UNIVERSIDADE FEDERAL DE UBERLÂNDIA

FACULDADE DE ODONTOLOGIA

LELIS DA COSTA NETO

# Machine learning application in sexual dimorphism analysis through zygomatic bone dimensions

UBERLÂNDIA

2025

LELIS DA COSTA NETO

# Machine learning application in sexual dimorphism analysis through zygomatic bone dimensions

Trabalho de conclusão de curso apresentado a Faculdade de Odontologia da UFU, como requisito parcial para obtenção do título de Graduado em Odontologia

Orientador: Prof. Dr. Thiago Leite Beaini

UBERLÂNDIA
2025

# SUMÁRIO

# Abstract

Forensic anthropology (FA) aims to study the human body to aid justice. The human skull is recognized as a structure that houses a series of characteristics subject to individual variability which can occur due to the biological sex of the individuals, genetic action of ancestry, or age. Volumetric examinations allow the observer to assess morphology in a differentiated manner, and volume is one of the ways to represent this topography. However, there is a gap in studies in this field, as various structures can aid in forensic analysis, especially in the context of searching for missing persons. To evaluate the individualizing capacity of the zygomatic bone in studying sexual dimorphism and human identification. The pilot study included 36 anonymous cone-beam computed tomography (CBCT), divided between 18 male and 18 female. The DICOM files were imported into Blender with OrthogOnBlender addon to reconstruct the bone tissues and export it into .stl files. In the Freeform software, using a haptic device to precisely select the zygomatic bone at its borders with the temporal, sphenoid, frontal, and maxillary bones. Tests showed normality in distribution. T-tests demonstrated a statistically significant difference (p=<0.05) in the variables Volume (p=0.033) and Height (z) (p=0.004). The volumetric analysis of the zygomatic bone proves to be a promising tool for sex estimation in human identification services.

**Key Words:** Forensic Anthropology, Zygomatic Bone, Sexual Dimorphism, Cone-Beam Computed Tomography

## Introduction

Forensic Anthropology (FA) is configured as a crucial area for criminal investigation, providing valuable information about human remains, such as sex determination, population affinity, and age estimation, assisting in the identification of victims, crime scene reconstruction, and understanding the causes of death [1]. Aiming to contribute to the development of new analytical techniques in FA, many studies aim to provide support for the search for missing persons and criminal investigation [2].

The analysis of skeletal remains, whether from fragmented or incomplete skeletons, is often the tool capable of providing evidence for human identification. Within the forensic context, the skull is the component most frequently found. The expertise of the professional, combined with the meticulous analysis of dental and bone elements, significantly contributes to the elucidation of crimes, giving families answers about the fate of their loved ones [3–5].

Although bone volumetric analysis holds great potential for FA, there is still a knowledge gap regarding the zygomatic bone, a facial structure with anatomical characteristics that may be relevant for human identification and the study of sexual dimorphism [6].

The digital realm has revolutionized the field of forensic and anthropological studies, particularly with the advent of technologies such as cone-beam computed tomography (CBCT) [7]. This technology enables the acquisition of high-resolution 3D images of skeletal remains and other bone materials, opening a range of new possibilities for more precise and detailed analyses [8]. It facilitates the identification of morphological characteristics relevant to human identification and the study of sexual dimorphism, for example [9,10].

CBCT enables the virtual reconstruction of fragmented or incomplete skeletal remains, which aids in the identification of victims in cases of mass disasters or crimes involving mutilation. CBCT images can be easily compared with digital databases of skeletal remains, enhancing the accuracy of human identification [7,9].

Thus, the present study aims to evaluate the zygomatic bone CBCT as a predictor of sexual dimorphism, contributing to the identification of individuals in forensic and archaeological contexts, as well as to the enhancement of knowledge about human anatomy and the differences between sexes. For this, we tested

Machine Learning methods (ML) to build a model capable of predicting sex from the provided measurements.

## Methodology

### Sample

Approved by the local Ethics Committee (CAAE 65383822.4.0000.5152), a database of 100 anonymous CBCT contains information on the sex and age of the individuals. The sample is equally divided between men and women, applying an inclusion criterion that selected adult individuals with no history of surgeries or facial deformities. Including exams with good visualization of the zygomatic bones, 36 CBCTs were selected, 18 from male and 18 from female individuals. A 3D model of the entire bone structure was generated from the CBCT exams, exported from Blender with the OrtogOnBlender addon (version 2.91.0). Models were exported into stereolithography (.STL) format and subsequently imported into Geomagic Freeform Plus® software (3D SYSTEMS Version 2023.1).

### Data acquisition

After importing the .STL files into Geomagic Freeform Plus® software (3D SYSTEMS Version 2023.1), the zygomatic bone was selected using a haptic device that allows digital interaction with virtual models, providing tactile feedback to the operator. Using the Select Clay tool, the zygomatic bone was segmented based on its anatomical structure, respecting its frontal, maxillary, sphenoid, and temporal boundaries. Next, the skull fragment is highlighted with the Reposition tool, found in the Select/move clay tab. Following this, the Box Select Tool is used to trim the fragment by removing the posterior portion. To clear any impurities that might affect the outcome, the Seed Selection Tool is employed. Finally, the Dimensional Bounding Box tool, located in the Analysis Tools tab, is used to obtain its spatial data, represented by the X, Y, and Z coordinates. In the tool's menu, the properties of the fragment are obtained through the Piece Properties function. The study in question analyzed the Volume, measured in cubic millimeters ($mm^3$).

### Statistical analysis

The data were then compiled into a Google Sheets spreadsheet (2024) including the individual's sex, registration number, Volume 1 (V1), and Volume 2 (V2), representing the intraobserver analysis, along with their respective dimensions in the x, y and z axis (x1, y1, z1; x2, y2, z2). Following this, the table was imported into Jamovi software (version 2.3.28, JASP) for statistical analysis.

Reproducibility was assessed in 10 examinations, comparing the data from the main evaluator with a repetition of the exams and with a second observer. The data were compared using the Intraclass Correlation Coefficient test (ICC).

The values in mm and mm³ were then compared between groups to analyze the zygomatic bone's capacity for sexual dimorphism. A normality test assessed the dispersion of measurements to select the appropriate statistical test for evaluating the difference between the groups.

Descriptive analysis was performed followed by the test of normal distribution. As all measurements returned normal distributed values, student t-test was made to search for differences between sex.

Once a statistical difference in each variable was established, a logistic regression model was applied to assess how accurately the pilot statistical model could explain the differences between sexes through the volumetric and linear analysis of the zygomatic bone.

Finally, a machine learning approach was tested using the software Orange (version 3.8) teste ML prediction models. The following tests were compared in search of the most efficient in the prediction of the biological sex using the significant variables. In this research were tested the following models of ML: decision tree (DT), Random Forest (RF), KNN, Adabost, Logistic regression (LR) and Artificial Neural Networks (ANN).

## Results

In the present study, 36 CBCT images were analyzed, consisting of 18 male and 18 female individuals. The descriptive data are presented in Table 1, showing slightly higher average volumes in men, but with few differences between the dimensions in the anteroposterior (y) and lateral (x) directions. However, the height (z) shows higher averages in men.

**Table 1. Descriptive statistics of data obtained from the zygomatic bone of men and women**

| | Sex | N | Mean | SE | Median | SD | Minimum | Maximum | Shapiro-Wilk W | Shapiro-Wilk p |
|---|---|---|---|---|---|---|---|---|---|---|
| **Vol** | F | 18 | 3385.72 | 162.77 | 3306.00 | 690.59 | 2020.00 | 4481.60 | 0.94 | 0.273 |
| | M | 18 | 3880.32 | 152.50 | 3766.90 | 646.99 | 2931.80 | 5181.50 | 0.95 | 0.450 |
| **X** | F | 18 | 29.73 | 0.72 | 29.65 | 3.05 | 25.20 | 37.50 | 0.95 | 0.497 |
| | M | 18 | 29.59 | 0.97 | 29.80 | 4.10 | 23.00 | 40.90 | 0.92 | 0.149 |
| **Y** | F | 18 | 24.34 | 0.58 | 24.85 | 2.45 | 20.30 | 29.50 | 0.97 | 0.703 |
| | M | 18 | 25.50 | 0.86 | 25.30 | 3.63 | 18.00 | 34.70 | 0.96 | 0.569 |
| **Z** | F | 18 | 43.30 | 0.92 | 43.45 | 3.92 | 33.20 | 49.10 | 0.93 | 0.197 |
| | M | 18 | 46.68 | 0.58 | 47.20 | 2.47 | 42.80 | 51.60 | 0.95 | 0.494 |

Regarding reproducibility, the intraobserver tests using the ICC indicated good correlation for Volume (0.87), Width (0.81), Depth (0.96), and Height (0.70). In the interobserver comparison, the correlation was also good for Volume (0.84), Depth (0.70), and Height (0.90); however, Width (0.33) showed a lower correlation index, meaning a more difficult reproduction of the measurements. Observing the pieces, it is noted that the second observer selected additional hard tissue fragments within the orbit.

The Shapiro-Wilk normality tests demonstrated adherence to normality in the W test, with values close to 1. The p-test did not reveal any data indicating a non-normal distribution.

For this reason, t-tests were used to assess the existence of differences between measurements in males and females. Considering a 95% confidence interval ($p<0.05$), the volume of the zygomatic bone showed statistically significant differences related to sex (Table 2), with a p-value of 0.03. The same was observed in the comparison of height measurements (z) with a p-value set at 0.004.

Independent Samples T-Test

| | | Statistic | df | p |
|---|---|---|---|---|
| Volume G2 | Student's t | -2.22 | 34.00 | 0.033 |
| X | Student's t | 0.11 | 34.00 | 0.913 |
| Y | Student's t | -1.12 | 34.00 | 0.271 |
| Z | Student's t | -3.10 | 34.00 | 0.004 |

*Note.* $H_a$ $\mu_F \neq \mu_M$

Given the existence of a significant relationship, a binomial logistic regression model was proposed. Two scenarios were considered: one presenting only the measurements with significant differences and another presenting all the combined measurements. Despite the demonstrated statistical relationship between the measurements and the biological sex of the individual, the model showed a higher level of accuracy when tested with all variables.

Including only volume and height (z) values, a $R^2N$, which measures how much the model explains the determination of sex in individuals, shows a level approaching 31% (0.31). The results of the Omnibus Likelihood Ratio test considered a significant relationship between the volume and the dependent variable, presenting a p-value of 0.7 and for z measurements a level of significance of 0.032.

Through a collinearity statistic, the interference of one variable on others with respect to the predictive outcome can be assessed. Considering that values close to 1 (VIF of 1.45 and tolerance 0.69) are ideal, it is observed that the variables indicate a good relationship among the measurements in the application of a joint predictive model.

The classification table (table 3) demonstrates a relevant accuracy rate, even considering that the test werer performed only on the sample used to produce the regression model.

Table 2. Classidication tabel showing a level of correct predictions

| | **Predicted** | | |
|---|---|---|---|
| **Observed** | **F** | **M** | **% Correct** |
| F | 11 | 7 | 61.11 |
| M | 6 | 12 | 66.67 |

*Note.* The cut-off value is set to 0.5

To measure the performance of the machine learning model, an AUC (Area Under the ROC Curve) graph was used. The obtained value of 0.77 indicates good performance of the model in binary classification, demonstrating that the model is capable of distinguishing between male and female sexes. The cutoff value set at 0.5 represents the point on the ROC curve where the sensitivity and specificity of the model are 0.67 and 0.61 respectivelly.

With this, we obtain a formula that in this pilot model can be used to estimate the sex of individuals based on the proposed measurements (prepared in MS Excel format):

**=1/(1+EXP(-(-15.73 + 0.002xVolume + 0.33xZ)))**

In the machine learning, a canvas containing the workflow of the data was set as represented on figure 1

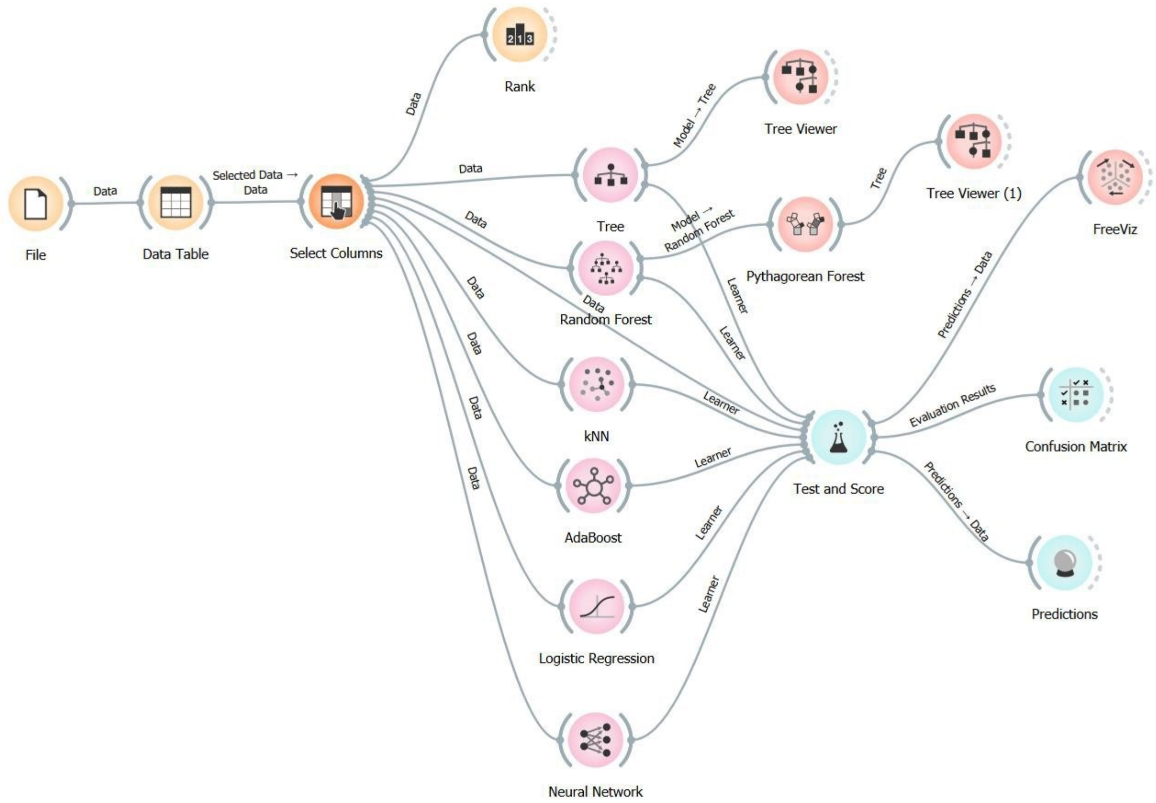


Figure 1. Canvas that visually demonstrate the path of the data as provided by the Orange software

The results included accuracy, AUC, specificity and sensitivity to predict biological sex with minimum false positive or negatives. In table 3, the results in order of best accuracy are shown.

| **Model** | **AUC** | **CA** | **Precision** | **Recall** | **MCC** | **Spec** |
|---|---|---|---|---|---|---|
| Neural Network | 0.653 | 0.647 | 0.644 | 0.647 | 0.276 | 0.623 |
| kNN | 0.568 | 0.588 | 0.583 | 0.588 | 0.154 | 0.563 |
| Logistic Regression | 0.674 | 0.559 | 0.551 | 0.559 | 0.087 | 0.525 |
| AdaBoost | 0.463 | 0.471 | 0.471 | 0.471 | -0.074 | 0.456 |
| Random Forest | 0.521 | 0.471 | 0.462 | 0.471 | -0.090 | 0.442 |
| Tree | 0.432 | 0.382 | 0.386 | 0.382 | -0.244 | 0.372 |

The classification table for the best ML model was selected and showed similar values, if compared to the statistical logistic regression. Males showed 66% of correct predictions and females 61%.

## Discussion

This study demonstrated promising characteristics for the application of the proposed methodology for sex determination. The analyses conducted highlighted height (Z) and volume as statistically significant variables in predicting sexual dimorphism. Height showed a high discriminatory power between the sexes, validating findings from previous studies [6].

Even though the prediction accuracy was not as high as other methods, in forensis anthropology it is necessary to develop numerous methods and different analysis considering parts of the human skull. This is due to the state of preservation of human remains that may only provide part or fragmented parts. These results could be used, with some caution in case of finding only zygomatic bones or part of the skull containing it.

Various authors agree on the importance of analyzing the human skull, in its measurements and shapes, as a tool of crucial importance in identifying individuals, as stated by Nunes & Gonçalves [4].

In the present study, the $R^2N$ obtained was 0.3, demonstrating a moderate level of explanation for the variability of sex by the metric variables analyzed. This result is close to the value found by Ceballos et al. in their qualitative study, which achieved an $R^2N$ of 0.118.

The tests indicate an adequate degree of reproducibility, although they suggest the need for training. The lack of correlation in the width of the models compared in the interobserver analysis shows that care must be taken in delineating the zygomatic bone in the segmented models. In addition to this factor, a poor recording of hard tissues in this region is observed in the CBCT scans. To avoid this bias, the main model was selected using the Seed selection tool to delete loose fragments and other bones that would impair the measurements.

The primary limitation of this study is the sample size. Studies with larger samples have shown similar results, indicating the zygomatic bone as a useful marker for sex determination in forensic anthropology contexts, corroborating the findings of Alves et al. [11] and Teles et al. [12]. It is likely that with an increase in the sample size assessed, a better condition for generalization can be evaluated.

A strong point of this study is the use of the haptic device, where tactile feedback contributed to better selection of the desired structure. Furthermore, the manipulation of the skull in a virtual environment, combined with the ease of use and rapid learning curve, provides greater safety and more efficient handling for the operator.

Considering the dimensions of the bone structure analyzed in this study, the hypothesis that the zygomatic bone can be used as a predictor of sexual dimorphism was accepted. This confirmation is because, generally, the male skull has greater dimensions and bone mass compared to the female skull [4]. This result reinforces the promising perspective of the zygomatic bone volume as a tool for sex identification in forensic anthropology contexts.

Furthermore, exploring the volume of other cranial bones, in addition to the zygomatic bone, presents itself as a promising tool for enhancing the assessment of sexual dimorphism.

## Conclusion

The volumetric analysis of the zygomatic bone proves to be a promising tool for sex estimation in human identification services using both volume and height as biological sexual discriminant. However, studies with larger samples are necessary to confirm its robustness and applicability across different populations.

## References


[1] Cattaneo C. Forensic anthropology: developments of a classical discipline in the new millennium. Forensic Science International 2007;165:185–93. https://doi.org/10.1016/j.forsciint.2006.05.018.

[2] Yaşar Işcan M. Rise of forensic anthropology. Am J Phys Anthropol 1988;31:203–29. https://doi.org/10.1002/ajpa.1330310510.

[3] Biancalana RC, Ortiz AG, De Araujo LG, Semprini M, Da Silva RHA, Galo R. Determinação do sexo pelo crânio: etapa fundamental para a identificação humana. Rev Bras Crimin 2015;4:38–43. https://doi.org/10.15260/rbc.v4i3.98.

[4] Nunes FB, Gonçalves PC. A importância da craniometria na criminalística: revisão de literatura. Rev Bras Crimin 2014;3:36–43. https://doi.org/10.15260/rbc.v3i1.69.

[5] Soares ATC, Guimarães MA. DOIS ANOS DE ANTROPOLOGIA FORENSE NO CENTRO DE MEDICINA LEGAL (CEMEL) DA FACULDADE DE MEDICINA DE RIBEIRÃO PRETO-USP. Medicina (Ribeirão Preto) 2008;41:7–11. https://doi.org/10.11606/issn.2176-7262.v41i1p7-11.

[6] Schlager S, Rüdell A. Sexual Dimorphism and Population Affinity in the Human Zygomatic Structure—Comparing Surface to Outline Data. The Anatomical Record 2017;300:226–37. https://doi.org/10.1002/ar.23450.

[7] Farhadian M, Salemi F, Shokri A, Safi Y, Rahimpanah S. Comparison of data mining algorithms for sex determination based on mastoid process measurements using cone-beam computed tomography. Imaging Sci Dent 2020;50:323. https://doi.org/10.5624/isd.2020.50.4.323.

[8] Scarfe WC, Farman AG. What is Cone-Beam CT and How Does it Work? Dental Clinics of North America 2008;52:707–30. https://doi.org/10.1016/j.cden.2008.05.005.

[9] Aguilar Campos Silva M, Freitas Fernandes AD, Ilídio Vespasiano Silva A, Villalobos MIDOEB. O USO DA TOMOGRAFIA COMPUTADORIZADA PARA IDENTIFICAÇÃO HUMANA EM ODONTOLOGIA LEGAL – REVISÃO DE LITERATURA. RBOL 2021;8. https://doi.org/10.21117/rbol-v8n12021-344.

[10] Arai Y, Tammisalo E, Iwai K, Hashimoto K, Shinoda K. Development of a compact computed tomographic apparatus for dental use. Dentomaxillofac Radiol 1999;28:245–8. https://doi.org/10.1038/sj.dmfr.4600448.

[11] Alves N, Deana NF, González J, Hernández P, Ceballos F. Sex Prediction by Analysis of the Morphological Characteristics of Macerated Skulls. International Journal of Morphology 2020;38:815–9.

[12] Teles HCC, Santos RA, Almeida Jr E, Sandes VA, Reis FP. Estimativa do Sexo e Idade por Meio de Mensurações Faciais em Crânios Secos de Adultos. BJFS 2020;9:292–307. https://doi.org/10.17063/bjfs9(3)y2020292.