MEDIRES
Symbol of Quality Research

Journal of Skeleton System
info@medireonline.org

ISSN: 2836-2284

RESEARCH ARTICLE

Determining The Sex of a Person by Odontometric Parameters Using Logistic Regression Equations

Anar Sh. Ibragimov[1*], Shakir M. Musayev[2], Namik N. Orudjov[2]

[1]*Khazar University, Assistant professor of the Department of Biology and Natural Sciences. Baku, Republic of Azerbaijan.*
[2]*Azerbaijan State Institute for Advanced Training of Medical Doctors named after A.Aliyev, Head of the Department "Forensic medicine and pathological anatomy, with courses of foundations of law". Baku, Republic of Azerbaijan.*
[3]*Azerbaijan State Institute for Advanced Training of Medical Doctors named after A.Aliyev, Senior researcher, MD. Baku, Republic of Azerbaijan.*

***Corresponding Author**: Anar Sh. Ibragimov, Khazar University, Assistant professor of the Department of Biology and Natural Sciences. Baku, Republic of Azerbaijan.

Received Date: 10 September 2022**; Accepted Date:** 23 September 2022**; Published date:** 28 October 2022

Citation: Anar Sh. Ibragimov, Shakir M. Musaye2, Namik N. Orudjov, (2022). Determining The Sex of a Person by Odontometric Parameters Using Logistic Regression Equations. Journal of Skeleton System. 1(1); DOI: 10.58489/2836-2284/001

Abstract

The development of personal identification methods based on bone remains, which are used to diagnose the population of Azerbaijan, is a priority task due to the unresolved issue. In this scenario, a new method for quickly determining sex by teeth is in demand, as it should increase the efficiency of the work of forensic experts. The purpose of this study was to develop modern methods for diagnosing human sex using odontometric studies. At the craniological collection of Azerbaijanis, the sizes of teeth were studied. On 7 teeth of the upper and 7 teeth of the lower jaw, 3 parameters were measured (in this case, symmetrical teeth were considered as a single array). Each of the 14 anatomical teeth was represented by 80 male and 80 female teeth. For odontometric procedures, a Mutitouyo digital caliper (accuracy 10 microns) was used. The results of odontometry were analyzed by statistical methods using ROC analysis. Mathematical modeling was carried out using the MATLAB statistical software package. With the help of ROC-analysis, the data obtained during odontometry were considered. 6 odontometric signs were established, which are interconnected with gender. The most correlated with gender were the bucco-lingual diameter of the upper canine, as well as the heights of the crowns – the second lower molar, lower incisors, lower canine and upper medial incisor. Using mathematical modeling, we have developed 6 logistic regression equations that can be used to determine the gender of a person by the size of his teeth. The predictive potential of the equations was also evaluated using ROC analysis procedures. As a result of the study, based on odontometric indicators and by using ROC analysis, diagnostic equations were obtained to determine the sex of a person by the size of the teeth. The equations have not yet been verified on independent samples, so they cannot yet be unconditionally recommended for practice. Nevertheless, the authors count on the high legitimacy of their own models due to the fact that they used a fairly powerful mathematical apparatus.

Keywords: *human identification; odontometrics; sex determination; ROC- analysis; prediction modeling.*

Introduction

A craniological collection of Azerbaijanis was studied (race, gender, age, height and some other features are known). This collection has been collected since 1970 in various regions of the country and is currently stored in the archive of the Person of Public Law "Association of FME and PA" of the Ministry of Health of the Republic of Azerbaijan. This collection (includes 205 skulls; male - 135; female - 70), which adequately reflects the craniological characteristics of the modern population of Azerbaijan, has been introduced into scientific discussion since 1995 and has repeatedly become material for various studies [3-5]. With the digital caliper "Mutitouyo" (accuracy 10 micrometers), the odontometric characteristics of the teeth (except for the third molars) were studied on these skulls. Crown height (Crh; for molars measured from the enamel-cement border to the level of the masticatory surface), vestibulo-lingual crown size (Vel; greatest thickness) and mesial-distal crown size (Mes; greatest width) were measured. For odontometry, standard procedures were used, according to the generally accepted recommendations of A.A. Zubov [6]. Only teeth that could be extracted from the alveoli were examined. Thus, the studied material represented a simple random sample. As a result, two samples (male and female) were completed, which met the objectives of our study and included 80 teeth each (sample sizes were not previously calculated). In our previous publications, we reported that in our previous studies we compared the parameters of identical right-sided and left-sided teeth with each other and did not reveal obvious differences between them in size [3]. Therefore, in both samples, data on symmetrical teeth were combined into one array. Thus, in general, summary odontometric indicators of 7 teeth of the upper and 7 teeth of the lower jaw were obtained. Abbreviation of the upper teeth: UI1-upper medial incisor, UI2-upper lateral incisor, UC-upper canine, UPm1-upper first premolar, UPm2-upper second premolar, UM1-upper first molar, UM2-upper second molar. Accordingly, if the first letter is "L" in the abbreviation, then this is a similar tooth on the lower jaw. As a result, a total of 1680 (3x7x80) odontometric features were studied on each sample of skulls. To analyze the results of odontometry, the methods of variation statistics and ROC analysis were used [7]. Mathematical modeling and calculations were performed using the MATLAB statistical software package (version 8.6).

Research results

Table1: Statistical characteristics of odontometric features in a sample of male skulls

Odontometric feature	N	Min.	Max.	X	X(r)	σ	Cv	As	Ek
Vel_UI1	80	5,70	8,18	7,03	0,08	0,74	0,54	-0,10	-1,19
Mes_UI1	80	6,94	10,20	8,45	0,11	0,99	0,98	0,25	-1,21
Crh_UI1	80	8,89	10,40	9,48	0,04	0,40	0,16	0,24	-1,10
Vel_LI1	80	4,80	7,05	6,15	0,06	0,56	0,31	-0,17	-1,06
Mes_LI1	80	4,59	7,19	5,74	0,08	0,75	0,56	0,14	-1,24
Crh_LI1	80	6,64	9,24	7,98	0,08	0,74	0,55	-0,17	-1,16
Vel_UI2	80	5,08	7,92	6,48	0,09	0,77	0,60	0,16	-1,16
Mes_UI2	80	5,22	8,36	6,85	0,10	0,89	0,79	0,01	-1,20
Crh_UI2	80	7,79	10,23	9,03	0,08	0,69	0,47	0,10	-1,31
Vel_LI2	80	4,81	7,72	6,31	0,09	0,82	0,67	0,04	-1,15
Mes_LI2	80	4,37	7,32	5,97	0,09	0,81	0,66	-0,05	-1,13
Crh_LI2	80	7,14	9,58	8,25	0,08	0,69	0,48	0,06	-1,27
Vel_UC	80	4,38	7,15	5,98	0,08	0,69	0,47	0,30	-1,09
Mes_UC	80	6,59	8,81	7,81	0,07	0,59	0,35	0,01	-0,97
Crh_UC	80	9,25	12,00	10,8	0,08	0,73	0,53	-0,09	-1,24
Vel_LC	80	6,09	8,63	7,29	0,08	0,73	0,53	0,06	-1,17
Mes_LC	80	5,76	8,31	6,91	0,09	0,78	0,60	0,10	-1,45
Crh_LC	80	8,88	11,59	10,1	0,08	0,68	0,46	-0,06	-0,89
Vel_UPm1	80	7,13	9,98	8,96	0,08	0,75	0,57	-0,48	-0,69
Mes_UPm1	80	6,09	9,53	7,04	0,08	0,70	0,49	0,92	1,04
Crh_UPm1	80	7,03	9,78	8,35	0,09	0,85	0,72	0,04	-1,31
Vel_LPm1	80	7,08	9,92	8,42	0,09	0,80	0,63	0,06	-1,18
Mes_LPm1	80	6,61	8,77	7,34	0,05	0,46	0,21	0,34	-0,41
Crh_LPm1	80	6,97	8,98	8,03	0,06	0,58	0,33	0,01	-1,21
Vel_UPm2	80	7,37	10,40	8,87	0,10	0,87	0,76	0,04	-1,04
Mes_UPm2	80	5,65	8,06	6,95	0,07	0,66	0,43	-0,02	-1,19
Crh_UPm2	80	6,28	8,62	7,47	0,08	0,72	0,52	0,01	-1,33
Vel_LPm2	80	6,92	9,68	8,35	0,09	0,84	0,71	-0,06	-1,34
Mes_LPm2	80	6,06	8,71	7,21	0,08	0,67	0,46	0,18	-1,02
Crh_LPm2	80	6,29	8,97	7,65	0,09	0,79	0,63	0,08	-1,12
Vel_UM1	80	9,81	12,67	11,1	0,09	0,78	0,61	0,01	-0,79
Mes_UM1	80	9,01	12,13	10,5	0,11	0,97	0,94	-0,03	-1,30
Crh_UM1	80	5,48	7,88	6,38	0,07	0,60	0,36	0,16	-1,10
Vel_LM1	80	8,65	11,81	10,4	0,10	0,94	0,88	-0,15	-1,25
Mes_LM1	80	9,15	12,66	11,0	0,12	1,04	1,09	-0,10	-1,29
Crh_LM1	80	6,03	7,98	6,78	0,05	0,45	0,21	0,36	-0,43
Vel_UM2	80	9,85	12,83	11,3	0,10	0,93	0,86	-0,02	-1,31
Mes_UM2	80	7,91	11,12	9,69	0,10	0,91	0,83	-0,01	-1,27
Crh_UM2	80	5,09	7,72	6,02	0,06	0,57	0,32	0,30	-0,35
Vel_LM2	80	8,52	11,45	10,1	0,09	0,85	0,71	-0,08	-1,13
Mes_LM2	80	9,39	12,41	10,8	0,10	0,90	0,81	-0,07	-1,21
Crh_LM2	80	4,78	7,16	5,97	0,07	0,66	0,43	0,07	-1,04

The total number of parameters, as we indicated, was 3360. It was also noted that the data for symmetrical teeth are presented in a combined array. The ratio of right-sided and left-sided teeth in our sample was very close to 1, which we noted in previous works [3]. The actual digital material collected in the odontometric blanks, after the grouping procedures described above, was digitized into an electronic database. For each gender, we obtained summary odontometric indicators separately. Then the corresponding statistical characteristics were calculated separately for the male and female samples. The results of the corresponding statistical analysis are presented below in tables №1 and №2.

Table2*:* Statistical characteristics of odontometric features in a sample of female skulls

Odontometric feature	N	Min.	Max.	$\bar{x}$	X(r)	σ	Cv	As	Ek
Vel_UI1	80	5,40	7,96	6,85	0,08	0,73	0,53	-0,06	-1,28
Mes_UI1	80	6,61	9,33	8,05	0,09	0,80	0,63	-0,03	-1,07
Crh_UI1	80	7,23	10,20	8,68	0,10	0,87	0,75	0,11	-1,22
Vel_LI1	80	4,22	6,89	5,48	0,09	0,76	0,58	0,13	-1,13
Mes_LI1	80	4,12	6,72	5,35	0,08	0,72	0,51	0,05	-1,14
Crh_LI1	80	6,03	8,31	7,05	0,07	0,66	0,44	0,18	-1,06
Vel_UI2	80	4,53	7,42	5,91	0,10	0,94	0,88	0,02	-1,55
Mes_UI2	80	5,02	8,08	6,52	0,10	0,92	0,84	0,09	-1,15
Crh_UI2	80	7,31	9,94	8,51	0,07	0,66	0,44	0,08	-0,92
Vel_LI2	80	4,48	7,50	5,72	0,09	0,78	0,61	0,23	-0,62
Mes_LI2	80	4,68	7,64	5,85	0,08	0,74	0,55	0,30	-0,75
Crh_LI2	80	6,26	8,78	7,45	0,08	0,71	0,50	-0,03	-1,12
Vel_UC	80	6,14	9,01	6,84	0,10	0,92	0,85	-0,11	-1,19
Mes_UC	80	6,02	9,18	7,62	0,11	0,98	0,97	-0,03	-1,39
Crh_UC	80	8,75	11,40	10,06	0,09	0,76	0,58	0,17	-1,09
Vel_LC	80	5,38	8,61	7,01	0,11	0,95	0,90	0,01	-1,07
Mes_LC	80	5,21	8,04	6,65	0,09	0,80	0,65	0,01	-1,10
Crh_LC	80	8,21	10,80	9,42	0,08	0,69	0,48	-0,09	-1,20
Vel_UPm1	80	7,14	10,07	8,62	0,10	0,92	0,85	0,18	-1,36
Mes_UPm1	80	6,39	8,98	7,68	0,09	0,77	0,60	0,01	-1,21
Crh_UPm1	80	6,61	9,13	7,87	0,08	0,74	0,55	0,06	-1,28
Vel_LPm1	80	6,15	8,91	7,47	0,09	0,83	0,69	-0,09	-1,17
Mes_LPm1	80	5,69	8,49	6,92	0,08	0,70	0,49	0,04	-0,79
Crh_LPm1	80	6,58	9,19	7,95	0,09	0,77	0,59	-0,12	-1,25
Vel_UPm2	80	7,02	9,96	8,46	0,10	0,86	0,74	0,01	-1,18
Mes_UPm2	80	5,23	7,87	6,56	0,09	0,78	0,60	0,01	-1,02
Crh_UPm2	80	5,95	8,56	7,06	0,08	0,67	0,45	0,07	-1,10
Vel_LPm2	80	6,33	9,46	7,85	0,10	0,90	0,80	0,01	-1,26
Mes_LPm2	80	6,01	8,58	7,12	0,08	0,69	0,48	0,19	-1,05
Crh_LPm2	80	6,11	8,82	7,41	0,08	0,74	0,55	0,06	-0,83
Vel_UM1	80	8,88	12,55	10,73	0,11	1,01	1,03	0,03	-1,13
Mes_UM1	80	8,57	11,53	10,08	0,10	0,89	0,80	-0,10	-1,26
Crh_UM1	80	4,61	7,01	5,87	0,07	0,67	0,45	0,01	-1,03
Vel_LM1	80	8,54	11,50	9,81	0,09	0,82	0,68	0,20	-1,12
Mes_LM1	80	9,03	12,21	10,68	0,10	0,91	0,83	0,04	-1,05
Crh_LM1	80	5,19	7,45	6,29	0,07	0,63	0,40	-0,03	-1,30
Vel_UM2	80	8,85	11,90	10,47	0,09	0,82	0,67	-0,11	-1,14
Mes_UM2	80	7,83	11,05	9,47	0,10	0,94	0,88	-0,03	-1,21
Crh_UM2	80	4,45	7,24	5,71	0,08	0,75	0,57	0,11	-1,22
Vel_LM2	80	8,27	11,23	9,77	0,09	0,83	0,69	-0,11	-1,11
Mes_LM2	80	8,34	11,72	10,03	0,11	1,01	1,02	-0,13	-1,24
Crh_LM2	80	4,28	6,91	5,16	0,06	0,57	0,32	0,52	0,01

The variation-statistical indicators of the studied samples did not reveal any significant features. The obtained database on odontometry of male and female populations is so far the first large-scale description of the odontometric features of the population of modern Azerbaijan.

However, with all this, some features were noted in the distribution of features, for example, regarding the indicators of asymmetry and kurtosis. The kurtosis coefficient for most odontometric features turned out to be negative and almost always exceeded the value of the standard error by more than two times. Statistical parameters for the primary sample showed its strong variability (the coefficient of variation (Cv) for only one parameter was 17%, and in other cases >20%). That is, the results of biometric tests showed that the implementation of diagnostic models requires a transition to nonparametric statistics. In addition, there were other restrictions. Thus, it was not possible to construct equations for gender prediction using canonical discriminant analysis – consideration of the corresponding correlation matrices and analysis of the predictor potential of the most appropriate odontometric signs did not reveal encouraging prospects (these results are not presented here). Based on these messages, the rational way out was to use another statistical classification method – logistic regression.

Taking into account the nuances described above, in the present work, we applied the mathematical techniques of ROC analysis. The digital data obtained by us related to odontometric signs were analyzed using the MATLAB software package. Analysis of the results of this analysis allowed us to identify six odontometric features that were most specifically related to gender. These signs turned out to be: Vel_UC, Crh_LM2, Crh_LI2, Crh_LC, Crh_LI2, Crh_UI1. Odontometric parameters are presented in descending order of their relationship to gender. That is, the bucco-lingual diameter of the upper canine and the height of the crown of the second lower molar fell into the category of the most gender-linked sizes, and the weakest relationship with gender was at the height of the crown of the upper medial incisor.

When using the ROC analysis method, each odontometric parameter can play the role of a predictive model. ROC-analysis in this case cuts off the most sensitive odontometric parameters in this context. Of course, many other tooth sizes also have some relationship to gender. But the classifier we use (in contrast to the correlation classifier) also takes into account the specifics of the characteristic feature for the predicted phenomenon. That is, to put it simply, ROC analysis also allows you to cut off the "reliability" of signs (odontometric dimensions) in the context of sex. The ROC-curves obtained in the course of the analysis take into account both the general trend in the behavior of the features (similarly as the correlation coefficients) and the number of objects in the matrix that truly meet the criteria for a specific segregation (that is how many teeth with a certain value of a particular odontometric indicator in the sample exactly match, for example, with the male sex). These interpretations of ours are very clearly demonstrated by the graph of the corresponding ROC curves in Figure 1.

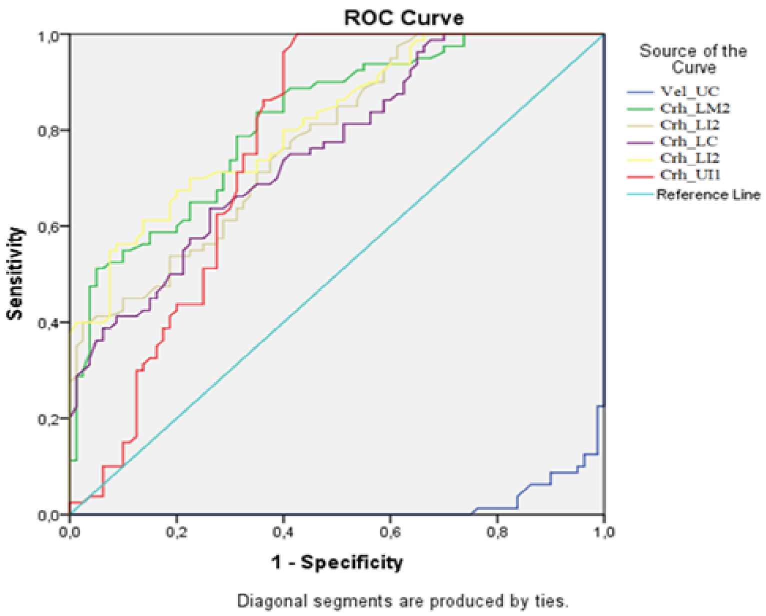

Fig 1: Graphic representation of the effects of prognostic classifier of some teeth sizes on the basis of ROC-analysis

The straight diagonal that divides our graph into two halves is called the line of indistinguishability of two classes, which allows us to divide (discriminate) points (our curves are a collection of points) into different classes. This line of indistinguishability between the two classes corresponds to the so-called "useless" classifier. The closer the points of our curves are to this line, the lower the predictor potential of the trait under consideration. Thus, we see that the six odontometric features listed above demonstrate fairly reliable predictive capabilities for gender prediction. Points on the line of indistinguishability of two classes are completely indifferent to the sex of the tooth and thus are, as it were, a barrier of differentiation in the general scalar field of the floor. As for the points from which the ROC-curves are formed, they, as shown above, are obtained while taking into account the sensitivity and specificity of the "behavior" of the odontometric sign.

From the foregoing, we can conclude that the optimal feature will be considered the one that will have such characteristics as 100% specificity and 100% sensitivity. However, on a practical biometric material, it is impossible to achieve a simultaneous increase in both sensitivity and specificity for some predictor. The problem in this case is solved by finding a compromise point, or the so-called cut-off threshold. The cutoff threshold is the point at which the specificity and sensitivity values are most effective in mutual accounting. Roughly speaking, the cutoff threshold is the largest indicator of the sum of the coordinate values along both axes of the classifier graph. The sensitivity and specificity coordinates of each point of any of our ROC curves for the above six odontometric features are shown in Table 3.

Table3: Sensitivity and specificity indices of gender-related odontometric features

Odontometric feature	Sensitivity	Specificity	Cutoff threshold	Feature size in mm	TDV *
Vel_UC	0,988	0,975	1,86	6,32	93,1%
Crh_LM2	0,838	0,650	1,49	5,32	73,8%
Crh_LI2	0,713	0,650	1,363	7,71	68,1%
Crh_LC	0,638	0,737	1,375	9,86	68,8%
Crh_LI2	0,700	0,775	1,475	7,60	73,8%
Crh_UI1	1,000	0,575	1,575	8,87	78,8%

***** – total diagnostic value

When analyzing using the MATLAB package, the program, implementing ROC analysis, simultaneously calculates the total diagnostic value (TDV) of the analyzed model. This characteristic reflects the quality level of the model and is expressed as a percentage. The higher it is, the more predictive power the model has. Table 3 shows that the best predictor of sex is the vestibulo-lingual size of the upper canine.

The next step in this work was to build prediction models using the most specifically sex-related odontometric parameters. The essence of the problem, in principle, came down to checking the adequacy of the developed dınamıc models as methods of diagnosis (sex prediction). There are a lot of ways to assess the reliability of various dynamic forecast and recognition models. But recently, various techniques from the theory of fuzzy logic and classification methods in machine learning have proven themselves well [8, 9]. The used ROC-analysis is one of the effective methods of such methodologies, and with its help we assessed the efficiency and reliability of the equations developed for diagnostics. Logistic regression models, unlike conventional linear regression equations, are not designed to predict (calculate) the value of any parameter based on several other indicators associated with it. That is, the regressors, which in this case are input variables, do not determine the output variable of the same class as the original ones. In this case, according to the proposed formulas, the probability of which gender (group, population, etc.) the tooth (or teeth) belongs to is established (more precisely calculated) based on the values of its parameters, which in fact are elements of the developed function. The essence of logistic regression is that the space of initial values (a set of indicators of odontometric features) can be divided into two areas, each of which corresponds to one of the gender dispositions, and the plane (or line) that delimits the matrix of odontometric parameters called is a linear discriminant. In the described models, the delimiting plane is given by the values of the coefficients of such a linear equation, the solution of which is obtained in the form of any single digit. This figure is compared with the center of the discriminant line, which is also some kind of figure, that is, a predetermined constant value. By the way, it is appropriate to note that this constant is set precisely with the help of ROC analysis procedures. Depending on whether this figure is greater or less than the one calculated using the logistic regression equation, the model recognizes the object's class (in this case, gender). It is obvious that the outcomes of the solution in the form of a result equal to the constant value itself are possible. In this case, it is considered that the model cannot classify the object based on the parameters of this object entered into it. A total of six logistic regression equations were developed, which are listed below in order of strength for their TDV:

1) D = -0,422 + Vel_UCx0,289;

2) D = 0,732 + Vel_UCx0,252 + Crh_LM2x (-0,164);

3) D = 1,719 + Crh_LI2x (-0,123) + Vel_UCx0,231 + Crh_LM2x (-0,143);

4) D = 2,94 + Crh_LI2x (-0,115) + Vel_UCx0,211 + Crh_LCx(-0,118) + Crh_LM2x(-0,142);

5) D = 3,837 + Vel_UCx0,183 + Crh_LM2x (-0,143) + Crh_LI2x (-0,105) + Crh_LCx (-0,119) + Crh_LI1x (-0,103);

6) D = 4,751 + Crh_LI2x (-0,102) + Vel_UCx0,162 + Crh_LCx(-0,116) + Crh_LM2x(-0,143) + Crh_LI1x(-0,096) + Crh_UI1x(-0,097);

* – Explanations on the equations are given below in the text.

To assess the level of confidence reliability of the developed equations in the described analysis, their specificity and sensitivity are determined. The sensitivity of the diagnostic model in this case was determined by the relative weight of positive results in the diagnosis of female. That is, the correct diagnosis of female sex accounted for the proportion of true positive test results. The specificity of the model was determined by the proportion of true negative results of a particular equation. In our case, it looked like a male diagnosis, when the gender was indeed male (i.e., teeth from a male skull were examined). Without delving into an even more detailed explanation of all ROC analysis procedures, we will present the ROC curves themselves directly. In our case, the characteristic ROC curves for the analysis of the equations were constructed based on the mentioned values of sensitivity and specificity. Each curve plotted the number of correctly diagnosed positive cases versus the number of misdiagnosed negative cases in our sex recognition models. In other words, a graphical display of the proposed binary classification test is carried out based on a matrix of odontometric indicators with different gender dispositions. We also note that the ideal ROC-curve has an “Г-shape” and the efficiency of the dynamic model is the higher, the greater the similarity of its characteristic curve with a graph of a similar configuration. Thus, with the help of a computer program, we obtained ROC curves for the developed equations. A comparative graphical representation of the ROC curves for the equations put forward by us is presented below in Figure 2.

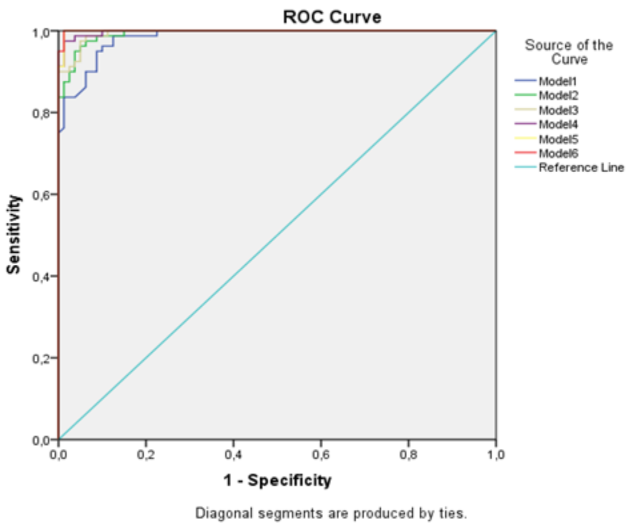

Fig 2: Characteristic ROC-curves of the diagnostic equations we created, which were used to evaluate their effectiveness

The presented graph clearly shows that the reliability of diagnostic models grows from the first to the sixth. This moment is quite simple to understand, because the numbering of our models was ranked taking into account the number of regressors in them. Thus, in the first model, one odontometric sign is used as a regressor, and in the sixth, six. It is obvious that the more predictors in the equation, the greater its predictive (predictive) ability. Therefore, the sixth equation, in which there are six regressors, turned out to be the most reliable for determining sex.

The first digit in each equation is a constant. Instead of the abbreviation of odontometric features, their numerical indicators for a particular tooth in millimeters are substituted. These values are then multiplied by their coefficient, which is indicated next to the sign with respect to signs. At the end, all the obtained indicators and the constant are summed up, which gives us the value D. The indicated indicator D is compared, as we said, with a given constant value – the central point of the discriminant separation plane. Our regression equations are designed in such a way that if the calculated value of «D» is greater than the given centroid, then the female gender will be diagnosed. Well, if the calculated value of «D» is less than the given centroid, then the male gender will be diagnosed. The numerical values of these centroids, as well as some of the statistical parameters that determine them, are presented in Table 4 below.

Table 4: Central values of discriminant lines for gender recognition based on the developed logistic regression equations

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Centroid	1,420	1,447	1,474	1,491	1,424	1,458
Sensitivity	95	95	97,5	97,5	100	100
Specificity	91,3	96,3	95	98,8	98,8	98,8
TDV	93,1%	95,6%	96,3%	98,1%	99,4%	99,4%

The centroid values themselves, which are presented in this table, were calculated using ROC analysis. Here, the same procedures were applied as when we extracted the most informative odontometric indicators for use in dynamic sex prediction models. The difference was that in this experiment, the linear discriminant of the indistinguishability of two classes served to determine the predictor potential of not one feature, but the entire equation. As we noted above, in this case, the computer program on the matrix of real data calculates variants of false and true results (both negative and positive), which are possible with the applied implementation of the constructed equations. The matrix is the entire set of our odontometric features, that is, the parameters of all 160 teeth (female + male). The chain of calculations shown allows the applied classifier to generate an ROC curve for each of the equations we have developed.

Discussion of the Research Results

Using the results of odontometry and with the help of Roc-analysis, in this study, it was possible to develop diagnostic equations for determining the sex of a person by the size of his teeth. In the course of the work, a sample that was quite representative of the target population was studied. As noted, the craniological collection of Azerbaijanis, which was studied in this work, has been repeatedly studied before. The techniques developed by various authors on the basis of these studies have found successful application in the daily work of forensic specialists in our country [3-5]. The equations obtained in the study are more reliable, the more predictors are associated in them. This is not difficult to see in Figure 2. Hence, it could be assumed that a subsequent increase in the number of predictors should improve the quality of the forecast, and it was necessary to continue compiling equations. However, our work with the following possible equations showed that the level of diagnostics does not improve further. Moreover, the reliability and reliability of diagnosis in some cases even worsens. The fact is that with an increase in the number of regressors in the equations, their sensitivity increases almost prohibitively. But at the same time, the specificity of the diagnostic model (it would seem that it is also quite high) nevertheless, starting from model No. 5, remains stable. At the same time, the extremely high sensitivity tends to unity (sensitivity 100%). In practical terms, this means that models with the specified characteristic (models No. 5, No. 6 and above) are no longer able to increase specificity, and when diagnosing with their help, in any scenario, truly false conclusions are possible, which, although minimal, are inevitable. In principle, today, there are no models with a 100% error-free result, and such a situation seems to be insignificant in terms of application. But there is still a moment with the so-called "hyperdiagnostics", which, according to theoretical information from statistical science, takes place in the described situation. The essence of this phenomenon is that, with a certain configuration, the constructed dynamic model for forecasting works better only in one direction. That is, the test works correctly where the diagnosis is doubtful and with other, slightly changed, parameters of the equation, it would not be confirmed, or there would be a motivated refusal to make a decision. In expert practice, a false positive conclusion about gender has no less catastrophic consequences than in the clinic. In this case, our position coincides, in general, with modern theoretical ideas that forecast models that seem to have no flaws should be treated critically [10]. This is clear even without mathematical calculations. After all, it is clear that a 100% reliable forecast in terms of quantitative indicators is already, as it were, a qualitative sign of some object or phenomenon (a physical characteristic like color, density, etc.). In view of these reservations, certain nuances in the creation of the class of diagnostic models we have chosen require increased attention. At the same time, we also note that the presented equations have not yet been verified on independent samples, therefore, they cannot yet be unconditionally recommended for applied implementation. A feature of our equations is the use of a smaller number of teeth for diagnosis (as a rule, the parameters of one tooth and one or two of the adjacent ones). True, in some literary sources [11, 12] there are data on the possibility of establishing the sex of an individual even by one odontometric sign. However, we still count on the high legitimacy of our own models due to the fact that we used a fairly powerful mathematical apparatus of ROC analysis. We also note that so far we have been able to compare our results with similar data from Iranian researchers [13]. These authors also presented data on the predictive value of the size of the upper canines and lower molars. True, these results were obtained on the basis of archaeological material from an era not close in chronology, and also differ in the methodology of odontometry. Similar information is still available in the work of a Russian author [14]. Of course, in this case, both works are the most interesting, as they show data on adjacent ethnic groups (regarding the material from the territory of Iran, one can even assume that we are talking about odontological data of some sub-ethnic group of Azerbaijanis).

Conclusion

The study of the odontometric indicators of Azerbaijanis in the craniological collection (museum exhibits) made it possible to develop 6 logistic regression equations, which are designed to determine the sex of a person by the size of his teeth. These equations were developed using the mathematical apparatus of ROC analysis, which is effective in this context. In the course of the work, it was possible to identify six odontometric predictors that were most specifically related to gender. Thus, we can say that the odontometric possibilities for recognizing sexual dimorphism of teeth have not yet been fully exhausted. However, in general, more detailed discussions are needed for each detected feature, and in this aspect, of course, I would like to know the opinion of respected colleagues on this issue.

References

1. Popov V.L., Musaev Sh.M. Forensic Medicine. Saint-Petersburg: «Yurtsentrpress»; 2009, 516 p. (In Russ).
2. İşcan, Mehmet Yaşar, and P. Sema Kedici. "Sexual variation in bucco-lingual dimensions in Turkish dentition." Forensic science international 137.2-3 (2003): 160-164.
3. Musaev Sh.M., Orudjov N.R. Diagnostics of gender identity using mathematical modeling of odontometric characteristics. Scientific reports of the Police Academy. Baku, 2017, vol.2(4), p. 26-33. (In Russ).
4. Ibragimov A.Sh. The ability to determine the length of the human body by the size of the skull. International Journal of Forensic Sciences, 2022, vol.7(2), p. 25-29.
5. Sh.M. Musayev. Medico-criminalistic identification populations of Caucasus on skull. Baku: Elm, 1999, 106 p.
6. Zubov A.A. Odontology. Methods of anthropological research. Moscow: Nauka, 2008, 198 p. (In Russ).
7. Glantz S.A. Primer of Biostatistics (7th Edition). McGraw-Hill Education – Europe, 2011, 320p.
8. https://basegroup.ru/community/articles/logistic
9. https://www.twirpx.com/file/1465664/
10. Kovalev, A. V., Shmarov, L. A., & Ten'kov, A. A. (2016). The reliable and plausible conclusions in the decisions of the forensic medical experts. Sudebno-meditsinskaia ekspertiza, 59(1), 60-64.
11. Narang, R., Manchanda, A., Malhotra, R., & Bhatia, H. (2014). Sex determination by mandibular canine index and molar odontometrics: A comparative study. Indian Journal of Oral Sciences, 5(1), 16-16.
12. Manashev Q.Q., Sharaykin P.N., Levenets A.A., Moiseev E.V. Variability of the dentoalveolar system depending on the sex of the individual. Scientific statements. 2000, vol.2, p. 176. (In Russ).
13. Kazzazi S.M., Kranioti E.F. Sex estimation using cervical dental measurements in an archaeological population from Iran. Archaeological and Anthropological Sciences. 2016, vol.23, p. 219-240.
14. Malsagov O.M. Sexual determinism and odontometric analysis of teeth (PhD thesis). Moscow, 2005, 22 p. (In Russ).