Title: CADOES: An interactive machine-learning approach for sex estimation with the pelvis

Authors: João d'Oliveira Coelho (Conceptualization) (Methodology) (Software) (Validation) (Formal analysis) (Data curation) (Writing - original draft) (Writing - review and editing) (Visualization), Francisco Curate (Validation) (Writing - original draft) (Writing - review and editing) (Supervision)

PII:	S0379-0738(19)30289-0
DOI:	https://doi.org/10.1016/j.forsciint.2019.109873
Article Number:	109873
Reference:	FSI 109873
To appear in:	*FSI*
Received date:	11 March 2019
Revised date:	25 June 2019
Accepted date:	28 June 2019

Please cite this article as: d'Oliveira Coelho J, Curate F, CADOES: An interactive machine-learning approach for sex estimation with the pelvis, *Forensic Science International* (2019), https://doi.org/10.1016/j.forsciint.2019.109873

CADOES: An interactive machine-learning approach for sex estimation with the pelvis

João d'Oliveira Coelho[1,2*]; Francisco Curate[2,3]

[1]Institute of Cognitive and Evolutionary Anthropology, University of Oxford, Oxford, United Kingdom

[2]Laboratory of Forensic Anthropology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal

[3]Research Centre for Anthropology and Health, Department of Life Sciences, University of Coimbra, Portugal

*Corresponding author: joaopedrocoelho@gmail.com

Abstract

The pelvis is consistently regarded as the most sexually dimorphic region of the human skeleton, and methods for sex estimation with the pelvic bones are usually very accurate. In this investigation, population-specific osteometric models for the assessment of sex with the pelvis were designed using a dataset provided by J.A. Serra (1938) that included 256 individuals (131 females and 125 males) from the Coimbra Identified Skeletal Collection and 38 metric variables. The models for sex estimation were operationalized through an online application and decision support system, CADOES. Different classification algorithms generated high accuracy models, ranging from 85% to 92%, with only three variables; and from 85.33% to 97.33%, with all 38 variables. CADOES conveys a probabilistic prediction of skeletal sex, as well as a suite of attributes with educational applicability in the fields of human skeletal anatomy and statistics. This study upholds the value of the pelvis for the estimation of skeletal sex and provides models for that can be applied with high accuracy and low bias.

Keywords

Forensic Anthropology Population Data, os coxa, sacrum, supervised learning, biological profile

Introduction

Sex diagnosis is a fundamental step for establishing the biological profile, and thus is of critical importance in the forensic task of identifying human skeletal remains [1]. In forensic anthropology and bioarcheology, most methods for sex estimation rely on statistical models generated through osteometric data collected from identified populations [2]. Among all the regions of the human skeleton, the pelvis has for long been consensually regarded as the most sexual dimorphic [3–5] and, as such, it is the most appropriate for the creation of such predictive models. Pelvic sexual dimorphism is strongly connected with the opposing selective pressures of bipedalism and parturition. Adaptative dissimilarities between sexes in the pelvis are likewise an outcome of sexual selection [6,7]. Furthermore, pelvic anatomy is contingent to developmental plasticity (as a consequence of ecological, climate and nutritional factors) and neutral demographic processes [6,8,9].

Sexual dimorphism in the pelvic region has been studied and documented for long [10–15] but, before the influential work by Phenice (1969), the methods for the estimation of sex based on pelvic morphology were scarce and subjective. Phenice's [5] technique involves the visual evaluation of three pelvic elements, namely the ischiopubic ramus, the subpubic concavity, and the ventral arc. The method is straightforward and precise, with published accuracy rates ranging from 83% to 96% [16–19]. Notwithstanding, accuracy rates are observer-dependent [17]. Other morphoscopic methods include, e.g., those by Brůžek [20] and Klales et al. [21]. Metric data from the pelvis has also been commonly used for sex assessment [2,22–28]. Comparably to the studies addressing pelvic morphology in the estimation of sex, metric studies also yield exceedingly high accuracy rates [29–32].

The manuscript *A Pelve nos Portugueses: Morfologia da Pelve no Homem* [4] is a classic anthropological description of the pelvic complex in a late 19th – early 20th centuries sample of Portuguese skeletal remains [33,34]. Unfortunately, *A Pelve nos Portugueses* is merely descriptive in a statistical sense, as it was typical for contemporary anthropometry works. Instead, our purpose is to use the raw data provided in the manuscript to create new approaches for sex estimation based on morphometric features of the pelvic bone complex. As such, the web application and decision support system CADOES [35] is presented. The acronym stands for *Classificação Automatizada de Dados Osteométricos para Estimar o Sexo*, or *sex estimation through automated classification of osteometric data*. CADOES is available at http://osteomics.com/CADOES and features a sex estimation framework that allows greater flexibility to the user enabling the selection of metric variables, as well as the statistical learning algorithms and cross-validation parameters, thus empowering users to iterate through available variables and statistical parameters until achieving a suitable model for any given samples or individuals.

Materials and Methods

The original paper by Serra [4] provided a dataset (Appendix, pp. 143 – 172, available at https://impactum.uc.pt/pt-pt/artigo/pelve_nos_portugueses_morfologia_da_pelve_no_homem; also obtainable at http://osteomics.com/CADOES/Dados-Serra-1938.zip) comprising a total of 256 individuals (131 females and 125 males) and 40 variables. All individuals were Portuguese nationals from the Coimbra Identified Skeletal Collection [CISC, 34]. The first two variables encompass ID and Sex, while the remaining 38 are osteometric variables, more precisely 32 Euclidean distances between anatomic landmarks and six angles of the pelvis (Table 1; Figure 1). The original measurements followed the operational definitions by Frassetto [36] and Martin [37]. Some data cleaning procedures had to be performed in order to prepare the dataset for a suitable statistical analysis [38].

Probably due to transcription errors, there were some repeated ID's in the original publication, with three ID's appearing twice in the original spreadsheet (individuals 76, 233 and 235). Since the original ID's from the incorrectly marked individuals could not be determined, an 'A' was added to the label of the second individual appearing by row order. So, to correct these inaccuracies, these individuals were labeled as 76A, 233A and 235A. Another necessary data cleaning step involved the variable *Altura Máxima da Bacia* (maximum pelvic height). It appears as a sole variable from page 143 to 157, but from page 158 until the end it is divided into two variables, by left and right side: *Altura Máxima da Bacia (esquerda)* and *Altura Máxima da Bacia (direita)*. An option for the most parsimonious solution was made, and the arithmetic mean from the left and right measurements was calculated in order to keep only one variable that represents the maximum pelvic height. All individuals (ID's 11, 33, 34, 40 and 108) containing missing values were removed. A reduction of less than 2% of the original dataset row-wise avoids adding error to the training set associated with the estimation of missing values. Thus, for generating models for sex estimation, a final dataset containing a total of 251 individuals was used (130 females and 121 males).

All data analyses were performed with R, and a web app was developed with the Shiny package for R [39,40]. Different modelling techniques can be applied to the same binary classification task, but error rates vary in distinct datasets, particularly those composed by real data. As such, users can select and test the results for the following classification models: k-Nearest neighbors, naive Bayes, partial least squares, linear discriminant analysis, flexible discriminant analysis, generalized additive model using splines, boosted logistic regression, penalized logistic regression, decision trees, random forests, stochastic gradient boosting and a simple classification neural network. The K-nearest neighbor is an instance-based learning classifier that stores the training data set and classifies new uncategorized records by comparing them to similar records in the training set. Naive Bayes algorithms are probabilistic classifiers grounded on Bayesian statistics featuring conditional independence assumptions. Partial least squares (for classification) is a supervised dimension reduction technique that was originally developed for regression problems. Linear discriminant analysis recognizes a linear combination of predictor variables that optimally splits mutually exclusive groups, and then creates a discriminant function that typifies the differences between groups and classifies new individuals with unspecified group membership. Flexible discriminant analysis is a nonparametric extension of the former method. Boosted logistic regression is an ensemble method that sequentially uses a generalized additive model and then applies the cost function of logistic regression. Penalized logistic regression enforces a penalty to the size of the L2 norm of the coefficients, decreasing the coefficients of less contributory variables towards zero. Decision trees are classification methods in the form of IF-THEN logical rules. Random forests are, fundamentally, an ensemble of multiple decision trees. Generalized additive models using splines are flexible methods used to identify and characterize nonlinear regression effects. Gradient boosting builds additive regression models by serially fitting a parameterized function to pseudo-residuals by least squares at each iteration and including randomization in the process. Finally, neural networks are adaptative models that imitate the nonlinear learning occurring in the neuronal networks found in animal brains [41–45]. These twelve machine learning algorithms were implemented using the caret package for R [46]. CADOES depends on caret to perform all calculations. While all algorithms available operate with different tuning parameters (automatically optimized during cross-validation), that information is indicated for each classification model in the app at Sex estimation > Predict > Model Information & Accuracy, after selecting and running a model (Figure 2). Hence, the web app allows end-users to perform different kinds of data exploration and analyses. CADOES requires at least two variables to generate a sex estimation model. The proficiency of the models was evaluated through the overall accuracy (a measure of total agreement between the real and the estimated sex) with corresponding 95% confidence intervals, Cohen's Kappa (an evaluation of the performance of a specified classifier as related to chance only), sensitivity (the ratio of females correctly classified), and specificity (the ratio of males properly classified), and the area under the curve (AUC) [42].

Results

Descriptive statistics, including group means, standard deviations, medians, and minimum and maximum values, are summarized in Table 2.

CADOES can generate a virtually endless number of models; as such, in order to present results concisely many of the parameters were fixed, as follow: data was split into training and testing samples containing respectively 70 and 30% of all individuals, and, as a preprocessing step, all variables were centered and scaled. The repeated k-fold cross-validation parameters were also kept fixed, as *folds* = 10 and *repeats* = 10. The use of a robust cross-validation strategy can mitigate overfitting. In order to enhance results reproducibility, the seed was set at 19920804 (first author's birth date) for pseudorandom number generators during cross-validation and model fitting. Results for all classification methods currently available in CADOES are presented. For each classification algorithm, two types of models were created, containing either three or all variables of our dataset. AMB (maximum pelvic height) + LB (bispiniatic width) + ASP (subpubic angle) were selected for the three-variable models, since these are highly dimorphic. All results presented are from confusion matrices of trained prediction scores against the testing dataset. Despite only using three measurements, six of the twelve models obtained overall accuracies superior to 90% on the holdout (testing) dataset. Among these, the best models were the partial least squares, LDA, and neural networks (Table 3). Parenthetically, if three more variables were added, for example PFIe + AAId + LAE, while keeping all the parameters just mentioned, the same neural network architecture would present an overall accuracy of 94.67% under the robust cross-validation scheme aforementioned.

When using all the 38 variables, overall accuracies ranged from 85.33 to 97.33%. In this case, the best models were a partial least squares model (AUC = 97.04%), a logistic regression with a quadratic penalization (AUC = 96.90%), and again a neural network (AUC = 96.73%). These three models had an overall accuracy of 97.33%, a sensitivity of 94.87%, a specificity of 100% and a kappa of 94.67%. Despite having the best results (Table 4), users might avoid using such models, since similar performance metrics can be obtained by just measuring three to six variables, as shown above. The expert ought to consider the time it takes to record measurements before starting its own protocol.

Discussion

The assessment of sex represents a key research task in the forensic and bioarcheological examination of unknown human skeletal remains [1]. Sex estimation methods have focused in different bones [3,47,48] but the pelvis prevails as the most sexually dimorphic skeletal region [3,5,8]. As such, the conception of new methods or the enhancement of preexisting techniques for sex estimation with the pelvis is justified.

CADOES stems from the recommendation of population specific standards for sex estimation and makes available novel features for the investigation of sexual dimorphism in the pelvic morphological complex, including (1) exploratory analyses of the original dataset through density plots, boxplots, scatterplots and correlation matrices; (2) the generation of virtually unlimited classification models based on the variables of the dataset selected by the user/expert, plus the implemented machine learning algorithms and their validation parameters; (3) the use of robust methods of cross-validation and partitioned testing in order to access accuracy (with 95% confidence intervals), no information rate, kappa, specificity and sensitivity, positive and negative predictive values, prevalence, detection rate, detection prevalence, balanced accuracy, area under the curve, precision, recall, F-1 value, and variable importance of the generated models; and (4) the prediction of sex with pelvic bones using the models generated and data inputted by the users. The web implementation of CADOES (http://osteomics.com/CADOES/) generates models for sex estimation based on different classical and machine-learning classifiers, as selected by the user, and offers a probabilistic determination of sex according to the Daubert guidelines [47].

Coinciding with other hip-based techniques for the assessment of sex in skeletal remains, CADOES exhibits high accuracy rates and low bias. A model using only three variables (maximum pelvic height, bispiniatic width and subpubic angle) achieves an accuracy of 92% under cross-validation, with a six-variable model attaining an overall accuracy of 95%, comparable to the results obtained with other pelvic methods in different populations [22,24,27,49–52]. Although morphoscopic methods are the most readily applied, metric techniques are acknowledged as less observer-dependent and more reliable [22,50,53], and depict the overall pattern of variability within dissimilar populations [2].

In agreement with previous studies, pelvic height [e.g. 22,54,55] and the subpubic angle [e.g. 14,50,56,57] appear as some of the most dimorphic variables in the models enacted by CADOES. Pelvic height is larger in males and suggests that size enacts an essential role in the dimorphic condition of the pelvic region. The broadest subpubic angles observed in female individuals are probably related with obstetrics [58], as obstetric problems, such as labor duration and risk of obstetric intervention due to poor progression, are inversely associated with the breadth of the subpubic angle [59]. Ilium blade length, bispiniatic width and the iliac fossa depth are also among the pelvic variables showing more sexual dimorphism. Ilium blade length, although defined differently from the iliac length or direct iliac length, shows the same pattern of sexual dimorphism, with males showing a longer ilium [e.g. 13,60,61]. Sexual dimorphism in ilium length starts at 15–16 years of age and is mainly a function of differences in size [60]. Bispiniatic width (also known as midpelvic breadth) is larger in female individuals. Similar results were observed by Torimitsu et al. [57] in a sample of contemporary Japanese. In fact, dimensions that are larger in females are usually related with the pelvic inlet [61,62]. Bispiniatic width, being an obstetric dimension, has rarely been measured in skeletal remains; notwithstanding, it is particularly dimorphic and easy to measure [57]. Finally, the iliac fossa depth is greater in males. The iliac fossa is the large concave surface on the ventral surface of the ilium, and it is the origin site for the iliacus muscle [63]. If variable selection is possible, e.g., in the case of a well-preserved pelvis, these five most dimorphic variables should be used to estimate sex with CADOES.

There are some limitations to CADOES. First, many measurements chosen by Serra [4] require the complete set of pelvic bones, including the sacrum, in order to be measured, and these measurements are not easy to perform. Unfortunately, this is an aspect of the original dataset that cannot be bypassed. To mitigate such problem, pelvis fragments and single bone measurements can be used for sex estimation, since the model generator can produce estimates with as few as two variables. In fact, univariable models are to be avoided as they usually are less accurate and more biased [22,50,64,65] and the number of variables required to yield the most accurate models generally range from two to eleven variables [54]. Another limitation is that metric methods are apparently population specific and tend to perform better within populations of similar height or general body proportions, since size in itself correlates more to these features than to sexual dimorphism [55,66,67]. Hence, methods to estimate sex from a skeletal remains ought to use fitting regional data [68]. However, additional research has suggested that population-specific methods may not be essential for pelvic data [29,69].

Several of the CADOES advantages stem from its - at first glance - limitations. The web app is intended to bring not just a functional sex estimation tool that provides quality metrics, but also a didactic implement. In a classroom context, it can work as an interactive and stimulating tool for students, simplifying the study of classic anthropometrics and pelvic anatomy while giving some insight into modern statistical thinking, data visualization and processing as well as predictive modelling under the machine learning paradigm. CADOES expands the set of available web applications designed to simplify forensic and bioarcheological procedures, such as age at death and sex estimation [64,70–73].

Conclusions

CADOES upgrades the descriptive nature of J.A. Serra [4] work and generates user-tailored models for the estimation of sex that can be used with high accuracy and low bias in Portuguese populations. It can be used in fragmented pelvic bones and conveys a probabilistic estimate of sex. Additionally, the web app provides a set of features with pedagogic relevance in the fields of human pelvic anatomy and statistics. CADOES must endure further verification in skeletal remains of different geographical origins to evaluate its generalization to independent datasets and to validate its reliability in both forensic and bioarcheological contexts.

Conflict of interest

The authors state that they do not have any conflict of interest to declare.

CRediT author statement

João d'Oliveira Coelho: Conceptualization, Methodology, Software, Validation, Formal Analysis, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization. **Francisco Curate:** Validation, Writing – Original Draft, Writing – Review & Editing, Supervision.

References

[1] J. Brůžek, P. Murail, Methodology and Reliability of Sex Determination From the Skeleton, in: A. Schmitt, E. Cunha, J. Pinheiro (Eds.), Forensic Anthropology and Medicine, Humana Press, 2006: pp. 225–242.
[2] P. Murail, J. Brůžek, F. Houët, E. Cunha, DSP: A tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements, Bulletins et mémoires de la Société d'Anthropologie de Paris. (2005) 167–176.

[3] K. Krishan, P.M. Chatterjee, T. Kanchan, S. Kaur, N. Baryah, R.K. Singh, A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework, Forensic Sci. Int. 261 (2016) 165. doi:10.1016/j.forsciint.2016.02.007.
[4] J.A. Serra, A pelve nos Portugueses: morfologia da pelve no Homem [The pelvic complex in the Portuguese: morphology of the human pelvis], Contribuições Para o Estudo Da Antropologia Portuguesa. 3 (1938) 1–174.
[5] T.W. Phenice, A newly developed visual method of sexing the os pubis, American Journal of Physical Anthropology. 30 (1969) 297–301. doi:10.1002/ajpa.1330300214.
[6] H.K. Kurki, Pelvic dimorphism in relation to body size and body size dimorphism in humans, Journal of Human Evolution. 61 (2011) 631–643. doi:10.1016/j.jhevol.2011.07.006.
[7] A.M. Christensen, N.V. Passalacqua, E.J. Bartelink, Forensic anthropology: Current methods and practice, Elsevier, 2014.
[8] L. Betti, Sexual dimorphism in the size and shape of the os coxae and the effects of microevolutionary processes, Am. J. Phys. Anthropol. 153 (2014) 167–177. doi:10.1002/ajpa.22410.
[9] A. Huseynov, C.P.E. Zollikofer, W. Coudyzer, D. Gascho, C. Kellenberger, R. Hinzpeter, M.S. Ponce de León, Developmental evidence for obstetric adaptation of the human female pelvis, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 5227–5232. doi:10.1073/pnas.1517085113.
[10] A.A. Mendes Correia, Osteometria Portuguesa: III cintura pélvica, Annaes Scientificos Da Academia Polytechnica Do Porto. XIV (1920) 56–71.
[11] Wm. Turner, The Index of the Pelvic Brim as a Basis of Classification, J Anat Physiol. 20 (1885) 125–143.
[12] D.E. Derry, Note on the Innominate Bone as a Factor in the Determination of Sex: with Special Reference to the Sulcus Præauricularis, J Anat Physiol. 43 (1909) 266–276.
[13] W.L. Straus, The human ilium: Sex and stock, American Journal of Physical Anthropology. 11 (1927) 1–28. doi:10.1002/ajpa.1330110102.
[14] M. Young, J.G. Ince, A radiographic comparison of the male and female pelvis, J. Anat. 74 (1940) 374–385.
[15] S.L. Washburn, Sex differences in the pubic bone, Am. J. Phys. Anthropol. 6 (1948) 199–207.
[16] N.C. Lovell, Test of Phenice's technique for determining sex from the os pubis, Am. J. Phys. Anthropol. 79 (1989) 117–120. doi:10.1002/ajpa.1330790112.
[17] L.D. Sutherland, J.M. Suchey, Use of the ventral arc in pubic sex determination, J. Forensic Sci. 36 (1991) 501–511.
[18] D.H. Ubelaker, C.G. Volk, A test of the phenice method for the estimation of sex, J. Forensic Sci. 47 (2002) 19–24.
[19] C. McFadden, M.F. Oxenham, Revisiting the Phenice technique sex classification results reported by MacLaughlin and Bruce (1990), Am. J. Phys. Anthropol. 159 (2016) 182–183. doi:10.1002/ajpa.22839.
[20] J. Brůžek, A method for visual determination of sex, using the human hip bone, Am. J. Phys. Anthropol. 117 (2002) 157–168. doi:10.1002/ajpa.10012.
[21] A.R. Klales, S.D. Ousley, J.M. Vollner, A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods, Am. J. Phys. Anthropol. 149 (2012) 104–114. doi:10.1002/ajpa.22102.
[22] M. Steyn, M.Y. İşcan, Metric sex determination from the pelvis in modern Greeks, Forensic Sci. Int. 179 (2008) 86. doi:10.1016/j.forsciint.2008.04.022.
[23] W.W. Howells, Détermination du sexe du bassin par fonction discriminante : étude du matériel du Dr. Gaillard, Bulletins et Mémoires de la Société d'Anthropologie de Paris. 7 (1965) 95–105. doi:10.3406/bmsap.1965.1308.

[24] F.P. Schulter-Ellis, D.J. Schmidt, L.-A. Hayek, J. Craig, Determination of Sex with a Discriminant Analysis of New Pelvic Bone Measurements: Part I, JFS. 28 (1983) 169–180. doi:10.1520/JFS12249J.
[25] F.P. Schulter-Ellis, L.C. Hayek, D.J. Schmidt, Determination of sex with a discriminant analysis of new pelvic bone measurements: Part II, J. Forensic Sci. 30 (1985) 178–185.
[26] J.V. Taylor, R. DiBennardo, Discriminant function analysis of the central portion of the innominate, Am. J. Phys. Anthropol. 64 (1984) 315–320. doi:10.1002/ajpa.1330640314.
[27] J. Albanese, A metric method for sex determination using the hipbone and the femur, J. Forensic Sci. 48 (2003) 263–273.
[28] P.N. Gonzalez, V. Bernal, S.I. Perez, Geometric morphometric approach to sex estimation of human pelvis, Forensic Sci. Int. 189 (2009) 68–74. doi:10.1016/j.forsciint.2009.04.012.
[29] M. Steyn, M.L. Patriquin, Osteometric sex determination from the pelvis--does population specificity matter?, Forensic Sci. Int. 191 (2009) 113.e1–5. doi:10.1016/j.forsciint.2009.07.009.
[30] M.Y. İşcan, M. Steyn, The Human Skeleton in Forensic Medicine, 3rd ed., Charles C Thomas Publisher, 2013.
[31] S. Mestekova, J. Brůžek, J. Veleminska, K. Chaumoitre, A Test of the DSP Sexing Method on CT Images from a Modern French Sample, J. Forensic Sci. 60 (2015) 1295–1299. doi:10.1111/1556-4029.12817.
[32] J. Brůžek, F. Santos, B. Dutailly, P. Murail, E. Cunha, Validation and reliability of the sex estimation of the human os coxae using freely available DSP2 software for bioarchaeology and forensic anthropology, Am. J. Phys. Anthropol. 164 (2017) 440–449. doi:10.1002/ajpa.23282.
[33] M.A. Rocha, Les collections ostéologiques humaines identifiées du Musée Anthropologique de l'Université de Coimbra, Antropologia Portuguesa. 13 (1995) 7–38.
[34] E. Cunha, S.N. Wasterlain, The Coimbra identified osteological collections, in: G. Grupe, J. Peters (Eds.), Skeletal Series and Their Socio-Economic Context, Verlag Marie Leidorf GmbH, Rahden/Westf, Germany, 2007: pp. 23–33.
[35] CADOES, Classificação Automatizada de Dados Osteométricos para Estimar o Sexo., 2018. http://osteomics.com/CADOES.
[36] F. Frassetto, Lezioni di Antropologia, Tipografica-Libraria Mareggiani, Bologna, 1923.
[37] R. Martin, Lehrbuch der Anthropologie in Systematischer Darstellung mit Besonderer Berücksichtigung der Anthropologischen Methoden für Studierende, Ärtze und Forschungsreisende. Zweiter Band: Kraniologie, Osteologie., 2nd ed., Gustav Fischer, Jena, 1928.
[38] H. Wickham, Tidy data, The Journal of Statistical Software. 59 (2014). http://www.jstatsoft.org/v59/i10/.
[39] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2019. http://R-project.org/.
[40] W. Chang, J. Cheng, J.J. Allaire, Y. Xie, J. McPherson, shiny: Web Application Framework for R, 2019. https://CRAN.R-project.org/package=shiny.
[41] D.T. Larose, C.D. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, 2nd ed., John Wiley & Sons, 2014.
[42] D.T. Larose, C.D. Larose, Data Mining and Predictive Analytics, John Wiley & Sons, 2015.
[43] J. Maroco, D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, A. de Mendonça, Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests, BMC Res Notes. 4 (2011) 299. doi:10.1186/1756-0500-4-299.
[44] D.L. Olson, D. Wu, Predictive Data Mining Models, Springer Singapore, 2017. https://www.springer.com/gb/book/9789811025426 (accessed May 6, 2019).

[45] J.H. Friedman, Stochastic gradient boosting, Computational Statistics & Data Analysis. 38 (2002) 367–378. doi:10.1016/S0167-9473(01)00065-2.
[46] M. Kuhn, Building predictive models in R using the caret package, Journal of Statistical Software. 28 (2008) 1–26.
[47] A.M. Christensen, N.V. Passalacqua, E.J. Bartelink, Chapter 8 - Sex Estimation, in: A.M. Christensen, N.V. Passalacqua, E.J. Bartelink (Eds.), Forensic Anthropology, Academic Press, San Diego, 2014: pp. 199–222. doi:10.1016/B978-0-12-418671-2.00008-2.
[48] Á. Azevedo, M.L. Pereira, S. Gouveia, J.N. Tavares, I.M. Caldas, Sex estimation using the mandibular canine index components, Forensic Sci Med Pathol. (2018). doi:10.1007/s12024-018-0051-2.
[49] P. Mahakkanukrauh, S. Praneatpolgrang, S. Ruengdit, P. Singsuwan, P. Duangto, D.T. Case, Sex estimation from the talus in a Thai population, Forensic Sci. Int. 240 (2014) 152.e1–8. doi:10.1016/j.forsciint.2014.04.001.
[50] D. Franklin, A. Cardini, A. Flavel, M.K. Marks, Morphometric analysis of pelvic sexual dimorphism in a contemporary Western Australian population, Int. J. Legal Med. 128 (2014) 861–872. doi:10.1007/s00414-014-0999-8.
[51] J.A. Gómez-Valdés, A. Menéndez Garmendia, L. García-Barzola, G. Sánchez-Mejorada, C. Karam, J.P. Baraybar, A. Klales, Recalibration of the Klales et al. (2012) method of sexing the human innominate for Mexican populations, Am. J. Phys. Anthropol. 162 (2017) 600–604. doi:10.1002/ajpa.23157.
[52] E.-K. Oikonomopoulou, E. Valakos, E. Nikita, Population-specificity of sexual dimorphism in cranial and pelvic traits: evaluation of existing and proposal of new functions for sex assessment in a Greek assemblage, Int. J. Legal Med. 131 (2017) 1731–1738. doi:10.1007/s00414-017-1655-x.
[53] J. Brůžek, P. Murail, F. Houët, E. Cleuvenot, Inter- and Intra-Observer Error in Pelvic Measurement and Its Implication for the Methods of Sex Determination, Anthropologie (1962-). 32 (1994) 215–223.
[54] P. Mahakkanukrauh, S. Ruengdit, S.M. Tun, D.T. Case, A. Sinthubua, Osteometric sex estimation from the os coxa in a Thai population, Forensic Sci. Int. 271 (2017) 127.e1-127.e7. doi:10.1016/j.forsciint.2016.11.043.
[55] M.L. Patriquin, M. Steyn, S.R. Loth, Metric analysis of sex differences in South African black and white pelves., Forensic Sci Int. 147 (2005) 119–127. doi:10.1016/j.forsciint.2004.09.074.
[56] H.M. Karakas, A. Harma, B. Alicioglu, The subpubic angle in sex determination: anthropometric measurements and analyses on Anatolian Caucasians using multidetector computed tomography datasets, J Forensic Leg Med. 20 (2013) 1004–1009. doi:10.1016/j.jflm.2013.08.013.
[57] S. Torimitsu, Y. Makino, H. Saitoh, A. Sakuma, N. Ishii, D. Yajima, G. Inokuchi, A. Motomura, F. Chiba, R. Yamaguchi, M. Hashimoto, Y. Hoshioka, H. Iwase, Morphometric analysis of sex differences in contemporary Japanese pelves using multidetector computed tomography, Forensic Sci. Int. 257 (2015) 530.e1-530.e7. doi:10.1016/j.forsciint.2015.10.018.
[58] R.G. Tague, Sexual dimorphism in the human bony pelvis, with a consideration of the Neandertal pelvis from Kebara cave, Israel, American Journal of Physical Anthropology. 88 (1992) 1–21. doi:10.1002/ajpa.1330880102.
[59] T. Ghi, A. Dall'Asta, A. Suprani, E. Aiello, A. Musarò, C. Bosi, G. Pedrazzi, A. Kiener, D. Arduini, T. Frusca, G. Rizzo, Correlation between Subpubic Arch Angle and Mode of Delivery in Large-for-Gestational-Age Fetuses, Fetal. Diagn. Ther. 44 (2018) 221–227. doi:10.1159/000481169.
[60] C. Rissech, A. Malgosa, Ilium growth study: applicability in sex and age diagnosis, Forensic Sci. Int. 147 (2005) 165–174. doi:10.1016/j.forsciint.2004.08.007.

[61] R. Segebarth-Orban, An evaluation of the sexual dimorphism of the human innominate bone, Journal of Human Evolution. 9 (1980) 601–607. doi:10.1016/0047-2484(80)90089-5.
[62] R.G. Tague, Big-bodied males help us recognize that females have big pelves, Am. J. Phys. Anthropol. 127 (2005) 392–405. doi:10.1002/ajpa.20226.
[63] L. Scheuer, S. Black, eds., Developmental Juvenile Osteology, Elsevier, 2000.
[64] F. Curate, J. d'Oliveira Coelho, D. Gonçalves, C. Coelho, M.T. Ferreira, D. Navega, E. Cunha, A method for sex estimation using the proximal femur, Forensic Science International. 266 (2016) 579.e1–7. doi:10.1016/j.forsciint.2016.06.011.
[65] E. Vacca, G. Di Vella, Metric characterization of the human coxal bone on a recent Italian sample and multivariate discriminant analysis to determine sex, Forensic Sci. Int. 222 (2012) 401.e1–9. doi:10.1016/j.forsciint.2012.06.014.
[66] S.M. MacLaughlin, M.F. Bruce, The sciatic notch/acetabular index as a discriminator of sex in European skeletal remains, J. Forensic Sci. 31 (1986) 1380–1390.
[67] P.L. Walker, Greater sciatic notch morphology: sex, age, and population differences, Am. J. Phys. Anthropol. 127 (2005) 385–391. doi:10.1002/ajpa.10422.
[68] D.H. Ubelaker, C.M. DeGaglia, Population variation in skeletal sexual dimorphism, Forensic Sci. Int. 278 (2017) 407.e1-407.e7. doi:10.1016/j.forsciint.2017.06.012.
[69] P.J. Macaluso, Sex determination from the acetabulum: test of a possible non-population-specific discriminant function equation, J Forensic Leg Med. 17 (2010) 348–351. doi:10.1016/j.jflm.2010.04.011.
[70] D. Gonçalves, J. d'Oliveira Coelho, M.A. Acosta, C. Coelho, F. Curate, M.T. Ferreira, M. Gouveia, C. Makhoul, D. Pinto, I. Oliveira-Santos, A.R. Vassalo, D. Navega, E. Cunha, One for all and all for one: Linear regression from the mass of individual bones to assess human skeletal mass completeness: Linear Regression of Human Skeletal Mass, American Journal of Physical Anthropology. 160 (2016) 427–432. doi:10.1002/ajpa.22979.
[71] G.R. Scott, M.A. Pilloud, D. Navega, J. d'Oliveira Coelho, E. Cunha, J.D. Irish, rASUDAS: A New Web-Based Application for Estimating Ancestry from Tooth Morphology, Forensic Anthropology. 1 (2018) 18–31.
[72] D. Navega, J. d'Oliveira Coelho, E. Cunha, F. Curate, DXAGE: A New Method for Age at Death Estimation Based on Femoral Bone Mineral Density and Artificial Neural Networks, J. Forensic Sci. 63 (2018) 497–503. doi:10.1111/1556-4029.13582.
[73] J.J. Lynch, C.N. Stephan, Computational Tools in Forensic Anthropology: The Value of Open-Source Licensing as a Standard, Forensic Anthropology. 1 (2018) 228–243.

Figures captions:

Figure 1 – A selection of metric variables that can be measured and used for sex estimation with CADOES. For a complete list see Table 1. For visual description of landmarks see help tabs in the web app (http://osteomics.com/CADOES/).

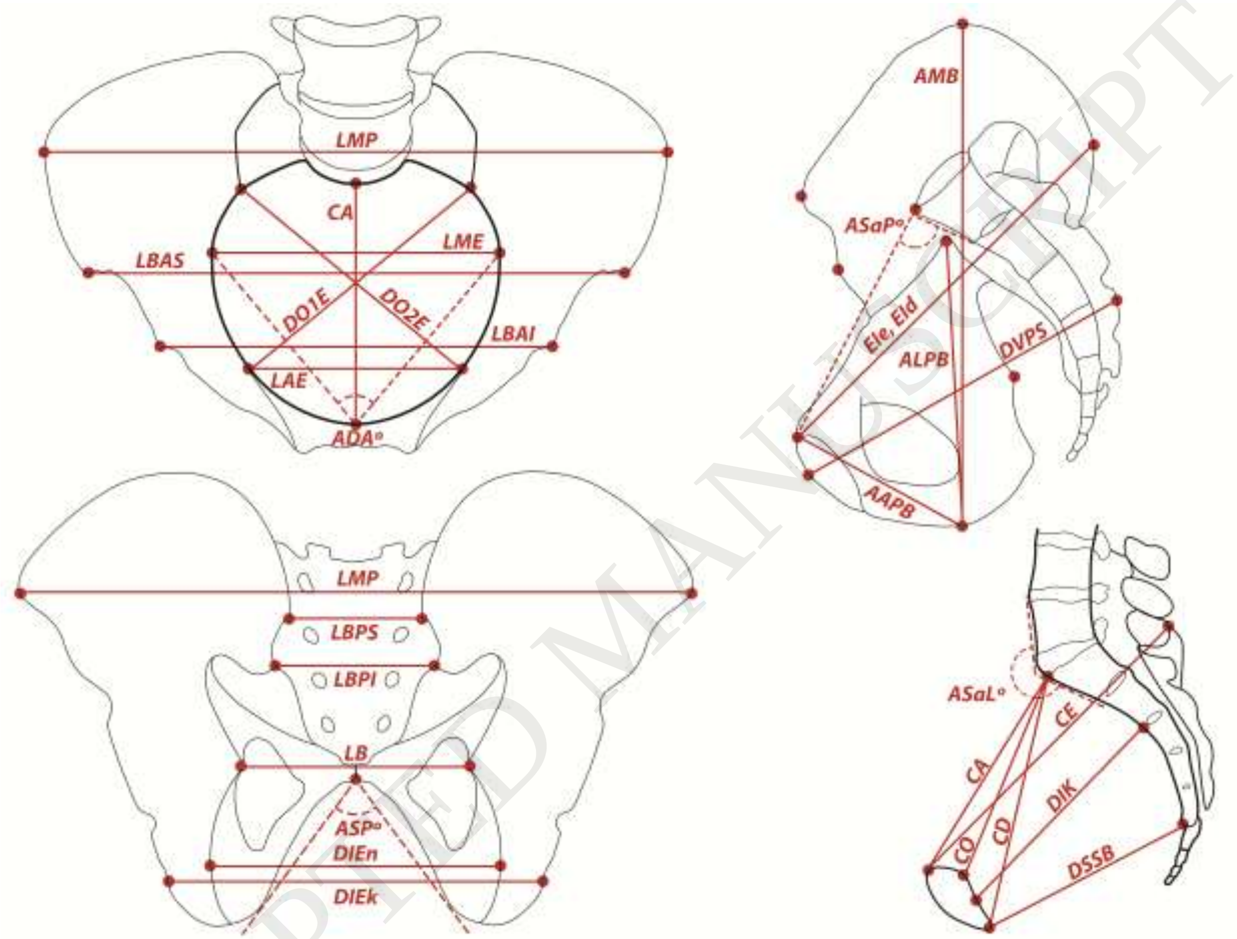

Figure 2 – CADOES (http://osteomics.com/CADOES/) allows for the generation of different models for sex estimation with metric features of the pelvis.

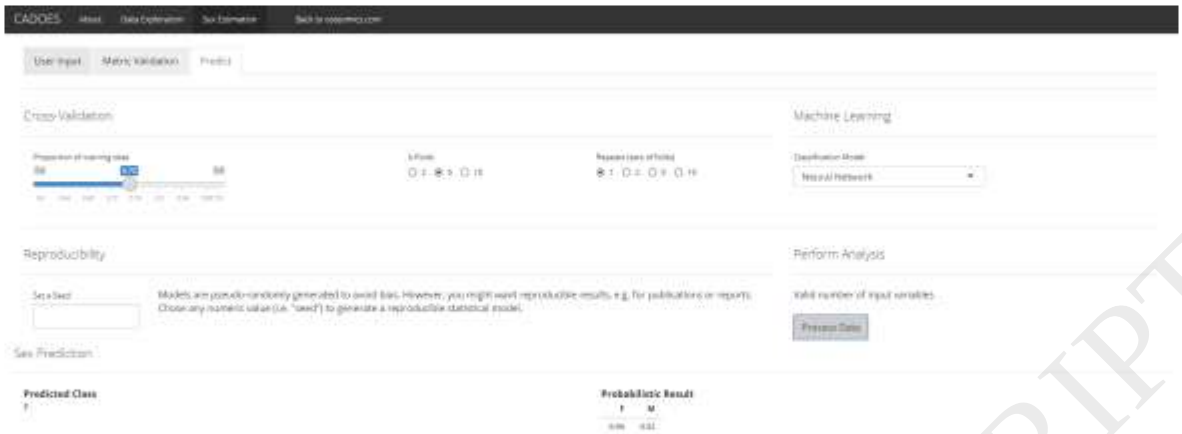
CADOES
User Input
Metric Validation
Predict
Cross-Validation
Machine Learning
Neural Network
Reproducibility
Perform Analysis
Process Data
Sex Prediction
Predicted Class
Probabilistic Result

Table 1: Variables available on CADOES that users can measure in order to create or test models, and to estimate sex of skeletal individuals.

Euclidean distances			
Codename	**Portuguese (original)**	**English**	**Definition**
CE	Conjugata externa	External conjugate	Diameter between the superior point of the pubic symphysis and the top of the spinous process of the 5th lumbar vertebra
CA	Conjugata anatómica	Anatomical conjugate	Distance between the sacral promontory and the anterosuperior margin of the pubic symphysis
CO	Conjugata obstétrica	Obstetric conjugate	Diameter from the sacral promontory to the posterosuperior point of the pubic symphysis
CD	Conjugata diagonalis	Diagonal conjugate	Diameter from the sacral promontory to the posteroinferior point of the pubic symphysis
DIK	Diâmetro inter-koilons	Inter-koilons diameter	Distance between the koilons (i.e., the deepest points) of the pubic symphysis (posterior symphysis koilon) and the anterior face of the sacrum (sacral koilon)
DSSB	Diâmetro sagital da saída da bacia	Inferior sacropubic diameter	Diameter from the sacral apex to the inferior point of the pubic symphysis
DVPS	Diâmetro venterpubes-sacral	External sagittal diameter	Distance from the most anterior point on the symphyseal surface to the median sacral crest (sacral spine)
LMP	Largura máxima da pelve	Maximum pelvic width	Maximum distance between the lateral margins of the iliac crests
LBAS	Largura bispinilíaca antero-superior	Macrospina width	Distance between the anterior superior iliac spines (left and right)
LBAI	Largura bispinilíaca antero-inferior	Microspina width	Distance between the anterior inferior iliac spines (left and right)
LBPS	Largura bispinilíaca postero-superior	Cryptospina width	Distance between the posterior superior iliac spines (left and right)
LBPI	Largura bispinilíaca postero-inferior	Metauricular width	Distance between the apex of the posterior inferior iliac spines (left and right)
LB	Largura bispinisquiática	Ischiatic spine width	Distance between the ischiatic spines (left and right)
DIEn	Diâmetro inter-endoischions	Inter-endoischions diameter	Distance between the endoischions; the *endoischion* is the point on the medial margin of the ischial tuberosity where it meets the sacrotuberous ligament
DIEk	Diâmetro inter-ektoischions	Inter-ektoischions diameter	Distance between the ektoischions; the *ektoischion* is the point on the lateral margin of the ischial tuberosity most distant to the sagittal/median plan of the pelvis.
DIKt	Diâmetro inter-kotilions	Inter-kotylions diameter	Distance between the kotylions; the *kotylion* is the mid-point in the acetabulum where the ilium, ischium and pubis bones converge.
DIKI	Diâmetro inter-koilons da incisura	Inter-koilon diameter	Minimum distance between the acetabular fossae. Measured by locating the deepest point (*koilon*) in the acetabulum.
DIP	Diâmetro inter-proobturatores	Inter-obturator diameter	Distance between the most superior points in the obturator foramen (*proobturatum*)

EIe	Espessura ilíaca (esquerda)	Iliac thickness (left)	Distance between the left posterior superior iliac spine (*cryptospina*) and the superior point of the pubic symphysis (*propubes*)
EId	Espessura ilíaca (direita)	Iliac thickness (right)	Distance between the right posterior superior iliac spine (*cryptospina*) and the superior point of the pubic symphysis (*propubes*)
LME	Largura máxima do estreito	Transverse diameter	Greatest mediolateral distance between the right and left arcuate lines
LAE	Largura anterior do estreito	Inter-pecten diameter	Inter-pecten distance, *pecten* is defined as the point where the arcuate line meets with the iliopubic eminence
DO1E	Diâmetro oblíquo I do estreito	Diagonal inlet I	Distance between the right ilio-auricular point (*proauricula*) and the point where the iliopectinal eminence meets the arcuate line on the left (*pecten*).
DO2E	Diâmetro oblíquo II do estreito	Diagonal inlet II	Distance between the left ilio-auricular point (*proauricula*) and the point where the iliopectinal eminence meets the arcuate line on the right (*pecten*).
AAIe	Altura da asa ilíaca (esquerda)	Iliac blade height (left)	Distance from the mid-point of the arcuate line (*arcuale*) to the most elevated point of the iliac (*epicrista*) on the left iliac
AAId	Altura da asa ilíaca (direita)	Iliac blade height (right)	Distance from the mid-point of the arcuate line (*arcuale*) to the most elevated point of the iliac (*epicrista*) on the right iliac
PFIe	Profundidade da fossa ilíaca (esquerda)	Iliac fossa depth (left)	Distance from the most elevated point of the iliac (*epicrista*) to the point in the arcuate line between the *arcuale* (arcuate line midpoint) and the *proauricula* (point where the arcuate line meets the sacrum)
PFId	Profundidade da fossa ilíaca (direita)	Iliac fossa depth (right)	
LS	Largura da saída	Outlet diameter	Greatest distance of the points located between the infero-posterior *obturator foramen* and the ischial spine, measured in the medial side of the hip bone.
AMB	Altura máxima da bacia	Os coxa height	Distance from the most inferior point in the ischial tuberosity (*ischion*) to the most superior point in the iliac (*epicrista*).
ALPB	Altura lateral da pequena bacia	Lateral height of the lesser pelvis	Distance from the most inferior point in the ischial tuberosity (*ischion*) to the ilio-auricular point (*proauricula)*
AAPB	Altura anterior da pequena bacia	Anterior height of the lesser pelvis	Distance from the most inferior point in the ischial tuberosity (*ischion*) to the point (*pecten*) where the iliopectinal eminence meets the iliopectinal line
Angles			
Codename	**Portuguese (original)**	**English**	**Definition**
ASP	Ângulo sub-púbico	Subpubic angle	Angular distance between the lines tangent to the inferior edge of the ischiopubic rami; vertex on the most inferior point where the pubic symphyses meet (*metapubes*).
AIPe	Ângulo de inclinação da pelve (esquerda)	Pelvic angle (left)	Avoid using this measurement.
AIPd	Ângulo de inclinação da pelve (direita)	Pelvic angle (right)	Avoid using this measurement.
ADA	Ângulo de divergência das asas	Iliac blades divergence angle	Angular distances between the left and right lines formed by the most lateral point of the iliac crest (*exocrista*) to the mid-point of the arcuate line (*arcuale*)

ASaP	Ângulo sacro-pélvico	Sacropelvic angle	Angle from the tangent of the ventral facet of the 1st sacral vertebra to the line defined by the anatomical conjugate.

Table 2: Descriptive statistics, including mean, standard deviation (SD), median, and minimum and maximum values, for pelvic measurements in both sexes; Coimbra Identified Skeletal Collection (CISC). All measurements in millimeters.

	Females					Males				
Measurement	**Mean**	**SD**	**Median**	**Min**	**Max**	**Mean**	**SD**	**Median**	**Min**	**Max**
CE	177.4	12.6	178.0	146	217	176.4	10.2	176.0	148.0	201.0
CA	112.6	10.0	113.0	87	139	105.6	9.6	105.0	83.0	136.0
CO	107.7	9.7	108.0	84	130	99.6	10.0	99.0	73.0	134.0
CD	122.9	9.8	122.5	100	145	117.9	10.3	117.0	92.0	150.0
DIK	127.9	9.3	128.0	103	150	121.1	7.6	121.0	101.5	145.0
DSSB	116.1	9.0	116.0	90	137	109.1	8.2	110.0	84.0	129.0
DVPS	162.5	11.7	162.5	133	195	159.9	9.4	160.0	140.0	189.0
LMP	262.5	17.3	261.5	220	306	261.8	13.4	261.0	232.0	299.0
LBAS	226.2	17.3	225.0	185	267	226.0	14.2	224.0	184.0	260.0
LBAI	186.6	12.0	187.5	137	213	186.8	12.2	188.0	130.0	220.0
LBPS	73.2	9.3	72.5	52	101	67.9	7.0	68.0	50.0	85.0
LBPI	88.2	6.42	88.0	72	109	87.5	5.2	87.0	75.0	102.0
LB	103.8	9.4	103.0	83	129	89.2	7.0	88.0	74.0	108.0
DIEn	122.9	12.2	122.0	102	159	110.4	10.6	110.0	85.0	134.0
DIEk	148.2	11.8	147.0	122	180	139.7	10.2	139.0	115.0	168.0
DIKt	138.7	10.2	139.5	113	164	136.0	10.2	137.0	107.0	161.0
DIKl	114.6	8.0	114.0	96	135	109.4	6.4	109.0	96.0	127.0
DIP	56.0	7.1	55.5	41	73	50.1	5.8	50.0	38.0	63.5
EIe	161.8	11.0	162.0	134	196	159.9	8.5	160.0	138.0	184.0
EId	163.6	11.5	163.0	135	199	161.3	8.2	161.0	136.0	182.0
LME	130.3	8.4	130.0	114	155	123.2	6.0	123.0	110.0	138.0
LAE	124.4	8.4	125.0	100	145	116.0	7.0	116.0	92.0	134.0
DO1E	124.6	7.4	124.0	108	145	118.3	6.0	118.0	106.0	131.0
DO2E	122.8	7.3	123.0	104	143	117.7	5.7	118.0	106.0	133.0
AAIe	99.8	6.1	100.0	80	114	106.2	6.1	106.0	92.0	119.0
AAId	98.6	6.0	99.0	78	113	104.9	5.6	104.0	92.0	118.0
PFIe	4.8	2.4	5.0	0	11	8.3	2.4	8.0	3.5	14.5
PFId	4.9	2.3	5.0	0	12	8.2	2.3	8.0	3.0	14.5
LS	111.6	9.3	112.0	88	134	99.5	7.3	99.0	86.0	116.0
AMB	190.7	9.5	191.0	167	211	206.5	9.9	207.0	181.0	229.0
ALPB	116.7	6.5	116.0	102	132	122.6	7.7	121.5	107.0	140.0
AAPB	110.9	7.8	111.0	88	129	115.1	5.8	115.0	101.0	128.0
ASP	79.3	8.6	80.0	52	97	66.5	8.7	67.0	43.0	90.0
AIPe	6.5	5.3	6.0	-8	25	6.2	5.9	6.0	-10.0	20.0
AIPd	5.1	5.3	5.0	-8	21	5.4	5.7	4.0	-8.0	19.0
ADA	98.9	10.1	100.0	77	126	98.6	9.3	98.0	78.0	123.0
ASaP	98.4	12.3	98.0	70	127	100.	14.3	101.0	67.0	138.0
ASaL	218.5	7.9	218.0	189	240	215.5	13.9	217.0	119.0	236.0

Table 3: The same three variables (AMB + LB + ASP) were selected in all models for sex estimation. Model accuracy for the training and testing (containing 30% of the data) samples. All other goodness-of-fit metrics were obtained from the testing set. All models generated using 19920804 as seed; 10 repeats of 10-fold cross-validation on a 70% data partition for training.

Model	Accuracy (Training)	Accuracy (Testing)	95%CI.lower	95%CI.upper	Kappa	Sensitivity	Specificity	AUC
kernelpls	0.9230	0.9200	0.8340	0.9701	0.8408	0.8462	1	0.9609
knn	0.9207	0.9067	0.8171	0.9616	0.8144	0.8205	1	0.5427
nb	0.9203	0.8667	0.7684	0.9342	0.7352	0.7692	0.9722	0.9467
lda	0.9218	0.9200	0.8340	0.9701	0.8408	0.8462	1	0.9610
fda	0.9060	0.8533	0.7527	0.9244	0.7084	0.7692	0.9444	0.9573
gamSpline	0.9275	0.9067	0.8171	0.9616	0.8140	0.8462	0.9722	0.9611
LogitBoost	0.8842	0.8533	0.7527	0.9244	0.7090	0.7436	0.9722	0.8971
plr	0.9207	0.9067	0.8171	0.9616	0.8144	0.8462	0.9722	0.9562
CART	0.8180	0.8533	0.7071	0.9244	0.7071	0.8205	0.8889	0.7803
rf	0.8784	0.8667	0.7684	0.9342	0.7357	0.7436	1	0.8055
gbm	0.9039	0.8667	0.7684	0.9342	0.7352	0.7692	0.9722	0.9507
nnet	0.9219	0.9200	0.8340	0.9701	0.8408	0.8462	1	0.9554

Partial Least Squares = kernelpls, k-Nearest Neighbors = knn, Naive Bayes = nb, Linear Discriminant Analysis = lda, Flexible Discriminant Analysis = fda, Generalized Additive Model using Splines = gamSpline, Boosted Logistic Regression = LogitBoost, Penalized Logistic Regression = plr, Decision Tree = CART, Random Forest = rf, Stochastic Gradient Boosting = gbm, Neural Network = nnet.

Table 4: Goodness-of-fit of the models when all 38 variables were used to estimate sex. Model accuracy for the training and testing (containing 30% of the data) samples. The other metrics were obtained from the testing set. All models generated using 19920804 as seed; 10 repeats of 10-fold cross-validation on a 70% data partition for training.

Model	Accuracy (Training)	Accuracy (Testing)	95%CI.lower	95%CI.upper	Kappa	Sensitivity	Specificity	AUC
kernelpls	0.9618	0.9733	0.9070	0.9968	0.9467	0.9487	1	0.9704
knn	0.9368	0.9600	0.8875	0.9917	0.9201	0.9231	1	0.7199
nb	0.9170	0.9200	0.8340	0.9701	0.8408	0.8462	1	0.9106
lda	0.9386	0.9467	0.8690	0.9853	0.8932	0.9487	0.9444	0.9642
fda	0.9128	0.9467	0.8690	0.9853	0.8932	0.9487	0.9444	0.9634
gamSpline	0.8989	0.9467	0.8690	0.9853	0.8932	0.9487	0.9444	0.2214
LogitBoost	0.8947	0.9067	0.8171	0.9616	0.8136	0.8718	0.9444	0.9017
plr	0.9108	0.9733	0.9070	0.9968	0.9467	0.9487	1	0.9690
CART	0.8019	0.8533	0.7527	0.9244	0.7071	0.8205	0.8889	0.7803
rf	0.9300	0.9600	0.8875	0.9917	0.9201	0.9231	1	0.9662
gbm	0.9131	0.9600	0.8875	0.9917	0.9200	0.9487	0.9722	0.9618
nnet	0.9630	0.9733	0.9070	0.9968	0.9467	0.9487	1	0.9673

Partial Least Squares = kernelpls, k-Nearest Neighbors = knn, Naive Bayes = nb, Linear Discriminant Analysis = lda, Flexible Discriminant Analysis = fda, Generalized Additive Model using Splines = gamSpline, Boosted Logistic Regression = LogitBoost, Penalized Logistic Regression = plr, Decision Tree = CART, Random Forest = rf, Stochastic Gradient Boosting = gbm, Neural Network = nnet.