, SD	0.0014	0.0007	Brown–Forsythe	**<.001**
L–R asymmetry, SD	0.025	0.026	Brown–Forsythe	>.99
AP asymmetry, SD	0.023	0.025	Brown–Forsythe	>.99

Panel A: lower BIC indicates better support; the minimum (bold) occurs at k = 1 in both sexes. Silhouette values near 0 indicate no separation between components, and bootstrap ARI values near 0 indicate partitions that are not reproducible under resampling (1 = identical partitions). Panel B: Procrustes distance is the root-mean-square distance of each aligned specimen from its sex-specific consensus. p values for the Brown–Forsythe tests are Bonferroni-corrected for eight comparisons; bold values indicate significance. Abbreviations: ARI, adjusted Rand index; BIC, Bayesian information criterion; CV, coefficient of variation; FM, foramen magnum; L–R, left–right; AP, anteroposterior; SD, standard deviation.

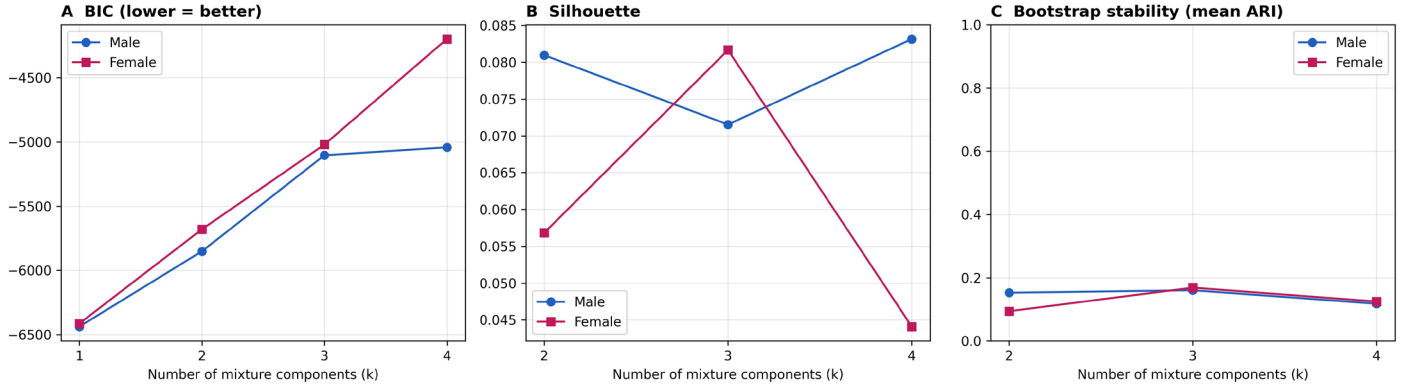

Figure 5. Selection of the number of Gaussian mixture components for within-sex shape structure (first 20 principal components of the Procrustes coordinates). (A) BIC for k = 1–4; lower values indicate better support, and BIC is minimal at k = 1 in both sexes. (B) Silhouette coefficient for k = 2–4. (C) Mean ARI between the full-sample partition and partitions refitted to 200 bootstrap resamples.

DOI: 10.5455/NOFOR.2026.08.012

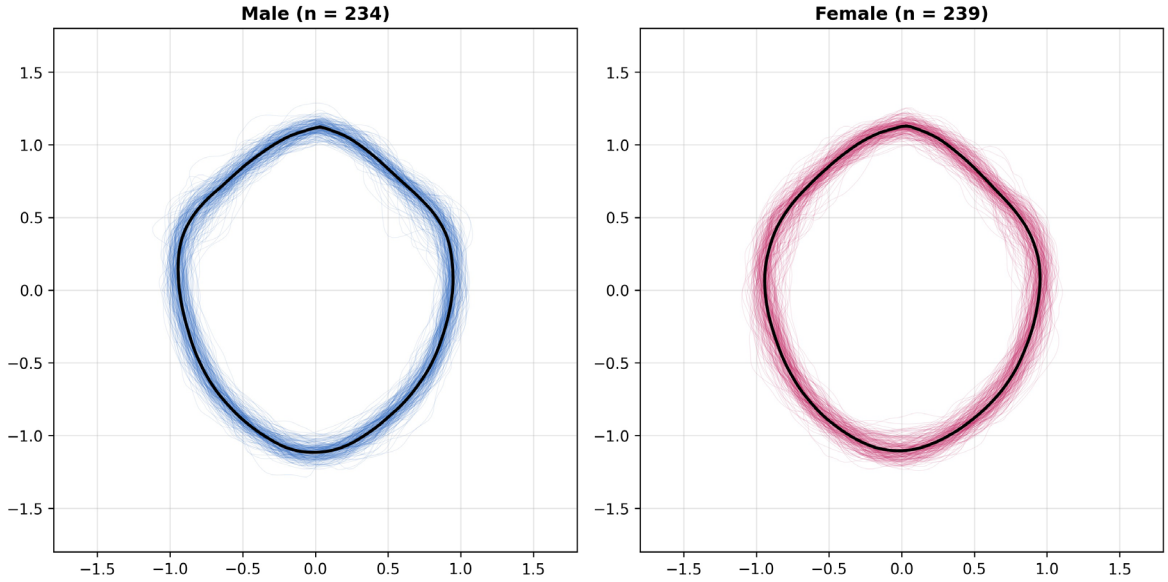

Figure 6. Within-sex variation of foramen magnum outline shape. Individual Procrustes-aligned contours (thin lines) and the sex-specific consensus outline (black) for males (left, n = 234) and females (right, n = 239). No visually distinct groups are apparent.

Sensitivity Analysis: Contour Quality Control

Without correction of the 21 self-intersecting contours, the size-based and combined classification results were essentially unchanged (S1 LDA 75.7%; FS-Size+Shape 76.3%). The forward-selected shape model was lower (52.3%), and the number of significant univariate descriptors was 76 of 209, again with roughness as the only significant scale-invariant descriptor (Table S1). Within-sex clustering was the analysis most affected. With the uncorrected contours, BIC no longer favoured a single component. In males it favoured three components: one contained two strongly distorted contours (mean circularity 0.52), and a second contained 19 irregular outlines (mean circularity 0.83), 7 of which had tracing loops. In females it favoured two components (204 and 35 specimens). The overall shape dispersion of males then appeared larger than that of females (0.0625 vs. 0.0580; p = .36). A handful of tracing artifacts can thus create an apparent "atypical" cluster, which underlines the need for topological quality control before unsupervised analyses.

Summary of Key Findings

The FM shows clear sexual dimorphism, but this dimorphism is carried by size. Males have FM perimeters approximately 12% longer and areas approximately 25% larger than females, yielding large effect sizes ($d \approx 1.2$–1.3). Once size is removed, male and female FM outlines are nearly indistinguishable. Only boundary roughness differed between the sexes, and scale-invariant shape descriptors classified sex with 58–64% accuracy under unbiased validation. Adding shape descriptors to size did not improve classification. In practical forensic terms, the six classical size measurements allow correct sex classification in about three of four individuals in this sample (77.1%; females slightly better classified than males). When size cannot be measured reliably, the FM outline alone offers little discriminative value.

DISCUSSION

Principal Findings and Novelty

This study presents a comprehensive computational geometric morphometric analysis of the FM that integrates multiple complementary outline-based approaches to jointly characterize FM shape and its relationship to biological sex: EFT, systematic polygon approximation, ellipse and circle fitting, radial distance profiling, and Procrustes-based density heatmaps. While conventional FM studies have relied on a small number of linear measurements and subjective shape classification [2,6,11], the present work extracted 209 quantitative descriptors per specimen from automatically detected and quality-controlled contour coordinates. The descriptors were explicitly classified as size, scale-dependent, or scale-invariant. This enabled an objective, reproducible, high-dimensional evaluation of FM morphology in which size and shape information are kept apart.

Our most consequential finding is the quantitative demonstration that FM sexual dimorphism is almost entirely size-mediated. All size descriptors differed strongly between the sexes, whereas only 1 of 116 scale-invariant shape descriptors did. Scale-invariant shape descriptors and Procrustes-based heatmaps classified sex only modestly above chance (58–64%), and adding shape information did not improve size-based classification. Procrustes alignment removes isometric size but not size-related (allometric) shape change. The near-superimposable male and female consensus outlines (Figure 1) therefore also indicate that any allometric component of FM shape is small.

Size Dominates FM Sexual Dimorphism

All six classical size measurements differed significantly between the sexes with large effect sizes (Cohen's $d = 1.13$–1.33), with perimeter showing the strongest effect ($d = 1.33$). Males had FM

DOI: 10.5455/NOFOR.2026.08.012

perimeters approximately 12% longer and areas approximately 25% larger than females. These findings are consistent with the broader FM literature. Gapert et al. [3] reported that FM width and area were among the best univariate discriminating parameters in a British sample, achieving a maximum multivariate accuracy of 70.3%. Madadin et al. [4] similarly found that all FM dimensions were significantly larger in males in a Saudi population, with width yielding the highest univariate accuracy of 65.0%. A recent systematic review and meta-analysis by Fernandes et al. [9] confirmed that FM dimensions are consistently larger in males across populations, although classification accuracy varies substantially by population. The size-only accuracy of 77.1% observed here lies within the reported range of 62.5–86.7% for conventional FM morphometry [3,47,11,36].

In contrast, shape indices showed negligible dimorphism in our sample. The FM index, the most widely used shape descriptor in the literature, did not differ between the sexes (d = 0.08), and circularity, compactness, and left–right and anteroposterior asymmetry showed only small, non-significant differences ($|d| \leq 0.23$). These results align with Toneva et al. [6], who found that FM shape did not provide substantial sex discrimination in a Bulgarian population, and with our previous study [11], which reported that two FM shape indices (the H/W index and the FM index) had no significant sex difference (p = .62 and p = .44, respectively) and yielded univariate accuracies of only 51.9% and 52.8%.

Scale-Invariant Shape Adds No Discriminative Value

Several aspects of the present analysis bear on the question of whether shape contributes to sex estimation, and they point in the same direction. First, the distinction between scale-invariant and scale-dependent descriptors matters. Descriptors such as the area-to-perimeter ratio or the RMSE of a fitted ellipse or polygon quantify outline irregularity but are expressed in millimetres. They therefore carry size information: the area-to-perimeter ratio, for example, differed between the sexes almost as strongly as the diameters (d = 1.19). Consistent with this, the 63 scale-dependent descriptors classified sex almost as well as the classical size measurements (71.7–74.4%). A feature set labelled "shape" that includes such descriptors can therefore appear discriminative because of size alone.

Second, the way performance is estimated after feature selection matters. When features are selected by cross-validated accuracy and the same cross-validation is then used to report performance, the estimate is optimistically biased, because the selection has already exploited chance associations in the evaluation data [29,30]. With selection nested within the outer cross-validation, this optimism amounted to approximately 5–8 percentage points in the present data (Table 4). The unbiased estimates were 58.3% for forward-selected shape features and 74.6–75.3% for forward-selected combined models, neither better than size alone. The instability of the selected features across outer folds, with no feature other than roughness or perimeter chosen in more than 3 of 10 folds, further indicates that the apparent gains of individual shape descriptors in the full sample reflected noise rather than a reproducible signal.

Third, the single scale-invariant descriptor that did differ between the sexes, boundary roughness, deserves a cautious interpretation. The higher roughness of male contours agrees with the classical qualitative description of a "rougher" male FM [2,12], and roughness was selected first in every outer fold of the shape-only model. However, roughness was also positively correlated with the size of the contour in pixels (Spearman ρ = 0.33), which suggests a partial dependence on image resolution. In an exploratory linear regression of log-roughness on sex and log contour perimeter in pixels, the sex coefficient was reduced by about 40% but remained significant (p = .001). A modest true sex difference in boundary texture is therefore plausible, but it cannot be separated completely from digitization effects in routine CT data, and on its own it provides little discrimination.

These findings have practical implications for forensic anthropology. The FM contributes to sex estimation through its dimensions, and computational shape analysis of its outline is unlikely to improve estimation when the dimensions can be measured. It has been suggested that shape could still contribute when size measurements are compromised, for example in fire-damaged or fragmentary remains [3,21]. The present results indicate that the FM outline alone would provide only weak evidence in such cases.

From Subjective Categories to Objective Shape Quantification

A particularly notable contribution of this work concerns the long-standing problem of FM shape classification. Numerous studies have attempted to characterize FM morphology using categorical descriptors such as "oval," "round," "tetragonal," "pentagonal," "egg-shaped," "hexagonal," and "irregular" [2,12,13]. As Zdilla et al. [14] noted in their comprehensive review, these shape categories rely on ambiguous terminology that is inconsistent between studies, observer-dependent, and difficult to reproduce. Indeed, visual shape classification is inherently subjective: the same foramen may be classified as "oval" by one observer and "egg-shaped" by another, and studies vary widely in how many and which shape categories they employ [14].

Our polygon approximation analysis offers a reproducible, quantitative alternative. Polygons of 3 through 13 sides were fitted to each FM contour, and the elbow point was identified at which additional sides yield diminishing improvements in fit. By this criterion, a six-sided polygon was the most parsimonious approximation of the typical FM outline (mode = 6 in both sexes, median = 6, mean = 5.62 ± 0.95). Individual optima spanned 4 to 8 sides (Figure 2). This result should be understood as an algorithmic approximation of outline complexity, not as an anatomical characterization. The polygon vertices are determined by angular sectoring from the centroid rather than by anatomical landmarks, and the optimal order depends on the fitting procedure and the elbow criterion. The alternating pattern of marginal gains up to eight sides, with larger improvements at the transitions to even-numbered polygons, also suggests that the roughly centrosymmetric FM outline is approximated more

DOI: 10.5455/NOFOR.2026.08.012

efficiently by polygons with opposing vertex pairs. Within these constraints, the absence of a sex difference in optimal polygon order (male mean = 5.66, female mean = 5.58; uncorrected p = .31) indicates that the geometric complexity of the FM outline, as captured by this approach, does not differ between the sexes.

Similarly, the EFT decomposed each FM contour into 20 harmonics, enabling objective quantification of boundary features at multiple spatial scales. None of the 89 scale-invariant EFT descriptors (normalized coefficients, harmonic power ratios, higher-harmonic energy ratio, and symmetry index) differed between the sexes. The only significant EFT feature, total spectral power, is a size measure. Thus, the frequency-domain description of the outline did not differ between the sexes; the only sex difference in boundary texture detected in this study was captured by the roughness index, as discussed above.

Heatmap-Based Classification: A Weak But Informative Signal

The density heatmap approach, particularly the KDE overlap scoring method, represents a conceptually distinct strategy that bypasses discrete feature extraction entirely, instead comparing continuous spatial distributions of boundary points. Because heatmap classification operates on Procrustes-aligned contours from which size has been removed, its modest performance (approximately 60%) provides an independent, feature-free estimate of the sex information contained in FM outline shape. This estimate was concordant with that obtained from the scale-invariant shape descriptors (58–64%). The two approaches therefore converge on the same conclusion: outline shape carries a weak sex signal that is small compared with the signal carried by size.

We note that this approach may prove more effective for anatomical structures where shape differences between groups are more pronounced. For example, EFA applied to the greater sciatic notch, a structure with well-documented qualitative shape dimorphism, has achieved classification rates exceeding 80% [37]. The heatmap methodology developed here could be readily extended to such structures, where it may provide useful discrimination precisely because the shape differences are larger.

Within-Sex Morphological Variation

Unsupervised mixture modelling of the aligned outlines provided no support for discrete FM subtypes in either sex. BIC favoured a single component, the silhouette coefficients were close to zero, and bootstrap partitions were not reproducible. Because a full-covariance mixture in 20 dimensions adds about 230 parameters per component, BIC penalizes additional components heavily at this sample size; the conclusion therefore rests equally on the near-zero silhouette coefficients and the lack of reproducibility under resampling. Within-sex variation of FM shape is therefore better described by continuous shape axes (for example, principal components) than by categorical types. This conclusion parallels the problems of the traditional visual shape categories discussed above. The sensitivity analysis further showed that a small number of contour tracing artifacts can generate an apparent minority cluster of irregular male outlines. Tracing artifacts are a practical hazard for automated outline studies, and unsupervised analyses are particularly vulnerable to them.

Regarding heterogeneity, males showed greater variance in indices of boundary regularity (circularity and the closely related compactness, and roughness) but not greater dispersion of overall outline shape, which was in fact marginally greater in females. Greater male variability in FM dimensions has also been reported in our previous study [11], where males showed higher SDs in most FM measurements (e.g., FM area SD: 121.5 mm^2 vs. 101.6 mm^2; perimeter SD: 8.15 mm vs. 6.8 mm), and the same pattern was seen for size in the present sample. For shape, the evidence of greater male heterogeneity is limited to boundary texture. Because roughness also shows a partial dependence on resolution, this finding should be regarded as preliminary. The slightly higher specificity than sensitivity of the size-based models (79.1% vs. 75.2% with LDA) is consistent with the wider male size distribution overlapping more with the female distribution.

Methodological Contributions and Broader Applicability

Beyond the specific findings regarding the FM, this study demonstrates the feasibility of applying a comprehensive, automated geometric morphometric pipeline to skeletal structures that lack well-defined anatomical landmarks. The FM is particularly challenging for traditional landmark-based morphometrics because, apart from basion and opisthion, its margin is a smooth, continuous curve without further discrete, reproducible points that would qualify as Type I or Type II landmarks [38]. Our approach treats the entire contour as the unit of analysis through resampling, Fourier decomposition, polygon approximation, and density mapping, and thereby circumvents this limitation.

Three methodological lessons from this study may be of general relevance for outline-based sex estimation. First, descriptors should be classified by their physical dimension before they are interpreted as "shape," because irregularity measures expressed in length units carry size. Second, any data-driven feature selection or tuning must be nested within the validation procedure; otherwise, performance estimates for high-dimensional outline descriptors are optimistic. Third, automated contours should undergo topological quality control, and the point correspondence (starting point and direction) must be standardized before coordinates are aligned or clustered.

Caple et al. [39] highlighted the potential of EFA in forensic anthropology, noting that outline-based methods can capture a large amount of shape information, in contrast to landmark approaches where information falling between landmarks fails to be acquired. Similarly, Kilmer and Garvin [37] demonstrated that EFA of skeletal outlines can support qualitative descriptions of sex differences with objective, quantitative data. Our work extends this paradigm by combining EFA with multiple additional geometric analysis methods, providing a more complete morphometric characterization than any single technique alone.

DOI: 10.5455/NOFOR.2026.08.012

The complete analysis pipeline, from contour quality control of the extracted coordinates through feature computation, statistical testing, classification, and clustering, is implemented as a single, self-contained, publicly available script. This design facilitates direct application to other skeletal structures where shape variation has traditionally been described only qualitatively, such as the obturator foramen [37,40] or piriform aperture. The objective quantification of shapes that are currently assessed visually (e.g., "oval vs. triangular obturator foramen") could enable more reproducible forensic assessments and contribute to a more precise anatomical vocabulary for morphological description.

Comparison with the Source Study

Our dataset derives from the same CT image collection used in our previous study [11], which reported an accuracy of 86.7% using linear discriminant function analysis with leave-one-out cross-validation on the full 720-case sample, and 88.2% using artificial neural networks. The lower size-based accuracy in the present study (77.1%) has several possible explanations. First, the measurements differ: the source study used manual measurements on three-dimensional reformations, whereas the present study used automatically extracted contours from a single axial slice. Second, 247 cases (34.3%) were excluded here because their contour quality was insufficient for automated edge detection. The proportion excluded did not differ between the sexes (35.0% vs. 33.6%), but exclusion was based on image quality and may not be independent of patient-related factors such as head positioning, motion, or age. If excluded and included cases differed systematically in FM size or in the separation between the sexes, the included subsample may be less dimorphic than the source population. The individual exclusion reasons were not recorded in a form that allows this to be examined, and the present estimates therefore apply to CT examinations of sufficient quality for automated boundary extraction. Third, the present study employed 10-fold rather than leave-one-out cross-validation, which can yield slightly different estimates depending on fold composition.

Limitations

Several limitations should be acknowledged. First, the study sample was drawn from a single Eastern Turkish population, and the population specificity of FM morphometry is well established [4,9]. As noted in our previous study [11], even within Türkiye, discriminant formulas from western Turkish samples showed variable performance when applied to the eastern Turkish sample. No external validation was performed. The cross-validated accuracies reported here estimate performance for new individuals from the same population and imaging protocol, and the classification models may not generalize to other populations or scanners. Second, the accuracy of the automated contour extraction was not validated against an independent reference such as manual expert tracing. Visual inspection and an automated topological check detected and corrected gross tracing errors, but subtler systematic deviations of the automated boundary from the true bony margin cannot be excluded. Third, contour extraction relied on 2D axial CT slices rather than 3D surface reconstructions, potentially missing three-dimensional shape information such as FM depth, curvature, and condylar morphology. The AP and transverse diameters and the asymmetry indices were also defined relative to the image axes and are therefore sensitive to head rotation in the axial plane. Fourth, the exclusion of 247 of 720 cases because of insufficient image quality introduces a potential selection bias, as discussed above. Fifth, age-related variation was not explicitly modeled. Although our previous study found no significant correlation between FM measurements and age apart from a weak trend in perimeter [11], age effects on shape features remain unexplored. Sixth, the contour points were treated as semilandmarks with a common starting point but without sliding, and fine-scale descriptors such as roughness are partly sensitive to image resolution. The reproducibility of the identification of the slice with the widest FM cross-section was also not assessed. Finally, the heatmap and KDE methods employed a fixed grid resolution, and the classifiers used fixed default hyperparameters. Neither choice was optimized, which avoids optimistic bias but may leave some performance unexploited.

Future Directions

The computational framework developed here opens several avenues for future research. Application to skeletal structures with more pronounced qualitative shape dimorphism, such as the greater sciatic notch, obturator foramen, or frontal sinus, could test whether the polygon approximation and heatmap methods achieve higher discriminative accuracy when shape differences are more substantial. Multi-population studies using the same pipeline, together with external validation on independent samples and scanners, would clarify which FM descriptors, if any, transcend population boundaries. Validation of the automated boundaries against expert manual tracing, and integration of 3D surface data enabled by modern CT reconstruction capabilities, could capture additional morphological information that is inaccessible from 2D contours.

CONCLUSION

This study demonstrates that the sexual dimorphism of the FM is mediated by overall size rather than by outline shape. Among 209 CT-derived descriptors, all 30 size descriptors differed strongly between the sexes, whereas only 1 of 116 scale-invariant shape descriptors did. Perimeter emerged as the single strongest discriminator (Cohen's $d = 1.33$), and six classical size measurements achieved 77.1% classification accuracy. Scale-invariant shape features achieved only 58–64% (all features, or forward-selected under nested validation), and adding them to size did not improve classification. Apparent gains from optimized combinations of "shape" features can arise from descriptors that retain scale information and from feature selection that is not separated from performance estimation. Both pitfalls should be avoided in future outline-based sex estimation studies.

From a methodological standpoint, this study replaces the subjective and non-reproducible categorical shape classification

DOI: 10.5455/NOFOR.2026.08.012

that has prevailed in the FM literature with an automated, deterministic pipeline, openly available from the extracted boundary coordinates onward, that computes objective, continuous descriptors of FM contours from standard axial CT images. The seven complementary analyses provide a multiscale characterization of the FM boundary that, for a given contour, is independent of the observer: classical morphometry; least-squares, inscribed, and enclosing ellipse fitting; polygon approximation; polygon elbow analysis; normalized EFT; radial distance profiling; and Procrustes-based density heatmaps. Polygon approximation indicated that a six-sided polygon is the most parsimonious geometric approximation of the FM outline in both sexes. This algorithmic result offers a reproducible alternative to the inconsistent subjective shape taxonomies. Unsupervised modelling provided no evidence of discrete within-sex FM subtypes.

In practical forensic terms, the FM contributes to sex estimation through its dimensions, correctly classifying approximately three of four individuals in this sample, whereas its outline shape adds little. Future studies should test the pipeline on external samples, validate automated boundaries against expert tracing, extend it to three-dimensional surface data, and apply it to anatomical structures in which shape dimorphism is expected to be more pronounced, such as the greater sciatic notch and the pelvic inlet.

Conflict of Interests
The authors declare that they have no competing interests. One of the authors (OC) is the Editor-in-Chief of the journal and was not involved in the editorial evaluation or peer-review process of this manuscript; this is disclosed on the title page in accordance with the journal's instructions for authors.

Financial Disclosure
The authors declare that they have received no financial support for the study.

Ethical Approval
The original study protocol received approval from the Non-Interventional Clinical Research Ethics Committee of Van Yüzüncü Yıl University Faculty of Medicine (decision number: 02, dated: 19.02.2017) and was conducted in accordance with the Declaration of Helsinki and the principles of Good Clinical Practice. Since the present study represents a secondary analysis of the same anonymized imaging dataset using computational geometric methods, no additional ethical approval was required. Informed consent was waived due to the retrospective design and full anonymization of all data.

Patient Consent
Informed consent was waived because of the retrospective design of the study and the full anonymization of all data. The manuscript contains no identifiable patient information, images, or personal data; therefore, consent for publication was not required.

References

1. Tubbs RS, Griessenauer CJ, Loukas M, Shoja MM, Cohen-Gadol AA. Morphometric analysis of the foramen magnum: an anatomic study. Neurosurgery. 2010;66:385-8.
2. Murshed KA, Çiçekcibaşi AE, Tuncer I. Morphometric evaluation of the foramen magnum and variations in its shape: a study on computerized tomographic images of normal adults. Turk J Med Sci. 2003;33:301-6.
3. Gapert R, Black S, Last J. Sex determination from the foramen magnum: discriminant function analysis in an eighteenth and nineteenth century British sample. Int J Legal Med. 2009;123:25-33.
4. Madadin M, Menezes RG, Al Saif HS, Abu Alola H, Al Muhanna A, Gullenpet AH, et al. Morphometric evaluation of the foramen magnum for sex determination: a study from Saudi Arabia. J Forensic Leg Med. 2017;46:66-71.
5. Aljarrah K, Packirisamy V, Al Anazi N, Nayak SB. Morphometric analysis of foramen magnum and occipital condyle using CT images for sex determination in a Saudi Arabian population. Morphologie. 2022;106:260-70.
6. Toneva D, Nikolova S, Harizanov S, Georgiev I, Zlatareva D, Hadjidekov V, et al. Sex estimation by size and shape of foramen magnum based on CT imaging. Leg Med (Tokyo). 2018;35:50-60.
7. Uthman AT, Al-Rawi NH, Al-Timimi JF. Evaluation of foramen magnum in gender determination using helical CT scanning. Dentomaxillofac Radiol. 2012;41:197-202.
8. Akay G, Güngör K, Peker İ. Morphometric analysis of the foramen magnum using cone beam computed tomography. Turk J Med Sci. 2017;47:1715-22.
9. Fernandes ALVC, Shetti A, Lagali-Jirge V, Keluskar V. Accuracy of sex estimation by morphometric evaluation of foramen magnum using computed tomography - a systematic review and meta-analysis. Forensic Sci Med Pathol. 2024;20:268-79.
10. Meral O, Toklu BB, Meydan R, Kaya A, Karadayı B, Acar T. Sex estimation from foramen magnum parameters in adult Turkish population: a computed tomography study. Leg Med (Tokyo). 2020;47:101775.
11. Kartal E, Etli Y, Asirdizer M, Hekimoglu Y, Keskin S, Demir U, et al. Sex estimation using foramen magnum measurements, discriminant analyses and artificial neural networks on an eastern Turkish population sample. Leg Med (Tokyo). 2022;59:102143.
12. Chethan P, Prakash KG, Murlimanju BV, Prashanth KU, Prabhu LV, Saralaya VV, et al. Morphological analysis and morphometry of the foramen magnum: an anatomical investigation. Turk Neurosurg. 2012;22:416-9.
13. Vinutha SP, Suresh V, Shubha R. Discriminant function analysis of foramen magnum variables in South Indian population: a study of computerised tomographic images. Anat Res Int. 2018;2018:2056291.
14. Zdilla MJ, Russell ML, Bliss KN, Mangus KR, Koons AW. The size and shape of the foramen magnum in man. J Craniovertebr Junction Spine. 2017;8:205-21.
15. Kuhl FP, Giardina CR. Elliptic Fourier features of a closed contour. Comput Graph Image Process. 1982;18:236-58.
16. Haines AJ, Crampton JS. Improvements to the method of Fourier shape analysis as applied in morphometric studies. Palaeontology. 2000;43:765-83.
17. Samuel OM, Casanova PM, Olopade JO. Elliptical Fourier descriptors of outline and morphological analysis in caudal view of foramen magnum of the tropical raccoon (Procyon cancrivorus) (Linnaeus, 1758). Morphologie. 2018;102:31-40.
18. Hayashizaki Y, Usui A, Hosokai Y, Sakai J, Funayama M. Sex determination of the pelvis using Fourier analysis of postmortem CT images. Forensic Sci Int. 2015;246:122.e1-9.
19. Gonzalez PN, Bernal V, Perez SI. Geometric morphometric approach to sex estimation of human pelvis. Forensic Sci Int. 2009;189:68-74.
20. Etli Y, Kartal E, Demir U, Etli E, Hekimoğlu Y, Aşirdizer M. Sex determination using image processing reveals elliptical features of the female pelvic inlet. Sci Rep. 2025;15:33800.
21. Krishan K, Chatterjee PM, Kanchan T, Kaur S, Baryah N, Singh RK. A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. Forensic Sci Int. 2016;261:165.e1-8.

DOI: 10.5455/NOFOR.2026.08.012

22. Tsiouris C, Triantafyllou G, Karangeli N, Botis GG, Papadopoulos-Manolarakis P, Kalamatianos T, et al. Morphometric assessment of occipital condyles and foramen magnum reveals enhanced sexual dimorphism detection via 3D imaging: a systematic review and meta-analysis utilizing classification and regression trees. Diagnostics (Basel). 2025;15:1359.

23. Etli Y, Kartal E, Asirdizer M, Hekimoglu Y, Keskin S, Demir U, et al. Analysis pipeline and anonymized boundary coordinate dataset for foramen magnum shape analysis (software and dataset). Zenodo; 2026. DOI: 10.5281/zenodo.23022266.

24. Fitzgibbon A, Pilu M, Fisher RB. Direct least square fitting of ellipses. IEEE Trans Pattern Anal Mach Intell. 1999;21:476-80.

25. Khachiyan LG. Rounding of polytopes in the real number model of computation. Math Oper Res. 1996;21:307-20.

26. Satopaa V, Albrecht J, Irwin D, Raghavan B. Finding a “Kneedle” in a haystack: detecting knee points in system behavior. 31st International Conference on Distributed Computing Systems Workshops, 20-24 June 2011. Minneapolis, MN, USA, 166-171. doi:10.1109/ICDCSW.2011.20.

27. Rohlf FJ, Slice D. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst Zool. 1990;39:40-59.

28. Gower JC. Generalized Procrustes analysis. Psychometrika. 1975;40:33-51.

29. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A. 2002;99:6562-6.

30. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics. 2006;7:91.

31. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika. 1947;12:153-7.

32. Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6:461-4.

33. Hennig C. Cluster-wise assessment of cluster stability. Comput Stat Data Anal. 2007;52:258-71.

34. Anderson MJ. Distance-based tests for homogeneity of multivariate dispersions. Biometrics. 2006;62:245-53.

35. Brown MB, Forsythe AB. Robust tests for the equality of variances. J Am Stat Assoc. 1974;69:364-7.

36. Atreya A, Shrestha R, Bhandari K, Malla SK, Acharya S, Menezes RG. Morphometric analysis of the foramen magnum in sex estimation: an additional 3DCT study from Nepal on a larger sample. Health Sci Rep. 2023;6:e999.

37. Kilmer K, Garvin HM. Outline analysis of sex and population variation in greater sciatic notch and obturator foramen morphology with implications for sex estimation. Forensic Sci Int. 2020;314:110346.

38. Bookstein FL. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med Image Anal. 1997;1:225-43.

39. Caple J, Byrd J, Stephan CN. Elliptical Fourier analysis: fundamentals, applications, and value for forensic anthropology. Int J Legal Med. 2017;131:1675-90.

40. Bierry G, Le Minor JM, Schmittbuhl M. Oval in males and triangular in females? A quantitative evaluation of sexual dimorphism in the human obturator foramen. Am J Phys Anthropol. 2010;141:626-31.