Egyptian Journal of Radiology and Nuclear Medicine

RESEARCH Open Access

Sex estimation from the sternum in Turkish population using various machine learning methods and deep neural networks

Hüseyin Uğur Bakan[1], Emel Emir Yetim[2*], Nevruz İlhanlı[3], Okan Kılınç[4], Melih Öz[5], Yaşar Mustafa Karagöz[6] and Sefa Sönmez[7]

Abstract

Objective(s) Sex estimation is an important initial step for personal identification of unknown skeletal remains in forensic examination. This study aims to determine sexual dimorphism in sternum measurements made with computed tomography (CT) in the Turkish population. The study also aims to compare the effectiveness of various machine learning techniques, including K-nearest neighbors (KNN), random forest, XGBoost, naive Bayes, logistic regression, and linear discriminant analysis (LDA), in sex prediction, and to assess the usability of deep neural networks (DNN) in sternum images.

Materials and methods CT images of 485 cases (248 males and 237 females) were used. From the sagittal plane manubrial length (M) and corpus sternal length (CSL), from the coronal plane manubrial width (MW), sternal body width at first sternebra (CSW1) and sternal body width at third sternebra (CSW3) were measured. From these measurements, 1 index (sternal index/SI), 1 length (corpus sterni length/CSL), and 1 area (sternal area/SA) were calculated. In addition, sternum images in the sagittal and coronal planes were recorded to train the deep neural network.

Results The difference between male and female cases was significant for all measurements. The best-performing model was LDA, followed by logistic regression, naïve Bayes, XGBoost, random forest, and KNN, respectively. The lowest performance was observed in the DNN model using sternum images. However, there was no statistically significant difference between the area under curves (AUCs) of DNN and LDA.

Conclusion This study shows that sternum measurements or images can be used for sex estimation in the Turkish population.

Keywords Anthropology, Deep neural networks, Forensic science, Machine learning, Sternum, Sex estimation

*Correspondence:
Emel Emir Yetim
emellemirr@gmail.com
[1] Council of Forensic Medicine, Tokat, Turkey, Tokat, Turkey
[2] Department of Radiology, Faculty of Medicine, Akdeniz University, Antalya, Turkey, Antalya, Turkey
[3] Department of Biostatistics and Medical Informatics, Faculty of Medicine, Akdeniz University, Antalya, Turkey, Antalya, Turkey
[4] Department of Forensic Medicine, Ankara Etlik Integrated Health Campus, Ankara, Turkey, Ankara, Turkey
[5] Department of Computer Engineering, Institute of Natural and Applied Sciences, Akdeniz University, Antalya, Turkey, Antalya, Turkey
[6] Department of Forensic Medicine, Faculty of Medicine, Akdeniz University, Antalya, Turkey, Antalya, Turkey
[7] Tokat Gaziosmanpaşa University, Faculty of Medicine, Department of Anatomy, Tokat, Turkey, Tokat, Turkey

Springer Open

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Introduction

In forensic anthropology, a crucial aspect involves identifying human skeletal remains by establishing a biological profile. This profile encompasses an individual's sex, age, stature, and ancestry. Among these components, sex estimation plays a pivotal role, as other biological characteristics—such as age, stature, and ancestry—are strongly influenced by sex [1, 2].

Nearly all parts of the human skeleton have been examined for the determination of sex. Among these, pelvic bones and the cranium exhibit the most pronounced sexually dimorphic characteristics, as widely documented in forensic anthropological studies [2, 3]. Consequently, the skull and pelvis are generally preferred to estimate sex. However, these bones may not always be available or well preserved in forensic cases, necessitating the examination of alternative skeletal elements. The sternum also exhibits its sexual dimorphism, with recent studies demonstrating an overall classification accuracy exceeding 80% across various populations [4, 5]. Early studies emphasized traditional morphometric approaches, with discriminant function analysis frequently used to assess dimorphic differences in sternal measurements. Although these classical methods achieved moderate to high accuracy rates, they often lacked generalizability due to population-specific variations [6–8]. In recent years, the application of machine learning techniques has shown promise in improving predictive accuracy [5]. Recent investigations confirm the potential of sternal metrics in diverse populations using both traditional and AI-based models [9–11]. In parallel, various machine learning and deep learning approaches have been increasingly applied to skeletal image analysis and classification tasks, demonstrating promising results in sex estimation and related areas [12]. Artificial intelligence has also shown great potential in forensic and medical imaging, particularly when integrated into morphometric analysis pipelines [13].

The estimation of sex relies on two methods: metric (morphometric) and non-metric (morphologic) [14]. Morphological methods necessitate expertise in the field of forensic anthropology, as they depend on the subjective evaluations of the examiner using terms like"large,"" small,""curved,"and"flat"[4, 15]. In contrast, metric methods are preferred due to their ease of repeatability, high accuracy, and deficiency of a requirement for particular expertise [14]. However, a major limitation of metric methods is that results from a particular population cannot be generalized to other populations and require population-specific studies [16].

Problem Statement: Despite the sternum's potential in sex estimation, limited studies have investigated its use in population-specific contexts using advanced machine learning methods. Moreover, traditional statistical models like linear discriminant analysis (LDA), although widely used, may not capture the complex, nonlinear relationships inherent in biological structures.

Key contributions: This study addresses the above gap by:

1. Evaluating the performance of various machine learning (ML) algorithms in sex estimation from the sternum in a contemporary Turkish population.
2. Introducing deep neural networks (DNNs) as a novel approach using sternum images to assess classification performance.
3. Comparing traditional models (e.g., LDA) with state-of-the-art ML and DNN models in terms of accuracy and reliability.
4. Providing a current population-specific model that can serve as a reference for future forensic applications in Turkey.

In addition, the population-specific nature of metric measurements and the change in populations over time necessitate separate studies for each population at certain time intervals for each bone. Thus, the requirement to access large data sets arises. However, collecting big data from human observers is costly and time-consuming. Therefore, another purpose of our study is to reveal the prediction accuracy of a deep neural network (DNN) using sternum images.

Materials and methods

Population data

This study was approved by the Clinical Research Ethics Committee of XXX University (Decision No: KAEK-453, Date: 20.07.2022) and conducted in accordance with the Declaration of Helsinki. Thorax CT images were retrospectively obtained from the radiology archive of XXX University Hospital. From an initial dataset of 2,168 cases scanned between June 3, 2020, and June 3, 2022, a total of 600 cases were randomly selected. To ensure the inclusion of only skeletally mature individuals, participants younger than 18 years were excluded, as the sternum undergoes significant ossification changes during adolescence. Additional exclusion criteria included the presence of trauma, tumors, surgical interventions, incomplete ossification, or foreign nationality. After applying these criteria, a final sample of 485 adult cases (248 males and 237 females), aged between 18 and 65 years, was included in the study.

CT protocol and measurements

The thorax CT examinations were performed with 16-detector or 128-detector multidetector CT (MDCT)

devices (Toshiba Medical Systems, Activion 16, Japan, or Siemens, Somatom Defnition Edge, Germany, respectively). The mean slice thickness on thorax CT images with and without contrast enhancement in randomly selected patients during this time period was 0.75 mm. MDCT has allowed isotropic data sampling, which has become the basis for implementing modern radiological examination methods in forensic medicine. Isotropic data describe the situation in which multiplanar reformation (MPR) can be performed in any plane with the same spatial resolution as the original sections. After 3D volume rendering, the 3D coordinates of 10 anatomical landmarks were obtained using Sectra IDS7 (version 23.2.3.5105, 2021), with a total of 5 linear interlandmark measurements calculated. All measurements were made using 3D volume rendering images with a 5 mm maximum intensity project (MIP) over the bone window. Sagittal and coronal plane sections of the obtained sternum images were used for measurement. In this study, the 5 measurements previously defined in the literature were used [6, 17, 18]. In addition, 1 index (sternal index), 1 length (combined length of the manubrium and body sternum), and 1 area (sternal area) were calculated. CT images of the sternum are shown in Fig. 1.

Anatomical landmarks description

To facilitate reproducibility, the anatomical landmarks used for sternal measurements are described and numbered as follows:

1. Incisura jugularis: Superior border of the manubrium (Landmark for manubrial length start).
2. Manubriosternal joint: Junction between manubrium and body of sternum (Landmark for both manubrial length end and corpus sterni start).
3. Xiphisternal joint: Inferior border of the sternal body (Landmark for corpus sterni length end).
4. Incisura costalis 1: Used to determine the manubrium width (MW).
5. Incisura costalis 2–3: Used to define corpus sterni width at first sternebra (CSW1).
6. Incisura costalis 4–5: Used to define corpus sterni width at third sternebra (CSW3).

Sternal measurements and calculated parameters

1. Manubrial length (M) indicates the distance between incisura jugularis and manubriosternal joint (sagittal plane).
2. Corpus sterni length (CSL) indicates the distance between manubriosternal joint and xiphisternal joint (sagittal plane).
3. Combined length of the manubrium corpus sterni (CL) shows sum of M and CSL (M + CSL).
4. Manubrium width (MW) indicates manubrium width at the level of the line passing through the midpoint of incisura costalis 1 on the right and left (coronal plane).
5. Corpus sterni width at first sternebra (CSW1) indicates sternal width at the level of the line passing

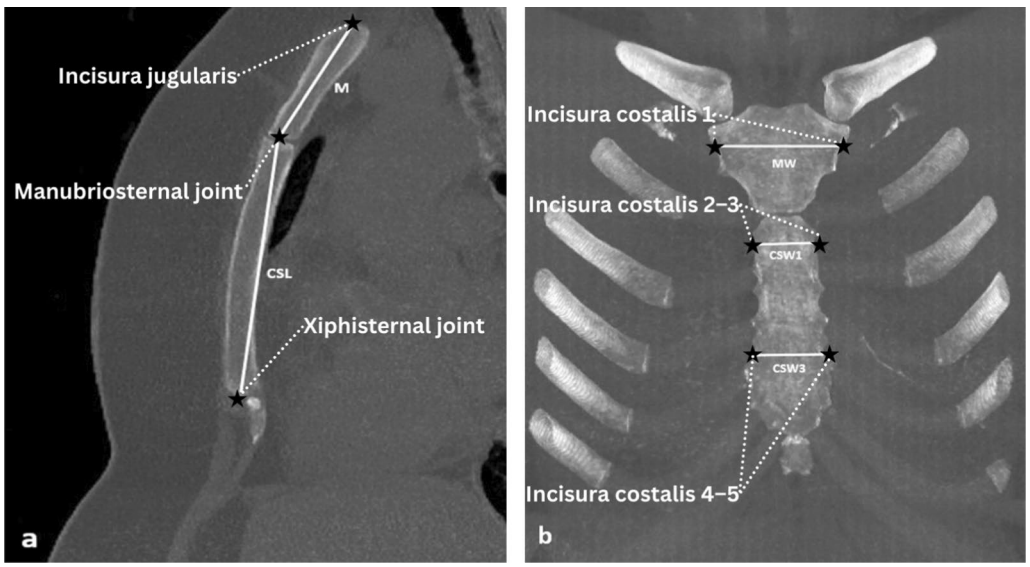


Fig. 1 CT images of the sternum. (**a**) Sagittal section the sternum and the 2 measurement performed on this section. (**b**) Coronal section parallel to the sternum long axis and the 3 measurements performed on this section

through the midpoint of incisura costalis 2 and 3 on the right and left (coronal plane).
6. Corpus sterni width at third sternebra (CSW3) indicates sternal width at the level of the line passing through the midpoint of incisura costalis 4 and 5 on the right and left (coronal plane).
7. Sternal index (SI) is calculated by dividing M by CSL, then multiplied by 100 [(M/CSL) × 100].
8. Sternal area (SA) is calculated as follows: The sum of M and CSL is multiplied by the sum of MW, CSW1 and CSW3 and divided by three [(M + CSL) x (MW + CSW1 + CSW3)/3].

All linear measurements were performed in millimeters (mm) using standard PACS software on sagittal and coronal CT images. A board-certified radiologist (Dr. XXX) with 10 years of experience in thoracic imaging conducted all measurements. Each measurement was repeated twice, and the average value was used. The radiologist was blinded to the sex of the cases to minimize bias. For DNN analysis, sagittal and coronal CT sections were exported from DICOM format and converted to JPEG format using the PACS system. All images were resized to 256 × 256 pixels and normalized prior to model training. This preprocessing ensured uniform input dimensions for the convolutional neural network.

Image processing

The images used for the study are of different sizes depending on the size of the individual and the way they were captured. Therefore, some image processing operations are applied to keep their spatial information intact and provide deep neural networks with the same-sized images.

Image padding

Image padding is a technique used in image processing to increase the overall size of an image by adding extra borders around the edge or edges of the original image. The purpose of padding is to change the image size without resizing or distorting the original image. Depending on the usage properties, padding can be applied to an image in multiple ways, including zero padding, reflection, or wrapping padding. In this study, a zero padding operation is applied to the images. Zero padding is the operation that adds 0 value pixels around the image, which are black and considered background in image segmentation problems. In this study, zero padding is applied to convert the image dimensions to 256 × 256 without changing the original aspect ratio of the image.

Image resizing

The image resizing technique is used to change the dimensions of an image during image processing. The resizing process of an image involves increasing or decreasing the number of pixels in the image. In this study, resizing was used to change the image dimensions to 256 × 256 to use in a deep neural network (DNN). In this study, bicubic interpolation is used to resize images. Bicubic interpolation uses 4 × 4 windows to calculate resized pixel values. This operation ensures more accurate results than other resizing algorithms.

Machine learning models

K-nearest neighbors (KNN), random forest, XGBoost, naïve Bayes, logistic regression, and LDA methods were used for sex prediction. In modeling, M, MW, CSL, and CSW1 measures were used as input variables. To evaluate multicollinearity, variance inflation factors (VIFs) were computed for input variables. The cutoff value for VIFs was 2.5, and all variables have VIFs below 2.5 [19]. Input variables are standardized by using StandardScaler in scikit-learn. To build and evaluate machine learning models, fivefold cross-validation approach was used. For all machine learning models, default hyperparameter settings that were defined in the scikit-learn library were used. The key hyperparameters of each model are represented in Supplementary Material. Python (version 3.7) and the scikit-learn library were used for modeling.

Deep neural network (DNN)

The DNN used in this study was created using Python (version 3.7), Tensorflow (version 2.7), and Keras API. The network takes front (coronal section) and side (sagittal section) views as input, producing sex probability for a given input. Moreover, the network processes two inputs simultaneously and combines their extracted information before producing an output. It also uses separable convolutions to reduce the computational requirement without sacrificing performance. Therefore, this study utilizes a multi-angle representation of the images. The DNN structure used in the study is shown in Fig. 2.

Evaluation

The test dataset was used to evaluate the performance of the prediction models. The area under the curve (AUC) and precision–recall area under the curve (PR AUC), Matthew's correlation coefficient (MCC), accuracy, and F1 score were used for the evaluation of prediction models. All performance metrics were reported as the mean

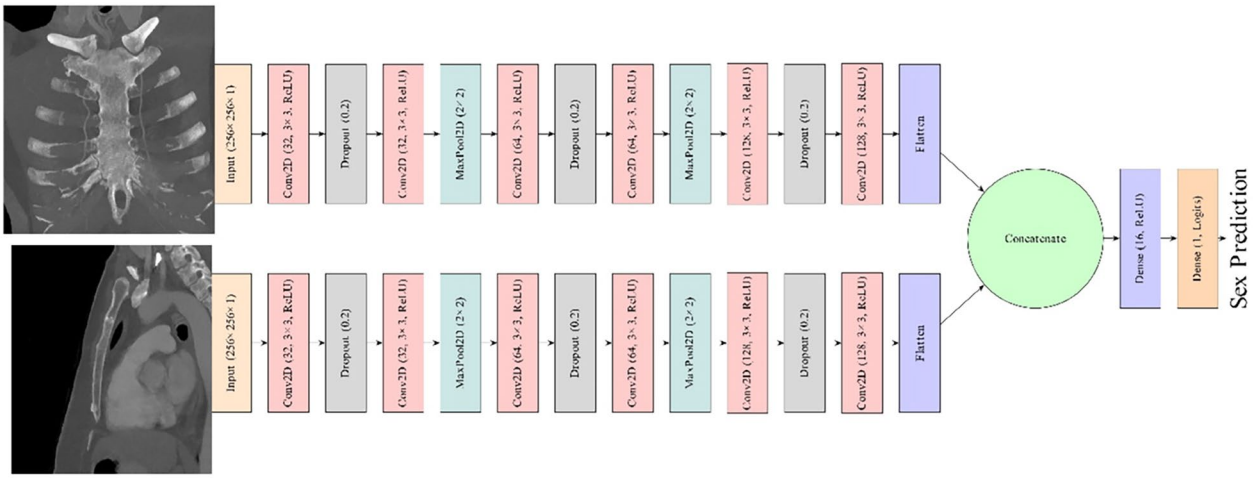


Fig. 2 Deep neural network structure used in the study

values from fivefold cross-validation. AUC was considered the main performance metric.

Statistical analysis

The statistical analysis was performed with IBM SPSS Version 21.0 (SPSS Inc., Illinois, USA) and R (R Core Team, 2020). Variables were expressed as the median (Q1–Q3). The Shapiro–Wilk test was used to show a normal distribution of the data. The Mann–Whitney U-test was used to compare the sternum measurements between males and females. To examine intra-observer reliability, the relative technical error of measurement (TEM) was calculated for randomly selected 30 cases by using the anthrocheckr library in R. For intra-observer error, a relative TEM value of less than 1.5% was reported to be acceptable [20]. To evaluate the association between sternal measurements and sex with females as the reference category, multivariable logistic regression analysis was used. SI, SA, and CL variables were excluded from the multivariable logistic regression because they were derived from other variables. The De Long test was performed to analyze differences between AUCs. A p value below 0.05 was used to define the significance of all statistical analyses.

Results

Of the 485 cases, 248 (%51.1) of the cases were male with a median age of 53 (44–60, Q1–Q3), and 237 (%48.9) of the cases were female with a median age of 53 (45–59, Q1–Q3). There was no statistically significant difference between the ages of male and female cases ($p=0.858$). Relative TEM for intra-observer error was less than %1.5 for all measurements (M: 1.15, CSL: 0.78, MW: 0.89, CSW1: 1.33, CSW3: 1.27). According to the univariate analysis results, the difference between male and female cases was significant in all variables. SA, CL, and MW had the largest effect sizes, respectively. When all measurements except SI taken from the sternum were assessed, it was shown that men had statistically significantly higher values than women ($p<0.001$). The median sternal index was found to be 55.87 in females and 51.09 in males. The sternal index was significantly higher in females ($p<0.001$). The characteristics of male and female cases are presented in Table 1.

Multivariate logistic regression analysis results showed that M (OR: 0.863, 95% CI: 0.798–0.934, $p<0.001$), CSL (OR: 0.882, 95% CI: 0.851–0.915, $p<0.001$), MW (OR: 0.805, 95% CI: 0.748–0.866, $p<0.001$), and CSW1 (OR: 0.887, 95% CI: 0.800–0.984, $p=0.024$) are the independent predictive variables of sex. In contrast, CSW3 was not significantly associated with sex (OR: 0.955, 95% CI: 0.887–1.027, $p=0.214$). Table 2 presents the results of the multivariate logistic regression analysis.

After confirming the predictors of sex by multivariate logistic regression analysis, machine learning models were built by using M, CSL, MW, and CSW1 variables as input variables. The AUCs of the prediction models were 0.937 for the DNN, 0.937 for the KNN, 0.942 for the random forest, 0.950 for the XGBoost, 0.959 for the naïve Bayes, 0.960 for the logistic regression, and 0.961 for the LDA. There was no statistically significant difference between the AUCs of DNN and LDA ($p>0.05$ in all folds). LDA also showed highest performance in MCC, accuracy, and F1 score, with values of 0.839, 0.920, and 0.920, respectively. In terms of PR AUC, naïve Bayes performed best with a score of 0.952. The performance evaluation of the sex prediction models is presented in Table 3. The ROC and PR curves for the LDA model,

Table 1 Statistical descriptive of the studied parameters in two sexes

Variables	Total	Sex		p^*
		Male ($n=248$)	Female ($n=237$)	
		Median (min–max)	Median (min–max)	
M(mm)	49.7(46.5–53.0)	52.3(48.8 55.7)	47.2(45.1–50.3)	< 0.001
CSL(mm)	93.7(83.7–103.1)	102.4(94.8–109.9)	85.1(79.2–91.3)	< 0.001
CL(mm)	143.3(132.6–156.0)	155.05(147.4–162.9)	133.1(126.8–139)	< 0.001
MW(mm)	56.7(51.5–62.9)	62.7 (58.5–68.4)	51.7(48.9–55.4)	< 0.001
CSW1(mm)	26.6(23.3–29.7)	28.9 (26.4–31.5)	23.9(22.0–26.8)	< 0.001
CSW3(mm)	31.8(27.9–36.6)	35.1 (31.4–38.4)	28.8 (25.3–31.9)	< 0.001
SI	52.84(48.19–58.69)	51.09(46.51–55.72)	55.87(50.45–61.49)	< 0.001
SA(mm^2)	5607.95(4665.45–6627.55)	6611.25(5982.83–7230.75)	4675.03(4263.60–5140.80)	< 0.001

* Mann–Whitney U test M: length of the manubrium, CSL: corpus sterni length, CL: combined length of the manubrium and corpus sterni, MW: manubrium width, CSW1: corpus sterni width at first sternebra, CSW3: corpus sterni width at third sternebra, SI: sternal index, SA: sternal area, $p < 0.05$

Table 2 Results of the multivariate logistic regression analysis

Variables (mm)	*β*	OR (95% CI)	*P*
M	−0.147	0.863 (0.798–0.934)	< 0.001
CSL	−0.125	0.882 (0.851–0.915)	< 0.001
MW	−0.217	0.805 (0.748–0.866)	< 0.001
CSW1	−0.120	0.887 (0.800–0.984)	0.024
CSW3	−0.046	0.955 (0.887–1.027)	0.214

OR: odds ratio, CI: confidence interval, M: length of the manubrium, CSL: corpus sterni length, MW: manubrium width, CSW1: corpus sterni width at first sternebra, CSW3: corpus sterni width at third sternebra

Table 3 Performance evaluation of sex prediction models

Prediction Models	AUC	PR AUC	MCC	Accuracy	F1 score
DNN with CT Images	0.937	0.935	0.759	0.878	0.874
KNN	0.937	0.901	0.749	0.870	0.860
Random forest	0.942	0.921	0.801	0.899	0.900
XGBoost	0.950	0.944	0.789	0.895	0.894
Naïve Bayes	0.959	0.952	0.814	0.907	0.905
Logistic regression	0.960	0.941	0.831	0.915	0.915
LDA	0.961	0.942	0.839	0.920	0.920

AUC: area under the curve, PR AUC: precision–recall area under the curve, MCC: Matthews correlation coefficient, DNN: deep neural network, CT: computed tomography, KNN: K-nearest neighbors (KNN), LDA: linear discriminant analysis

which showed the highest performance, are presented in Fig. 3. The ROC and PR curves for the other models are provided in Supplementary Material.

Discussion

Sternum length measurement techniques vary across studies. Some researchers utilized bone collections to measure the sternum [1, 6, 21], while others used measurements made at autopsy [7, 8]. In addition, several studies have been conducted using chest radiographs [22, 23] and computed tomography of the sternum [5, 21]. In this present study, we opted to measure sternum length using thorax CT.

Consistent with previous studies, the difference between sternum measurements according to sex was found to be statistically significant in our study [7, 21, 22]. In this study, assuredly, the female sternum was generally smaller than the male sternum. This finding is consistent with previous studies conducted in the Indian, Croatia, and Japanese populations, where male sternal dimensions were reported to be significantly larger than those of females [22–24]. These sex differences are probably due to body size differences between the two sexes or different endocrine factors affecting the growth of the sternum in the two genders.

In our study, only SI among the variables was greater in women than in men. This inverse relationship was also observed in studies by Ramadan et al. [14] and Das et al. [1], suggesting that SI may be a population-sensitive indicator of sex.

The study results showed that deep neural networks (DNN) applied to CT images yielded the lowest performance among the evaluated models, while K-nearest neighbor (KNN), random forest, XGBoost, naïve Bayes, and logistic regression performed better. The best performance was observed with linear discriminant analysis (LDA), consistent with previous studies [18, 21, 22, 25]. However, when comparing the area under the curve (AUC) values of the lowest (DNN) and highest (LDA) performing models, the difference was not statistically significant.

Although the finding that LDA outperforms DNN is not unexpected, it highlights the critical importance of

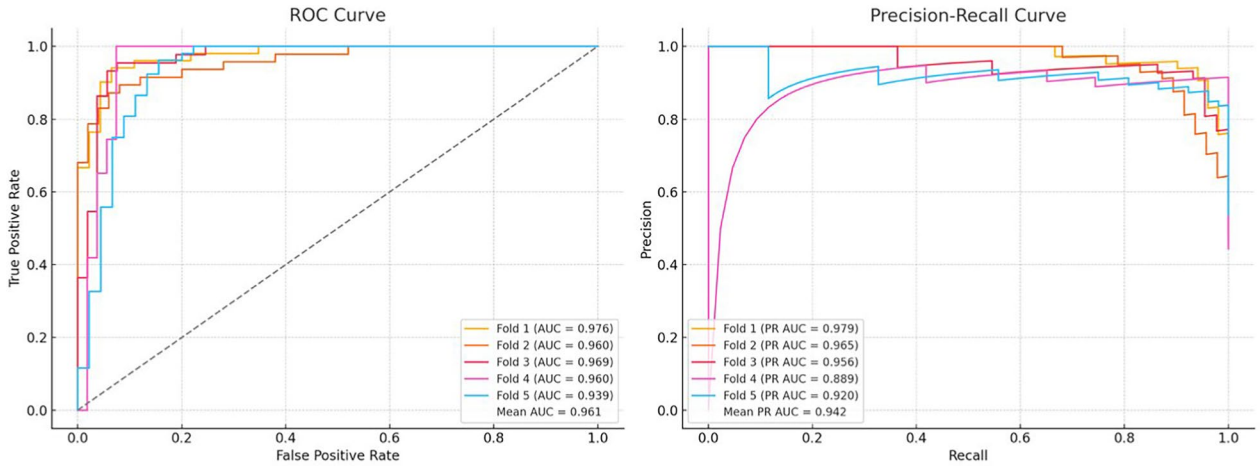


Fig. 3 Receiver operating characteristic (ROC) curve (left) and precision–recall (PR) curve (right) for the linear discriminant analysis (LDA) model

aligning model selection with dataset characteristics. Deep learning models generally require large, diverse, and high-dimensional datasets to achieve optimal performance. The relatively small sample size and structured nature of sternal morphometric data likely limited the learning capacity of the DNN in this study.

Despite its lower relative performance in this context, DNNs offer practical advantages that are valuable in forensic applications. They can operate directly on raw CT images without manual measurement or landmarking, enabling fully automated workflows. This reduces operator bias and saves time—especially in cases with limited expert availability or requiring rapid analysis. In addition, DNNs are scalable and adaptable to other bones or populations. As more annotated datasets and higher-resolution imaging become available, their predictive performance is expected to improve.

Similar to other recent studies employing deep learning for sex classification using skeletal imaging [13, 26], our findings confirm that the performance of DNNs is strongly influenced by dataset size, image quality, and anatomical complexity. Despite its lower relative performance in this context, the DNN approach remains a promising complementary tool in forensic anthropology, particularly for future applications involving large imaging archives and minimal human intervention.

They can analyze raw imaging data directly, without requiring manual landmarking or measurements, which enables fully automated workflows. This can be particularly advantageous in time-sensitive forensic scenarios or settings where expert availability is limited. Similar to other recent studies using deep learning for sex classification from skeletal imaging [13, 26], our findings confirm that the performance of DNNs is highly dependent on factors such as image resolution, anatomical region, and dataset scale. Therefore, despite its lower comparative accuracy in this context, the DNN approach remains promising as a supplementary tool—especially as larger annotated datasets and higher-resolution imaging become more widely available.

Importantly, the novelty of our study lies in its broader context. To our knowledge, this is the first investigation applying a DNN to sternum-based sex estimation using CT images in a Turkish population. Moreover, we conducted a comprehensive comparison of multiple predictive models using the same dataset, providing practical insights into their relative strengths and limitations.

Compared to previous studies that utilized traditional statistical methods such as discriminant analysis and logistic regression on sternal measurements [5, 17, 22], our study is among the first to systematically evaluate and compare the performance of both classical machine learning models and deep learning approaches (DNN) within a single dataset. While prior research often focused on single-method evaluation or non-imaging data, we incorporated both morphometric and image-based data to assess prediction accuracy. Furthermore, the application of DNN to CT-derived sternal images in a Turkish population has not been previously documented, adding novelty to our approach. This comprehensive comparison across multiple models provides a more holistic understanding of sex estimation performance in forensic settings.

Although no prior study has applied DNN to sternal images, several studies have used deep learning approaches for sex estimation based on other skeletal parts. For instance, Lee et al. applied convolutional neural networks to pelvic CT scans and achieved an AUC of 0.96 [27], while in our study, the DNN model achieved an AUC of 0.937—competitive despite being

applied to a relatively small and anatomically complex bone such as the sternum. The lower performance of the DNN model may be attributed to limited training data and variability in image quality. Adams et al. similarly emphasized the need for large and diverse datasets to fully exploit the predictive potential of deep learning in skeletal analysis [10].

Despite its relatively lower performance, DNN-based analysis offers practical advantages such as full automation, elimination of manual measurements, and potential usability in time-sensitive or resource-constrained forensic contexts. These findings support the relevance of DNN as a supplementary tool and underscore the importance of contextualizing model performance within the specific constraints of forensic applications.

Although we suppose that the present study will contribute to the literature, it has limitations. In this study, sex was assumed to be binary (male–female), whereas sex actually exists along a spectrum and includes additional sex categorizations and gender identities such as people who are intersex/have differences of sex development. The dataset in our study was relatively small, which may have affected the performance of our prediction models. Furthermore, our study lacked external validation, which may have limited the generalizability and reliability of our results. Our study contains up-to-date data on men and women belonging to a certain part of the modern Turkish population. However, the data obtained were collected only from the cases who admitted to our hospital. Since physical characteristics may differ in different populations, our results cannot be generalized to the entire Turkish population. Finally, the use of default hyperparameters may limit the predictive performance of our models. The application of hyperparameter tuning on larger datasets may provide improved performance in future studies.

In conclusion, this study indicated that the sternum of Turkish subjects showed high sexual dimorphism. KNN, random forest, XGBoost, naïve Bayes, logistic regression, and LDA using the measurements of the sternum and DNN using the sternum images provided sex classification accuracy rates of approximately 85–95%. The equations obtained in this study may be useful in forensic contexts, particularly in cases where the pelvis or skull is not available for analysis or as an adjunct to other information.

Future studies may benefit from the development of larger and more diverse datasets to improve the performance and generalizability of deep learning models for sex estimation. Moreover, combining sternal imaging with data from other skeletal parts may yield higher predictive accuracy. Cross-population validation studies using standardized imaging and AI protocols are also needed to assess the robustness of existing models.

Conclusion

This study demonstrated that sternum-based morphometric data and imaging can be effectively used for sex estimation in a contemporary Turkish population. Traditional models such as LDA and logistic regression outperformed other machine learning algorithms and deep neural networks in this context. Despite its relatively lower performance, the DNN model using CT images showed comparable accuracy and may offer a viable alternative in cases where manual measurements are not feasible. Future research should focus on increasing sample size, applying multi-bone models, and validating deep learning frameworks in diverse populations. The integration of explainable AI approaches may also help forensic experts better understand and trust model predictions in real-world forensic scenarios.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43055-025-01583-1.

Additional file 1.

Acknowledgements
No acknowledgements made.

Author contributions
HU B involved in conceptualization, methodology, investigation, resources, writing—original draft, project administration. E EY took part in investigation, writing—review & editing. N I involved in conceptualization, methodology, software, validation, formal analysis, writing—original draft. O K involved in resources, writing—original draft. M O took part in software, formal analysis, writing—original draft, visualization. YM K involved in writing—review & editing. S S took part in writing and revised the manuscript. All authors read and approved the final manuscript.

Funding
Not applicable.

Data availability
Data are available on request due to privacy/ethical restrictions. The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Declarations

Ethics approval and concent to participate
Not applicable.

Concent for publication
Not applicable.

Competing interests
The authors declare that there are no conflicts of interest. In selecting the literature reviewed in this study, we focused on peer-reviewed articles published in indexed scientific journals that investigated sex estimation using

the sternum or thoracic structures. We included studies that employed both traditional morphometric techniques (e.g., discriminant analysis) and modern machine learning or deep learning methods. Priority was given to research involving CT imaging, metric data, and population-specific findings, especially those focusing on Turkish or comparable populations. We excluded case reports, reviews without original data, and studies that did not report performance metrics or validation methods. Additionally, studies focusing solely on age or stature estimation without sex differentiation were not considered relevant for our primary research objective.

Received: 18 November 2024 Accepted: 29 August 2025
Published online: 29 September 2025

References

1. Das S, Ruengdit S, Singsuwan P, Mahakkanukrauh P (2015) Sex determination from different sternal measurements: a study in a Thai population. J Anat Soc India 64:155–161
2. Buikstra JE, Ubelaker DH (1994). Standards for data collection from human skeletal remains.
3. Walker PL (2005) Greater sciatic notch morphology: sex, age, and population differences. Am J Phys Anthropol. https://doi.org/10.1002/ajpa.10422
4. Etli Y, Asirdizer M, Hekimoglu Y, Keskin S, Yavuz A (2019) Sex estimation from sacrum and coccyx with discriminant analyses and neural networks in an equally distributed population by age and sex. Forensic Sci Int 303:109955
5. Oner Z, Turan MK, Oner S, Secgin Y, Sahin B (2019) Sex estimation using sternum part lenghts by means of artificial neural networks. Forensic Sci Int 301:6–11
6. Bongiovanni R, Spradley MK (2012) Estimating sex of the human skeleton based on metrics of the sternum. Forensic Sci Int 219:290.e1-290.e7
7. Singh J, Pathak R (2013) Morphometric sexual dimorphism of human sternum in a north Indian autopsy sample: sexing efficacy of different statistical techniques and a comparison with other sexing methods. Forensic Sci Int 228:174e1-174e10
8. Mukhopadhyay PP (2010) Determination of sex from adult sternum by discriminant function analysis on autopsy sample of Indian Bengali population: a new approach. J Indian Acad Forensic Med 32:321–324
9. Saha S, Datta P, Bandyopadhyay M (2022) Sexual dimorphism of sternum in Eastern Indian population using multidetector CT and machine learning techniques. Egypt J Forensic Sci 12:1–9
10. Adams J, Rhoads J, Moore J (2021) Deep learning-based sex estimation from 3D models of the sternum. Forensic Imaging 25:200468
11. Bataineh A, Al-Bayyari N (2023). Accuracy of sex estimation using sternum and deep learning: a cross-population approach. International Journal of Legal Medicine. (If accepted or in press)
12. Kumar I, Rawat J (2024) Segmentation and classification of white blood SMEAR images using modified CNN architecture. Discov Appl Sci 6:587
13. Elsheikh TM, Wahba M (2020) Artificial intelligence in forensic medicine. In: Barh D (ed) Artificial Intelligence in Precision Health, 1st edn. Academic Press, NY, pp 171–185
14. Ramadan SU, Türkmen N, Dolgun NA, Gökharman D, Menezes RG, Kacar M (2010) Sex determination from measurements of the sternum and fourth rib using multislice computed tomography of the chest 19. Forensic Sci Int 197(1–3):120e1-120e5
15. Rogers TL (2005) Determining the sex of human remains through cranial morphology. J Forensic Sci. https://doi.org/10.1520/JFS2003385
16. Turan MK, Oner Z, Secgin Y, Oner S (2019) A trial on artificial neural networks in predicting sex through bone length measurements on the first and fifth phalanges and metatarsals. Comput Biol Med 115:103490
17. Cunha E, Cattaneo C (2006) Forensic anthropology and forensic pathology. In: Schmitt A, Cunha E, Pinheiro J (eds) Forensic Anthropology and medicine. Humana Press, NJ, pp 39–53
18. Koşar Mİ, Gençer CU, Tetiker H, Yeniçeri İÖ, Çullu N (2022) Sex and stature estimation based on multidetector computed tomography imaging measurements of the sternum in Turkish population. Forensic Imaging 28:200495
19. Johnston R, Jones K, Manley D (2018) Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. Qual Quant 52:1957–1976
20. Perini TA, Oliveira GLd, Ornellas JdS, Oliveira FPd (2005) Technical error of measurement in anthropometry. Rev Bras Med Esporte 11:81–85
21. Darwish RT, Abdel-Aziz MH, El Nekiedy A-AM, Sobh ZK (2017) Sex determination from chest measurements in a sample of Egyptian adults using Multislice Computed Tomography. J Forensic Leg Med 52:154–158
22. Torimitsu S, Makino Y, Saitoh H, Sakuma A, Ishii N, Inokuchi G (2015) Estimation of sex in Japanese cadavers based on sternal measurements using multidetector computed tomography. Leg Med 17:226–231
23. Selthofer R, Nikolić V, Mrčela T, Radić R, Lekšan I (2006) Morphometric analysis of the sternum. Coll Antropol 30:43–47
24. Mahajan A, Batra APS, Khurana BS, Sharma SR (2009) Sex determination of human sterna in North Indians.
25. Franklin D, Flavel A, Kuliukas A, Cardini A, Marks MK, Oxnard C (2012) Estimation of sex from sternal measurements in a Western Australian population. Forensic Sci Int 217:230.e1–5
26. Kim D, Lee H, Lee J (2020) Automatic gender classification of skull CT images using deep learning. Multimed Tools Appl 79(39):29245–29263
27. Lee JH, Kim YS, Park HJ (2020) Sex classification using deep convolutional neural networks on pelvic CT images. J Digit Imaging 33(2):426–432

Publisher's Note

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