PAMUKKALE ÜNİVERSİTESİ
Mühendislik Bilimleri Dergisi

PAMUKKALE UNIVERSITY
Journal of Engineering Sciences

Araştırma makalesi | Research paper

Cilt | Vol: XX • Sayı | No: X • ss | pp: XXXX-XXXX • Yıl | Year: 20XX

Sex determination from cephalometric radiography using data mining

Gizemnur Erol Doğan [1], Betül Uzbaş [2], Hatice Kök [3]

ARTICLE INFO

Dates:
Received: 07.04.2026
Accepted: 21.07.2026

doi:
10.65206/pajes.1924886

Corresponding author:
Gizemnur Erol Doğan
(gerol@ktun.edu.tr)

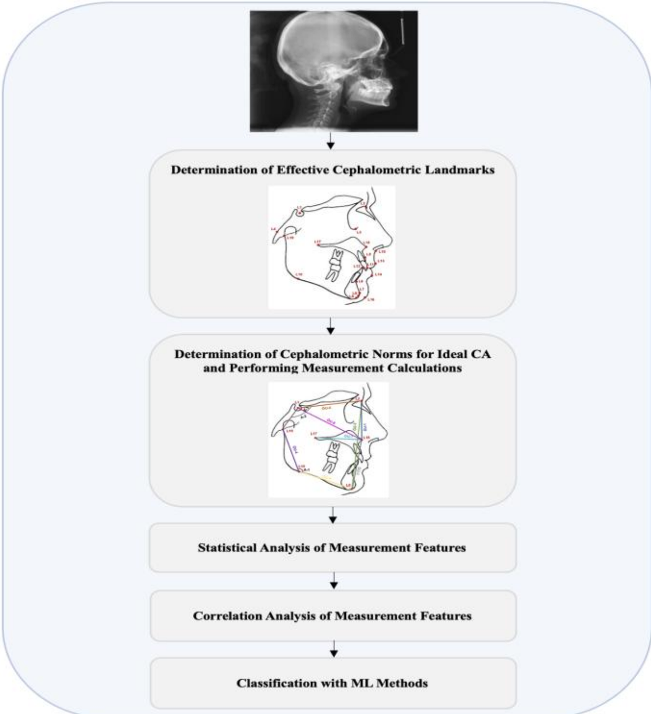
Author addresses:
[1] Software Engineering Department, Faculty of Computer and Information Sciences, Konya Technical University, Konya, 42250, Türkiye (gerol@ktun.edu.tr)

[2] Computer Engineering Department, Faculty of Computer and Information Sciences, Konya Technical University, Konya, 42250, Türkiye (buzbas@ktun.edu.tr)

[3] Orthodontics Department, Faculty of Dentistry, Selçuk University, Konya, 42130, Türkiye (hatice.kok@selcuk.edu.tr)

ABSTRACT

Context—The skull is a crucial reference structure in skeletal analysis due to its hard and durable nature, which enables it to resist environmental degradation and preserve its morphological characteristics over extended periods. Because of these properties, craniofacial structures are widely used in forensic science, anthropology, and medical studies for identification purposes, including sex determination.
Objective—This study aims to perform sex determination using cephalometric radiography (CR) and to evaluate the effectiveness of statistical and correlation-based feature selection methods in improving classification performance. Additionally, the study investigates whether a reduced set of informative cephalometric features (CFs) can achieve comparable accuracy to the full feature set, thereby increasing model efficiency and interpretability.
Method—A dataset was created by extracting the x and y coordinates of 19 cephalometric landmarks (CLs) from CR images of 300 patients (150 females and 150 males). Based on these landmarks, 8 Euclidean distance (8-ED), 3 angular (3-A), and 3 frontal sinus (3-FS) measurements were calculated, resulting in a total of 14 CFs. Initially, statistical analyses were conducted to determine the relationship between each feature and the sex variable, identifying the most discriminative parameters. Following this, correlation analysis was performed to examine inter-feature relationships and to detect redundant information. Feature combinations with high correlation and strong association with the target variable were generated using predefined threshold values. To evaluate the effectiveness of these feature sets, comparative classification experiments were carried out using machine learning (ML) algorithms, with a particular focus on performance differences between full and reduced feature subsets.



Results—The experimental results showed that using the Support Vector Machine (SVM) classifier with 4 CFs yielded an accuracy of 89.33%. Notably, a comparable level of classification performance was achieved using only 4 CFs selected based on strong statistical significance and high correlation with the sex class. This finding suggests that many of the original CFs contain redundant or less informative data, and that careful feature selection can simplify the model without compromising its predictive capability.
Conclusion—The results demonstrate that effective feature selection based on statistical and correlation analyses can significantly reduce the number of CFs required while maintaining high classification performance in sex determination from cephalometric radiographs. This reduction not only improves computational efficiency but also enhances model interpretability and applicability in real-world scenarios. The proposed approach provides a robust framework for optimizing feature sets in similar biomedical classification problems.

Key Words—Cephalometric analysis, Correlation analysis, Data mining, Machine learning, Statistical analysis.

I. INTRODUCTION

The analysis of human skeletal remains plays a crucial role in both forensic applications and archaeological investigations. A primary objective of this process is to determine biological profile parameters, particularly sex and age. Among skeletal structures, the skull is frequently available for examination due to its resistance to postmortem degradation and its relatively high preservation rate [1]. In forensic anthropology, although the pelvis provides the highest accuracy for sex estimation, the skull is widely accepted as the next most informative structure [2]. Accordingly, numerous studies have focused on cranial characteristics for this purpose. CR has become a preferred imaging modality in such analyses, as it allows simultaneous observation of multiple anatomical landmarks on a single image, facilitating detailed cranial evaluation.

CR originates from anthropometry, which focuses on the measurement of human body structures. Within this discipline, craniometry is widely applied to examine the structural features of the skull and facial region. This approach is based on the quantitative evaluation of craniofacial dimensions as well as angular relationships between anatomical points. Such measurements can be obtained either directly from individuals or indirectly using facial models, photographs, or radiographic data; however, radiographic methods are generally favored, as they provide a more comprehensive representation of both hard and soft tissues [3]. As a dental imaging modality, CR allows detailed observation of craniofacial structures from different perspectives, including anterior, posterior, and lateral views. Due to these capabilities, it has become a standard tool in orthodontic practice, particularly for diagnosis and treatment planning [4], [5]. The systematic evaluation of these radiographic images is defined as CA.

Lateral cephalometric images, which are standardized cranial radiographs used in CA, offer a wide range of anatomical landmarks that can support sex estimation. Numerous CLs and measurement strategies have been introduced in the literature over time for this purpose [6]. However, the large number of proposed landmarks and analytical methods has led to increased methodological complexity. To overcome this limitation, several studies have focused on a reduced set of 19 CLs that enable the extraction of essential clinical information while limiting the number of required measurements in CA [4], [7], [8]. The use of this predefined set of landmarks has been reported to yield more consistent and accurate results in sex determination compared to conventional morphological and metric approaches based on linear, angular, and proportional evaluations [9].

Numerous studies have explored the effectiveness of CR in sex estimation using various cephalometric variables and analytical approaches. In an early study [1], ten variables extracted from lateral CR images were analyzed using discriminant function analysis, achieving a classification accuracy of 99%. Similarly, in [10], sex estimation in a pediatric population was investigated using lateral CR images of 50 males and 50 females. Twenty-two cephalometric measurements exhibited significant sexual dimorphism, and a stepwise discriminant analysis identified seven key variables, yielding an overall accuracy of 95%. In another study [11], nine cephalometric parameters derived from lateral CR images of 100 individuals achieved classification accuracies of 88% for males and 84% for females. Furthermore, in [12], both sex and stature estimation were investigated using ten cephalometric measurements, resulting in an overall accuracy of 73.10%. Mastoid height (Ma-Ht) showed the highest discriminatory ability for sex estimation (71.80%), whereas the Glabella–Opisthocranium (G–Op) distance was the most informative parameter for stature estimation. Similarly, in [13], eight linear and two angular cephalometric measurements obtained from 111 subjects revealed that effective maxillary length, effective mandibular length, and lower anterior facial height were significantly greater in males. Likewise, in [14], eleven cephalometric variables extracted from lateral CR images of 60 young adults achieved an overall classification accuracy of 93%, with four variables identified as the most reliable predictors of sex. Moreover, in [15], analysis of twenty-four anatomical landmarks demonstrated significantly higher linear and angular measurements in males than in females. In addition, in [2], sex estimation using CR-based skull images achieved 100% classification accuracy on 80 cases, highlighting Frontal Bone Prominence (G–Sg–M), Supraorbital Ridges (G–Sg–N), and the GPI index as highly reliable indicators. Using a different methodology, in [16], nine anatomical landmarks were used to derive twenty-two linear, angular, and areal measurements from uncalibrated lateral CR images, achieving an overall accuracy of 82.40%. The angular parameters DN–S–V, DN–S–G, and DG–N–S, together with selected areal ratios, exhibited the greatest discriminatory power. Similarly, in [17], fourteen anatomical landmarks were analyzed to calculate eleven linear and one angular cephalometric parameter. Facial height (N–Me), mandibular ramus height (AR–Go), mandibular plane length (Me–Go), and frontal sinus width (FsWd) were identified as the most discriminative features, resulting in an overall accuracy of 87.60%, with higher performance in females (91.10%) than males (84.20%). Linear measurements were consistently greater in males, whereas the angular parameter showed no significant sex-related difference. More recently, in [18], 190 CBCT images from individuals aged 18–70 years were analyzed using nine anatomical landmarks. Most variables demonstrated significant sex-related differences, except for several mastoid-related measurements, including the lowest point of the mastoid triangle, mastoid, and the mastoid–mastoid connecting angle. A Random Forest (RF) classifier achieved a classification accuracy of 97%.

FSs are commonly utilized in sex estimation research due to their marked variations between males and females in terms of size, shape, and structural characteristics. In this regard, lateral CR images are evaluated by incorporating parameters such as FS area, maximum height, maximum width, and the FS index into an ANN-based model, achieving an accuracy rate of 73.3% [19]. In a similar vein, it is reported that FS index values derived from lateral cephalometric radiographs differed significantly between male and female groups [20]. Moreover, it is highlighted in [21] that integrating FS morphology with cephalometric measurements improves the robustness of sex estimation. Their results indicated that the combined use of FS area and FS index offers a reliable framework for distinguishing sex. The main contributions of the present study are outlined below:

- In the literature, sex determination studies based on CR are largely focused on dentistry and forensic medicine and predominantly rely on statistical methods, while data mining (DM)-based approaches remain limited. In this study, alternative methodologies for CA were explored, a topic of long-standing interest in dentistry. Measurement features were derived through detailed calculations using CLs identified by specialist dentists from patient CR images. These extracted features (CFs) were then used to classify sex using various DM techniques.
- Statistical analyses were first conducted to assess the significance of the extracted measurement features with respect to the sex class, and the most effective feature combination with strong discriminative power was identified. Subsequently, correlation analysis was applied between these selected strong features and the sex class. Based on the correlation results, highly correlated feature subsets were identified using predefined thresholds. Classification experiments were then performed using these feature combinations, enabling sex estimation with an optimal number of features that exhibited both statistical significance and strong correlation with the target class.Based on the DM analyses conducted, an optimal subset

of effective CFs was identified for use preliminary analytical studies in dentistry and orthodontics.

- It was demonstrated that 4 CFs derived from six CLs for CA are sufficient to achieve successful sex determination.
- An explainable artificial intelligence (AI) analysis based on SHapley Additive exPlanations (SHAP) was incorporated to improve model transparency and identify the most influential cranial measurements contributing to sex classification. The SHAP analysis demonstrated that larger values of the selected cranial measurements generally increased the likelihood of male classification, providing additional insight into the biological relevance of the model predictions.

II. METHOD

An overview of the proposed workflow is presented in the graphical abstract. The framework consists of CL extraction, feature computation, statistical and correlation analyses, feature subset generation, ML-based classification. After, SHAP used to determine the relative importance of features and to explain how each feature contributes to the predictions generated by ML models, thus enhancing the transparency and interpretability of the proposed framework.

In this study, CLs were manually identified by an expert dentist Hatice Kök with more than 20 years of clinical expertise, following standard CA procedures. All radiographs were evaluated under standardized conditions, and the coordinates of the identified landmarks were subsequently used to calculate cephalometric measurements. Since the proposed framework relies on quantitative measurements derived from landmark coordinates rather than directly processing raw radiographic images, image-based data augmentation techniques were not employed.

A. Cephalometric analysis

Cephalometry is a subdivision of anthropometry, that focuses on quantitative measurement of the human body, and it is strongly connected to craniometry, a method widely used in physical anthropology for analyzing craniofacial structures. This approach is based on the numerical evaluation of cranial and facial dimensions together with angular relationships among anatomical landmarks. These measurements may be collected directly from subjects or indirectly via facial casts, photographs, and radiographic records. Nevertheless, radiographic-based methods are generally preferred since they provide more detailed information on both hard and soft tissue structures. CR is a dental imaging modality that enables the visualization of cranial bones and adjacent soft tissues in multiple projections, such as anterior, posterior, and lateral views. Due to these advantages, it is extensively utilized in dentistry, especially in orthodontic practice, for diagnostic assessment and treatment planning [4], [5]. The interpretation of these radiographic data is defined as CA. In clinical settings, CA is applied for various purposes, including evaluation of mandibular and lower facial anatomy, support in the diagnosis and management of sleep-related breathing disorders such as sleep apnea, and assessment of facial soft tissue characteristics [22]–[25].

Lateral CR play a key role in orthodontic diagnosis and treatment planning [26]. As early as 1921, several anthropological CLs such as Turcicon, Akousticon, Gonion, Pogonion, Nasion, and Spina Nasalis Anterior were introduced, with the idea that these points could be useful in evaluating human growth, developmental patterns, and related abnormalities [27]. Since that time, many additional CLs have been proposed for lateral CR analysis in orthodontics; however, around 30 of the landmarks defined by [6] are still commonly used in practice. The choice of specific landmarks in cephalometric studies is often influenced by clinician experience and preference, which has led to variability in methodology and encouraged efforts to standardize landmark selection procedures [28]–[30]. As a result, numerous CL sets and measurement strategies have been developed over time. These studies have helped identify the most informative parameters for CA and have supported the development of reduced landmark sets that preserve diagnostic information while reducing measurement complexity [4]. At present, 19 cephalometric landmarks are generally accepted as the most frequently used set in CA applications [7], [8].

In CA studies, there are 19 commonly used CLs [7], [8], [31] and the locations of these points on a lateral CR are shown in Fig. 1. These points are used in many diagnostic and treatment studies in orthodontics.

B. Determination of cephalometric norms

In anthropological research, failure to accurately establish sex can limit the effectiveness of identification procedures such as facial reconstruction. For this reason, the detection, evaluation, and quantification of sexually dimorphic traits constitute an important part of skeletal assessment and are considered fundamental in both forensic and anthropological contexts [2]. Biological sex is generally classified into two categories, female and male. Although distinct morphological and metric differences are present between the skeletal features of these groups, reliable differentiation requires detailed and systematic analysis of the relevant anatomical characteristics.

CR play an important role in sex estimation by providing detailed morphological information of the cranial structure and enabling the assessment of multiple anatomical landmarks [32]. Previous research has shown that CR-based approaches for sex determination can achieve relatively high reliability, with reported accuracy values ranging between 80% and 100%, and have contributed to the formulation of cephalometric standards [10]. In this study, previously reported cephalometric standards used for sex differentiation in the literature were first systematically examined. These parameters were subsequently evaluated through statistical analysis in cooperation with an experienced dentist, and study-specific reference values were derived for the investigated sample. For comparison purposes, the cephalometric standards obtained from earlier studies together with their corresponding significance levels are presented in Table 1.

C. Performing measurement calculations in cephalometric analysis

In clinical CA, the first step consists of identifying CLs, and a standard set of 19 landmarks is commonly used in practice [7], [8], [31]. These landmarks serve as the basis for many orthodontic diagnostic evaluations and treatment planning procedures. In this study, x- and y-coordinates of the 19 CLs were manually annotated on CRs by an expert dentist with an expert dentist Hatice Kök with more than 20 years of clinical expertise.

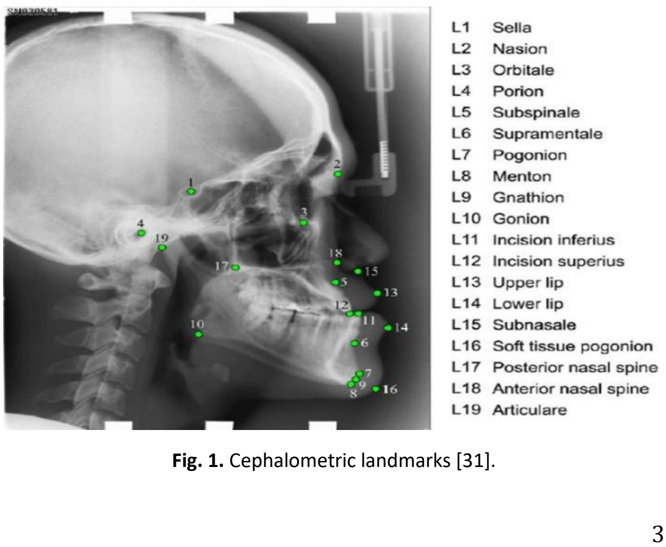


Fig. 1. Cephalometric landmarks [31].

Table 1. Cephalometric norms and statistical significance values presented in the literature for sex detection.

Ref.	Number of patients	Cephalometric norms			p-value		
		Linear	Angle	FS	Linear	Angle	FS
		G-Op	GM-BaN	FsHT	<0.001*	<0.001*	0.002
		Ba-Br	GM-SN	FsWd	<0.001*	<0.001*	<0.001*
		Ba-O	GM-FH		<0.001*	<0.001*	
[10]	100	Ma Length	Op-BaN		<0.001*	<0.001*	
	(50 Female-50 Male)	Ma-SN	Op-SN		<0.001*	0.002	
		Ma-FH	G-Sg- M		<0.001*	0.010	
		Sg-GM			<0.001*		
		N-S	S-N-A		<0.001*	0.369	
	60	Me-Go	S-N-B		0.041	0.71	
[14]	(30 Female -30 Male)	N-ANS	A-N-B	-	0.003	0.6	-
		ANS-Me	Me-Go-Ar		0.746	0.017	
		Co-Gn	SN-Mand pl		0.003	0.025	
		G-Op	N-S-V			0.001	
		V-Ba	V-S- Op				
		G-V	Op-S-Ba				
		V-Op	Ba-S-ANS				
	135	Op-Ba	ANS-S-N				
[16]	(67 Female -68 Male)	Ba-G	N-S-G	-	-	0.005	-
		N-Ba	G-Ba-Op			0.507	
		S-ANS	G-N-S			<0.001*	
		N-S					
		S-Ba					
		Ba-ANS					
		ANS-N					
		N-ANS	S-N-Ar	FsHT	<0.001*	0.187	<0.001*
		N-Me		FsWd	<0.001*		<0.001*
		ANS-Me			<0.001*		
	202	Id-Me			<0.001*		
[17]	(101 Female -101 Male)	Ar-Go			<0.001*		
		Me-Go			<0.001*		
		S-N			<0.001*		
		Cond-Gn			<0.001*		
		ANS- PNS			<0.001*		
	255			FsHT			0.01
[19]	(132 Female -123 Male)	-	-	FsWd	-	-	0.001*
				FS-Index			0.001*

As a result, the dataset was constructed by collecting the x and y coordinate values of the 19 predefined landmarks obtained from the CRs of orthodontic patients.

Following the identification of CLs reported in the literature, distances and angles between connecting lines are calculated. Based on these measurements, qualitative and quantitative evaluations of craniofacial characteristics and potential anatomical variations are performed. Through these analyses, key linear, angular, and FS-related cephalometric parameters are obtained [33]. In this context, cephalometric norms for sex determination reported in previous studies were examined in detail, and population-specific reference values were defined in collaboration with an expert dentist. All required measurement computations for the identified cephalometric parameters were performed, and the resulting CFs were then extracted for use in the study.

In CA studies for sex estimation, three FS measurements are commonly used: FS length (FS-1), FS width (FS-2), and the FS index (FS-Index) [19]–[21]. As shown in Fig. 2, four anatomical reference points are defined for these measurements: point A represents the most superior point of the frontal sinus border, point B the most inferior point, point E the most posterior point, and point F the most anterior point. FS-1 is defined as the linear distance between points A and B, while FS-2 corresponds to the distance between points E and F. The FS-Index is calculated as the ratio of AB to EF. Based on these measurements, together with other cephalometric parameters, a total of 8-ED, 3-A, and 3-FS features were extracted, which together constitute the Cephalometric Norm Set (CNS).

D. Dataset

The dataset for this study was established following approval from the Ethics Committee of Selçuk University Faculty of Dentistry, Konya, Turkey (Approval No. 2023/22). The study cohort comprised 300 individuals from the Turkish population, including 150 females and 150 males aged 17–20 years. For each participant, x and y coordinates of 19 CLs were extracted from lateral cephalometric radiographs. These landmarks were subsequently used to calculate 8-ED, 3-A, and 3-FS CFs. A specialist dentist identified the precise coordinates of the points from which these features were derived. The resulting CNS developed in this study is summarized in Table 2.

E. Data mining

DM focuses on discovering hidden patterns and relationships in data using methods such as statistical analysis, association rule mining, regression, and correlation-based techniques [34]–[36]. By identifying these relationships, DM supports dimensionality reduction and improves the efficiency of predictive models. Within this context, two main approaches, feature selection and feature extraction, are used. Feature selection involves choosing the most informative variables while removing irrelevant or redundant ones [37]. Feature extraction, on the other hand,

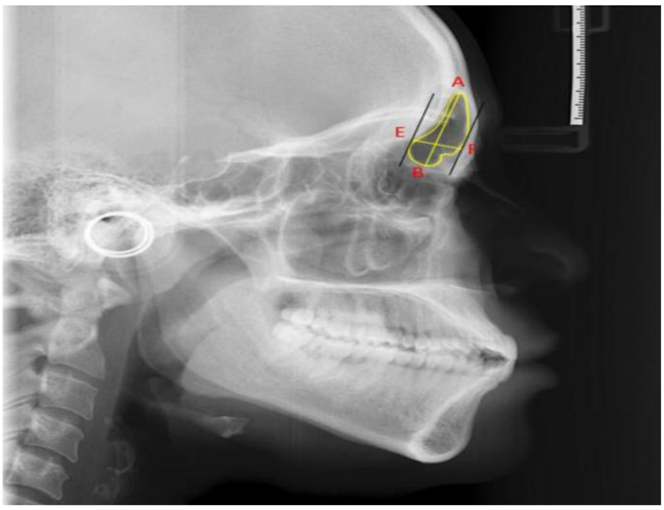


Fig. 2. 3-FS measurement on CR [19].

Table 2. CNS dataset.

CFs	Explanation	Related CLs	Relevant Coordinate Values of CLs
ED-1	Euclidean distance between 2 and 18 CL	Nasion-ANS	2-x, 2-y and 18-x, 18-y
ED-2	Euclidean distance between 2 and 8 CL	Nasion-Menton	2-x, 2-y and 8-x, 8-y
ED-3	Euclidean distance between 8 and 18 CL	ANS-Menton	8-x, 8-y and 18-x, 18-y
ED-4	Euclidean distance between 10 and 19 CL	Articular-Gonion	10-x, 10-y and 19-x, 19-y
ED-5	Euclidean distance between 8 and 10 CL	Menton-Gonion	8-x, 8-y and 10-x, 10-y
ED-6	Euclidean distance between 1 and 2 CL	Cella-Nasion	1-x, 1-y and 2-x, 2-y
ED-7	Euclidean distance between 17 and 18 CL	ANS-PNS	17-x, 17-y and 18-x, 18-y
ED-8	Euclidean distance between 1 and 18 CL	Cella-ANS	1-x, 1-y and 18-x, 18-y
A-1	Angle between 8, 10 and 18 CL	Menton-Gonion- Articular	8-x, 8-y, 10-x, 10-y and 19-x, 19-y
A-2	Angle between 2, 1 and 19 CL	Cella-Nasion- Articular	2-x, 2-y, 1-x, 1-y and 19-x, 19-y
A-3	Angle between 18, 1 and 2 CL	ANS-Cella-Articular	18-x, 18-y, 1-x, 1-y and 2-x, 2-y
FS-1	Frontal Sinus Length	FS-A and FS-B	A-x, A-y and B-x, B-y
FS-2	Frontal Sinus Width	FS-E and FS-F	E-x, E-y and F-x, F-y
FS-Index	Frontal Sinus Index	FS-1 / FS-2	A-x, A-y, B-x, B-y, E-x, E-y and F-x, F-y

transforms original variables into a smaller set of new features that preserve the essential information in the data [38].

1. Statistical analysis

A key step in statistical analysis is selecting an appropriate test for the dataset, primarily determined by whether the data follow a normal distribution. The normal (Gaussian) distribution is symmetric and bell-shaped, with most observations clustered around the mean and fewer extreme values. Since many statistical methods assume normality, assessing this assumption is essential. Accordingly, normality tests are applied, and based on the results, parametric methods are used when normality is met, while non-parametric methods are preferred otherwise.

2. Correlation analysis

Correlation analysis is a statistical method used to evaluate the relationship between two or more variables in terms of direction and strength [39]. When a relationship exists between variables, changes in one variable are associated with changes in another. This relationship is measured using a correlation coefficient, which indicates both the strength and direction of the association. Correlation can be positive, negative, or zero. A positive correlation means both variables increase together, while a negative correlation indicates one increases as the other decreases. A zero correlation suggests no linear relationship. The correlation coefficient, r, ranges from −1 to 1, where 1 represents a perfect positive correlation, −1 a perfect negative correlation, and 0 no linear association. The value of the coefficient may also be influenced by the data's distribution properties.

F. Classification algorithms

ML is a subfield of AI that enables systems to learn from data and improve performance over time without explicit programming after training. Its main objective is to identify patterns in data using knowledge gained during training, enabling automated prediction and decision-making [40]. In this study, several ML classifiers were used, including Naive Bayes (NB), Logistic Regression (LR), SVM, K-Nearest Neighbors (KNN), Bagging, AdaBoost, RF, and Artificial Neural Networks (ANN).

1. Naive Bayes

NB is a probabilistic classification method based on Bayes' theorem, which assumes conditional independence among features given the class label [41]. Bayes' theorem provides a way to calculate probabilities by combining prior knowledge with observed evidence. In the NB framework, class membership probabilities are computed using this theorem. The model then compares the resulting posterior probabilities for all classes and assigns the sample to the class with the highest probability.

2. Logistic Regression

LR is commonly used in binary classification problems in which the dependent variable is categorical. The method models the relationship between a binary outcome and a set of independent predictor variables [42]. These predictors are then used to estimate the probability of each outcome, where the target variable is typically encoded as 1 for the presence of a class and 0 for its absence [43].

3. Support Vector Machine

SVM, originally introduced in [44] within the framework of statistical learning theory, is widely recognized for its effectiveness in handling complex and nonlinear datasets. It constructs an optimal hyperplane by maximizing the margin between classes, ensuring robust data separation [44], [45]. For linearly separable data, this hyperplane is determined directly, whereas for nonlinearly separable data, SVM uses kernel functions to implicitly map the input space into a higher-dimensional feature space, enabling a linear decision boundary [45].

4. K-Nearest Neighbors

KNN is a distance-based classification method in which the class label of a new observation is determined according to the labels of its k nearest neighbors. Here, k represents the number of closest samples considered, and its appropriate selection is a critical factor affecting model performance [46]. In this approach, similarity between observations is typically measured using distance metrics, with Euclidean and Manhattan distances being the most commonly used criteria [47], [48].

5. Bagging

Bagging, introduced by [49], is an ensemble learning technique that generates multiple bootstrap subsets from the original dataset. A selected base learner is then trained independently on each subset, and the resulting predictions are combined to obtain the final output. In regression problems, this combination is typically performed by averaging the individual predictions, whereas in classification tasks, majority voting is commonly used [50]. By aggregating multiple models in this way, bagging reduces variance and improves the stability and robustness of the overall predictive performance.

6. AdaBoost

AdaBoost, introduced by Yoav Freund and Robert Schapire, is an ensemble learning method that iteratively focuses on misclassified samples during training [51]. It initially assigns equal weights to all training instances [52], then increases the weights of misclassified samples while decreasing those of correctly classified ones. During prediction, the weighted outputs of multiple base learners are aggregated, improving predictive accuracy and robustness by reducing errors and emphasizing difficult instances [53].

7. Random Forest

RF algorithm, proposed by [54], is an ensemble learning method composed of multiple decision trees. Instead of relying on a single model, RF builds a large number of trees using randomly selected subsets of the training data and combines their outputs to generate the final prediction. The method can be applied to both classification and regression problems. In classification tasks, the final class label is determined through majority voting across all trees in the ensemble [55]. RF has two key hyperparameters: the number of features considered at each split and the total number of trees in the forest [56], both of which influence model performance and computational efficiency. In general, increasing the number of trees improves the stability of the model's generalization error and helps reduce the risk of overfitting.

8. Artificial Neural Network

Artificial Neural Networks (ANNs) are computational models inspired by the structure and function of the human brain. They are widely applied in areas such as image and speech recognition, signal processing, natural language processing, decision support systems, and biomedical applications [57]. ANNs consist of interconnected neurons that receive inputs, apply weights and a bias, and generate outputs through an activation function [58]. The Multilayer Perceptron (MLP) is a commonly used ANN architecture with input, hidden, and output layers. During training, the network compares predicted and actual outputs to compute an error, and iteratively adjusts weights to minimize this error until convergence is achieved [59].

G. Performance evaluation metrics

The classification success of the ML algorithms was evaluated by calculating the accuracy, sensitivity, precision, and F-Score metrics, as expressed in (1)-(4), respectively, [60]:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}, \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}}, \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}, \tag{3}$$

$$\text{F}-\text{Score} = 2\left(\frac{\text{Precision}\times\text{Sensitivity}}{\text{Precision}+\text{Sensitivity}}\right). \tag{4}$$

AUC used as a summary of the ROC curve was also used for performance evaluation.

III. RESULTS and DISCUSSION

In this study, effective cephalometric norms for sex determination using CR were first identified in collaboration with an expert dentist. Based on 19 CLs from 300 patients, measurements were computed and used to construct a dataset named CNS. Statistical analyses were then performed to evaluate the selected norms and their consistency with the data, identifying features strongly associated with sex. ML classifiers were applied to assess the predictive performance of the CNS dataset. Correlation analysis was subsequently conducted, and feature combinations with high correlation to the sex class were derived using predefined thresholds. Finally, classification experiments were repeated using these optimized feature sets, and sex determination was performed using the most informative CFs that were both strongly associated and highly correlated with the target class.

A. Performing measurement calculations and creating CNS

In clinical CA, 19 commonly used CLs are first identified [7], [8], [31]. In this study, an expert dentist extracted the x- and y-coordinates of these CLs and 3-FS measurements from the CRs of 300 orthodontic patients. Population-specific cephalometric norms for sex determination were then established based on the literature and expert consultation. Using these norms, 8-ED, 3-A, and 3-FS features were calculated and compiled to construct the CNS shown in Table 2.

B. Statistical analysis of the CNS

For each participant, the x- and y-coordinates of the 19 CLs were used to compute the 8-ED, 3-A, and 3-FS measurements constituting the CNS. The relevance of these features for sex classification was evaluated through statistical analysis. As the sample size exceeded 50, normality was assessed using the Kolmogorov–Smirnov test [61], [62], which indicated that none of the measurements were normally distributed. Therefore, group differences were analyzed using the non-parametric Mann–Whitney U test [63]. Statistical significance was set at $p < 0.05$, with $p < 0.001$ indicating a strong association [64]. The results are summarized in Table 3.

C. Classification with CNS and strongly associated features

To evaluate the predictive performance of the models, eight classifiers—NB, LR, SVM, KNN, Bagging, AdaBoost, RF, and ANN—were employed without performing hyperparameter optimization, and the default hyperparameter values provided by the implementation library were retained. Model validation was conducted using a 10-fold cross-validation approach, ensuring that each subset of the data was utilized for both training and testing. The hyperparameters for all classifiers are presented in Table 4.

Table 3. Significance levels of 8-ED, 3-A, and 3-FS measurements.

CFs	Explanation	Relevant Coordinate Values of CLs	Significance
ED-1	Euclidean distance between 2 and 18 CL	2-x, 2-y and 18-x, 18-y	Strong
ED-2	Euclidean distance between 2 and 8 CL	2-x, 2-y and 8-x, 8-y	Strong
ED-3	Euclidean distance between 8 and 18 CL	8-x, 8-y and 18-x, 18-y	Strong
ED-4	Euclidean distance between 10 and 19 CL	10-x, 10-y and 19-x, 19-y	Strong
ED-5	Euclidean distance between 8 and 10 CL	8-x, 8-y and 10-x, 10-y	Strong
ED-6	Euclidean distance between 1 and 2 CL	1-x, 1-y and 2-x, 2-y	Strong
ED-7	Euclidean distance between 17 and 18 CL	17-x, 17-y and 18-x, 18-y	Strong
ED-8	Euclidean distance between 1 and 18 CL	1-x, 1-y and 18-x, 18-y	Strong
A-1	Angle between 8, 10 and 18 CL	8-x, 8-y, 10-x, 10-y and 19-x, 19-y	Insignificant
A-2	Angle between 2, 1 and 19 CL	2-x, 2-y, 1-x, 1-y and 19-x, 19-y	Strong
A-3	Angle between 18, 1 and 2 CL	18-x, 18-y, 1-x, 1-y and 2-x, 2-y	Strong
FS-1	Frontal Sinus Length	A-x, A-y and B-x, B-y	Strong
FS-2	Frontal Sinus Width	E-x, E-y and F-x, F-y	Strong
FS-Index	Frontal Sinus Index	A-x, A-y, B-x, B-y, E-x, E-y and F-x, F-y	Strong

Table 4. Hyperparameters for all classifiers.

Classifier	Hyperparameters
NB	var_smoothing=1e−9
LR	C = 1.0, penalty = L2, solver = lbfgs, max_iter = 5000
SVM	Kernel = RBF, C = 1.0, gamma = scale, probability = True
KNN	k = 5, metric = Minkowski (p = 2)
Bagging	n_estimators = 10
AdaBoost	n_estimators = 50, learning_rate = 1.0
RF	n_estimators = 100, criterion = gini, max_features = sqrt
ANN	hidden_layer_sizes = (100), activation = relu, solver = adam, max_iter = 1000

The resulting mean accuracies obtained with the CNS dataset, which includes all 14 original features, are reported in Table 5. As shown in Table 5, the CNS features, consisting of 8-ED, 3-A, and 3-FS measurements, were statistically analyzed to evaluate their relevance for sex classification. The measurements were classified as significant ($p < 0.05$), highly significant ($p < 0.001$), or non-significant based on the obtained p-values. According to the results, the relationship between each feature and sex determination was assessed. Among the evaluated variables, A-1 did not show a meaningful association with sex, while the remaining thirteen features—ED-1 to ED-8, A-2, A-3, FS-1, FS-2, and FS-Index—exhibited strong statistical associations. The classification results for these statistically significant features are presented in Table 5. As shown in Table 5, excluding the non-significant feature A-1 from the dataset improved performance across several classifiers using the remaining 13 strongly associated features. The highest classification accuracy obtained with the CNS dataset (Table 5) was 87.33%, achieved by LR. Similarly, LR also yielded the best performance with the strongly associated features, achieving an accuracy of 87.33%. Moreover, a slight decrease in performance was observed for some models. These findings suggest that the removed feature, although not statistically significant on its own, may still contribute complementary information when combined with other features. Therefore, further correlation analysis was conducted to investigate the relationships in more detail.

D. Correlation analysis and classification according to certain correlation thresholds

A correlation analysis was performed to investigate the relationships between the features previously identified as strongly associated with sex in the statistical analyses and the sex categories. The resulting correlation coefficients were visualized using a heat map, which illustrates both the strength and direction of the associations between each feature and sex classification, as shown in Fig. 3. According to this figure, the A-1 feature, previously identified as statistically non-significant, also exhibited the weakest correlation with the sex class, further supporting the statistical findings. Based on the calculated correlation values, feature subsets were generated using predefined correlation thresholds, as summarized in Table 6. This table presents the feature combinations obtained using the defined correlation threshold values of 0.3, 0.4, and 0.5. Based on these thresholds, 11 features exhibited correlation values of 0.3 and above with the sex class,while 7 features met the threshold of 0.4 and above, and 4 features satisfied the threshold of 0.5 and above. Separate classification experiments were conducted for each of these three feature sets, and the corresponding results are reported in Table 7.

The comparison of results in Table 7 shows that using ED-2, ED-4, ED-6, and ED-8 at a 0.5 correlation threshold improves the performance of several classifiers. Compared with the 14-feature CNS in Table 5, this four-feature subset achieved overall performance gains, with SVM reaching the highest accuracy of 89.33%. These results demonstrate that statistical and correlation-based feature selection provides a more compact CNS representation while maintaining, or even improving, predictive performance.SHAP analysis was conducted to investigate the contribution of individual features to the SVM classifier. As shown in Fig. 4a, ED-8 was identified as the most influential feature, followed by ED-4, ED-2, and ED-6. As shown in Fig. 4b, higher values of ED-8 and ED-4 were generally associated with positive SHAP values, indicating that these features increased the likelihood of the positive class (male) prediction. In contrast, lower feature values tended to contribute negatively to the model output. The broader SHAP value distributions observed for ED-8

Table 5. Classification results obtained with CNS and strongly associated features.

Dataset	Classifier	Accuracy (%)	Precision (%)	Sensitivity (%)	F-score (%)	AUC (%)
CNS	NB	78.67	71.56	97.33	82.24	92.98
	LR	**87.33**	88.82	86.00	86.97	94.44
	SVM	85.67	86.88	84.67	85.12	95.02
	KNN	86.67	86.51	87.33	86.72	92.96
	Bagging	83.33	86.04	81.33	83.08	91.07
	AdaBoost	81.67	81.42	83.33	82.07	90.36
	RF	84.67	84.97	86.00	84.83	94.49
	ANN	86.67	87.11	86.67	86.75	93.42
Strongly associated features	NB	**80.33**	73.20	98.00	83.57	93.51
	LR	**87.33**	88.24	86.67	87.02	93.73
	SVM	**85.67**	86.94	84.67	85.21	94.80
	KNN	86.00	86.41	86.00	85.98	92.27
	Bagging	83.00	84.44	82.00	82.76	90.67
	AdaBoost	**82.33**	80.42	86.00	82.88	89.73
	RF	**85.00**	85.56	85.33	84.89	94.29
	ANN	86.33	87.19	85.33	86.09	91.82

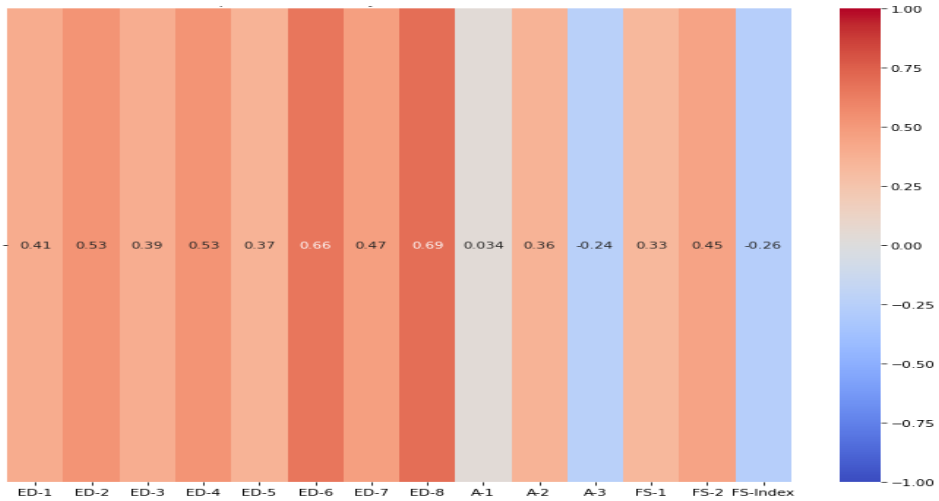


Fig. 3. Correlation heat map.

Table 6. Correlation threshold values and feature combinations.

Threshold	Feature Combinations
0.3	ED-1, ED-2, ED-3, ED-4, ED-5, ED-6, ED-7, ED-8, A-2, FS-1, FS-2
0.4	ED-1, ED-2, ED-4, ED-6, ED-7, ED-8, FS-2
0.5	ED-2, ED-4, ED-6, ED-8

Table 7. Classification results with threshold values of {0.3, 0.4, 0.5} > correlated features.

Threshold	Classifier	Accuracy (%)	Precision (%)	Sensitivity (%)	F-score (%)	AUC (%)
0.3	NB	84.33	79.28	94.00	85.79	93.38
	LR	86.33	86.52	86.67	86.19	93.60
	SVM	85.33	86.79	84.00	84.73	94.44
	KNN	84.00	83.18	86.00	84.29	92.60
	Bagging	80.67	82.54	79.33	80.35	90.24
	AdaBoost	85.00	85.14	86.00	85.10	89.96
	RF	85.67	86.24	86.00	85.68	93.93
	ANN	86.33	87.29	85.33	86.11	92.76
0.4	NB	86.33	86.67	86.67	86.16	94.27
	LR	86.67	87.84	86.00	86.40	93.42
	SVM	87.00	88.86	85.33	86.57	95.07
	KNN	86.67	86.62	87.33	86.72	93.49
	Bagging	81.67	82.62	82.00	81.61	90.11
	AdaBoost	84.00	84.83	84.00	83.93	91.60
	RF	84.67	84.85	86.00	84.85	93.96
	ANN	86.00	87.88	85.33	85.75	91.33
0.5	NB	**87.00**	87.30	87.33	86.66	94.80
	LR	86.67	88.14	86.00	86.50	93.78
	SVM	**89.33**	91.38	87.33	88.94	95.07
	KNN	84.33	84.89	84.00	84.08	92.04
	Bagging	**85.00**	87.39	82.67	84.47	92.60
	AdaBoost	**84.33**	83.56	87.33	84.95	90.84
	RF	84.33	85.80	83.33	84.12	93.78
	ANN	86.00	87.37	85.33	85.75	92.98

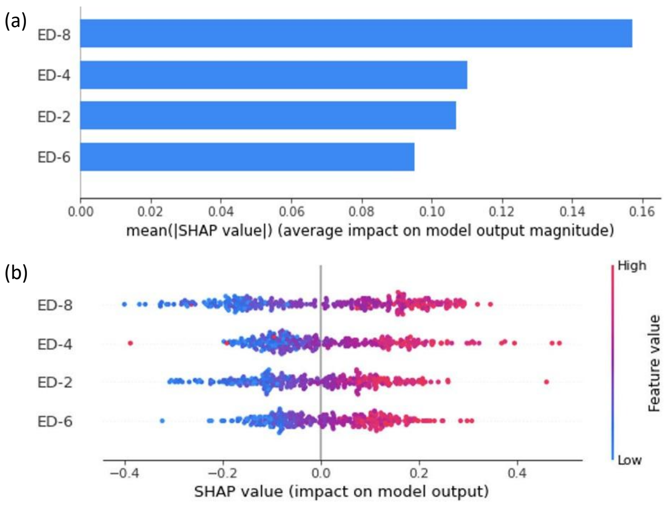


Fig. 4. Plots of the four-feature CNS model using the SVM classifier: (a) SHAP summary plot, (b) SHAP bar plot.

and ED-4 further suggest that these features exert a stronger influence on the classifier's decisions than the remaining features.

To provide a more meaningful evaluation of the proposed framework, a comparative analysis with previous studies is presented in Table 8. Unlike image-based approaches that perform sex classification directly from radiographic images using deep learning models, the studies included in this comparison employ CL information and measurements derived from landmark coordinates. Since the proposed method is also based on quantitative CFs extracted from CLs, comparing it with studies that utilize similar coordinate-derived measurements provides a more appropriate assessment of its effectiveness. Table 8 presents a comparison between the proposed framework and previous studies that performed sex determination using CFs derived from landmark coordinates. When Table 8 is examined, the reviewed studies primarily employed DFA and LR using between 7 and 21 CFs extracted from lateral cephalograms. Reported classification accuracies ranged from 74.70% to 95.60%, depending on population characteristics, sample size, and the cephalometric measurements selected. Although some studies achieved higher accuracies, these investigations were generally conducted on relatively small datasets consisting of approximately 100–202 individuals. In contrast, the proposed study was evaluated on a larger and balanced dataset comprising 300 Turkish individuals (150 females and 150 males). More importantly, the proposed framework achieved an accuracy of 89.33% using only 4 CFs selected based on statistical significance and correlation analyses. This result demonstrates that a substantial reduction in feature dimensionality can be achieved without a considerable loss in classification performance. Therefore, the proposed approach offers a favorable balance

Table 8. Comparison of the proposed method with previous studies on sex determination using CL-derived measurements.

Ref.	Method	Dataset	Result (Accuracy, %)
[10]	7 CFs and DFA	100 Taiwanese children (50 female and 50 male)	95.00
[11]	8 CFs and DFA	100 Indian individuals (50 female and 50 male)	86.00
[16]	21 CFs and LR	135 Caucasian individuals (67 female and 68 male)	82.40
[17]	12 CFs and DFA	202 Iranian individuals (101 female and 101 male)	87.60
[65]	15 CFs and Discriminant Function Analysis (DFA)	100 Indian individuals (50 female and 50 male)	84.00
[66]	18 CFs and DFA	114 European individuals (59 female and 55 male)	95.60
[67]	12 CFs and LR	150 Greek individuals (75 female and 75 male)	74.70
This Study	4 CFs and SVM	300 Turkish individuals (150 female and 150 male)	89.33

between predictive accuracy, model simplicity, and interpretability, highlighting the effectiveness of selecting a small set of highly informative CFs for sex determination.

IV. CONCLUSION

Lateral CRs, as standardized cranial imaging modalities, enable the detection of multiple anatomical landmarks informative for sex classification. In this study, measurement features for CA were initially defined, and subsequent statistical and correlation analyses on the CNS dataset reduced the feature set from 14 to 4. Using these four selected features, effective classification performance was achieved even with a relatively limited sample of 300 patients. This result supports the suitability of the measurement features determined through literature review and expert dentist evaluation for sex estimation. Moreover, the findings indicate that these four features are sufficient to preserve classification accuracy, suggesting that instead of extracting all 19 CLs for CA, it is adequate to identify only six CLs corresponding to the selected features. Accordingly, the results demonstrate that semi-automatic systems can achieve comparable performance with lower complexity and reduced computational effort.

AUTHOR STATEMENT

Plagiarism Check—The article was scanned with iThenticate and found to be compliant with the journal's plagiarism policy.

Conflict of Interest—This article does not have any conflict of interest with any person or organization.

Ethics Committee Approval—The study was approved by the Ethics Committee of Selcuk University Faculty of Dentistry (protocol code 2023/22, approval date: 4 May 2023).

Use of Artificial Intelligence Tools—This study utilized an AI tool (ChatGPT, OpenAI) for language editing and grammar checking.

Funding—No institutional/financial support was received for this study.

Data availability—The datasets generated and analyzed during the current study are not publicly available due to ethical restrictions and patient and institutional privacy policies.

CRediT Author Contribution—Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision **(Gizemnur Erol Doğan)**; Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision **(Betül Uzbaş)**; Conceptualization, Data acquisition, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision **(Hatice Kök)**.

REFERENCES

[1] K. R. Patil, R. N. Mody, "Determination of sex by discriminant function analysis and stature by regression analysis: a lateral cephalometric study", *Forensic Science International*, 147(2–3), 175–180, 2005. https://doi.org/10.1016/j.forsciint.2004.09.071.

[2] A. Mathur, A. Sande, M. Risbud, P. Ramdurg, S. R. Ashwinirani, "Determination of sex by discriminant function analysis of lateral radiographic cephalometry using angular, linear and proportional cephalometric variables in Western Maharashtrian population", *Journal of Oral Medicine, Oral Surgery, Oral Pathology and Oral Radiology*, 3(3), 153–157, 2017, https://joooo.org/archive/volume/3/issue/3/article/16223/pdf.

[3] M. Ülgen, *Ortodonti: anomaliler, sefalometri, etyoloji, büyüme ve gelişim, tanı*, Ankara Üniversitesi Diş Hekimliği Fakültesi Yayınları, Ankara, Türkiye, 2000.

[4] A. E. Athanasiou, *Orthodontic Cephalometry*, Mosby-Wolfe, London, 1995.

[5] B. Phulari, An atlas on cephalometric landmarks, JB Medical Publishers Ltd, 2013.

[6] T. Rakosi, An atlas and manual of cephalometric radiography, Wolfe Medical Publications, 1982.

[7] T. C. Niño-Sandoval, S. V. G. Pérez, F. A. González, R. A. Jaque, C. Infante-Contreras, "Use of automated learning techniques for predicting mandibular morphology in skeletal class I, II and III", *Forensic Science International*, 281, 187.e1, 2017. https://doi.org/10.1016/j.forsciint.2017.10.004.

[8] A. K. Subramanian, Y. Chen, A. Almalki, G. Sivamurthy, D. Kafle, "Cephalometric analysis in orthodontics using artificial intelligence—A comprehensive review", *BioMed Research International*, 2022, 1880113, 2022, https://doi.org/10.1155/2022/1880113.

[9] V. G. Naikmasur, R. Shrivastava, S. Mutalik, "Determination of sex in South Indians and immigrant Tibetans from cephalometric analysis and discriminant functions", *Forensic Science International*, 197(1-3), 2010. https://doi.org/10.1016/j.forsciint.2009.12.052.

[10] T. Hsiao, S. Tsai, S. Chou, J. Pan, Y. Tseng, H. Chang, H. Chen, "Sex determination using discriminant function analysis in children and adolescents: a lateral cephalometric study," *International Journal of Legal Medicine*, 124(2), 155–160, 2010. https://doi.org/10.1007/s00414-009-0412-1.

[11] A. Binnal, B. K. Devi, "Identification of sex using lateral cephalogram: role of cephalofacial parameters", *Journal of Indian Academy of Oral Medicine and Radiology*, 24(4), 280–283, 2012.

[12] M. IP, P. M. David, "Determination of sexual dimorphism and stature using lateral cephalogram", *Journal of Indian Academy of Oral Medicine and Radiology*, 25(2), 1–5, 2013.

[13] A. Yağcı, S. K. Büyük, "Dengeli yüz oranlarına ve normal oklüzyona sahip genç Türk erişkinlerin McNamara sefalometrik normları", *Journal of Health Sciences*, 22(1), 1–6, 2013. https://izlik.org/JA25PP46KG.

[14] R. U. Mathur, A. M. Mahajan, R. C. Dandekar, R. B. Patil, "Determination of sex using discriminant function analysis in young adults of Nashik: a lateral cephalometric study", *Journal of Advanced Medical and Dental Sciences Research*, 2(1), 21–25, 2014. https://jamdsr.com/pdf1/DeterminationofSexusingDiscriminantFunction.pdf

[15] D. D. Divakar, J. John, A. A. Al Kheraif, S. Mavinapalla, R. Ramakrishnaiah, S. Vellappally, M. I. Hashem, M.H. N. Dalati, B. H. Durgesh, R. A. Safadi, S. Anil, "Sex determination using discriminant function analysis in Indigenous (Kurubas) children and adolescents of Coorg, Karnataka, India: a lateral cephalometric study", *Saudi Journal of Biological Sciences*, 23(6), 782–788, 2016, https://doi.org/10.1016/j.sjbs.2016.05.008.

[16] R. Qaq, S. Mânica, G. Revie, "Sex estimation using lateral cephalograms: a statistical analysis", *Forensic Science International: Reports*, 1, 100034, 2019. https://doi.org/10.1016/j.fsir.2019.100034.

[17] A. Bagherpour, N. Anbiaee, S. Motaghi, A. Jahanbin, "Gender determination using digital lateral cephalograms: a discriminant function analysis", *Journal of Dental Materials and Techniques*, 9(4), 221–230, 2020. https://doi.org/10.22038/jdmt.2021.48358.1371.

[18] M. Farhadian, F. Salemi, A. Shokri, A. Safi, S. Rahimpanah, "Comparison of data mining algorithms for sex determination based on mastoid process measurements using cone-beam computed tomography", *Imaging Science in Dentistry*, 50(4), 323–330, 2020. https://doi.org/10.5624/isd.2020.50.4.323.

[19] C. P. Alves, C. Costa, E. Michel-Crosato, M. G. H. Biazevic, "Use of the frontal sinus to evaluate sexual dimorphism in a Brazilian sample", *Forensic Imaging*, 33, 200548, 2023. https://doi.org/10.1016/j.fri.2023.200548.

[20] S. Kenawy, M. Ellabban, A. Fadel, "A novel frontal sinus index protocol for gender determination: a retrospective lateral cephalometric study for Egyptians", *Egyptian Dental Journal*, 67(3), 2199–2204, 2021. https://doi.org/10.21608/edj.2021.76583.1639.

[21] B. Zheng, Y. Zhong, N. A. Al-Worafi, Y. Liu, "The dimensional and morphological assessment of the frontal sinus in sex estimation among different populations", *Head & Face Medicine*, 19(1), 8, 2023, https://doi.org/10.1186/s13005-023-00355-4.

[22] J. McDonnell, R. McNeill, R. West, "Advancement genioplasty: a retrospective cephalometric analysis of osseous and soft tissue changes", *Journal of Oral Surgery*, 35(8), 640–647, 1977. https://europepmc.org/article/med/267188.

[23] W. Hochban, U. Brandenburg, "Morphology of the viscerocranium in obstructive sleep apnoea syndrome cephalometric evaluation of 400 patients", *Journal of Cranio-Maxillofacial Surgery*, 22(4), 205–213, 1994. https://doi.org/10.1016/S1010-5182(05)80559-1.

[24] R. S. Nanda, R. M. Merill, "Cephalometric assessment of sagittal relationship between maxilla and mandible", *American Journal of Orthodontics and Dentofacial Orthopedics*, 105(4), 328–344, 1994. https://doi.org/10.1016/S0889-5406(94)70127-X.

[25] M. Y. Hajeer, A. F. Ayoub, D. T. Millett, "Three-dimensional assessment of facial soft-tissue asymmetry before and after

orthognathic surgery", *British Journal of Oral and Maxillofacial Surgery*, 42(5), 396–404, 2004. https://doi.org/10.1016/j.bjoms.2004.05.006.

[26] B. H. Broadbent, "A new X-ray technique and its application to orthodontia", *Angle Orthodontist*, 51(2), 93–114, 1981.

[27] A. J. Pacini, "Roentgen ray anthropometry of the skull", *Journal of Radiology*, 3, 230–238, 1922.

[28] A. Kamoen, L. Dermaut, R. Verbeeck, "The clinical significance of error measurement in the interpretation of treatment results", *European Journal of Orthodontics*, 23(5), 569–578, 2001. https://doi.org/10.1093/ejo/23.5.569.

[29] H. L. D. Da Silveira, H. E. D. Silveira, "Reproducibility of cephalometric measurements made by three radiology clinics", *Angle Orthodontist*, 76(3), 394–399, 2006.

[30] A. P. R. Durao, A. Morosolli, P. Pittayapat, N. Bolstad, A. P. Ferreira, R. Jacobs, "Cephalometric landmark variability among orthodontists and dentomaxillofacial radiologists: a comparative study", *Imaging Science in Dentistry*, 45(4), 213–220, 2015. https://doi.org/10.5624/isd.2015.45.4.213.

[31] C. Lindner, C. W. Wang, C. T. Huang, C. H. Li, S. W. Chang, T. F. Cootes, "Fully automatic system for accurate localization and analysis of cephalometric landmarks in lateral cephalograms", *Scientific Reports*, 6, 33581, 2016. https://doi.org/10.1038/srep33581.

[32] R. H. Biggerstaff, "Craniofacial characteristics as determinants of age, sex, and race in forensic dentistry", *Dental Clinics of North America*, 21(1), 85–97, 1977. https://doi.org/10.1016/S0011-8532(22)00892-8.

[33] S. Ö. Arık, B. Ibragimov, L. Xing, "Fully automated quantitative cephalometry using convolutional neural networks", *Journal of Medical Imaging*, 4(1), 014501, 2017. https://doi.org/10.1117/1.JMI.4.1.014501.

[34] R. Agrawal, T. Imielinski, A. Swami, "Mining association rules between sets of items in large databases", *ACM SIGMOD Record*, 22(2), 207–216, 1993. https://doi.org/10.1145/170036.170072.

[35] D. A. Freedman, *Statistical models: theory and practice*, New York, NY, USA, Cambridge University Press, 2009.

[36] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Second Ed., New York, NY, USA, Springer, 2009. https://doi.org/10.1007/978-0-387-84858-7.

[37] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection", *The Journal of Machine Learning Research*, 3, 1157–1182, 2003.

[38] I. T. Jolliffe, *Principal component analysis*, Second Ed., New York, NY, USA, Springer-Verlag, 2002. https://doi.org/10.1007/b98835.

[39] J. Han, M. Kamber, Data mining: concepts and techniques, 2nd ed., San Francisco, CA, USA, Morgan Kaufmann Publishers, Elsevier 2006.

[40] A. Vellido, J. Martín-Guerrero, P. J. G. Lisboa, "Making machine learning models interpretable", *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, 25-27 April 2012, 163-172.

[41] E. Frank, L. Trigg, G. Holmes, I. H. Witten, "Technical note: naive Bayes for regression", *Machine Learning*, 41, 5–25, 2000.

[42] A. Verma, "Evaluation of classification algorithms with solutions to class imbalance problem on bank marketing dataset using WEKA", *International Research Journal of Engineering and Technology*, 6(3), 54–61, 2019.

[43] S. Bharati, M. A. Rahman, P. Podder, "Breast cancer prediction applying different classification algorithm with comparative analysis using WEKA", *Proceedings of the 4th International Conference on Electrical Engineering and Information and Communication Technology (iCEEiCT)*, Dhaka, Bangladesh, 13-15 September 2018, 580-583. https://doi.org/10.1109/CEEICT.2018.8628084.

[44] V. N. Vapnik, "An overview of statistical learning theory", *IEEE Transactions on Neural Networks*, 10(5), 988–999, 1999. https://doi.org/10.1109/72.788640.

[45] Y. Huang, S. Meng, X. Li, W. Fan, "A classification method for wood vibration signals of Chinese musical instruments based on GMM and SVM", *Traitement du Signal*, 35(2), 137-151, 2018. https://doi.org/10.3166/TS.35.137-151.

[46] T. Cover; P. Hart, "Nearest neighbor pattern classification", *IEEE Transactions on Information Theory*, 13(1), 21–27, 1967. https://doi.org/10.1109/TIT.1967.1053964.

[47] Y. Z. Ayık, A. Özdemir, U. Yavuz, "Lise türü ve lise mezuniyet başarısının, kazanılan fakülte ile ilişkisinin veri madenciliği tekniği ile analizi", Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 10(2), 441–454, 2010. https://izlik.org/JA32RJ25TZ.

[48] J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*, 4th ed., Morgan Kaufmann, 2022.

[49] L. Breiman, "Bagging predictors", *Machine Learning*, 24(2), 123–140, 1996. https://doi.org/10.1007/bf00058655.

[50] S. A. A. Shah, W. Aziz, M. Arif, M. S. A. Nadeem, "Decision trees based classification of cardiotocograms using bagging approach", *Proceedings of the 13th International Conference on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, 14-16 December 2015. https://doi.org/10.1109/FIT.2015.14.

[51] Y. Freund, R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences*, 55(1), 119–139, 1997. https://doi.org/10.1006/jcss.1997.1504.

[52] O. Sagi, L. Rokach, "Ensemble learning: a survey", *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, 8(4), e1249, 2018. https://doi.org/10.1002/widm.1249.

[53] V. A. Dev, M. R. Eden, "Formation lithology classification using scalable gradient boosted decision trees", *Computers & Chemical Engineering*, 128, 392–404, 2019. https://doi.org/10.1016/j.compchemeng.2019.06.001.

[54] L. Breiman, "Random forests", *Machine Learning*, 45, 5–32, 2001.

[55] C. Avcı, A. Akbaş, "Sleep apnea classification based on respiration signals by using ensemble methods", *Biomedical Materials and Engineering*, 26, S1703–S1710, 2015. https://doi.org/10.3233/BME-151470.

[56] B. Akkoç, A. Arslan, H. Kök, "Automatic gender determination from 3D digital maxillary tooth plaster models based on the random forest algorithm and discrete cosine transform", *Computer Methods and Programs in Biomedicine*, 143, 59–65, 2017. https://doi.org/10.1016/j.cmpb.2017.03.001.

[57] W. H. Chen, S. H. Hsu, H. P. Shen, "Application of SVM and ANN for intrusion detection", *Computers & Operations Research*, 32(10), 2617–2634, 2005. https://doi.org/10.1016/j.cor.2004.03.019.

[58] A. K. Dwivedi, "Artificial neural network model for effective cancer classification using microarray gene expression data", *Neural Computing and Applications*, 29(12), 1545–1554, 2018. https://doi.org/10.1007/s00521-016-2701-1.

[59] A. Ahad, A. Fayyaz, T. Mehmood, "Speech recognition using multilayer perceptron", *Proceedings of the IEEE Students Conference (ISCON'02)*, Lahore, Pakistan, 16-17 August 2002. https://doi.org/10.1109/ISCON.2002.1215948.

[60] F. Z. Solak, "COVID-19 detection using variational mode decomposition of cough sounds", *Konya Journal of Engineering Sciences*, 11(2), 354–369, 2023. https://doi.org/10.36306/konjes.1110235.

[61] P. Mishra, C.M. Pandey, U.Singh, A. Gupta, C. Sahu, A. Keshri, "Descriptive Statistics and Normality Tests for Statistical Data", Annals of Cardiac Anaesthesia., 22(1), 67–72, 2019. https://doi.org/10.4103/aca.ACA_157_18.

[62] N. Smirnov, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples", *Moscow University Mathematics Bulletin*, 2, 3–26, 1939.

[63] H. B. Mann, D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other", *Annals of Mathematical Statistics*, 18(1), 50–60, 1947. https://doi.org/10.1214/aoms/1177730491.

[64] H. C. Jan, W. H. Yang, C. H. Ou, "Combination of the preoperative systemic immune-inflammation index and monocyte-lymphocyte ratio as a novel prognostic factor in patients with upper-tract urothelial carcinoma", *Annals of Surgical Oncology*, 26(2), 669–684, 2019. https://doi.org/10.1245/s10434-018-6942-3.

[65] R. K. Badam, M. Manjunath, M. S. Rani, "Determination of sex by discriminant function analysis of lateral radiographic cephalometry", *Journal of Indian Academy of Oral Medicine and Radiology*, 23(3), 179–183, 2011, https://www.researchgate.net/publication/284833858_Determination_of_Sex_by_Discriminant_Function_Analysis_of_Lateral_Radiographic_Cephalometry.

[66] S. A. Veyre-Goulet, C. Mercier, O. Robin, C. Guérin, "Recent human sexual dimorphism study using cephalometric plots on lateral teleradiography and discriminant function analysis", *Journal of Forensic Sciences*, 53(4), 786–789, 2008. https://doi.org/10.1111/j.1556-4029.2008.00759.x.

[67] A. Mitsea, I. Pouliezou, N. Christoloukas, I. Sifakakis, "Is lateral cephalometric analysis useful to sex estimation in forensic science? A cross-sectional study", *Journal of Forensic and Legal Medicine*, 118, 103077, 2026. https://doi.org/10.1016/j.jflm.2026.103077.