Sex estimation: a comparison of techniques based on binary logistic, probit and cumulative probit regression, linear and quadratic discriminant analysis, neural networks, and naïve Bayes classification using ordinal variables

Efthymia Nikita1 · Panos Nikitas2

Received: 22 April 2019 / Accepted: 15 August 2019 © Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The performance of seven classification methods, binary logistic (BLR), probit (PR) and cumulative probit (CPR) regression, linear (LDA) and quadratic (QDA) discriminant analysis, artificial neural networks (ANN), and naïve Bayes classification (NBC), is examined in skeletal sex estimation. These methods were tested using cranial and pelvic sexually dimorphic traits recorded on a modern documented collection, the Athens Collection. For their implementation, an R package has been written to perform cross-validated (CV) sex classification and give the discriminant function of each of the methods studied. A simple algorithm that combines two discriminant functions is also proposed. It was found that the differences in the classification performance between BLR, PR, CPR, LDA, QDA, ANN, and NBC are overall small. However, LDA is simpler and more flexible than CPR, QDA, and ANN and has a small but clear advantage over BLR, NBC, and PR. Consequently, LDA may be preferred in skeletal sex estimation. Finally, it is striking that the combination of pelvic and cranial traits via their discriminant functions, determined either by BLR or LDA, removes practically any population-specificity and yields much better predictions than the individual functions; in fact, the prediction accuracy increases above 97%.

Keywords Forensic anthropology · Sex estimation · Cranium · Pelvis · Statistical methods

Introduction

Sex estimation is a key aspect of any forensic anthropological analysis as sex is one of the three key parameters used in the identification of an unknown individual, along with age and stature [1–3]. Even though sexual dimorphism in humans is small, there are morphological and metric differences between male and female skeletons, which have been used in sex estimation. Morphological methods focus primarily on the pelvis and secondarily on the cranium, and they assess the degree of expression of specific traits. Based on the qualitative assessment of this degree, they assign the individual to the male or female sex [4–6]. More recently, statistical analysis has started being employed, whereby the degree of expression of selected pelvic and cranial traits, recorded in an ordinal scale, is used as input variable in sex prediction using binary logistic regression analysis and/or discriminant analysis [7, 8]. This statistical approach allows not only the prediction of sex but also the estimation of the probability of an individual being male or female.

Discriminant analysis and, recently, machine learning techniques, such as artificial neural networks (ANN) and naïve Bayes classification (NBC), have also been used in metric sex estimation, whereby the variables are measurements that capture both size and shape differences between males and

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00414-019-02148-4) contains supplementary material, which is available to authorized users.

Efthymia Nikita
efi.nikita@gmail.com

Panos Nikitas
nikitas@chem.auth.gr

1 Science and Technology in Archaeology and Culture Research Center, The Cyprus Institute, 20 Konstantinou Kavafi Street, 2121 Aglantzia, Nicosia, Cyprus

2 Department of Chemistry, Aristotle University of Thessaloniki, University Campus, 54124 Thessaloniki, Greece

Published online: 23 August 2019
females [9–19]. A restriction of metric methods compared with morphological ones is that they are of limited applicability to fragmented partially preserved remains.

A comparison of different techniques used for sex classification based on ordinal variables was carried out by Walker [8]. Walker [8] used cranial traits and examined the performance of four multivariate techniques: k-nearest neighbor, binary logistic regression (BLR), linear (LDA), and quadratic (QDA) discriminant analysis. He found that QDA, and especially k-nearest neighbor analysis, performed poorly in terms of the sex bias criterion since they produced discriminant functions the classification error rates of which were much greater for one sex than the other. In contrast, BLR and LDA had both low misclassification and low sex bias rates. From these two techniques, Walker [8] chose BLR as the best technique, mainly because it relies on fewer assumptions than LDA. Comparisons between artificial neural networks (ANN) and the classical methods LDA, QDA, and BLR have been recently carried out and found that ANN perform better than the other methods. However, all these comparisons concern metric and not ordinal data [15–17].

The current paper examines the performance of the above classification methods, BLR, LDA, QDA, ANN, and NBC, as well as two alternatives of BLR, the probit (PR) and the cumulative probit (CPR) regression, in sex estimation using ordinal predictors. From these methods, the most popular is the first one, since it allows the direct development of mathematical relationships between sex traits and sex. These relationships can then be easily applied for sex estimation in different assemblages. In contrast, discriminant analyses, neural networks, and, naïve Bayes classification require that the training sample be used each time they are employed in sex estimation and this is a major disadvantage. For this reason, in the present paper, apart from the evaluation of these seven methods, we develop simple mathematical relationships for each of them that can be used for sex estimation, as is already the case for binary logistic regression. In addition, we explore the performance of a new approach that combines two different discriminant functions for sex prediction.

### Classification methods

#### Binary logistic and probit regression

Logistic and probit models are common statistical tools for the analysis of dichotomous data. In the context of sex prediction, binary logistic regression (BLR) presumes that the sex traits (predictors) have been properly coded, usually in an ordinal scale. Similarly, the sex of the individual may also be coded as a binary variable, say 0 for male and 1 for female. Consider the sex traits \(X_1, X_2, \ldots\), binary logistic regression establishes a relationship among these variables and the probability \(P(\text{sex} = 1)\) that an individual is female (\(\text{sex} = 1\)) based on the following expression [20]:

\[
\ln \frac{P(\text{sex} = 1)}{1 - P(\text{sex} = 1)} = c_0 + c_1X_1 + c_2X_2 + \cdots \tag{1}
\]

where \(c_0, c_1, c_2, \ldots\) are (adjustable) parameters usually determined using the maximum likelihood estimation provided that a training sample with documented sex vs. \(X_1, X_2, \ldots\) data is available. When \(c_0, c_1, c_2, \ldots\) have been calculated, the above equation can be used to calculate \(P(\text{sex} = 1)\). Thus, if \(y = c_0 + c_1X_1 + c_2X_2 + \cdots\)

\[
\text{then }
\]

\[
P(\text{sex} = 1) = \frac{1}{1 + e^{-y}} \tag{3}
\]

Therefore, if \(P(\text{sex} = 1) > 0.5\), the individual is female; otherwise, the individual is male with a probability equal to \(P(\text{sex} = 0) = 1 - P(\text{sex} = 1)\).

Tables with \(c_0, c_1, c_2, \ldots\) values or the expression \(y = c_0 + c_1X_1 + c_2X_2 + \cdots\) can be found in the literature and they can be used to compute \(P(\text{sex} = 1)\) by means of Eq. (3) and, therefore, estimate sex from specific traits. Such a well-known table is the one given by Walker [8] with \(y\) equations for estimating sex from cranial traits based on American/English and Native American assemblages. In this table, as well as in all relevant tables, the percentage of correctly classified males/females is also presented. This is an important piece of information because it evaluates the performance of the regression equations when applied to a certain sample. This information is obtained by means of leave-one-out cross-validation (LOOCV). According to this technique, we remove from the dataset the first case, apply binary logistic regression to the remaining cases, and, based on the obtained regression equation, we estimate the sex of the held-out case. This procedure is repeated for the second case and so on until the last case. Then we count the correct predictions and express them as a percentage of correctly classified cases.

In probit regression (PR), the inverse of the standard normal cumulative distribution \(\Phi^{-1}\) of the probability \(P(\text{sex} = 1)\) is modeled as a linear combination of the predictors [20, 21]:

\[
\Phi^{-1}(P(\text{sex} = 1)) = c_0 + c_1X_1 + c_2X_2 + \cdots \tag{4}
\]

where the adjustable parameters \(c_0, c_1, c_2, \ldots\) may also be determined using the maximum likelihood method as in BLR. When \(c_0, c_1, c_2, \ldots\) are known, the above equation can be used to calculate \(P(\text{sex} = 1)\) since:

\[
P(\text{sex} = 1) = \Phi(y) = \Phi(c_0 + c_1X_1 + c_2X_2 + \cdots) \tag{5}
\]

where \(\Phi\) is the standard normal cumulative distribution function. This function may be easily calculated using for example
the NORMSDIST function in Excel or the pnorm function of R.

BLR and PR are not subject to strict assumptions. Thus, they do not require a linear relationship between the dependent and independent variables; the predictors do not need to be normally distributed; they can handle both categorical and continuous variables, whereas homoscedasticity is not required [20]. In addition, both methods are sensitive to outliers but this is not a serious issue when using ordinal data. The difference between logistic and probit models lies in the assumptions about the distribution of the errors in order to use maximum likelihood estimation. In PR, the error terms should be normally distributed, whereas in BLR, they should have a logistic distribution.

**Cumulative probit regression**

The use of binary logistic and probit models in sex assessment and in particular the validity of the estimated probabilities \( P(\text{sex} = 1) \) was criticized by Konigsberg and Hens [21] within the frames of Bayes’ Theorem. These authors proposed as a better alternative the cumulative probit model/regression (CPR). Here, we applied the cumulative probit model as follows. First, consider the simple case of just one sex indicator, \( X_1 \), which can be coded using a J-point ordinal scale. That is, it can take the integer values \( i = 1, 2, ..., J \). If we model the ordinal variable \( X_1 \) as a function of the binary variable sex (independent variable), the fitting probit model is expressed not by just one equation but from \((J-1)\) equations of the form:

\[
\Phi^{-1}(P_{ci}) = a_i + b \times \text{sex}, \quad i = 1, 2, ..., J-1
\]  

(6)

where \( P_{ci} \) is the cumulative probability of category \( i \), that is, the probability of \( X_1 \) falling in category \( i \) or below. Therefore,

\[
P_{ci} = P_1 + P_2 + \cdots + P_i
\]  

(7)

where \( P_1, P_2, ..., P_i \) is the probability of \( X_1 \) falling exactly in category \( 1, 2, ..., i \), respectively. In fact, \( P_1, P_2, ..., P_i \) are conditional probabilities and, in particular, \( P_i \) is the probability conditional on sex that an individual is in the \( i \)th category of the sex indicator \( X_1 \). When we fit the probit model to the ordinal response \( X_1 \) using sex as an independent variable, we obtain the adjustable parameters \( a_1, a_2, ..., a_{J-1}, b \) and then the conditional probabilities may be calculated from the following system of equations:

\[
P_{c1} = P_1 = \Phi(a_1 + b \times \text{sex})
\]

\[
P_{c2} = P_1 + P_2 = \Phi(a_2 + b \times \text{sex})
\]

\[
P_{c(i-1)} = P_1 + P_2 + \cdots + P_{i-1} = \Phi(a_{i-1} + b \times \text{sex})
\]  

(8)

Therefore, for \( i > 1 \), we have

\[
P_i = \Phi(a_i + b \times \text{sex}) - \Phi(a_{i-1} + b \times \text{sex})
\]  

(9)

Now, if the observed score of \( X_1 \) for an individual is \( i \), the probability that the individual is a female may be estimated by applying Bayes’ Theorem [21].

\[
P(\text{sex} = 1|X_1 = i) = \frac{P_i(\text{sex} = 1)P_F}{P_i(\text{sex} = 0)P_M + P_i(\text{sex} = 1)P_F}
\]  

(10)

where \( P_M \) is the prior probability that an individual is male and \( P_F \) is the prior probability that an individual is female. These quantities may be estimated, as in discriminant analysis, from \( P_M = n_M/n, P_F = n_F/n \) where \( n_M, n_F \) is the number of males and females, respectively, and \( n \) is the total number of cases. Alternatively, they may be estimated via an optimization procedure that searches for the \( P_M, P_F \) values that yield the optimum prediction performance in the training sample.

At this point, we should clarify the following. In the original paper by Konigsberg and Hens [21], function \( \Phi(x) \) is defined as the standard normal integral from \( x \) to infinity, whereas in the present paper, \( \Phi(x) \) is the standard normal integral from -infinity to \( x \). Thus, their sum is equal to 1. If we take this into account, as well as the different notation used by Konigsberg and Hens [21], Eqs. (9) and (10) are identical to Eqs. (8) and (9) proposed in [21]. Therefore, up to this point, i.e., for the simple case of just one sex indicator, the method described above is close to the original approach proposed by Konigsberg and Hens. For many sex indicators, the following variation has been adopted. When there are many sex traits, \( X_1, X_2, ..., \) Eq. (10) is still valid but now \( P_i = P_i \) is the probability conditional on sex that an individual is in the state \( I \) composed of the \( i \)th category of \( X_1 \), \( i_j \)th category of \( X_2 \), and so on. In this case, we may calculate the conditional probabilities \( P_1, P_2, ..., P_J \) for each sex indicator \( X_1, X_2, ... \) and then \( P_i \) may be estimated under the assumption of conditional independence. This is the main assumption of this approach. Note that the assumption of conditional independence is widely used in statistical studies, like in the naive Bayes classification as well as in the Bayesian age estimation. In particular, we should stress that the naive Bayes classification uses also the extension of Eq. (10) for many sex traits, \( X_1, X_2, ..., \) to estimate sex probabilities (see below).

The above approach is simple and applicable using any software that fits the probit model. An alternative approach, closer to the Konigsberg-Hens study, could be the following. Instead of fitting a univariate probit model, Eq. (8), we fit a multivariate cumulative probit model. This fitting gives the adjustable parameters \( a_1, a_2, ..., a_{J-1}, b \) of all traits \( X_1, X_2, ... \) and the probability \( P(\text{sex} = 1|X_1 = i_1, X_2 = i_2, \ldots) \) is estimated after the generalization of Eqs. (9) and (10) that now contain multivariate normal integrals in place of the univariate standard normal integral \( \Phi \).
Discriminant analysis

An alternative approach for sex classification and prediction is discriminant analysis (DA). DA is based on the Mahalanobis distance between \( x = (X_1, X_2, \ldots, X_m) \) and the mean vectors \( \mu_0 \) and \( \mu_1 \), where \( x \) is a vector of sex traits (variables) and \( \mu_0 \) and \( \mu_1 \) are the mean vectors of the male and female subsamples.

There are two main variants of DA: quadratic (QDA) and linear (LDA) discriminant analysis. In QDA, the probability \( P(\text{sex} = 1) \) is estimated via the following relationships [22, 23]:

\[ P(\text{sex} = 1) = \frac{e^{f_1(x)}}{e^{f_1(x)} + e^{f_0(x)}} \]  

(11)

\[ d_k(x) = -\frac{1}{2}(x-\mu_k)'\Sigma_k^{-1}(x-\mu_k) - \frac{1}{2}\ln|\Sigma_k| + \ln P_k \]  

(12)

where \( k = 0 \) or 1 are the classes of male and female, respectively, \( \Sigma_k \) is the covariance matrix for class \( k \), and \( P_k \) are the prior probabilities used. The last quantity is estimated from \( P_k = n_k/n \), where \( n_k \) is the number of cases \( k \) and \( n \) is the total number of cases. Note that the first term in Eq. (12) is the so-called Mahalanobis distance between \( x \) and \( \mu_k \). It can be shown that Eqs. (11) and (12) may be expressed as Eqs. (2) and (3) where \( y \) is given by:

\[ y = a_0 + \sum_{i=1}^{m} a_i X_i + \sum_{i=1}^{m} b_i X_i^2 + \sum_{i<j}^{m} c_{ij} X_i X_j \]  

(13)

QDA is reduced to LDA if we assume \( \Sigma_k = \Sigma \), where \( \Sigma \) is the pooled sample covariance matrix. In this case, Eq. (13) is simplified to

\[ y = a_0 + \sum_{i=1}^{m} a_i X_i \]  

(14)

Equations (13) and (14) can be used like Eq. (2) for sex estimation. That is, based on a training dataset, the probabilities \( P = P(\text{sex} = 1) \) are estimated for each case using linear or/and quadratic discriminant analysis. Then the adjustable parameters \( a_0, a_i, b_i, c_{ij} \) may be calculated by means of linear or multilinear regression using \( \log(P/(1-P)) \) as a dependent variable and \( X_i \) or \( X_i, X_i^2, X_i X_j \) as independent variables. Sex prediction in the target sample is carried out via Eqs. (3) and (14) or/and (13).

Discriminant analysis is a parametric multivariate technique and, therefore, it is subject to several assumptions:

- The size of the smallest group should exceed the number of sex traits (independent variables); the variables should follow the multivariate normal distribution; the variance/covariance matrices of variables should be homogenous across groups, whereas multi-collinearity among variables should be excluded. Note that the equality of covariance assumption is not required in QDA and this is the basic reason that LDA is a much less flexible classifier than QDA. Nonetheless, as Tabachnick and Fidell [20] point out, if classification is the primary goal, then most of the above requirements do not affect the classification performance, provided that there are no outliers. Deviation from these assumptions may distort only tests of statistical significance, just as in MANOVA, if such tests have been carried out.

**Neural networks**

An artificial neural network (ANN) is a system of interconnected neurons, which is used for classification via processes that mimic the way the human brain works [24, 25]. ANN is one of the main tools used in machine learning and it can be used for sex classification and prediction. A simple neural network is shown in Fig. 1. When the neural network is used for sex classification, the leftmost layer of the network (input layer) consists of \( m \) input units that represent the traits \( X_1, X_2, \ldots, X_m \), whereas the output layer has only one node, which is related to the sex variable. Note that this variable, i.e., the output of the network, takes values between 0 and 1 and, therefore, it is related to the probability \( P(\text{sex} = 1) \). In the example of Fig. 1, there is just one hidden layer with three nodes. It is called hidden layer because its values are not observed in the training set. The circles labeled “1” are called bias units.

In every neural network, the input values \( X_1, X_2, \ldots, X_m \) are transformed to an output value via proper weights, \( w_i \). Thus, a neural network with \( m \) input nodes and one hidden layer with \( p \) nodes represents the following function provided that the numbering of the weights is as in Fig. 1 [25, p., 143]:

\[ \text{sex} = f \left( w_{(m+1)p+1} + \sum_{i=1}^{p} w_{(m+1)p+i} H_i \right) \]  

(15)

where

\[ H_i = f \left( w_{(i-1)(m+1)+1} + \sum_{j=1}^{m} w_{(i-1)(m+1)+j} X_j \right) \]  

(16)

Here, \( f \) is the activation function of the network and \( \text{sex} \approx P(\text{sex} = 1) \), i.e., we may at least as a first approximation adopt that the ANN output is the probability \( P(\text{sex} = 1) \). In what concerns the function \( f \), we have adopted the logistic function \( f(x) = e^x/(1+e^x) \).

It is seen that if we know the number of inputs and the hidden nodes as well as the weights, the estimation of sex is straightforward via Eqs. (15) and (16). There are several algorithms to train a network and compute weights. The most interesting point is that all these algorithms do not assume any underlying pattern for the data.
Naïve Bayes classification

Naïve Bayes classification (NBC) also belongs to the field of machine learning. It is a simple probabilistic classification method based on Bayes' theorem with the “naïve” assumption of conditional independence. Within the frames of the present study, NBC may be considered to be related to the cumulative probit regression method, CPR, described above, because it uses the same relationship, Eq. (10), and its extension for many sex traits, \( X_1, X_2, \ldots \), to estimate sex probabilities. However, in CPR, the conditional probabilities are computed via probit regression, whereas in NBC, they are calculated directly from the counts table of the training data.

Statistical methods

Fifteen R functions have been written to perform cross-validated sex estimation based on the above classification methods: CVBLR, CVPR, CVCPR, CVLDA, CVQDA, CVNN, CVNBC, CPRopt, CPRpm, CPRkh, sdiscr, sdiscr2, discrNN, discrQDA, and discrCPR. The first five functions perform cross-validated binary logistic and probit regression, cumulative probit regression, and linear and quadratic discriminant analysis. The CVNN function is used to implement artificial neural networks and the CVNBC function implements naïve Bayes classification. The CPRopt function searches for optimum priors that may be used for the cumulative probit regression using predefined priors, and the CPRkh performs cumulative probit regression without cross-validation and stepwise variable reduction based on the Konigsberg-Hens [21] method. The next function, sdiscr, may be used to estimate the sex and the corresponding probability for one or more individuals using an equation of the form of Eqs. (2), (3), (5), (13), and (14), i.e., equations that may be determined from the CVBLR, CVPR, CVLDA, and CVQDA functions. The sdiscr2 function can combine two discriminant functions that arise either from the BLR or LDA methods. Finally, the discrNN function estimates the sex and the corresponding (approximate) probability by means of weighting factors obtained from the CVNN function using Eq. (15); the discrQDA function uses Eqs. (11) and (12) instead of Eq. (13) for predictions based on QDA, and the last function, discrCPR, can be used for predictions based on the CPR method. The last function may also be used to make predictions based on the NBC method.

Binary logistic and probit regression are implemented using the basic glm function in combination to the CVbinary function of the DAAG library to perform leave-one-out cross-validation (LOOCV). In addition, the stepAIC function of the MASS library is used for backward stepwise model selection, i.e., for variable reduction, which is a crucial step for building a simpler model without losing the predictive power of the data. Linear and quadratic discriminant analyses are performed via the lda and qda functions of the MASS library. These functions can also perform LOOCV. For variable
reduction, the stepclass function of the klaR library is used adopting the backward technique and the accuracy (AC) as performance measure. For the CPR method, we used the polr function of the MASS library. The CVCPR function may perform LOOCV but it does not have the option for variable reduction. Note that in the CVCPR function, the probability \( p(\text{sex} = 1) \) for many sex indicators \( X_1, X_2, \ldots \) is estimated under the assumption of conditional independence. To test the alternative approach of fitting a multivariate cumulative probit model, i.e., the Konigsberg-Hens [21] method, we used the mvord function of the mvord library in the CPRkh function. In this function, the cumulative distribution function of the multivariate normal distribution was calculated using the pmvnorm function of the mvtnorm library. The basic code is similar to that proposed by Konigsberg-Frankenberg [26]. Note that, due to the mvord function, this approach may be very slow at certain datasets and this is the reason we have not included the option of cross-validation.

For the implementation of ANN, the R function nnet of the homonymous library was adopted. This function is based on the feed-forward algorithm, which is a rather fast algorithm. However, we should point out that, irrespective of the algorithm used for the estimation of the weights, different runs of an ANN algorithm may give different weights, especially if their number is large enough. For this reason, for the neural network training, we used 20 repetitions and the networks with the optimum performance were selected for sex prediction. It is evident that the ANN performance depends upon the number of the hidden nodes. This number was varied from 1 to 40. An important consideration when using ANN is overfitting issues and, for this reason, it is important to perform LOOCV. Note that nnet does not have the option to perform LOOCV, and for this reason in the present study, we have written the relevant code. In what concerns variable reduction, this may be achieved qualitatively via Garson’s plot, which shows the relative importance of the input variables. However, in the present study, we instead chose to use the input layer the variables obtained from BLR/LDA/QDA.

The problem of variable reduction appears also in NBC and it was approached as in ANN. For the implementation of NBC, the naive_bayes function of the naivebayes library was used.

Finally, we attempted to combine two discriminant functions based on different skeletal traits, and in particular one based on cranial sex traits and another based on pelvic traits, to test if this combination would improve the performance of the individual functions. It is known that models/discriminant functions based on cranial traits exhibit strong population-specificity; thus, they have a rather poor prediction performance especially if the discriminant function has been obtained from a different population. In contrast, discriminant functions based on pelvic traits give much better predictions (see the “Results and discussion” section). If two such functions give the same prediction for the sex of an individual, then we can be quite confident for the estimated sex. The problem occurs when the two functions/models give different predictions. For example, consider that the discriminant function based on cranial traits predicts that the sex of an individual is male with a probability say \( P_1(\text{sex} = 0) = 0.9 \), whereas the discriminant function based on pelvic traits predicts that the sex of this individual is female with a probability say \( P_2(\text{sex} = 1) = 0.6 \). In this case, for sex estimation, we need a selection rule. There are several choices: We may select the sex based exclusively on the probability of the best model, i.e., the model based on pelvic traits. However, this choice excludes completely the model based on cranial traits. Alternatively, we may disregard the performance of the discriminant functions and make the selection based exclusively on the predicted probabilities. Thus, in the above example, we select the prediction of the discriminant function that uses cranial traits because simply the probability of the individual being male is greater than that of being female (0.9 > 0.6). To take into account the performance of the discriminant functions, we may multiply the above probabilities by proper weighting factors that come from the performance of the discriminant functions at the target population.

In the present study, we found that the simple selection rule that focuses exclusively on the predicted probabilities is very effective when one function is based on cranial traits and the other on pelvic traits. In addition, we found that this selection rule gives results identical to those obtained from the formula for the joint probability of the individual being a female (as suggested by an anonymous reviewer): \( P_1(\text{sex} = 1) \times P_2(\text{sex} = 1)/(P_1(\text{sex} = 1) \times P_2(\text{sex} = 1) + P_1(\text{sex} = 0) \times P_2(\text{sex} = 0)) \). It is evident that the above two options to combine two discriminant functions give identical results about the classification but not about the probabilities since the first option selects the maximum probability, whereas the second option computes a (new) joint probability. Although these two selection options can be applied manually, an R script has been written and is implemented in the sdiscr2 function.

More details and the code of the R functions are given as Online Resources 1 and 2.

**Materials**

The Athens Collection, housed at the Department of Animal and Human Physiology at the National and Kapodistrian University of Athens, Greece, consists of 225 skeletons. Most of this material has documented age, sex, occupation, and cause of death and represents individuals who lived in the second half of the twentieth century and were buried in cemeteries in the area of Athens [27]. Of this collection, 191 skeletons (106 males and 85 females) were studied in the context
of this paper, excluding subadults, individuals with insufficient documentation, and individuals with pathological lesions or taphonomic damage that could inhibit the correct recording of the sex traits under examination. Note that the Athens Collection is divided in two parts; the skeletons that formed the original core of this collection have the coding WLH, while the skeletons that were added in the collection subsequently received a code of ABH. From these data, the 132 skeletons with ABH code were used for training and the 59 skeletons with WLH code were used for prediction.

The pelvic traits recorded included the ventral arc (VA), subpubic concavity (SPC), and medial aspect of the ischiopubic ramus (MAR) and their scoring followed the 5-grade scheme proposed by Klales et al. [7]. The cranial traits included the mental eminence (ME), supraorbital margin (ORB), supraorbital ridge/glabella (GL), nuchal crest (NC), and mastoid process (MA), recorded as described by Walker [8] using also a 5-grade scheme. Part of this dataset (120 cases) is provided as Online Resource 3 in order to practice using the provided R code. The entire dataset is available upon request by the authors.

Results and discussion

The results obtained are presented in Tables 1, 2, 3, 4, 5, 6, and 7 as well as in the tables and figures given as Online Resource 4. Tables 1, 2, and 3 show the percentage of correctly classified individuals by means of BLR, PR, LDA, and QDA, whereas Tables 4 and 5 present selected results concerning the classification performance of the neural networks when the number $p$ of the nodes of the hidden layer is equal to 1 and when $p$ gets its optimum value, i.e., the value that yields optimum sex classification. Table 6 shows the discriminant functions of LDA that result in the optimum sex classification for the Athens Collection and, finally, Table 7 shows the percentage of correctly classified individuals when two discriminant functions, one using pelvic and the other cranial traits, are combined. Note that some small differences in the percentages observed between Table 7 and Tables 1 and 2 are due to the different way that the missing values are eliminated in the relevant functions. The results concerning the CPR and NBC methods are presented in Tables S1 to S4 and Fig. S1. In addition, Online Resource 4 contains the complete results obtained from the neural networks (Tables S5 to S7 and Fig. S2). Note that in all ANN results, the value 0.5 was used for the weights decay.

From Tables 1, 2, and 3, we obtain that the differences between BLR, PR, LDA, and QDA are overall small; LDA gives slightly better results than BLR in 36% of the cases given in these tables, the opposite is valid in 26% of the cases, and the two methods give identical results in the remaining 38% of the cases. However, if we examine the sex prediction on the target sample, we observe that LDA generally gives better and more balanced results between sexes. For the differences between LDA and QDA, the above figures are 24%, 18%, and 58%, respectively. It is seen that although QDA includes many more terms than LDA in its discriminant function, Eqs. (13) and (14), it does not give better results. In fact, given this large number of parameters, we readily conclude that QDA does not offer a better alternative to either LDA or BLR.

### Table 1 Percentage of correctly classified individuals by means of BLR, PR, LDA, and QDA based on pelvic traits

| Dataset     | Accuracy | BLR/PR | LDA | QDA |
|-------------|----------|--------|-----|-----|
|             |          | VA, SPC, MAR | VA, SPC | VA, SPC, MAR | VA, SPC | VA, SPC |
| Original    | CV (T)   | 94.74  | 96.32 | 96.84  | 96.32  | 96.84  | 95.79 |
| Original    | CV (M)   | 96.19  | 97.14 | 99.05  | 97.14  | 99.05  | 97.14 |
| Original    | CV (F)   | 92.94  | 95.29 | 94.12  | 95.29  | 94.12  | 95.29 |
| Original    | Prediction (T) | 96.32  | 96.32 | 96.84  | 96.32  | 96.84  | 95.79 |
| Original    | Prediction (M) | 96.19  | 97.14 | 99.05  | 97.14  | 99.05  | 97.14 |
| Original    | Prediction (F) | 96.47  | 95.29 | 94.12  | 95.29  | 94.12  | 95.29 |
| Training    | CV (T)   | 96.18  | 96.95 | 98.48  | 96.95  | 98.48  | 96.84 |
| Training    | CV (M)   | 96.97  | 98.48 | 95.38  | 95.38  | 95.38  | 95.38 |
| Training    | CV (F)   | 95.38  | 95.38 | 96.95  | 96.95  | 96.95  | 96.95 |
| Training    | Prediction (T) | 96.97  | 96.95 | 96.95  | 96.95  | 96.95  | 96.95 |
| Training    | Prediction (M) | 96.97  | 98.48 | 98.48  | 98.48  | 98.48  | 98.48 |
| Training    | Prediction (F) | 96.92  | 95.38 | 95.38  | 95.38  | 95.38  | 95.38 |
| Target      | Prediction (T) | 91.53  | 91.53 | 94.87  | 94.87  | 94.87  | 94.87 |
| Target      | Prediction (M) | 92.31  | 92.31 | 90.00  | 90.00  | 90.00  | 90.00 |
| Target      | Prediction (F) | 90.00  | 90.00 | 90.00  | 90.00  | 90.00  | 90.00 |
| Target      | Sex bias  | 2.31   | 2.31  | −0.13  | −0.13  | −0.13  | −0.13 |

Key: $T$ total, $M$ male, $F$ female; sex bias is the difference between male and female percentage correct predictions.
From these tables we also observe that the logit model produces results very similar to probit regression. Therefore, the choice of probit versus logit depends largely on individual preferences. The same holds for the cumulative probit model as well as for NBC (Fig. S1, Tables S1 to S4). Although there are differences between the results obtained from BLR, CPR, and NBC, these differences are rather random and do not indicate that any of these models provides better predictions. Note that the results of the cumulative probit model are not improved, at least in most of the cases examined in this study, if we optimize the prior probabilities, \(P_M\), \(P_F\) (Table S2). In addition, they are not improved even if we use the original, Konigsberg-Hens method (Table S3).

In what concerns the ANN method, its performance strongly depends upon the number \(p\) of the nodes in the hidden layer. Best results are usually obtained for \(p\) ranging from 10 to 40. This means that the discriminant functions expressed via Eq. (15) include a great number of terms, which makes their use difficult, especially if we have to publish such equations for use by other scholars. If, in addition, we compare the prediction performance of ANN with that of LDA using the combination of traits for which LDA gives the best results, we readily conclude that LDA should be preferred over ANN.

Therefore, from the classification methods studied, LDA has a small but clear advantage over the other approaches and it may be preferred in sex estimation. However, one critical question concerning the applicability of LDA is the fulfillment of the various assumptions that should be met for the method to give reliable results. The most important of these assumptions are the normality, equality of variances/covariances, and the lack of multi-collinearity and outliers. Note that for BLR, the only assumptions that should be taken into account are the lack of multi-collinearity of the sex traits and the lack of outliers. The use of ordinal variables minimizes the existence of outliers in all methods. In R, multi-collinearity may be tested using the vif function of the car library. The application of this function showed that there is no multi-collinearity among the sex traits except for ORB when it is examined in combination with pelvic traits. Therefore, only the results concerning the eight variables in Table 3 should be treated cautiously. In what concerns normality and equality of variances/covariances across groups, both are not valid when the predictors (sex traits) are ordinal variables. The reason is that ordinal data cannot be drawn from a multivariate Gaussian distribution and the Box’s M test that is usually adopted to test the homogeneity of variance-covariance matrices is very sensitive to violations of normality, leading to the rejection of this assumption. Note that in R, the Box’s M test may be performed by means of the boxM function of the biotools library and its application to the datasets under study showed that, indeed, the homogeneity assumption is violated. Thus, two basic assumptions underlying LDA application are violated. However, the improved prediction performance of LDA is a strong indication that the method is fairly robust to violations of these assumptions when classification is the primary goal. This is in line with Tabachnick and Fidell [20], who also point out that if a 95% accuracy in classification is achieved, “you hardly worry about the shape of distributions” (20; p. 381). Thus, the fact that the accuracy of LDA is comparable with and often better than that of BLR shows that this method can be applied for sex classification without tests for the underlying pattern of the data, as in the BLR and ANN methods.

### Table 2 Percentage of correctly classified individuals by means of BLR, PR, LDA, and QDA based on cranial traits

| Dataset | Accuracy | BLR/PR | LDA | QDA |
|---------|----------|--------|-----|-----|
|         |          | All 5 vars | GL, MA, NU | GL, MA, ORB | All 5 vars | GL, MA, NU | GL, MA, ORB | All 5 vars | GL, MA, NU | GL, MA, ORB |
| Original | CV (T) | 88.70/89.83 | 87.43/89.01 | 89.53 | 89.83 | 89.01 | 89.84 | 89.27 | 89.01 | 89.01 |
| Original | CV (M) | 87.50/88.54 | 86.79 | 86.79 | 87.50 | 86.79 | 84.91 | 87.50 | 86.79 | 85.85 |
| Original | CV (F) | 90.12/91.36 | 88.24/91.76 | 92.94 | 92.59 | 91.76 | 92.94 | 91.36 | 91.76 | 92.94 |
| Original | Prediction (T) | 90.40 | 89.01 | 89.53 | 89.83 | 89.01 | 88.48 | 90.96 | 90.05 | 89.53 |
| Original | Prediction (M) | 89.58 | 86.79 | 86.79 | 87.50 | 86.79 | 84.91 | 88.54 | 86.79 | 86.79 |
| Original | Prediction (F) | 91.36 | 91.76 | 92.94 | 92.59 | 91.76 | 92.94 | 93.83 | 94.12 | 92.94 |
| Training | CV (T) | 90.76/91.6 | 90.15/91.67 | 90.15 | 92.44 | 92.44 | 91.67 | 90.76 | 90.91 | 92.42 |
| Training | CV (M) | 89.66 | 86.57/88.06 | 88.06 | 87.93 | 87.93 | 86.57 | 89.66 | 88.06 | 88.06 |
| Training | CV (F) | 91.80/93.44 | 93.85/95.38 | 92.31 | 96.72 | 96.72 | 96.92 | 91.80 | 93.85 | 96.92 |
| Training | Prediction (T) | 94.12 | 91.67 | 91.67 | 92.44 | 92.44 | 92.42 | 92.44 | 92.42 | 92.42 |
| Training | Prediction (M) | 91.38 | 88.06 | 88.06 | 87.93 | 87.93 | 88.06 | 90.66 | 88.06 | 88.06 |
| Training | Prediction (F) | 96.72 | 95.38 | 95.38 | 96.72 | 96.72 | 96.92 | 95.08 | 96.92 | 96.92 |
| Target | Prediction (T) | 84.48 | 84.75 | 83.05 | 86.21 | 84.75 | 81.36 | 86.21 | 83.05 | 83.05 |
| Target | Prediction (M) | 89.47 | 87.18 | 87.18 | 88.48 | 84.62 | 82.05 | 89.47 | 84.62 | 84.62 |
| Target | Prediction (F) | 75.00 | 80.00 | 75.00 | 85.00 | 85.00 | 80.00 | 80.00 | 80.00 | 80.00 |
| Target | Sex bias | 14.47 | 7.18 | 12.18/14.74 | 1.84 | −0.38 | 2.05 | 9.47 | 4.62 | 4.62 |
Table 3: Percentage of correctly classified individuals by means of BLR, PR, LDA, and QDA based on combinations of cranial and pelvic traits

| Dataset | Accuracy | BLR/PR | LDA | QDA |
|---------|----------|--------|-----|-----|
|         |          | GL, VA, MAR | GL, VA, SPC | GL, VA, SPC |
| All 8 vars | GL, VA, SPC, MAR | GL, VA, SPC | MA, VA, SPC | All 8 vars |
| Original CV (T) | 97.16 | 98.95 | 97.89 | 96.8/96.3 | 98.86 | 98.95 | 97.89 | 96.84 | 98.30 | 99.47 | 97.37 | 96.84 |
| Original CV (M) | 97.89 | 99.05 | 98.10 | 97.1/96.2 | 100.0 | 100.0 | 98.10 | 97.14 | 98.95 | 100.0 | 98.10 | 97.14 |
| Original CV (F) | 96.30 | 98.82 | 97.65 | 96.47 | 100.0 | 97.65 | 97.65 | 96.47 | 97.53 | 98.82 | 96.47 | 96.47 |
| Original Prediction (T) | 100.0 | 100.0 | 97.89 | 96.84 | 98.86 | 98.95 | 97.89 | 96.84 | 100.0 | 99.47 | 97.37 | 96.84 |
| Original Prediction (M) | 100.0 | 100.0 | 98.10 | 97.14 | 100.0 | 100.0 | 98.10 | 97.14 | 100.0 | 100.0 | 98.10 | 97.14 |
| Original Prediction (F) | 100.0 | 100.0 | 97.65 | 96.47 | 97.53 | 97.65 | 97.65 | 96.47 | 100.0 | 98.82 | 96.47 | 96.47 |
| Training CV (T) | 98.3/97.5 | 99.24 | 100/99.2 | 100.0 | 100.0 | 98.97 | 97.98 | 99.24 | 98.47 | 98.47 | 97.98 |
| Training CV (M) | 98.25 | 98.48 | 100/98.5 | 100.0 | 100.0 | 96.97 | 100.0 | 98.47 | 99.24 | 100.0 | 96.97 | 98.47 |
| Training CV (F) | 98.4/96.7 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Training Prediction (T) | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.24 | 98.47 | 99.24 | 100.0 | 100.0 | 96.97 | 100.0 |
| Training Prediction (M) | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 98.48 | 96.97 | 100.0 | 100.0 | 100.0 | 96.97 | 100.0 |
| Training Prediction (F) | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Target Prediction (T) | 87.93 | 94.92 | 94.92 | 91.53 | 96.55 | 96.61 | 96.61 | 91.53 | 96.55 | 98.31 | 96.61 | 91.53 |
| Target Prediction (M) | 89.47 | 100.0 | 100.0 | 92.31 | 100.0 | 100.0 | 100.0 | 92.31 | 100.0 | 100.0 | 100.0 | 92.31 |
| Target Prediction (F) | 85.00 | 85.00 | 85.00 | 90.00 | 90.00 | 90.00 | 90.00 | 90.00 | 90.00 | 90.00 | 90.00 | 90.00 |
| Target Sex bias | 4.47 | 15 | 15 | 2.31 | 10 | 10 | 10 | 2.31 | 10 | 5 | 10 | 2.31 |
Table 4  Percentage of correctly classified individuals using pelvic and cranial variables by means of ANN when \( p = 1 \) and \( p \) gets its optimum value

| Dataset | Accuracy | Pelvic variables | | Cranial variables | |
|---------|----------|-----------------|-----------------|-----------------|-----------------|
|         |          | VA, SPC, MAR \((p = 1/ p = 10)\) | VA, SPC \((p = 1 / p = 20)\) | GL, MA, NU, ME, ORB \((p = 1/ p = 15)\) | GL, MA, NU \((p = 1/ p = 30)\) | GL, MA, ORB \((p = 1/ p = 10)\) |
| Original | CV (T) | 96.8/97.4 | 93.7/95.8 | 87.8/90 | 88.1/89.2 | 89.2/89.2 |
| Original | CV (M) | 90/100 | 97.1/97.1 | 86.5/88.5 | 87.7/87.7 | 87.7/87.7 |
| Original | CV (F) | 92.9/94.1 | 89.4/94.1 | 89.3/91.7 | 86.9/90.9 | 90.9/90.9 |
| Original | Prediction (T) | 97.4/97.4 | 94.7/96.3 | 90.9/11.1 | 89.2/90.2 | 89.2/89.2 |
| Original | Prediction (M) | 100/100 | 99.1/97.1 | 88.5/89.6 | 87.7/87.7 | 87.7/87.7 |
| Original | Prediction (F) | 94.1/94.1 | 89.4/95.3 | 91.7/92.9 | 90.9/92.1 | 90.9/90.9 |
| Training | CV (T) | 96.2/97 | 97/97 | 90.1/91.7 | 86.9/91 | 88.9/91.8 |
| Training | CV (M) | 98.5/98.5 | 98.5/98.5 | 86.2/86.2 | 85.1/86.6 | 86.6/88.1 |
| Training | CV (F) | 93.9/95.4 | 95.4/95.4 | 93.7/96.8 | 94/95.5 | 91.9/95.5 |
| Training | Prediction (T) | 97/97 | 97/97 | 90.9/93.4 | 91/91.8 | 91.9/91.8 |
| Training | Prediction (M) | 98.5/98.5 | 98.5/98.5 | 87.9/89.7 | 86.6/88.1 | 86.6/88.1 |
| Training | Prediction (F) | 95.4/95.4 | 95.4/95.4 | 93.7/96.8 | 95/95.5 | 95.5/95.5 |
| Target | Prediction (T) | 94.9/94.9 | 94.9/94.9 | 74.6/84.8 | 83.3/86.7 | 73.3/81.7 |
| Target | Prediction (M) | 94.9/94.9 | 94.9/94.9 | 65.8/84.2 | 87.2/89.7 | 69.2/84.6 |
| Target | Prediction (F) | 95/95 | 95/95 | 90.5/85.7 | 76.2/81 | 81/76.2 |
| Target | Sex bias | −0.1 | −0.1 | −1.5 | 8.7 | 8.4 |

Table 6 summarizes the discriminant functions of LDA that result in optimum sex classification for the Athens Collection. As expected, pelvic traits result in higher correct sex classification percentages in relation to cranial traits. It is also interesting to observe that the combination of pelvic and cranial traits improves the predictions, especially the total sex prediction, but it gives rather unbalanced results per sex as it overestimates males and underestimates females.

The combination of different sex traits is straightforward by means of the methods under study. The combination of discriminant functions already published in the literature may be achieved by means of the algorithm described in the “Statistical methods” section. Some of the results obtained are presented in Table 7. In particular, this table shows the percentage of correctly classified individuals by means of the Klales et al. [7] BLR discriminant function, various Walker’s [8] BLR discriminant functions, their combination, and the corresponding LDA discriminant functions of the present work. We observe that Walker’s [8] equations have a very poor prediction performance, which in turn shows that cranial sex traits exhibit very strong population-specificity, as supported by previous studies [28, 29]. The Klales et al. [7]
A discriminant function gives better predictions; the total sex classification rate is rather high, 93.2%, but sex bias is also high, −10.35%, indicating unbalanced results per sex, suggesting some population-specificity. Similar results have been obtained from Oikonomopoulou et al. [30]. However, the most interesting finding of Table 7 is the results obtained from the combination of discriminant functions. It is seen that in most cases, the total sex classification rate is greater than 98%, whereas the sex bias is decreased below 4% and these figures are practically the same as those obtained from the LDA.

Table 6: Linear discriminant equations and percentage of correctly classified individuals

| Dataset | LDA discriminant equations                        | Accuracy | Total | Males | Females | Sex bias |
|---------|---------------------------------------------------|----------|-------|-------|---------|----------|
| Pelvis  | $y = 22.1513 - 3.4643 \times VA - 2.6502 \times SPC - 1.7547 \times MAR$ | Original | 96.84 | 99.05 | 94.12 | 4.93     |
|         |                                                   | Cross-validated | 96.84 | 99.05 | 94.12 | 4.93     |
|         | $y = 18.5549 - 3.6464 \times VA - 3.1649 \times SPC$ | Original | 96.32 | 97.14 | 95.29 | 1.85     |
|         |                                                   | Cross-validated | 96.32 | 97.14 | 95.29 | 1.85     |
| Cranium | $y = 11.7197 - 2.3981 \times GL - 0.7422 \times MA - 0.6611 \times NU - 0.1273 \times ORB - 0.0594 \times ME$ | Original | 89.83 | 97.14 | 95.29 | 5.09     |
|         |                                                   | Cross-validated | 89.83 | 97.14 | 95.29 | 5.09     |
|         | $y = 10.9611 - 2.3077 \times GL - 0.7849 \times MA - 0.6586 \times NU$ | Original | 89.01 | 96.79 | 91.76 | 4.97     |
|         |                                                   | Cross-validated | 89.01 | 96.79 | 91.76 | 4.97     |
|         | $y = 10.0796 - 2.4159 \times GL - 0.8941 \times MA - 0.1559 \times ORB$ | Original | 88.48 | 94.89 | 92.94 | 8.09     |
|         |                                                   | Cross-validated | 88.48 | 94.89 | 92.94 | 8.09     |

Table 7: Percentage of correctly classified individuals when two discriminant functions, one based on pelvic and the other on cranial traits, are combined

| No. | Function | Total | Males | Females | Sex bias |
|-----|----------|-------|-------|---------|----------|
| 1   | $y = 22.1513 - 3.4643 \times VA - 2.6502 \times SPC - 1.7547 \times MAR + 0.0594 \times ME$ | 96.59 | 98.95 | 93.83 | 5.12     |
| 2   | $y = 10.3422 - 2.5346 \times GL - 0.9158 \times MA - 0.0373 \times ME$ | 89.77 | 87.37 | 92.59 | 5.22     |
| 3   | $y = 7.3023 - 2.7024 \times GL + 0.0104 \times ME$ | 98.86 | 100 | 97.53 | 2.47     |
| 4   | $y = 1+2$ | 98.86 | 100 | 97.53 | 2.47     |
| 5   | $y = 7.3823 - 2.4469 \times GL - 3.7121 \times VA - 2.9414 \times SPC$ | 98.42 | 100 | 96.47 | 3.53     |
| 6   | $y = 4+4$ | 98.42 | 100 | 96.47 | 3.53     |
| 7   | $y = 5.459 - 1.3538 \times MA - 0.2269 \times ME$ | 97.89 | 98.1 | 97.65 | 0.45     |
| 8   | $y = 10+6$ | 97.89 | 98.1 | 97.65 | 0.45     |
| 9   | $y = 5.459 - 1.3538 \times MA - 0.2269 \times ME$ | 97.89 | 98.1 | 97.65 | 0.45     |

BLR functions by Kales et al. [7] and Walker [8]

| No. | Function | Total | Males | Females | Sex bias |
|-----|----------|-------|-------|---------|----------|
| 10  | $y = 16.312 - 2.726 \times VA - 1.073 \times SPC - 1.214 \times MAR$ | 93.18 | 88.42 | 98.77 | −10.35   |
| 11  | $y = 9.128 - 1.375 \times GL - 1.185 \times MA - 1.151 \times ME$ | 77.27 | 95.79 | 55.56 | 40.23    |
| 12  | $y = 10 + 11$ | 98.3 | 100 | 96.3 | 3.7      |
| 13  | $y = 7.372 - 1.525 \times GL - 1.485 \times ME$ | 75 | 94.74 | 51.85 | 42.89    |
| 14  | $y = 10 + 13$ | 98.86 | 100 | 97.53 | 2.47     |
| 15  | $y = 7.434 - 1.568 \times GL - 1.459 \times MA$ | 81.58 | 97.14 | 62.35 | 34.79    |
| 16  | $y = 10 + 15$ | 98.42 | 100 | 96.47 | 3.53     |
| 17  | $y = 7.382 - 1.629 \times ME - 1.415 \times MA$ | 58.52 | 88.42 | 23.46 | 64.96    |
| 18  | $y = 10 + 17$ | 97.73 | 100 | 95.06 | 4.94     |
discriminant functions derived from the Athens Collection. That is, the combination of pelvic and cranial traits via their discriminant functions removes practically any population-specificity and yields much better predictions than the individual functions, especially those obtained from the literature. This is very clearly shown in case 14, Table 7, whereby we combined the Klales et al. [7] equation with Walker’s [8] equation using only the glabella and mental eminence. It can be seen that this combination achieved a correct sex classification of 98.9% in total, while the application of the individual equations achieved a correct classification in 93.2% for the pelvic traits and 75% for the cranial ones. At the same time, the sex bias for these two functions is −10% and 43%, respectively, whereas the proposed method reduces the bias to 2.5%. In addition, the results of the combined function are identical to those obtained when we derive equations for sex prediction based on the Athens Collection itself (case 5, Table 7).

To summarize the above results, the differences in the classification performance between BLR, PR, CPR, LDA, QDA, ANN, and NBC are overall small. However, LDA is more simple and flexible than CPR, QDA, and ANN and has a small but clear advantage over BLR, PR, and NBC. Despite being a parametric multivariate technique, LDA is fairly robust to violations of relevant assumptions, and may be preferred in sex classification problems. These results concern cases where pelvic and cranial traits are examined independently. The most striking result of the current study is that the proposed approaches to combine pelvic and cranial traits via their discriminant functions, either from LDA or BLR, yield better predictions than the individual functions (correct classification rates over 98% for pooled sexes and sex bias below 3%), free from population-specificity issues.

Acknowledgments We would like to thank two anonymous reviewers for their suggestions, especially with regard to combining pelvic and cranial traits via their discriminant functions as well as implementing the Konigsberg-Hens method of cumulative probit regression. Ethlyna Nikita’s contribution to this research was supported by the H2020 Promised project (Grant Agreement 811065), and the European Regional Development Fund and the Republic of Cyprus through the Research Promotion Foundation (People in Motion project: EXCELLENCE/1216/0023).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Dirkmaat D (2014) A companion to forensic anthropology. Wiley, New York
2. İsand MY (2005) Forensic anthropology of sex and body size. Forensic Sci Int 147:107–112. https://doi.org/10.1016/j.forsciint.2004.09.069
3. Spalding MK, Jantz RL (2011) Sex estimation in forensic anthropology: skull versus postcranial elements. J Forensic Sci 56:289–296. https://doi.org/10.1111/j.1556-4029.2010.01635.x
4. Buikstra JE, Ubelaker DH (1994) Standards for data collection from human skeletal remains. Arkansas Archaeological Survey Research Series 44, Fayetteville
5. Phemice TW (1969) A newly developed visual method of sexing the os pubis. Am J Phys Anthropol 30:297–301. https://doi.org/10.1002/ajpa.1330300214
6. Williams BA, Rogers TL (2006) Evaluating the accuracy and precision of cranial morphological traits for sex determination. J Forensic Sci 51:729–735. https://doi.org/10.1111/j.1556-4029.2006.00177.x
7. Klales AR, Osley SD, Vollner JM (2012) A revised method of sexing the human innominate using Phemice’s nonmetric traits and statistical methods. Am J Phys Anthropol 149:104–114. https://doi.org/10.1002/ajpa.22102
8. Walker PL (2008) Sexing skulls using discriminant function analysis of visually assessed traits. Am J Phys Anthropol 136:39–50. https://doi.org/10.1002/ajpa.20776
9. Ali Z, Cox C, Stock MK, Zandee vanRilland EE, Rubio A, Fowler DR (2018) Estimating sex using metric analysis of the scapula postmortem computed tomography. J Forensic Sci 63:1346–1349. https://doi.org/10.1111/1556-4029.13751
10. Aluni-Perret V, du Jardin P, Nogueira L, Buchet L, Quatrehomme G (2015) Comparing discriminant analysis and neural network for the determination of sex from femur head measurements. Forensic Sci Int 253:81–87. https://doi.org/10.1016/j.forsciint.2015.05.023
11. Bidmos MA, Steinberg N, Kuykendall KL (2005) Patella measurements of South African whites as sex assessors. Homo 56:69–74. https://doi.org/10.1016/j.jchb.2004.10.002
12. Blake KA, Hartnett-McCann K (2018) Metric assessment of the pubic bone using known and novel data points for sex estimation. J Forensic Sci 63:1472–1478. https://doi.org/10.1111/1556-4029.13732
13. Clavero A, Salicrú M, Turbón D (2015) Sex prediction from the femur and hip bone using a sample of CT images from a Spanish population. Int J Legal Med 129:373–383. https://doi.org/10.1007/s00414-014-1069-y
14. Nikita E (2017) Osteoarchaeology: a guide to the macroscopic study of human skeletal remains. Academic Press, San Diego
15. Darmawan MF, Yusuf SM, Abdul Kadir MR, Haron H (2015) Comparison on three classification techniques for sex estimation from the bone length of Asian children below 19 years old: an analysis using different group of ages. Forensic Sci Int 247:130.e1–130.e11. https://doi.org/10.1016/j.forsciint.2014.11.007
16. Du Jardin P, Ponsaillé J, Aluni-Perret V, Quatrehomme G (2009) A comparison between neural network and other metric methods to determine sex from the upper femur in a modern French population. Forensic Sci Int 192:e1–e6. https://doi.org/10.1016/j.forsciint.2009.07.014
17. Mahfouz M, Badawi A, Merkt B, Fatah EEA, Pritchard E, Kesler K, Moore M, Jantz R, Jantz L (2007) Patella sex determination by 3D statistical shape models and nonlinear classifiers. Forensic Sci Int 173:161–170. https://doi.org/10.1016/j.forsciint.2007.02.024
18. Coelho JO, Curate F (2019) CADOES: an interactive machine-learning approach for sex estimation with the pelvis. Forensic Sci Int. https://doi.org/10.1016/j.forsciint.2019.108973
19. Navega D, Vicente R, Vieira DN, Ross AH, Cunha E (2015) Sex estimation from the tarsal bones in a Portuguese sample: a machine learning approach. Int J Legal Med 129:651–659. https://doi.org/10.1007/s00414-014-1070-5
20. Tabachnick B, Fiddell L (2012) Using multivariate statistics, 6th edn. Pearson Education Limited, Boston
21. Konigsberg LW, Hens SM (1998) Use of ordinal categorical variables in skeletal assessment of sex from the cranium. Am J Phys

© Springer
22. Johnson RA, Wichern DW (1988) Applied multivariate statistical analysis, 2nd edn. Prentice Hall International, London
23. Narsky I, Porter FC (2014) Statistical analysis techniques in particle physics: fits, density estimation and supervised learning, 1st edn. Wiley, Hoboken
24. Haykin S (2009) Neural networks and learning machines, 3rd edn. Prentice Hall, London
25. Ripley BD (1996) Pattern recognition and neural networks. Cambridge University Press, Cambridge
26. Konigsberg LW, Frankenberg SR (2019) Multivariate ordinal probit analysis in the skeletal assessment of sex. Am J Phys Anthropol 169:385–387. https://doi.org/10.1002/ajpa.23832
27. Eliopoulos C, Lagia A, Manolis S (2007) A modern, documented human skeletal collection from Greece. Homo 58:221–228. https://doi.org/10.1016/j.jchb.2006.10.003
28. Klales AR, Cole SJ (2017) Improving nonmetric sex classification for Hispanic individuals. J Forensic Sci 62:975–980. https://doi.org/10.1111/1556-4029.13391
29. Krüger GC, L’Abbé EN, Stull KE, Kenyhercz MW (2015) Sexual dimorphism in cranial morphology among modern South Africans. Int J Legal Med 129:869–875. https://doi.org/10.1007/s00414-014-1111-0
30. Oikonomopoulou EK, Valakos E, Nikita E (2017) 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 131:1731–1738. https://doi.org/10.1007/s00414-017-1655-x

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.