Comparison of Binary Classifiers in Forensic Dentistry for Sex Determination

Ramon Alvarez-Vaz[1,2*], Sassi Carlos[3], Gargano Verónica[2] and Picapedra Alicia[1,2]

[1]*Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay*

[2]*Facultad de Odontología, Universidad de la República, Uruguay*

[3]*Facultad de Odontologia de Piracicaba. Universidad Estadual de Campinas, Sao Paulo, Brasil*

*Corresponding author
Ramon Alvarez-Vaz, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay.

Submitted: 2025 Jul 21; **Accepted**: 2025 Aug 14; **Published**: 2025 Aug 22

Abstract
In forensic odontology, teeth are essential components of the stomatognathic system, since they constitute excellent material for research, thanks to their almost indestructible structure and therefore the information linked to their size and characteristics, is extremely useful in determining sex and age. In this study, which corresponds to a sample of 524 lower plaster models (286 male subjects and 238 females) of patients assisted in an orthodontic clinic in Montevideo, Uruguay, the mesiodistal diameter (width), and gingivo-incisal height of the 2 canines and the intercanine distance are measured for sex identification based on odontometric measurements and their ratios. Different binary classifiers from what is known as supervised classification methods are applied in the statistical learning paradigm. Several methods are proposed and their performance is evaluated by working with test and learning samples and evaluating different metrics to measure the accuracy of the models, showing a performance greater than 65%. The details of this project are on the OSF platform in the project. Relaciones Odontom´etricas en Odontolog´ıa Forense (Odontometric Relationships in Forensic Dentestry) en https://osf.io/javru/

1. Introduction

There has long been a need and desire for human identification, given their sociopolitical nature and ambition to differentiate themselves from others. Indeed, the dynamics and changing collective organizations are characterized by the desire to individualize and hold their members civilly, administratively, commercially, and5 criminally liable [1-3]. In the forensic field, a positive identification must be based on the intervention of a multidisciplinary team capable of performing an appropriate reconstruction of the individual's biological profile, using four essential components: age, sex, height, and ancestry [2,4-7]. Dental organs exhibit taphonomic resistance, being composed of almost indestructible structures, thus enabling post-mortem odontometric, comparative, and reconstructive analyses [8,9]. The most commonly used measurements, obtained from plaster models, are the mesiodistal, vestibulopalatine, mesiovestibular-distopalatine, and distovestibular-mesiopalatine diameters, the gingivoincisal height, and the intercanine distance, which constitute several indices [3,6,7,10-13]. Teeth, in general, and permanent canines in particular, are indicated to demonstrate sexual dimorphism by several authors in different populations around the world [14-16]. In these studies, taking size into account has guided the process of sex determination, taking into account the fact that most teeth develop before skeletal maturation, making these measurements a valuable indicator of sex, mainly in subadult individuals, in the absence or emergence of secondary sexual characteristics [17]. Legal and/or forensic dentistry has benefited from the invaluable contribution of statistics, developing predictive models tailored to the conditions of each case [7]. In view of the above, this work seeks to find methods that allow identifying the sex with the highest degree of accuracy, using statistical models, as parsimonious as possible, based on the measurements of lower canines, and the relationships between these, for a sample of plaster models of patients assisted in an orthodontic clinic in the city of Montevideo, Uruguay.

For this purpose, the work is divided into 3 parts, the first one presents some binary classifiers in the Methodology section that can be used for the previously proposed objective, then we see how they work empirically as shown in the Materials section and then in the Conclusion section we propose some alternatives to always improve the models for sex identification based on odontometric measurements.

2. Methodology

There are several binary classifiers that can be used in the field of statistical learning, more specifically what is known as supervised classification, where there is an input data matrix X and one or more variables Y with labels. From X, each observation xi is classified,

assigning it a label. For this purpose, some binary classifiers are presented, which are based on different types of input variables.

3. Discriminant Analysis Method

Discriminant Analysis (DA) is a multivariate statistical technique for supervised classification but has two purposes, the description, where it is of interest to analyze whether there are differences between a series of groups into which a population is divided with respect to a set of variables and, if so, to find out why. On the other hand, DA seeks to make predictions through a systematic procedure for classifying new observations of unknown origin into some of the groups considered [18-20].

Although this work focuses on binary classifiers (i.e., there are two groups), there are k samples of size ng (g = 1, 2, . . . k) from k populations from which p quantitative characteristics are measured. Using this information, we want to determine which of those k populations a new observation is most likely to have been randomly selected from. Each element of the population is assigned to one of the groups according to a specific decision rule, trying to make the smallest possible errors. The analysis has a certain" predictive" power because, in a way, the criteria used to classify a current population can be used for new elements that are incorporated into it. The disadvantage of this method is that it is limited to quantitative explanatory variables that also follow a varied Normal distribution. Functions are sought that best discriminate between groups, which is why it is necessary to define decision rules, ensuring the fewest errors possible.

When classifying, three types of distance are of interest:
- Distance between units.
- Distance between populations.
- Distance between unit and population.

4. Distance between Individuals

The Mahalanobis distance between individuals i and l is:

$$D_{il} = \sqrt{(\mathbf{x_i} - \mathbf{x_l})'\Sigma^{-1}(\mathbf{x_i} - \mathbf{x_l})} \; \forall i \neq l. \tag{1}$$

where $\mathbf{x_i'} = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ observed vector of individual i, Σ variance and covariance matrix.

Distance between populations or groups

$$D_{gg'} = \sqrt{(\mu_{\mathbf{g}} - \mu_{g'})'\Sigma^{-1}(\mu_g - \mu_{\mathbf{g'}})} \; \forall g' \neq g. \tag{2}$$

where $\mu_{\mathbf{g}}' = (\mu_{1g}, \ldots, \mu_{pg})'$ centroid, vector of means of the p variables. Σ matrix of variances and covariances.

Distance between individual i and centroid of each group

$$D_{ig} = \sqrt{(\mathbf{x_i} - \mu_{\mathbf{g}})'\Sigma_g^{-1}(\mathbf{x_i} - \mu_{\mathbf{g}})}. \tag{3}$$

where $\mathbf{x_i'} = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ observed vector of individual i. Σ_g variance-covariance matrix of group g.

Unit i is classified into group g if $D_{ig} < D_{ig'} \; \forall g' \neq g$.

In generic form, $P(g'|g)$ is the error of classifying an observation from group g into group g', where the probability of error of classifying an observation from group g' into group g is:

$$P(g'|g) = \int_{R_{g'}} f_g(x)dx. \tag{4}$$

The probability of misclassifying all observations belonging to group g:

$$P(g) = \sum_{g'=1, \; g' \neq g}^{k} P(g'|g) = 1 - P(g|g). \tag{5}$$

The total probability of misclassification is:

$$P(R, f) = \sum_{g=1}^{k} \pi_g P(g). \tag{6}$$

The decision rule is the one that minimizes the total probability of error.

5. Principle of Maximum Likelihood
The principle consists in assigning observation i to the population where the observed vector x′ = (xi1, xi2, . . . , xip)′ has the greatest likelihood of occurring, that is, i is assigned to group g if:

$$f(\mathbf{x_i}|g) > f(\mathbf{x_i}|g') \ \forall g' \neq g \tag{7}$$

$$\Rightarrow \frac{f(x_i|g)}{f(x_i|g')} > 1. \tag{8}$$

6. Principle of a Posteriori Probability
The principle is to assign observation i to the population with the highest posterior probability (the probability that i belongs to g conditioned on the observed vector xi). The posterior probability, according to Bayes' Theorem, is stated as follows:

Principle of a posteriori probability

The principle is to assign observation i to the population with the highest posterior probability (the probability that i belongs to g conditioned on the observed vector x_i). The posterior probability, according to Bayes' Theorem, is stated as follows:

$$P(i \in g|\mathbf{x} = \mathbf{x_i}) = \frac{\pi_g P(\mathbf{x_i}|g)}{P(\mathbf{x_i})} = \frac{\pi_g f(\mathbf{x_i}|g)}{\sum\limits_{g'=1}^{k} \pi_{g'} f(\mathbf{x_i}|g')}. \tag{9}$$

where $\pi_g = P(i \in g)$ is the prior probability of belonging to group g.
Unit i is classified into group g if:

$$P(i \in g|\mathbf{x} = \mathbf{x_i}) > P(i \in g'|\mathbf{x} = \mathbf{x_i}) \ \forall \ g' \neq g. \tag{10}$$

The proposed rules are optimal if the joint distribution is known, with all parameters known.

Classification Rules applied to 2 groups

Given $\mathbf{x_i} \sim \mathbf{N}_p(\mu, \Sigma)$ the density function is:

$$f(\mathbf{x_i}|\mathbf{g}) = \frac{1}{(2\pi)^{p|2} |\Sigma_\mathbf{g}|^{1|2}} \exp\left\{-\frac{1}{2}(\mathbf{x_i} - \mu_\mathbf{g})'\Sigma_\mathbf{g}^{-1}(\mathbf{x_i} - \mu_\mathbf{g})\right\}. \tag{11}$$

$$\frac{\hat{\pi}_g |\mathbf{S_g}|^{-1|2} \exp\left\{-\frac{1}{2}\hat{D}_{ig}^2\right\}}{\sum\limits_{g'=1}^{k} \hat{\pi}_{g'} |\mathbf{S_{g'}}|^{-1|2} \exp\left\{-\frac{1}{2}\hat{D}_{ig'}^2\right\}}. \tag{12}$$

Two cases are distinguished:

1. In the case of equality of variance matrices, observation i is assigned to group g if:

$$\hat{\pi}_g \exp\left\{-\frac{1}{2}\hat{D}_{ig}^2\right\} > \hat{\pi}_{g'} \exp\left\{-\frac{1}{2}\hat{D}_{ig'}^2\right\} \ \forall g' \neq g. \tag{13}$$

2. In the case of matrices of different variances, observation i is assigned to group g if:

$$\hat{\pi}_g |\mathbf{S_g}|^{-1|2} \exp\left\{-\frac{1}{2}\hat{D}_{ig}^2\right\} > \hat{\pi}_{g'} |\mathbf{S_{g'}}|^{-1|2} \exp\left\{-\frac{1}{2}\hat{D}_{ig'}^2\right\} \ \forall g' \neq g. \tag{14}$$

When the cost of error is not considered

- i is assigned to *group* 1 if: $\pi_1 f(\mathbf{x_i}|\mathbf{1}) > \pi_2 \mathbf{f}(\mathbf{x_i}|\mathbf{2})$, the classification rule is usually presented in the following form

$$\frac{f(\mathbf{x_i}|\mathbf{1})}{f(\mathbf{x_i}|\mathbf{2})} > \frac{\pi_2}{\pi_1}. \tag{15}$$

Regarding a priori probabilities, they are usually estimated based on the proportion that each group represents in the sample.

- If equal a priori probabilities are also assumed, i is assigned to *group* 1 if:

$$\frac{f(\mathbf{x_i}|\mathbf{1})}{f(\mathbf{x_i}|\mathbf{2})} > 1. \tag{16}$$

Considering different costs and prior probabilities for each group, i is assigned to *group* 1 if: $\frac{f(\mathbf{x_i}|\mathbf{1})}{f(\mathbf{x_i}|\mathbf{2})} > \frac{\pi_2 C(1|2)}{\pi_1 C(2|1)}$ where $C(1|2)$ is the cost of classifying a unit in group 1 when it comes from group 2 and $C(2|1)$ is the cost of classifying a unit as 2 when it comes from group 1.

6.1 Discriminant Function for Linear Discriminant Analysis
The classification rules from which the linear discriminant function is derived are hypothesized as follows:
• Joint normal distribution in each group.
• Equal variance and covariance matrix ($\Sigma_1 = \Sigma_2 = \ldots = \Sigma_k$).

According to what is stated in maximizing $P(i \in g|x = x_i)$ is equivalent to maximizing $\pi_g \exp\{-\frac{1}{2}D_{ig}^2\}$. Applying logarithm and replacing μ with $\bar{x}$ and Σ with S:

$$\log_e \hat{\pi}_g - \frac{1}{2}\hat{D}_{ig}^2 = \log_e \hat{\pi}_g - \frac{1}{2}\{(x_i - \bar{x}_g)'S^{-1}(x_i - \bar{x}_g)\} =$$
$$\log_e \hat{\pi}_g - \frac{1}{2}x_i'S^{-1}x_i + \frac{1}{2}\bar{x}_g'S^{-1}x_i + \tag{17}$$
$$\frac{1}{2}x_i'S^{-1}\bar{x}_g - \frac{1}{2}\bar{x}_g'S^{-1}\bar{x}_g.$$

The *"score"* of observation i in the **linear discriminant function** estimated for *group g* is:

$$L_{ig} = \log_e \hat{\pi}_g + \bar{x}_g'S^{-1}x_i - \frac{1}{2}\bar{x}_g'S^{-1}\bar{x}_g \quad g = 1\ldots k. \tag{18}$$

or otherwise the discriminant function is

$$L_g = Xb + C$$

$$L_{ig} = b_g'x_i + c_g = b_{1g}x_{1ig} + b_{2g}x_{2ig} + \ldots + b_{pg}x_{pig} + c_g. \tag{19}$$

con $b_g' = \bar{x}_g'S^{-1}$ y $c_g = \frac{1}{2}\bar{x}_g'S^{-1}\bar{x}_g + \log \hat{\pi}_g$

The estimated *classification function* of observation i in *group g* is:

$$L_{igg'} = \frac{1}{2}\bar{x}_g'S^{-1}x_i + \frac{1}{2}x_i'S^{-1}\bar{x}_g$$
$$-\frac{1}{2}\bar{x}_g'S^{-1}\bar{x}_g - \frac{1}{2}\bar{x}_{g'}'S^{-1}x_i - \frac{1}{2}x_i'S^{-1}\bar{x}_{g'} = (\bar{x}_g - \bar{x}_{g'})S^{-1}[x_i - \frac{1}{2}(\bar{x}_g + \bar{x}_{g'})]. \tag{20}$$

The classification region is

$$R_{gg'} : L_{igg'} > \log(\frac{\hat{\pi}_{g'}}{\hat{\pi}_g}) \tag{21}$$

In the case of equal a priori probabilities the rule reduces to Ligg′ > 0. Another way to122 put it is: observation i is classified in the group g if $L_{ig} > L_{ig}'$ *forall g'*⊨ *g.*

6.2 Logistic Regression Analysis

The response variable Yi is a Bernoulli random variable, with possible outcomes:
success, failure encoded as {0, 1}, probability distribution:

$P(Y_i = 1) = \pi_i$, $P(Y_i = 0) = 1 - \pi_i$ and expected value, [19]. [1]: $E(Y_i|X) = \pi_i$.

$$P(Y = 1|X) = \pi = \frac{e^{X\beta}}{1 + e^{X\beta}}. \tag{22}$$

$$P(Y_i = 1|X_i) = \pi_i = \frac{e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi}}}. \tag{23}$$

The model proposed in (23) can be linearized by performing the following transformation:

$$\log_e \left(\frac{\pi}{1 - \pi} \right). \tag{24}$$

This transformation is called a Logit transformation. The quotient $\frac{\pi}{1-\pi} = e^{X\beta}$ is what is called odds, which is a probability ratio, the probability of success over the probability of failure, and it is considered a "measure of risk". For example, when it is said that the odds are 4 to 1 (odds=4) that a person x will suffer from a disease, it is equivalent to saying that the probability of disease is 0.8.

Predicting the value of the response variable based on certain values of other explanatory variables involves determining the critical value above which the estimates of the expected value imply a value of 1 for the response variable. Large values of πi imply Yi = 1 while small values of πi imply Yi = 0, so it is a problem for the researcher to determine when an estimate is considered a very large or very small value.

7. K-Neighbor Methods

Another binary classifier that can be used and that, unlike AD and ARL, is what is called the k nearest neighbors (kNN), which is non-parametric, very easy to implement and based on a very intuitive notion to understand and is based on the Bayesian classifier, which as stated by, is the one with the lowest error rate, on average [21]. The Bayesian classifier assigns each observation to the most probable class, conditioned on its predictor values, that is, a test observation, conditioned by the predictor vector x0 to the class j where it is verified

$$P(Y = j|X = x_o). \tag{25}$$

is maximum. For the binary classifier case, the Bayesian classifier that if there are 2148 classes g1 and g2 , P (Y = 1|X = x0) ≤ 0.5 ends up classifying in g1.149 Fig shows the Bayesian decision boundary for simulated data, where values that fall150 on the orange side of the boundary will be assigned the orange class
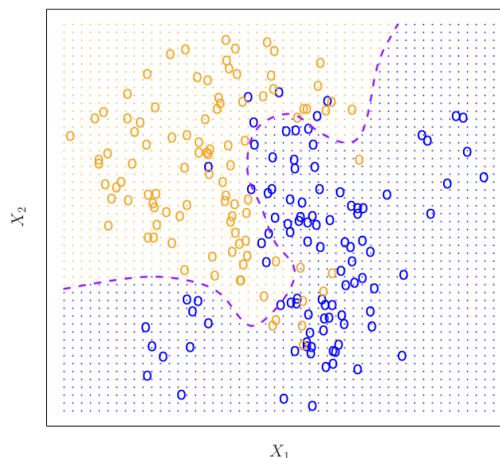

Figure 1: Example of a Bayesian Classifier with 2 groups and 2 variables (taken from An introduction to statistical learning: with applications in R, page 38).

The Bayes classifier will always choose the class for which (25) is largest, so the error 152 rate will be $1 - max_j Pr\,(Y = j \mid X = x_0)$ at $X = x_0$ state that the global Bayesian 153 error rate would be given by [21]:

$$1 - E\,(\,max_j P(Y = j|X))\,. \tag{26}$$

which should be interpreted as the average expected probability value over all possible values of X.156 As an idea, this seems like something that could work very well; however, it is very difficult to know the conditional distribution of $Y \mid X$ for real data. For this reason, one alternative is to estimate the conditional distribution and classify a given observation into the group with the highest estimated probability. One method that works with this principle is the k nearest neighbors, kNN.161 Taking an integer k and an observation x_0, the kNN identifies the k points closest to x_0 and estimates the conditional probability for cluster j as the fraction of the nearest neighbor set N_0, where the answer is j

$$Pr(Y = j|X = x_0) = \frac{1}{K} \sum_{i \in N_0} I(y_i = j). \tag{27}$$

Therefore, the classifier's performance will depend on the buffer size k at each point, where for the same data in Figure 1, changing k changes the decision boundary in solid black for the same Bayesian decision boundary in violet dotted line. It can be seen that the change in k is important since for k = 1 the boundary adapts locally with a low bias but large variance , while when k = 100 the classifier has the inverted properties, and the boundary becomes linear.
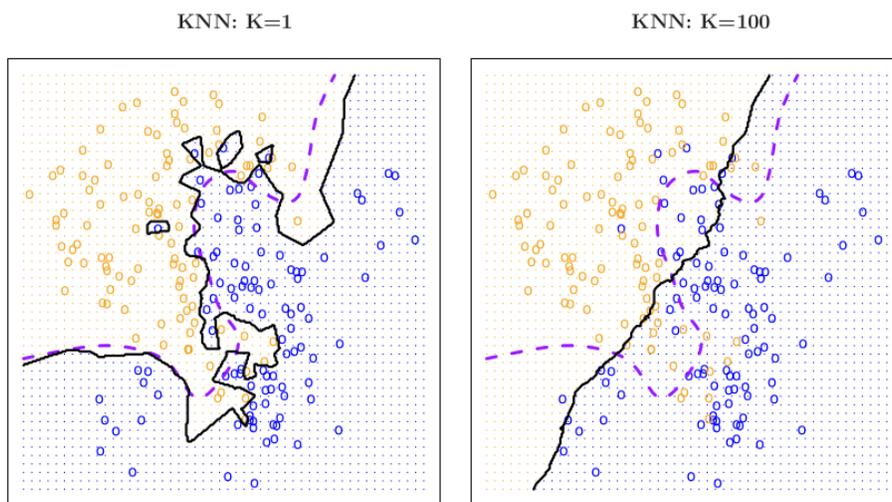

Figure 2: Example of a Bayesian Classifier with 2 groups and 2 variables (extracted from An introduction to statistical learning : with applications in R, page 41).

7.1 Confusion Matrices and Performance Metrics

Usually, as a way of measuring the performance of the different classifiers, in this case binary, what is called confusion matrix is used, which allows to see different dimensions of the success capacity for each one [22-24].

For this, as a result of the prediction, a table with the following structure is created

		Reference	
		Event	Not event
Prediction	Event	A	B
	Not evento	C	D

Table 1: Confusion matrix

From Table 1, a series of metrics are developed that allow the classifier's different176 performance characteristics to be evaluated on this data.

$$Sen = \frac{A}{A+C} \ (Sensitivity)$$

In this case, sensitivity measures the model's ability to detect true events, which in the case of a health condition (patient), i.e., the ability to detect the vast majority of patients.

$$Esp = \frac{D}{D+B} \ (Specificity)$$

In the case of specificity, it is the true negative rate, i.e., the model's ability to detect healthy individuals in the face of a health condition. Therefore, it is desirable to have both metrics with maximum values, which is not possible since they have a negative correlation. A change is any of the metrics; it involves changing the cutoff threshold, which changes the prediction, and in doing so, the confussion matrix is reconfigured.

$$Prev = \frac{A+C}{A+B+C+D} \ (Prevalence)$$

Continuing with the possible situation of the event being a case of disease, it is also important to know the fraction of people observed to have the condition of the event, which is known as prevalence, a term derived from epidemiology.

The positive and negative predictive values (Pred Value (+) and Pred Value (-)) respectively) are the proportions of positive and negative results generated by a classifier that are actually true positives and true negatives, respectively. High values for both metrics can be interpreted as an indicator of the accuracy of the result generated by the classifier. The Predictive Value (+) and Predictive Value (-) are not intrinsic to the classifier, as is the case with the true positive rate and the true negative rate, since the Predictive Values depend on the prevalence, as shown in the formula. It should be noted that the Predictive Value (+) and Predictive Value (-), like the NPV, can be derived using Bayes' theorem. A first expression for Pred Value (+) is

$$\text{Pred Value (+)} = \frac{A}{A+B}$$

This can be rewritten through sensitivity, specificity, and prevalence

$$\text{Pred Value (+)} = \frac{Sen * Prev}{(Sen * Prev) + ((1 - Sp) * (1 - Prev))}$$

For the case of Pred Value (-)

$$\text{Pred Value (-)} = \frac{C}{C+D}$$

Using sensitivity, specificity, and prevalence for Pred Value (-) we have

$$\text{Pred Value (-)} = \frac{Sp * (1 - Prev)}{((1 - Sen) * Prev) + ((Sp) * (1 - Prev))}$$

The Detection Rate can be viewed as the classifier's ability to detect as events, cases that actually constitute events, out of the total number of cases:

$$\text{Detection Rate} = \frac{A}{(A+B+C+D)}$$

While the Detection Prevalence can be viewed as the proportion of events that the classifier classifies as such out of the total number of cases:

$$\text{Detection Prevalence} = \frac{(A+B)}{(A+B+C+D)}$$

Balanced Precision is an important metric as it shows the classifier's performance by averaging the ability to detect the true positive rate and the true negative rate.

$$\text{Balanced Accuracy} = \frac{Sen + Sp}{2}$$

There are two other metrics that are also widely used in clinical diagnostic tests: the positive likelihood ratio LR(+) (in English, LR(+)).

$$RV(+)\frac{P(T+|D+)}{P(T+|D-)} = \frac{Sen}{1 - Sp}$$

where the expression P (T + |D+) can be seen as the probability that a given person with the disease will be classified by the classifier as sick (since it is an event, it will be classified as an event). This can be expressed through sensitivity and specificity.

The same can be said for the positive likelihood ratio LR(-) (in English, English LR(-))

$$RV(-)\frac{P(T-|D+)}{P(T-|D+)} = \frac{1 - Sen}{Esp}$$

Combining both likelihood ratios produces another metric called ORD (diagnostic odds ratio), which measures the effectiveness of the classifier and expresses the relationship between the probabilities of the classifier being positive, if the subject is really positive (has the disease) in relation to the probabilities of the classifier being positive, if the subject does not have the disease.

$$ODR = \frac{LR(+)}{LR(-)} = \frac{Sen * Esp}{(1 - Sen)(1 - Esp)}$$

In the statistical analysis of information retrieval and binary classification systems, the F-score or F-measure is a measure of predictive performance. It is calculated from the test's accuracy and sensitivity, where accuracy is the number of true positive results divided by the number of all samples predicted to be positive, including those not correctly identified.

The F1 metric is the harmonic mean of the precision (which is the Pred Value (+)) and the sensitivity (called recall). Therefore, it symmetrically represents both precision and recall in a single metric. For other β values, differential weights are applied, valuing one or recall more than the other.

$$F_\beta = \frac{(1 + \beta^2) * precision * Sen}{((\beta^2 * precision) + Sen)}$$

This expression has a direct link to the confusion matrix seen from the Type I error and Type II error in the context of hypothesis testing.

8. Materials
In this work, only the measurements of 524 lower plaster models (286 male subjects and 238 females) from patients treated at an orthodontic clinic in Montevideo, Uruguay, are taken into account. The mesiodistal diameter (width), gingivo-incisal height of the 4 canines, and the intercanine distance were measured for sex identification based on odontometric measurements and their ratios, from a larger study published in [7,20].

To evaluate the performance of the binary classifiers presented in previous sections, the metrics presented in section about Confusion matrices and performance metrics are used. The work is implemented in R language and used several libraries like MASS, klar, heplots, rsample, class libraries [18,25-28].

For the graphics, ggplot2 , GGally are used, and for the different metrics, caret [24,29-31]. The computational code is available at https://osf.io/javru/

9. Results and Discussion
After evaluating which measures had anomalous values, the sample is reduced to 511 by removing the values that have DMDd ≥ 8.5 or AGId ≥ 11.5 and also for DMDi ≥ 8.5 or AGIi ≥ 11.5. Although in Figure 5 there are almost 30 anomalous values, which are
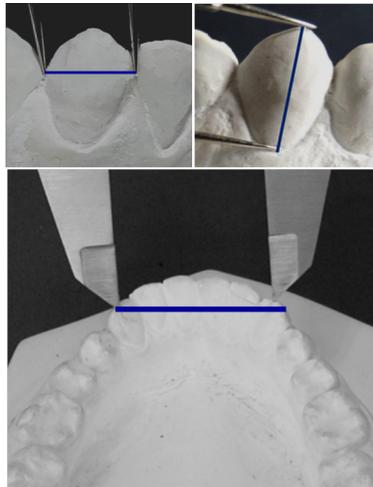
Figure 3: Measures considered in the study (own elaboration)/

Measures	Statistic	M (N=264)	F (N=260)	p
DMDd (Messiodistal diameter for right side)	x ± DE	7.1 ± 0.5	6.7 ± 0.5	< .001
AGId (Gingivo-Incisal Height for right side)	x ± DE	9.3 ± 1.1	8.9 ± 0.9	< .001
DMDi (Messiodistal DIameter for left side)	x ± DE	7.1 ± 0.5	6.7 ± 0.5	< .001
AGIi (Gingivo Incisal Height for lefty side)	x ± DE	9.3 ± 1.1	9.0 ± 0.9	< .001
DIC (Intercanine Distance)	x ± DE	26.5 ± 2.4	25.5 ± 2.0	< .001

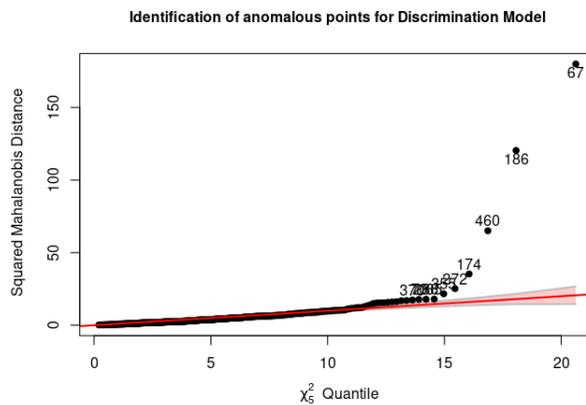**Table 2: Odontometric Measurements by Sex**

Figure 4: Multivariate anomalous observations using Mahalanobis distance, with complete data
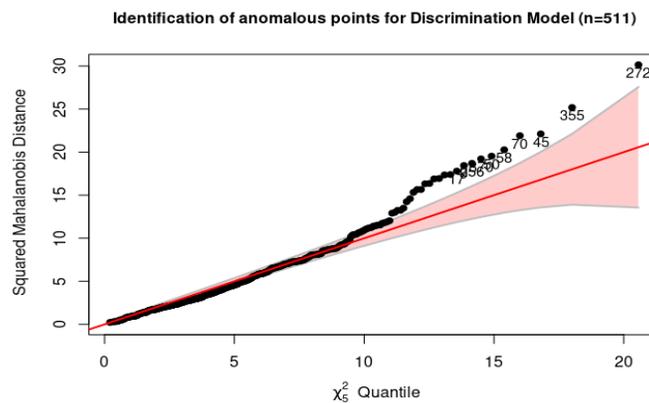
Figure 5: Multivariate anomalous observations using Mahalanobis distance

Measures	Statistic	M (N=255)	F (N=256)	p
DMDd	x ± DE	7.1 ± 0.4	6.7 ± 0.4	< .001
AGId	x ± DE	9.2 ± 1.0	8.9 ± 0.9	< .001
DMDi	x ± DE	7.1 ± 0.5	6.7 ± 0.4	< .001
AGIi	x ± DE	9.3 ± 1.0	9.0 ± 0.9	< .001
DIC	x ± DE	26.4 ± 2.3	25.5 ± 2.0	< .001

Table 3: Odontometric Measurements by Sex Without Outliers.

outside the confidence band, but in Table 3, they do not essentially change the mean values but which by themselves can be "lever" points. With the data filtered, the territorial maps are presented using the 2-by-2 variables and considering linear and quadratic boundaries. These maps show areas with a significant mix of observations that are misclassified and where there is significant overlap. For the DA, LRA, and kNN methods, it is proposed to work with a 75% learning sample and a 25% testing sample.

Measures	M (N=255)	F (N=256)	OR (multivariate)
DMDd	7.1 ± 0.4	6.7 ± 0.4	0.25 (0.11-0.54, $p < .001$)
AGId	9.2 ± 1.0	8.9 ± 0.9	0.78 (0.53-1.17, $p = .234$)
DMDi	7.1 ± 0.5	6.7 ± 0.4	0.55 (0.27-1.15, $p = .114$)
AGIi	9.3 ± 1.0	9.0 ± 0.9	1.16 (0.77-1.75, $p = .468$)
DIC	26.4 ± 2.3	25.5 ± 2.0	0.92 (0.84-1.01, p = .088)

Table 4: Logistic Regression Without Outliers.

Looking at the results obtained in Table 4, only one measurement remains as part of the model, and it is one-sided, after applying a sequential ANOVA to remove variables. This means that, from a practical point of view, having a model that only uses canine diameter or only measurements from one side does not make sense for forensic specialists, which is why it is decided to develop two models, only with measurements for each side.

Sensitivity	0.741	0.661
Specificity	0.636	0.606
(+) Valor Pred	0.657	0.611
(-) Valor Pred	0.724	0.655
Precision	0.657	0.611
Recall	0.741	0.661
F_1	0.697	0.635
Prevalence	0.484	0.484
Detection Rate	0.359	0.320
Detection Prevalence	0.546	0.523
Balanced Accuracy	0.689	0.633

Table 5: Metrics for ADL and ADC for testing sample.

Measures	M (N=255)	F (N=256)	OR (multivariate)
DMDd	7.1 ± 0.4	6.7 ± 0.4	0.16 (0.09-0.26, $p < .001$)
AGId	9.2 ± 1.0	8.9 ± 0.9	0.89 (0.72-1.11, $p = .299$)
DIC	26.4 ± 2.3	25.5 ± 2.0	0.92 (0.83-1.01, $p = .065$)

Table 6: Logistic Regression Without Outliers for Right Side.

Measures	M (N=255)	F (N=256)	OR (multivariate)
DMDi	7.1 ± 0.5	6.7 ± 0.4	0.20 (0.12-0.32, $p < .001$)
AGIi	9.3 ± 1.0	9.0 ± 0.9	0.88 (0.71-1.09, p = .234)
DIC	26.4 ± 2.3	25.5 ± 2.0	0.90 (0.82-0.99, $p = .022$)

Table 7: Logistic Regression without Outliers for Left Side.
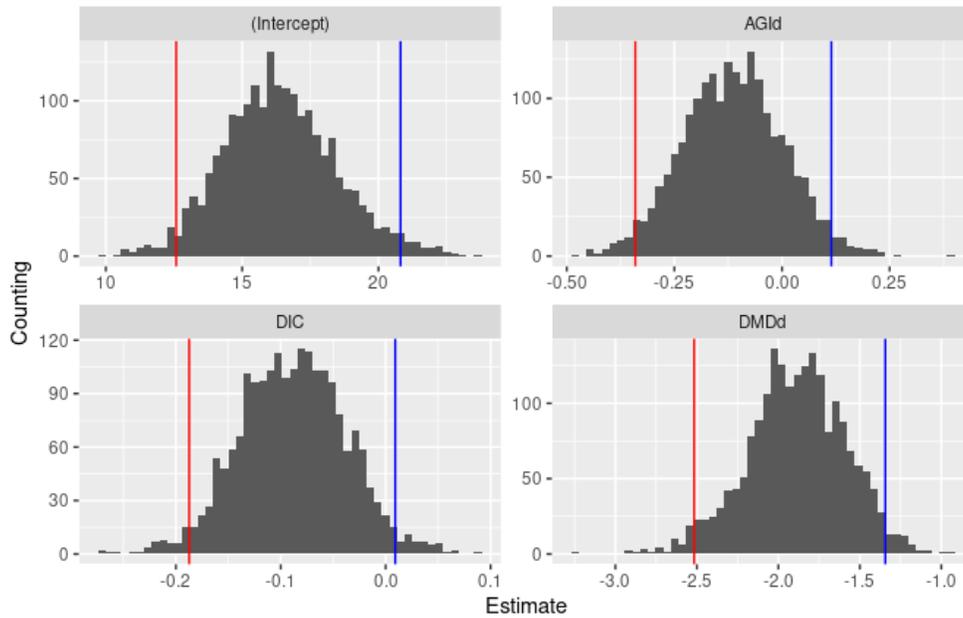
Figure 6: Bootstrap Distribution Logistic Model, Right side, 500 samples.

Both figures 6 and 7 show that the LRA only considers a single measure, which is why it is ruled out as a parametric model, as it only considers a single measure in each model, even though they may have good performance in classification metrics. Finally, for comparison purposes, the k − neighbors method is applied to the same training sample used for the LDA and QDC, changing the parameter k to see how the metrics vary.
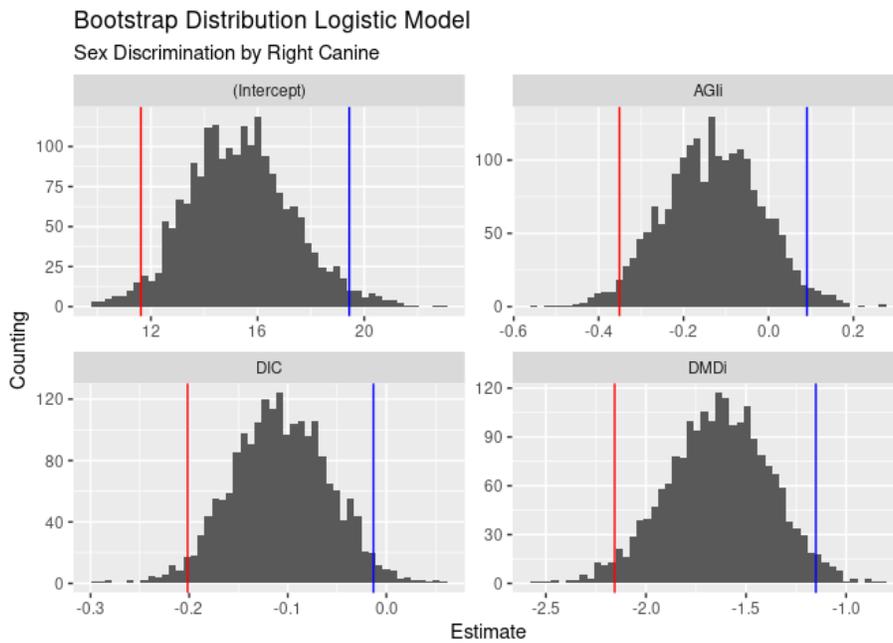

Figure 7: Bootstrap Distribution Logistic Model, Left side, 500 samples.
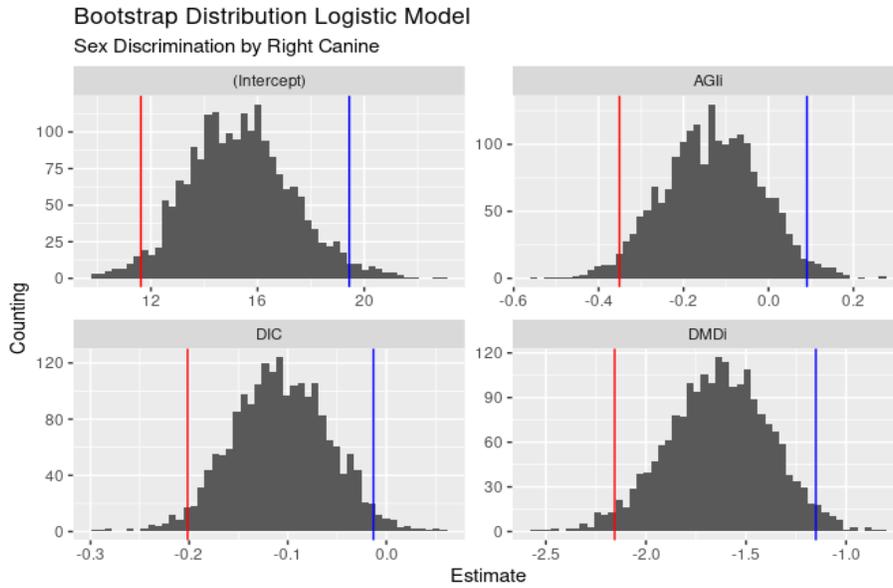
Figure 7: Bootstrap Distribution Logistic Model, Left side, 500 samples.

K	3	5	7	9	11
Sensitivity	0.532	0.532	0.645	0.677	0.661
Specificity	0.606	0.591	0.561	0.606	0.652
(+) Valor Pred	0.559	0.550	0.580	0.618	0.641
(-) Valor Pred	0.580	0.574	0.627	0.667	0.672
Precision	0.559	0.550	0.580	0.618	0.641
Recall	0.532	0.532	0.645	0.677	0.661
F1	0.545	0.541	0.611	0.646	0.651
Prevalence	0.484	0.484	0.484	0.484	0.484
Detection Rate	0.258	0.258	0.312	0.328	0.320
Detection Prevalence	0.461	0.469	0.539	0.531	0.500
Balanced Accuracy	0.569	0.562	0.603	0.642	0.656

Table 8: Metrics for k = 3 a k = 11.

K	13	15	17	19	21
Sensitivity	0.661	0.694	0.661	0.710	0.726
Specificity	0.636	0.636	0.621	0.636	0.636
(+) Valor Pred	0.631	0.642	0.621	0.647	0.652
(-) Valor Pred	0.667	0.689	0.661	0.700	0.712
Precision	0.631	0.642	0.621	0.647	0.652
Recall	0.661	0.694	0.661	0.710	0.726
F1	0.646	0.667	0.641	0.677	0.687
Prevalencia	0.484	0.484	0.484	0.484	0.484
Detection Rate	0.320	0.336	0.320	0.344	0.352
Detection Prevalence	0.508	0.523	0.516	0.531	0.539
Balanced Accuracy	0.649	0.665	0.641	0.673	0.681

Table 9: Metrics for k = 13 a k = 21.

Distribution of prediction metrics
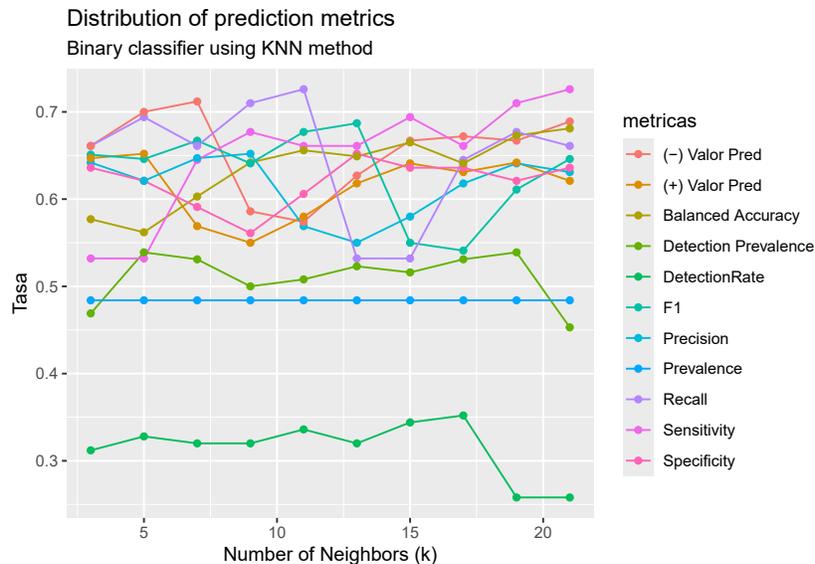Binary classifier using KNN method

Figure 8: Performance of the classification model.

When comparing the three classifiers (DA, LRA, and kNN), it is observed that the data present anomalous values that must be filtered out.However, despite this, multinormality is not verified for DA, which should justify its discarding, despite its good performance as a classifier. When using an alternative, also parametric, classifier such as LRA, without the restrictions of distribution in the two groups, the measurements appear to be correlated, which prevents obtaining a model with all five variables. By discarding nonsignificant variables, two models are chosen that consider the two dimensions of each canine (height and diameter), considering each side.

However, the idea persists that the only relevant measurement is the mediodistal diameter for both sides, which, strictly speaking, makes it unusable for forensic specialists. Bootstrap resampling corroborates that this is a characteristic of the data and that a potential learning sample does not depend on it. Finally, the third binary classifier used is the kNN, using several values of k, where it can be observed that since the reference for classification is the male sex, the sensitivity (the rate of correctly classified men) vs. the specificity (the rate of correctly classified women) changes, showing a slight bias depending on k, with a behavior that is not monotonous and that if compared with the DA, the bias is maintained and, in addition, for the CDA the metrics are impoverished.

10. Conclusion
The conclusions so far show that the classifiers' performance does not differ drastically when changing, and in particular for the LRA, the fact that only one measure is significant suggests that another attribute should be considered, which was not used because it is a qualitative attribute: crowding (crowded teeth) or diastema (spaced teeth). This attribute was not used to allow for binary classifiers for any type of teeth, and not if those data are removed, which would lead to binary classifiers for normal teeth.
For this reason, three paths to follow are proposed as a continuation of the work.
• Filter the data for teeth without crowding or diastema pathologies and evaluate the three proposed classifiers.
• For the case of the LRA, not only evaluate the relevance of joint or side-by-side models using the two measurements of each canine, but also the performance of the classification metrics (which was not done in this work).
• For crowded or diastema teeth, evaluate the three classifiers separately.
• For all teeth, test an LRA considering the crowding variable as a regressor, which would perhaps allow for a global model, but which would not be comparable with the other classifiers.

Furthermore, as alternatives to testing other classifiers, it is proposed to see how support vector machines that only work with quantitative regressor variables work, [21, cap 9], and classification trees that support any type of variable, being a non-parametric method that makes it comparable to LRA and that also allows extension through techniques such as random forests and boosting and bagging methods, [21, cap 8].

References
1. Buchner, A. (1985). The identification of human remains. *International dental journal, 35*(4), 307-311.
2. Paulete, J. Borborema, M. 2011 Vade Macum de Medicina Legal e Odontologia, Legal, 2 ed.

3. Sassi, C., Picapedra, A., Lima, L. N. C., Júnior, L. F., Daruge, E., & Júnior, E. D. (2012). Sex determination in Uruguayans by odontometric analysis.

4. Clark, D. H. (1994). An analysis of the value of forensic odontology in ten mass disasters. *International Dental Journal, 44*(3), 241-250.

5. Campos, Neto., M. Paulete., Vanrell, J. (2014). Atlas de medicina legal. Guia pra´tico para m´edicos e operadores do direito, Vol. tomo 1., LEUD.

6. Nahidh, M., Ahmed, H. M. A., Mahmoud, A. B., & Murad, S. M. (2013). The role of maxillary canines in forensic odontology. *Journal of Baghdad College of Dentistry, 325*(2209), 1-5.

7. ´Álvarez-Vaz, R., & Sassi, C. (2020). DETERMINACIÓN DEL SEXO MEDIANTE TÉCNICAS DE CLASIFICACIÓN SUPERVISADA. Revista de la Facultad de Ciencias, 9(1), 6-24.

8. Harvey, J. 1975. Dental identification and Forensic Odontology, Bristol edn, John Wright &Sons, pp. 140–157.

9. Acharya, A. B., Prabhu, S., & Muddapur, M. V. (2011). Odontometric sex assessment from logistic regression analysis. *International journal of legal medicine, 125*, 199-204.

10. Rao, N. G., Rao, N. N., Pai, M. L., & Kotian, M. S. (1989). Mandibular canine index—a clue for establishing sex identity. *Forensic science international, 42*(3), 249-254.

11. Eboh, D. E. O., & Etetafia, M. O. (2010). Maxillary canine teeth as supplement tool in sex determination. *Annals of Biomedical Sciences, 9*(1).

12. Picapedra, A., Sassi, C., Massa, F., Francesquini Jr, L., Daruge, E., & Daruge Jr, E. (2012). Odontometric analysis of maxillas: a device for sex determination. *Inter J Dental Anthropol, 21*, 01-16.

13. Gargano, V., Picapedra, A., Sassi, C., Lima, L., Alvarez, R., & Francesquini Jr, L. D. (2014). ¿ Son los índices caninos mandibular y maxilar herramientas fidedignas para la determinación del sexo. *Actas odontológicas, 11*(1), 22-34.

14. Kaushal, S., Patnaik, V. V. G., & Agnihotri, G. (2003). Mandibular canines in sex determination. *J Anat Soc India, 52*(2), 119-24.

15. Rai, B., & Anand, S. C. (2007). Gender determination by diagonal distances of teeth. *The Internet Journal of Biological Anthropology, 1*(1).

16. Srivastava, P. C. (2010). Correlation of odontometric measures in sex determination. *Journal of Indian Academy of Forensic Medicine, 32*(1), 56-61.

17. Narang, R. S., Manchanda, A. S., Malhotra, R., & Bhatia, H. S. (2014). Sex determination by mandibular canine index and molar odontometrics: A comparative study. *Indian Journal of Oral Sciences*, Vol, 5(1).

18. Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S, Springer, New York: ISBN 0-387-95457-0.

19. Blanco Gónzalez, J. 2006. Introducción al Análisis Multivariado, IESTA.

20. ´ Álvarez-Vaz, R. (2020). Sistematización y creación de indicadores e índices para la vigilancia epidemiológica en salud bucal: uso de técnicas estadísticas multivariantes y de análisis espacio-temporal.

21. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, No. 1). New York: springer.

22. Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. BMJ: British Medical Journal, 308(6943), 1552.

23. Altman, D. G., & Bland, J. M. (1994). Statistics Notes: Diagnostic tests 2: predictive values. *Bmj, 309*(6947), 102.

24. Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of statistical software, 28*, 1-26.

25. R Core Team 2024. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

26. Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). klaR analyzing German business cycles. Data analysis and decision support, 335-343.

27. Friendly, M. (2010). HE plots for repeated measures designs. *Journal of Statistical Software, 37,* 1-40.

28. Frick, H., Chow, F., Kuhn, M., Mahoney, M., Silge, J. Wickham, H. (2023). rsample: General Resampling Infrastructure. R package version 1.2.0.

29. Wickham, H. (2016). Data analysis. In ggplot2: Elegant graphics for data analysis (pp. 189-201). Cham: Springer International Publishing.

30. Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A. Crowley, J. (2023). GGally: Extension to 'ggplot2'. R package version 2.2.0.

31. Álvarez Vaz, R., Sassi, C., Gargano, V., & Picapedra, A. (2025). Comparison of binary classifiers in forensic dentistry for sex determination. *medRxiv*, 2025-05.