



OPEN

## A study on sex estimation by using machine learning algorithms with parameters obtained from computerized tomography images of the cranium

Seyma Toy<sup>1</sup>✉, Yusuf Secgin<sup>1</sup>, Zual Oner<sup>2</sup>, Muhammed Kamil Turan<sup>3</sup>, Serkan Oner<sup>4</sup> & Deniz Senol<sup>5</sup>

The aim of this study is to test whether sex prediction can be made by using machine learning algorithms (ML) with parameters taken from computerized tomography (CT) images of cranium and mandible skeleton which are known to be dimorphic. CT images of the cranium skeletons of 150 men and 150 women were included in the study. 25 parameters determined were tested with different ML algorithms. Accuracy (Acc), Specificity (Spe), Sensitivity (Sen), F1 score (F1), Matthews correlation coefficient (Mcc) values were included as performance criteria and Minitab 17 package program was used in descriptive statistical analyses.  $p \leq 0.05$  value was considered as statistically significant. In ML algorithms, the highest prediction was found with 0.90 Acc, 0.80 Mcc, 0.90 Spe, 0.90 Sen, 0.90 F1 values as a result of LR algorithms. As a result of confusion matrix, it was found that 27 of 30 males and 27 of 30 females were predicted correctly. Acc ratios of other MLs were found to be between 0.81 and 0.88. It has been concluded that the LR algorithm to be applied to the parameters obtained from CT images of the cranium skeleton will predict sex with high accuracy.

The main purpose of forensic anthropology is to reconstruct the biological profile of deceased individuals; that is, to predict sex, age of death, lineage and height based on the remains of skeletons<sup>1</sup>. Forensic sex prediction has taken a large place in literature since the late 1960s and identification of sex from human skeleton has been described as an important factor, even a key element in both forensic medicine and bio-archaeological context<sup>2-4</sup>. Sex prediction is an indispensable part of biological profile. Anthropologist uses the biomarkers of the skeletal system that vary between sexes to determine sex<sup>5,6</sup>.

It is noteworthy that studies have been conducted in literature for the estimation of sex almost with all bones of the human skeleton and that the accuracy of gender determination has been researched frequently by comparing with different populations. It can be seen that various bones such as femur<sup>2,3</sup>, patella<sup>7,8</sup>, mandible<sup>9</sup>, calcaneus<sup>10</sup>, metatarsal bone and phalanx<sup>11,12</sup>, occipital condyle<sup>13</sup>, hand bones<sup>14,15</sup> and sternum<sup>16</sup> are used in sex prediction. It has been reported in a large number of studies in literature that cranium and pelvis bones, which are considered to be the most dimorphic areas according to skeletal parts, can be used in sex prediction by using different assessment methods<sup>4,10,16-19</sup>.

Identification of sex includes some inherent limitations that are affected by different factors such as ethnicity, socio-economic status, diet and geographic location. The inability to generalize the results obtained from a specific population, especially in skeletal parts such as cranium, to other populations and the need for population-specific studies increase the interest in cranium and mandible in sex determination<sup>4,20</sup>. For these reasons, all techniques reported for identifying sex are specific to related studies and they may not be applicable to different samples or data sets<sup>3</sup>.

<sup>1</sup>Department of Anatomy, Faculty of Medicine, Karabük University, Karabük, Turkey. <sup>2</sup>Department of Anatomy, Faculty of Medicine, İzmir Bakırçay University, İzmir, Turkey. <sup>3</sup>Department of Medical Biology, Faculty of Medicine, Karabük University, Karabük, Turkey. <sup>4</sup>Department of Radiology, Faculty of Medicine, İzmir Bakırçay University, İzmir, Turkey. <sup>5</sup>Department of Anatomy, Faculty of Medicine, Düzce University, Karabük, Turkey. ✉email: seymatoy@karabuk.edu.tr

It can be seen that methods such as discriminant analysis, machine learning algorithms (ML), support vector machine and artificial neural network are commonly used in sex prediction in which these bones are examined<sup>2,3,7</sup>.

ML is a modern classifier that is used extensively in the field of engineering, and it is now gradually integrated in the field of health. These algorithms are classified as supervised, unsupervised and reinforcement. Supervised learning is algorithms that match the relationship between input and output, unsupervised learning is algorithms that match the characteristics of the data about which there is no information and reinforcement learning is the algorithms that match the input data with desired characteristics<sup>20</sup>. Decision Tree (DT) algorithm is one of the simple, powerful, fast and frequently used data mining classification algorithms that processes the inputs by dividing them continuously<sup>8,21–23</sup>. Logistic regression (LR) is a classification algorithm that uses the sigmoidal curve function to classify the relationship between output probability and parameters. Random Forest (RF) is an ensemble algorithm that can derive more than one decision tree within the system<sup>24</sup>. Extra Tree Classifier (ETC) is a superior method to RF, and this advantage is due to the random division of nodes and using all data as a training set<sup>25</sup>. Linear discriminant Analysis (LDA) is a classification algorithm that reveals the difference and relationship between classes<sup>26</sup>. Quadratic Discriminant Analysis (QDA) is a superior method to LDA and is a second-order parametric classifier<sup>27</sup>.

Computerized tomography (CT) is an imaging method that can show all tissues, especially bone tissue with sharp borders. In case of thin section, image orientation can be changed in three dimensions and can be taken to orthogonal plane. In this way, length and angle measurements can be calculated in a way that is less affected by orientation. With all these aspects, it provides superior results compared to studies carried out with more conventional osteometric devices<sup>16</sup>.

The aim of this study is to show the success of sex prediction by using ML with parameters obtained from CT images of cranium and mandible skeleton.

## Results

Of the 25 parameters determined, 20 (NVIC, NSVC, NNL, PC, NIVA, PNIC, VIC, NIC, RML, CML, GHGA, HML, COL, CMHA, HGGC, COIC, HGGMC, HGGMA) were found to be statistically significant between males and females ( $p \leq 0.05$ ). In 18 of these parameters which were found to be statistically significant, the average of the parameter used was higher in males, while the average of the parameter used was higher in females in 2 parameters (GHGA, CMHA) (Tables 1, 2).

ROC analysis was performed with the IBM SPSS (Version 21) package program to reveal the discriminative power of the parameters in distinguishing between male and female individuals, and the highest AUC ratio was obtained with the CGL parameter (Fig. 1). AUC, cut-off, p, Sen, Spe values of all parameters are given in Table 3. In addition, ROC curves and AUC values for each algorithm are given in Fig. 2.

0.90 Acc, 0.80 Mcc, 0.90 Spe, 0.90 Sen and 0.90 F1 values were found as a result of the LR algorithm. As a result of the confusion matrix performed, 27 of 30 males and 27 of 30 females were predicted correctly (Fig. 3). Of the MLs, the highest Acc, Mcc ratio was found as 0.90, 0.80 with LR algorithm. Acc ratios of the other MLs were between 0.81 and 0.88. The coefficient of each parameter according to the LR algorithm, respectively – 5.33, 1.45, 1.05, 1.01, – 6.10, – 5.30, – 5.29, 2.84, – 4.94, 5.80, – 7.77, – 1.73, – 1.50, 1.61, – 2.28, 8.12, 1.50, 1.10, 1.22, – 2.90, 7.4, – 5.59, 4.03, 4.20, – 3.01 as was found, and HGGMA, PC, BIC HGGGA, CMHA, HGGC parameters were statistically significant in terms of gender.

0.88 Acc, 0.77 Mcc, 0.88 Spe, 0.88 Sen, 0.88 F1 values were found as a result of LDA algorithm and 26 of 30 males and 27 of 30 females were predicted correctly as a result of confusion matrix. 0.83 Acc, 0.67 Mcc, 0.83 Sep, 0.83 Sen, 0.83 F1 values were found as a result of QDA algorithm and 24 of 30 males and 26 of 30 females were predicted correctly as a result of confusion matrix. 0.88 Acc, 0.77 Mcc, 0.88 Spe, 0.88 Sen, 0.88 F1 values were found as a result of RF algorithm and 24 of 30 males and 27 of 30 females were predicted correctly as a result of confusion matrix. 0.85 Acc, 0.70 Mcc, 0.85 Spe, 0.85 Sen, 0.85 F1 values were found as a result of ETC algorithm and 24 of 30 males and 27 of 30 females were predicted correctly as a result of confusion matrix. 0.81 Acc, 0.67 Mcc, 0.81 Spe, 0.81 Sen, 0.81 F1 values were found as a result of DT algorithm and 24 of 30 males and 23 of 30 females were predicted correctly as a result of confusion matrix.

In addition, in terms of the reliability of our study, the tenfold cross-validation estimation values of the algorithms are also included. As a result of tenfold cross validation, Acc ratio of  $87.766 \pm 0.819$  with LR algorithm, Acc ratio of  $87.733 \pm 0.410$  with LDA algorithm, Acc ratio of  $86.533 \pm 0.592$  with QDA algorithm, Acc ratio of  $85.766 \pm 1.045$  with RF algorithm, Acc ratio of  $77.200 \pm 1.970$  with ETC algorithm, Acc ratio of  $80.266 \pm 1.396$  was obtained with the DT algorithm (Table 4).

In our study, the SHAP explanatory model of the RF algorithm was used to reveal the contribution of the parameters to the general algorithm, and it was found that the first five contributions were found to be with the parameters HGGMC, PC, GGL, HGGGA, HGGC (Fig. 4).

## Discussion

The aim of this study is to test whether sex identification can be made by using ML with the parameters obtained from cranium and mandible CT images taken to orthogonal plane. In the statistical analysis performed, NVIC, NSVC, NNL, PC, NIVA, PNIC, VIC, NIC, RML, CML, HML, COL, HGGC, COIC, HGGMC, HGGMA parameters were found to be statistically significant in distinguishing between sexes ( $p \leq 0.05$ ). Of the MLs tested, 0.90 Acc, 0.80 Mcc, 0.90 Spe, 0.90 Sen, 0.90 F1 values were found as a result of LR algorithm. It was found that 27 of 30 males and 27 of 30 females were predicted correctly as a result of confusion matrix. Acc ratios of other MLs were found to be between 0.81 and 0.88. Working in small datasets, lack of external validation, and not working in different populations are the limitations of our study.

Parameters	Sex	Mean $\pm$ std (cm)	p value
NNL	Male	2.46 $\pm$ 0.34	<b>&lt; 0.01</b>
	Female	2.30 $\pm$ 0.31	
NNZA	Male	120.68 $\pm$ 11.12	0.35
	Female	119.55 $\pm$ 9.92	
HGGA	Male	120.22 $\pm$ 6.90	0.92
	Female	120.30 $\pm$ 6.78	
RML	Male	5.29 $\pm$ 0.51	<b>&lt; 0.01</b>
	Female	4.54 $\pm$ 0.41	
CML	Male	7.34 $\pm$ 0.57	<b>&lt; 0.01</b>
	Female	7.06 $\pm$ 0.49	
GHGA	Male	37.05 $\pm$ 4.60	<b>&lt; 0.01</b>
	Female	38.99 $\pm$ 4.87	
HML	Male	8.24 $\pm$ 0.56	<b>&lt; 0.01</b>
	Female	7.55 $\pm$ 0.58	
COLI	Male	106.61 $\pm$ 6.98	0.06
	Female	105.07 $\pm$ 7.23	
CMHA	Male	110.68 $\pm$ 10.41	<b>&lt; 0.01</b>
	Female	114.82 $\pm$ 10.04	
HGGC	Male	12.56 $\pm$ 0.79	<b>&lt; 0.01</b>
	Female	11.44 $\pm$ 0.69	
COIC	Male	8.61 $\pm$ 0.73	<b>&lt; 0.01</b>
	Female	7.93 $\pm$ 0.67	
HGGMA	Male	31.15 $\pm$ 4.10	<b>&lt; 0.01</b>
	Female	25.57 $\pm$ 3.73	

**Table 1.** Comparison of parametric data of males and females. NNL: Nasion–nasal end point length, NNZA: Nasal end point–nasion–zygomatic angle, HGGA: Head of mandible–gonion–gnathion angle, RML: Ramus of the mandible length, CML: Corpus of the mandible length, GHGA: Gnathion–head of mandible–gonion angle, HML: Head of mandible–mental foramen length, COLI: Coronoid process–obliqua line–infradental angle, CMHA: Coronoid process–mandibular notch–head of mandible angle, HGGC: Head of mandible–gonion–gnathion curvature length, COIC: Coronoid process–obliqua line–infradental curvature length, HGGMA: Head of mandible–gonion–gnathion–mandibular notch area. Significant values are in bold.

Forensic anthropologists constantly try to improve skeletal identification methods by using various methods in various parts of the skeleton or by developing new methods to determine gender<sup>4</sup>. Pelvis and cranium are known as the most dimorphic skeletal parts and they form the basis of sex determination researches<sup>4,10,17–19</sup>. Bertsatos et al.<sup>19</sup> reported that they predicted sex with an Acc ratio of 0.71–0.90 in total according to the results of the discriminant function analysis they carried out with the parameters taken from the cranium. Franklin et al.<sup>28</sup> and Dayal et al.<sup>29</sup> reached Acc ratios of 0.88–0.90 and 0.80–0.85, respectively according to the results of the discriminant function analysis they carried out with the parameters taken from the cranium. In this study, 0.90 Acc, 0.80 Mcc, 0.90 Spe, 0.90 Sen, 0.90 F1 results were found as a result of LR algorithm. Since the ML results included Mcc value which can evaluate Acc, Spe, Sen values together and which shows the reliability of algorithm, it is thought that reliability and accuracy were tested with various methods and reliable results were found in the study<sup>12</sup>.

While discriminant function analysis is one of the most widely used methods in forensic and archaeological cases for the determination of sex in literature, it is known that error rates are always different from 0%<sup>2</sup>. The fact that the MLs used in the present study were trained as 80% training and 20% test set increases the prediction reliability of the study and makes it more advantageous when compared with discriminant analysis.

CT is preferred for providing advantage in the measurement of missing and damaged parts by making bone measurements very close to original and allowing for the reconstruction of each bone part, unlike conventional osteometry devices (calliper, odontometer, digital distance meter)<sup>16,22</sup>. As far as we know, studies that associate parameters taken from cranium and mandible on orthogonal plane with ML based sex prediction are very limited. Even if CT is used in current studies, the results can show differences because the orientation of the image is not converted to the orthogonal plane since especially angular measurements are parameters affected by orientation.

In their study they predicted sex from cranium by using CT, Gillet et al.<sup>30</sup> used geometric morphometric model in their study and reported that they reached 0.90 Acc ratio for skull model. Zaafrane et al.<sup>31</sup> reported that they estimated sex with an Acc ratio of 0.90 from parameters of cranium in CT images they analysed by using basic statistical methods. These differences in results can be explained with the fact that the evaluation of sexually dimorphic features depend on group specific standards and skeletal characteristics differ among different populations, as well as the methodological methods used and differences in statistical analyses.

Parameters	Sex	Median (min–max), (cm)	p value
NVIA	Male	76.65 (69.54–86.03)	0.89
	Female	76.66 (43.77–86.43)	
NVIC	Male	32.41 (29.04–40.50)	< <b>0.01</b>
	Female	30.84 (27.59–33.60)	
NSVC	Male	14.49 (12.20–18.14)	< <b>0.00</b>
	Female	13.49 (11.43–38.97)	
ZA	Male	77.86 (62.76–95.62)	0.21
	Female	77.22 (43.65–92.02)	
PC	Male	4.67 (3.15–5.53)	< <b>0.01</b>
	Female	4.14 (3.06–5.53)	
NIVA	Male	43.99 (14.18–73.93)	<b>0.03</b>
	Female	43.73 (37.23–51.16)	
PNIC	Male	39.47 (35.32–43.93)	< <b>0.01</b>
	Female	37.13 (29.01–41.44)	
VIC	Male	17.78 (14.02–22.53)	< <b>0.0</b>
	Female	17.28 (13.14–29.72)	
NFIA	Male	70.88 (45.18–81.06)	0.78
	Female	70.60 (60.79–80.64)	
NIC	Male	34.90 (31.12–38.12)	< <b>0.01</b>
	Female	32.59 (23.45–37.39)	
GA	Male	125.31 (18.59–143.71)	0.18
	Female	125.58 (111.16–148.95)	
COL	Male	3.11 (2.13–7.36)	< <b>0.01</b>
	Female	2.81 (1.94–3.85)	
HGGMC	Male	29.86 (17.62–33.83)	< <b>0.01</b>
	Female	27.37 (23.08–32.79)	

**Table 2.** Comparison of non-parametric data of males and females. NVIA: Nasion–vertex–inion angle, NVIC: Nasion–vertex–inion curvature length, NSVC: Nasion–superciliary arch–vertex curvature length, ZA: Zygomatic angle, PC: Piriform aperture curvature length, NIVA: Nasion–inion–vertex angle, PNIC: Piriform aperture–Nasal end point–inion curvature length, VIC: Vertex–inion curvature length, NFIA: Nasion–frontal tuber–inion angle, NIC: Nasal end point–inion curvature length, GA: Gonial angle, COL: Coronoid process–obliqua line length, HGGMC: Head of mandible–gonion–gnathion–mandibular notch curvature length. Significant values are in bold.

Imaizumi et al.<sup>32</sup> They used the support vector machine in their study in which they examined 100 skull skeletons and obtained a gender prediction rate of over 90% with 10% cross validation. In this study, we use image-based CNN, SVM, etc. We did not choose algorithms. The reason for this is due to the selection of only anthropometric points, not the entire cranium skeleton. Anthropometric points were measured manually using the Horos Project program and the results were used as ML algorithm input. Because image-based algorithms will produce a result by learning all the points of the given cranium skeleton.

It has been reported in literature that the possibility of removing the mandible intact is high<sup>33</sup>. The reason for this is the fact that the presence of a dense compact bone layer in the mandible makes it durable and therefore more likely to be found intact<sup>34</sup>. It is reported in literature that the measurements taken from the mandible are generally obtained from panoramic radiography images and that these images are affected by orientation<sup>35</sup>. According to the results of studies in which only the measurements taken from mandible are evaluated, an Acc ratio between 0.60 and 0.88 seems to be a reliable structure for sex prediction<sup>29,35–37</sup>. In this study, combining the parameters taken from the mandible with the cranium strengthened gender prediction. RML, CML, GCGA, CFL, PLL, PICA, CGC, PLIC, CGGIC, CGGIA parameters taken from the mandible were found to be statistically significant in sex identification.

Since the identity of individuals should be predicted quickly and accurately in events such as war, natural disasters and fire, which deeply affect the society, the CT technology and MLs used in the present study show that prediction time can be minimized and high accuracy can be obtained. Considering the high Acc ratio found as a result of LR algorithm, it is thought that the present study will strengthen and contribute to studies related with sex prediction.

## Materials and methods

**Image set and population.** The study was conducted at Karabük University Training and Research Hospital, Department of Radiology after 2020/363 coded approval of Karabük University Faculty of Medicine non-interventional clinical research ethics committee.

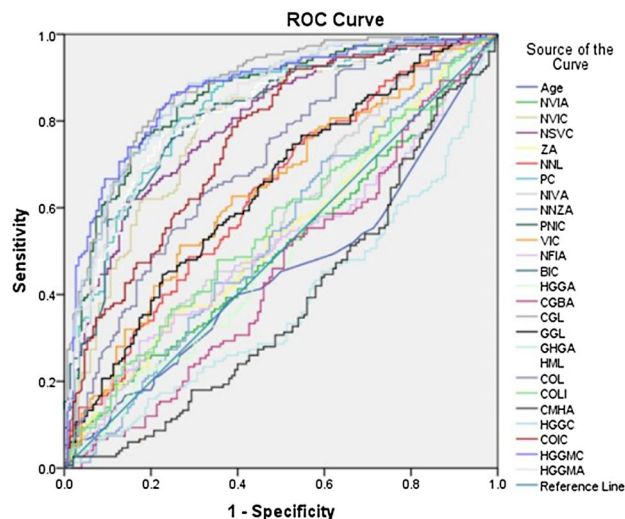
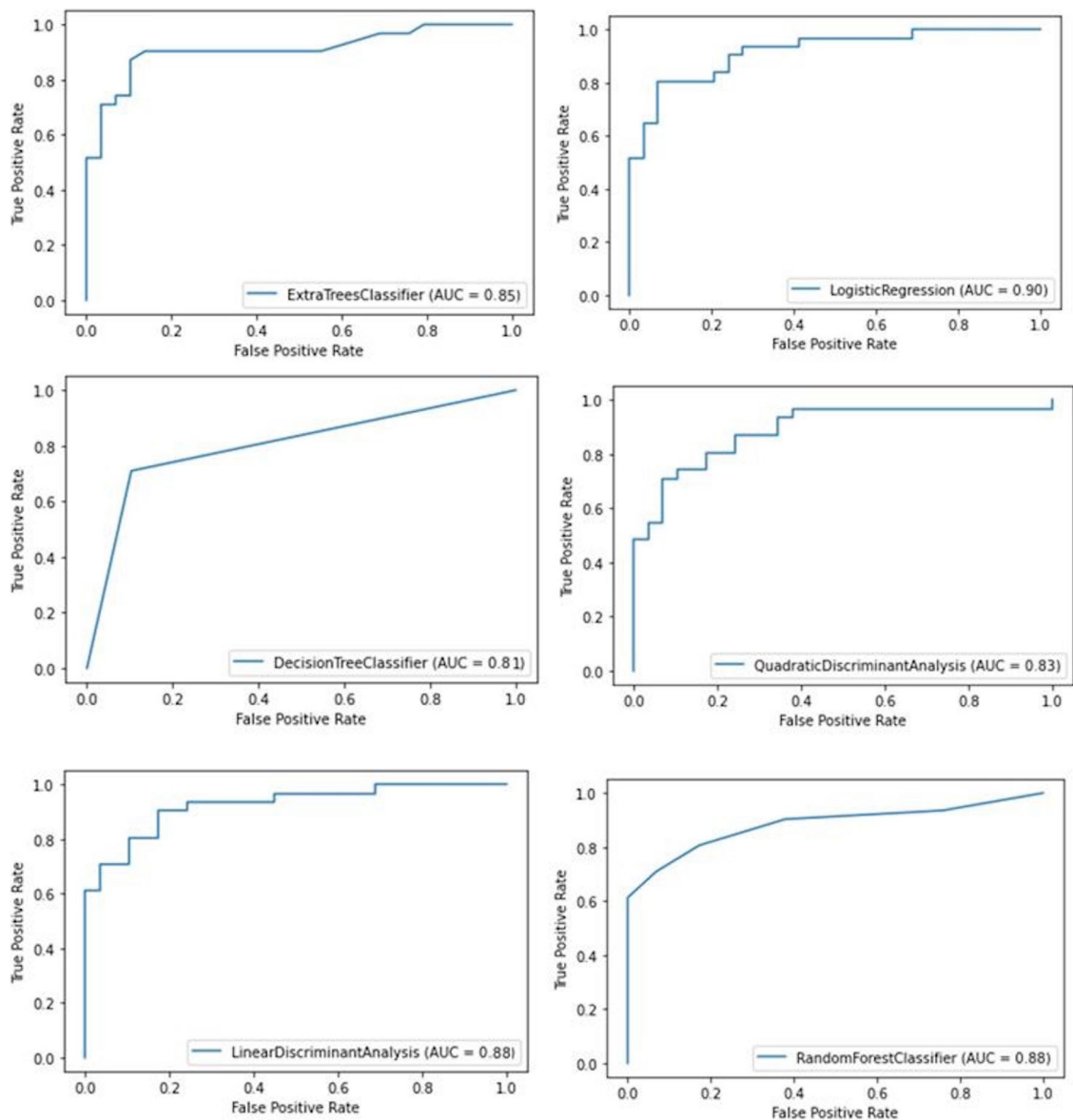


Figure 1. ROC curve.

Parameters	AUC (%95 CI)	Cutt-off	p	Sen	Spe
Age	0.451 (0.385–0.516)	59.5	0.141	0.467	0.54
NVIA	0.504 (0.439–0.570)	76.7	0.896	0.5	0.5
NVIC	0.791 (0.741–0.841)	31.56	0.000	0.72	0.28
NSVC	0.794 (0.744–0.845)	13.99	0.000	0.713	0.287
ZA	0.541 (0.476–0.606)	77.55	0.219	0.533	0.467
NNL	0.632 (0.570–0.695)	2.38	0.000	0.587	0.413
PC	0.823 (0.775–0.871)	4.36	0.000	0.753	0.247
NIVA	0.569 (0.504–0.634)	43.79	0.038	0.513	0.487
NNZA	0.536 (0.470–0.601)	120.77	0.286	0.527	0.473
PNIC	0.854 (0.810–0.897)	38.35	0.000	0.78	0.22
VIC	0.639 (0.576–0.701)	17.55	0.000	0.613	0.387
NFIA	0.509 (0.443–0.575)	70.71	0.784	0.513	0.487
BIC	0.812 (0.763–0.860)	33.59	0.000	0.753	0.247
HGGA	0.501 (0.435–0.566)	120.07	0.987	0.48	0.52
CGBA	0.456 (0.391–0.521)	125.55	0.189	0.493	0.507
CGL	0.877 (0.838–0.915)	4.86	0.000	0.8	0.2
GGL	0.642 (0.580–0.704)	7.18	0.000	0.587	0.413
GHGA	0.378 (0.315–0.442)	38.34	0.000	0.427	0.573
HML	0.815 (0.767–0.864)	7.91	0.000	0.767	0.233
COL	0.706 (0.648–0.764)	2.95	0.000	0.648	0.764
COLI	0.557 (0.492–0.622)	105.97	0.089	0.533	0.467
CMHA	0.383 (0.370–0.447)	112.88	0.000	0.42	0.58
HGGC	0.863 (0.822–0.904)	12.03	0.000	0.787	0.22
COIC	0.754 (0.700–0.808)	8.25	0.000	0.667	0.333
HGGMC	0.869 (0.827–0.910)	28.60	0.000	0.787	0.213
HGGMA	0.841 (0.797–0.884)	28.28	0.000	0.767	0.233

Table 3. ROC result table.

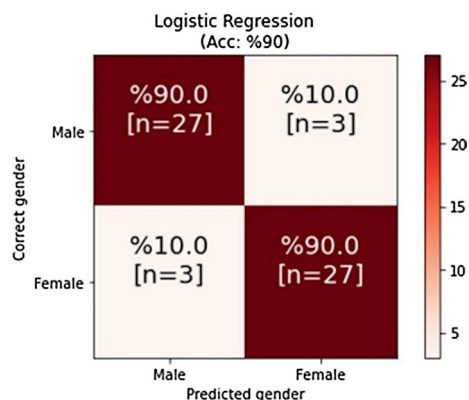


**Figure 2.** ML ROC curve.

The image set in the study consisted of the CT images of 150 male and 150 female individuals whose ages ranged between 20 and 65. Individuals with any surgical operation or pathology of the cranium skeleton were excluded from the study. Average age of the males was 54 (min 20, max 65), while average age of the females was (min 21, max 65). No statistically significant difference was found between the average ages of males and females ( $p = 0.395$ ).

**Multidetector CT (MDCT) protocol.** Radiological images used in the study were obtained from CT images with a section thickness of 5 mm taken in supine position by using a 16-row MDCT scanner (Aquilion 16; Toshiba Medical Systems, Otawara, Japan) in the department of radiology of a Karabük University Training and Research Hospital. Scanning protocol values were tube voltage: 120 kV, gantry rotation: 0.75 s and pitch: 1.0 mm.

**Image analysis.** The images obtained were transferred to Horos Medical Image Viewer (Version 3.0, USA) program, which is a personal workstation in Digital Imaging and Communications in Medicine (DICOM) format. Images in sagittal, transversal and coronal planes were obtained from the transferred images by using 3D Curved Multiplanar Reconstruction (MPR). The line passing through the nasion and inion points of the images in these three planes was determined and all images were brought to the orthogonal plane (Fig. 5A). Later, CT images brought to orthogonal plane were overlapped by increasing their section thicknesses (Fig. 5B).



**Figure 3.** LR confusion matrix.

Testing set	LR	LDA	QDA	RF	ETC	DT
1	87	87.33	85.33	85.33	75.33	78.67
2	86.33	87.67	86.67	85	75.67	81
3	86	87	86.33	86.67	74	83.67
4	87.67	88.33	87.33	85.33	78	80.67
5	86.33	87.67	86.67	87.67	80.33	80.33
6	85.33	88	86	84.33	78	79.33
7	87.67	88	86.67	85.33	77.33	80
8	86.33	87.33	87.33	87	78.67	80.33
9	87.67	88	86.67	86	75.67	79.33
10	87.33	88	86.33	85	79	79.33
Mean	87.766	87.733	86.533	85.766	77.200	80.266
Std	0.819	0.410	0.592	1.045	1.970	1.396

**Table 4.** Tenfold cross validation results (%Acc).

Length, angle, area and curvature length measurements of the anatomic points of the overlapped images were performed. These parameters and their abbreviations are listed below in Tables 5, 6 and 7. Demonstration of all evaluated parameters is shown in Fig. 6.

**Machine learning algorithms.** In this study, scikit-learn model (Version 0.20.0) in Python programming language (Version 3.7.1) was used to make ML modelling<sup>38</sup>. ML modelling was performed by using i7, 8 GbHp-Folio 1040 model computer. Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Extra Tree Classifier (ETC) algorithms were used. The dataset was mixed by shuffling, and the first 80% (240 measurements) was designated as the training set, while the last 20% (60 measurements) was designated as the test set. In addition, tenfold cross validation accuracy values are also included in terms of the reliability of our study.

**Performance criteria.** Accuracy (Acc), Specificity (Spe), Sensitivity (Sen), F1 score (F1), and Matthews correlation coefficient (Mcc) values were included as performance criteria.

$$\begin{aligned}
 \text{Acc} &= \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \\
 \text{Sen} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\
 \text{Spe} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \\
 \text{Mcc} &= \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \\
 \text{F1} &= 2 \frac{\text{Specificity} \times \text{Sensitivity}}{\text{Specificity} + \text{Sensitivity}}
 \end{aligned}
 \tag{1}$$

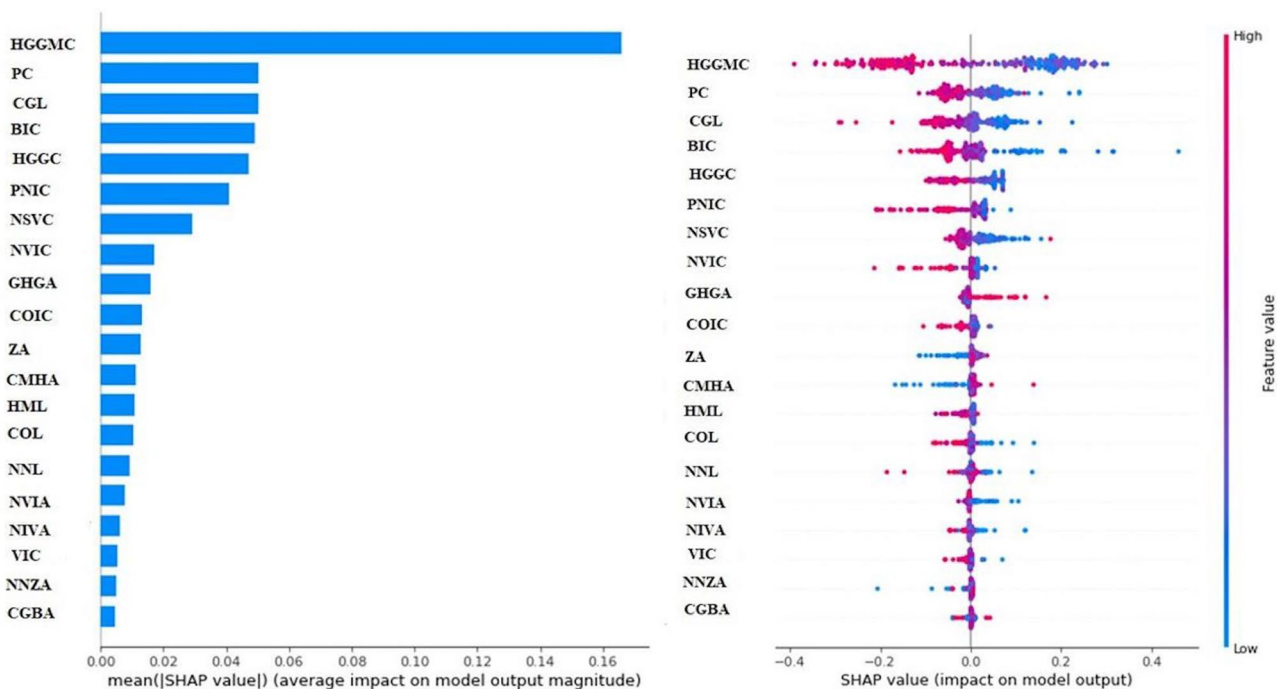


Figure 4. RF algorithm SHAP explanatory image.

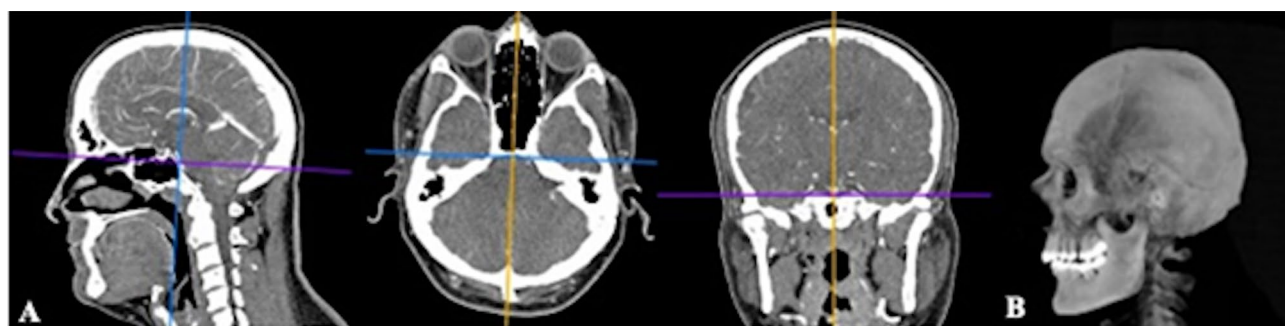


Figure 5. (A) Sagittal, transversal and coronal images brought to orthogonal plane, (B) Overlapped image.

Parameters	Abbreviations
Ramus of the mandible length	RML
Corpus of the mandible length	CML
Head of mandible–mental foramen length	HML
Coronoid process–obliqua line length	COL
Nasion–nasal end point length	NNL

Table 5. Length parameters and abbreviations.

TP: True positive, TN: True negative, FP: False positive, FN; False negative.

**Statistical analysis.** Mean, standard deviation, minimum and maximum values were included in the descriptive statistics of each data according to gender groups. Normality test Anderson Darling test was applied to each parameter and it was checked whether the data were normally distributed. Two simple T test was applied to parametric data and Mann–Whitney U test was applied to nonparametric data and  $p \leq 0.05$  value was considered as statistically significant. In order to reveal the differences of the parameters in terms of gender, ROC analysis was performed and the ROC curve was included. Minitab 17 and IBM SPSS (Version 21) package program was used in analyses.

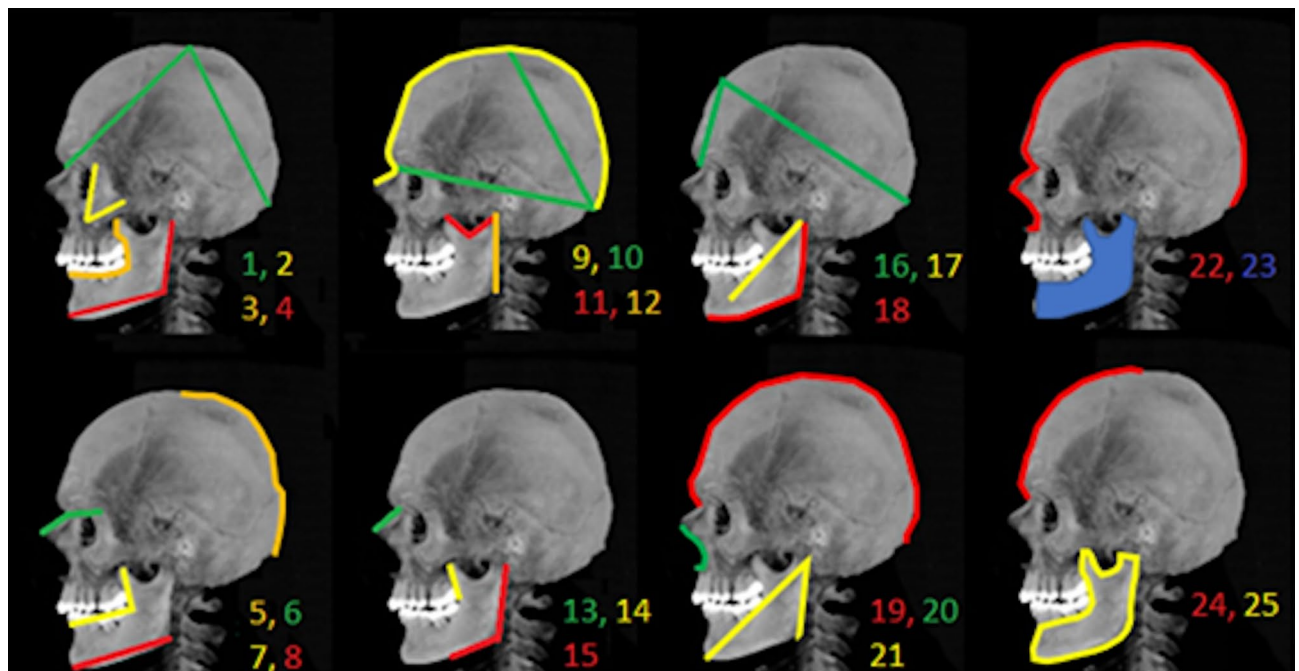


Parameters	Abbreviations
Head of mandible–gonion–gnathion angle	HGGA
Gonial angle	GA
Gnathion–head of mandible–gonion angle	GHGA
Coronoid process–obliqua line–infradental angle	COLI
Coronoid process–mandibular notch–head of mandible angle	CMHA
Nasion–vertex–inion angle	NVIA
Zygomatic angle	ZA
Nasion–inion–vertex angle	NIVA
Nasal end point–nasion–zygomatic angle	NNZA
Nasion–frontal tuber–inion angle	NFIA

**Table 6.** Angle parameters and abbreviations.

Parameters	Abbreviations
Head of mandible–gonion–gnathion curvature length	HGGC
Coronoid process–obliqua line–infradental curvature length	COIC
Head of mandible–gonion–gnathion–mandibular notch curvature length	HGGMC
Nasion–vertex–inion curvature length	NVIC
Nasion–superciliary arch–vertex curvature length	NSVC
Piriform aperture curvature length	PC
Piriform aperture–nasal end point–inion curvature length	PNIC
Vertex–inion curvature length	VIC
Nasal end point–inion curvature length	BIC
Head of mandible–gonion–gnathion–mandibular notch area	HGGMA

**Table 7.** Curve length–area parameters and abbreviations.



**Figure 6.** Demonstration of parameters (1: NVIA, 2: ZA, 3: COIC, 4: HGGA, 5: VIC, 6: NNZA, 7: COLI, 8: CML, 9: NIC, 10: NIVA, 11: CMHA, 12: RML, 13: NNL, 14: COL, 15: GA, 16: NFIA, 17: HML, 18: HGGC, 19: NVIC, 20: PC, 21: GCGA, 22: PNIC, 23: HGGMA, 24: NSVC, 25: HGGMC).

**Ethical considerations.** This retrospective study was initiated with the 2020/363 decision of the Karabük University Faculty of Medicine non-interventional clinical research ethics committee.

**Ethical approval.** The present study was approved by Karabük University Faculty of Medicine Local Non-Interventional Clinical Trials Ethics Committee with the protocol number 2020/363. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent.** This study is retrospective and based on images taken from the hospital archive system. Therefore, the requirement for informed consent for the study was waived by the Karabük University Faculty of Medicine Local Non-Interventional Clinical Trials Ethics Committee.

**Conference presentation.** This study was presented as an oral presentation at the 21st National Anatomy Congress in Turkey.

Received: 23 August 2021; Accepted: 10 February 2022

Published online: 11 March 2022

## References

- Nikita, E. & Nikitas, P. On the use of machine learning algorithms in forensic anthropology. *Legal Med.* **47**, 101771 (2020).
- du Jardin, P., Ponsaillé, J., Alunni-Perret, V. & Quatrehomme, G. A comparison between neural network and other metric methods to determine sex from the upper femur in a modern French population. *Forensic Sci. Int.* **192**, 127.e121–127.e126 (2009).
- Darmawan, M., Yusuf, S. M., Kadir, M. A. & Haron, H. Comparison on three classification techniques for sex estimation from the bone length of Asian children below 19 years old: An analysis using different group of ages. *Forensic Sci. Int.* **247**, 130.e131–130.e111 (2015).
- Steyn, M. & İşcan, M. Y. Metric sex determination from the pelvis in modern Greeks. *Forensic Sci. Int.* **179**, 86.e81–86.e86 (2008).
- Plavcan, J. M. Sexual dimorphism in primate evolution. *Am. J. Phys. Anthropol.* **116**, 25–53 (2001).
- Dawson, C., Ross, D. & Mallett, X. *Forensic Anthropology: 2000 to 2010* (CRC Press, 2011).
- Mahfouz, M. *et al.* Patella sex determination by 3D statistical shape models and nonlinear classifiers. *Forensic Sci. Int.* **173**, 161–170 (2007).
- Serkan, Ö., Turan, M. & Zülal, Ö. Estimation of gender by using decision tree, a machine learning algorithm, with patellar measurements obtained from MDCT images. *Med. Rec.* **3**, 1–9 (2021).
- Lin, C. *et al.* Sex determination from the mandibular ramus flexure of Koreans by discrimination function analysis using three-dimensional mandible models. *Forensic Sci. Int.* **236**, 191.e191–191.e196 (2014).
- Kim, D.-I., Kim, Y.-S., Lee, U.-Y. & Han, S.-H. Sex determination from calcaneus in Korean using discriminant analysis. *Forensic Sci. Int.* **228**, 177.e171–177.e177 (2013).
- Akhlaghi, M. *et al.* Sex determination based on radiographic examination of metatarsal bones in Iranian population. *Int. J. Med. Toxicol. Forensic Med.* **7**, 203–208 (2017).
- Turan, M. K., Oner, Z., Secgin, Y. & Oner, S. 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 (2019).
- Gapert, R., Black, S. & Last, J. Sex determination from the occipital condyle: Discriminant function analysis in an eighteenth and nineteenth century British sample. *Am. J. Phys. Anthropol.* **138**, 384–394 (2009).
- El Morsi, D. A. & Al Hawary, A. A. Sex determination by the length of metacarpals and phalanges: X-ray study on Egyptian population. *J. Forensic Leg. Med.* **20**, 6–13 (2013).
- Ozsoy, T., Oner, Z. & Oner, S. An attempt to gender determine with phalanx length and the ratio of phalanxes to whole phalanx length in direct hand radiography. *Medicine* **8**, 692–697 (2019).
- Oner, Z., Turan, M. K., Oner, S., Secgin, Y. & Sahin, B. Sex estimation using sternum part lengths by means of artificial neural networks. *Forensic Sci. Int.* **301**, 6–11 (2019).
- Franklin, D., O'Higgins, P., Oxnard, C. E. & Dadour, I. Determination of sex in South African blacks by discriminant function analysis of mandibular linear dimensions. *Forensic Sci. Med. Pathol.* **2**, 263–268 (2006).
- Fawzy, I. A. & Kamal, N. N. Stature and body weight estimation from various footprint measurements among Egyptian population. *J. Forensic Sci.* **55**, 884–888 (2010).
- Bertsatos, A., Chovalopoulou, M.-E., Brůžek, J. & Bejdová, Š. Advanced procedures for skull sex estimation using sexually dimorphic morphometric features. *Int. J. Leg. Med.* **134**, 1927–1937 (2020).
- Krems, R. Bayesian machine learning for quantum molecular dynamics. *Phys. Chem. Chem. Phys.* **21**, 13392–13410 (2019).
- Erickson, B. J., Korfiatis, P., Akkus, Z. & Kline, T. L. Machine learning for medical imaging. *Radiographics* **37**, 505–515. <https://doi.org/10.1148/rg.2017160130> (2017).
- Secgin, Y., Oner, Z., Turan, M. K. & Oner, S. Gender prediction with parameters obtained from pelvis computed tomography images and decision tree algorithm. *Med. Sci. Int. Med. J.* **10**, 356–361 (2021).
- Deo, R. C. Machine learning in medicine. *Circulation* **132**, 1920–1930. <https://doi.org/10.1161/CIRCULATIONAHA.115.001593> (2015).
- Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F. & Campbell, J. P. Introduction to machine learning, neural networks, and deep learning. *Transl. Vis. Sci. Technol.* **9**, 14. <https://doi.org/10.1167/tvst.9.2.14> (2020).
- Mishra, G., Sehgal, D. & Valadi, J. K. Quantitative structure activity relationship study of the anti-hepatitis peptides employing random forests and extra-trees regressors. *Bioinformatics* **13**, 60–62. <https://doi.org/10.6026/97320630013060> (2017).
- Curate, F. *et al.* Sex determination from the femur in Portuguese populations with classical and machine-learning classifiers. *J. Forensic Leg. Med.* **52**, 75–81. <https://doi.org/10.1016/j.jflm.2017.08.011> (2017).
- Chikara, R. K. & Ko, L. W. Prediction of human inhibition brain function with inter-subject and intra-subject variability. *Brain Sci.* **10**, 726. <https://doi.org/10.3390/brainsci10100726> (2020).
- Franklin, D., Cardini, A., Flavel, A. & Kuliukas, A. Estimation of sex from cranial measurements in a Western Australian population. *Forensic Sci. Int.* **229**(158), e151–158. <https://doi.org/10.1016/j.forsciint.2013.03.005> (2013).
- Dayal, M. R., Spocter, M. A. & Bidmos, M. A. An assessment of sex using the skull of black South Africans by discriminant function analysis. *Homo* **59**, 209–221 (2008).
- Gillet, C. *et al.* Sex estimation in the cranium and mandible: A multislice computed tomography (MSCT) study using anthropometric and geometric morphometry methods. *Int. J. Leg. Med.* **134**, 823–832 (2020).

31. Zaafrane, M. *et al.* Sex determination of a Tunisian population by CT scan analysis of the skull. *Int. J. Leg. Med.* **132**, 853–862 (2018).
32. Imaizumi, K. *et al.* Development of a sex estimation method for skulls using machine learning on three-dimensional shapes of skulls and skull parts. *Forensic Imaging* **22**, 200393 (2020).
33. Franklin, D., O'Higgins, P., Oxnard, C. E. & Dadour, I. Discriminant function sexing of the mandible of indigenous South Africans. *Forensic Sci. Int.* **179**, 84.e81–84.e85 (2008).
34. Saini, V. *et al.* Mandibular ramus: An indicator for sex in fragmentary mandible. *J. Forensic Sci.* **56**, S13–S16 (2011).
35. Indira, A. P., Markande, A. & David, M. P. Mandibular ramus: An indicator for sex determination—A digital radiographic study. *J. Forensic Dent. Sci.* **4**, 58 (2012).
36. Steyn, M. & İşcan, M. Y. Sexual dimorphism in the crania and mandibles of South African whites. *Forensic Sci. Int.* **98**, 9–16 (1998).
37. Marinescu, M., Panaitescu, V. & Rosu, M. Sex determination in Romanian mandible using discriminant function analysis: Comparative results of a time-efficient method. *Rom. J. Leg. Med.* **21**, 305–308 (2013).
38. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

## Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Author contributions

S.T., Y.S. Z.O., M.K.T., S.O., and D.S. conceived and designed this research work. M.K.T., S.O., and Y.S. image collection and data analysis. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to S.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**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/>.

© The Author(s) 2022