Sex determination from lateral cephalometric radiographs using an automated deep learning convolutional neural network

Maryam Khazaei[ID][1], Vahid Mollabashi[ID][2], Hassan Khotanlou[ID][3], Maryam Farhadian[ID][4,*]

[1]*Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran*
[2]*Department of Orthodontics, Faculty of Dentistry, Dental Research Center, Hamadan University of Medical Sciences, Hamadan, Iran*
[3]*Department of Computer Engineering, Bu-Ali Sina University, Hamadan, Iran*
[4]*Department of Biostatistics, School of Public Health, Research Center for Health Sciences, Hamadan University of Medical Sciences, Hamadan, Iran*

ABSTRACT

Purpose: Despite the proliferation of numerous morphometric and anthropometric methods for sex identification based on linear, angular, and regional measurements of various parts of the body, these methods are subject to error due to the observer's knowledge and expertise. This study aimed to explore the possibility of automated sex determination using convolutional neural networks (CNNs) based on lateral cephalometric radiographs.

Materials and Methods: Lateral cephalometric radiographs of 1,476 Iranian subjects (794 women and 682 men) from 18 to 49 years of age were included. Lateral cephalometric radiographs were considered as a network input and output layer including 2 classes (male and female). Eighty percent of the data was used as a training set and the rest as a test set. Hyperparameter tuning of each network was done after preprocessing and data augmentation steps. The predictive performance of different architectures (DenseNet, ResNet, and VGG) was evaluated based on their accuracy in test sets.

Results: The CNN based on the DenseNet121 architecture, with an overall accuracy of 90%, had the best predictive power in sex determination. The prediction accuracy of this model was almost equal for men and women. Furthermore, with all architectures, the use of transfer learning improved predictive performance.

Conclusion: The results confirmed that a CNN could predict a person's sex with high accuracy. This prediction was independent of human bias because feature extraction was done automatically. However, for more accurate sex determination on a wider scale, further studies with larger sample sizes are desirable. *(Imaging Sci Dent 2022; 52: 239-44)*

KEY WORDS: Sex Determination Analysis; Deep Learning; Radiography; Cephalometry; Cervical Vertebrae

Introduction

Sex determination is an important topic in various fields, from anthropology and archeology to forensics. Sex determination is the first step in the identification of skeletal remains, as it reduces the population under analysis by half.

After death, with the loss of soft tissue in the body, skeletal remains are available for research and investigation. As the most common part of the human skeleton, the pelvis has been extensively used in research to determine sex. However, the weakness and fragility of pelvic bones, the complex shape of the pelvis, and the separation of its components often hinder the availability of pelvic bones for examinations. The skull contains useful information, making it an acceptable reference for sex determination other than the pelvis. The skull, which has a rigid structure, is the most important preserved part of a skeleton, and it is the only component available for forensic research in many situations.[1-3]

Despite the proliferation of numerous morphometric and anthropometric methods for sex determination based on linear, angular, and regional measurements of various parts of the body, especially the skull, these methods are subject to error due to the knowledge and expertise of the observer. These methods are also often very complex and time-consuming.[4-7]

Lateral cephalometric radiographs, in contrast, have long been used as an affordable, reliable, and cost-effective tool to help diagnose skeletal and dental patterns in orthodontics. Measurements based on lateral cephalograms have been of interest to researchers, since they provide detailed information about cranial morphology, as an efficient method of sex determination.[8-10]

In recent years, traditional machine learning techniques based on manual feature engineering from various parts of the body in various fields, including forensics, have shown very acceptable performance in identifying victims. A limitation of these approaches is that unknown and potentially sex-related features may not be properly identified.[11,12]

More recently, significant progress has been made in the use of deep learning and convolutional neural networks (CNNs), as branches of artificial intelligence, for image processing. In medical imaging, the utilization of deep learning with CNNs to process various types of images has been actively researched, with promising performance. CNNs, in which several layers are trained, are one of the most important and powerful deep learning methods. This method is very efficient and is one of the most common methods in the field of computer vision. A CNN is a kind of feed-forward artificial neural network created to process and classify multidimensional data such as images.[13,14]

Deep learning algorithms, especially CNNs, are rapidly becoming an efficient method for analyzing medical images. The advantage of CNNs is that they automatically discover the features of an image that are particularly useful for the classification task. Deep learning eliminates the need to determine these features manually and independently extract features from the image that are useful for determining the response class (in this context, sex), and by assigning more weight to more important features, deep learning models also automatically exclude parts of the image that are not useful. On this basis, they perform the act of classification (sex determination).[15-17]

Since limited studies have explored this issue, the present study aimed to explore the possibility of automated sex determination using CNNs based on lateral cephalometric radiographs.

Materials and Methods

Data set

In this study, 1476 lateral cephalometric radiographs in the archives of the orthodontics department of the School of Dentistry of Hamadan University of Medical Sciences were used. The images belonged to patients aged 18 to 49 years (794 women and 682 men) who were referred for orthodontic treatment. The mean age of the subjects was 23.15 ± 5.61 years. The population originated from western Iran. Cephalometric radiographs were taken using a Scara2 device (Planmeca, Helsinki, Finland). The Department of Orthodontics indicated that all the images were derived in a standard position with teeth in centric occlusion and lips relaxed. The images were downloaded in PNG format. A sample of lateral cephalometric radiographs is presented in Figure 1. These patients also had no relevant symptoms, skeletal abnormalities on their panoramic radiographs, or history of previous orthognathic surgery or trauma.

The images were used without mentioning the names and details of the subjects and kept confidential by the researchers. The present study was approved by the ethics committee of Hamadan University of Medical Sciences with the code IR.UMSHA.REC.1398.1048.

Fig. 1. A sample of a lateral cephalometric radiograph that was considered as the network input.

CNN

CNNs, like "vanilla" neural networks, are made up of neurons, layers, and weights. In addition to the input and output layers, a CNN consists of 3 main layers: the convolution layer, the pooling layers, and the fully connected layer. Each of these layers performs different tasks. The convolution layer uses filters that perform convolution operations on the input image to become an abstract feature map. The pooling layers reduce data dimensions by combining the output of neuron clusters in one layer into a single neuron in the next layer. There are 2 special kinds of pooling: max and average. Max pooling uses the maximum value of each local neuron cluster in the feature map, while average pooling takes the average value. In the fully connected layer, each node in the output layer connects directly to a node in the previous layer. This layer performs the task of classification based on the features extracted through the previous layers and their different filters. While convolutional and pooling layers tend to use ReLu functions, fully-connected layers usually apply a softmax activation function to classify inputs appropriately and then produce a probability from 0 to 1.

Training CNNs is similar to a feed-forward network including 2 stages: feed-forward and back-propagation. After defining a loss function such as cross-entropy, the kernels of the model are fit using a stochastic gradient descent algorithm by using a back-propagation algorithm to compute gradients. Hyperparameters, such as the number of layers, kernel dimensions, or pooling layer sizes, are typically tuned by cross-validation.[15-17] Figure 2 shows a schematic view of different layers of a CNN. The classification of images in the CNN is such that, in the training stage, the image matrix is fed layer by layer to the network, and going through all the layers, a probability is finally reported as the output. After the image enters the network, the inner multi-plication operation is performed between the input and the parameters of each neuron, and finally, after applying the convolution operation in each output layer of the network, it is calculated as the probability of belonging to each sex class. This output is compared with the actual label of each image (sex in the individual's medical record), and in the back-propagation stage, the weight of each layer is updated according to the error of this step. This process is repeated to achieve acceptable final accuracy.[18,19]

Transfer learning is a popular approach in deep learning where knowledge and experience gained in solving a certain problem are utilized to analyze and tackle other related problems. In this regard, a pre-trained network for a given task is used as the starting point model for a second task with different data sets. The advantages of transfer learning include a reduced training time, improved neural network performance, and the absence of a requirement for a large amount of data. Since the model is already pre-trained, a good classification model with relatively little training data can be achieved using transfer learning.

One method of transfer learning is to use an existing pre-trained model. Many of these models are currently available. In the present study, the DenseNet, ResNet, and VGG architectures were used.

Network training

Lateral cephalometric radiographs were considered as the network input, and the output layer included 2 classes (male and female). The networks used in this study accepted images with fixed sizes. Therefore, before entering the images into the network, the size of all images became the same (200×200). Standardization was also performed for the images, for which each pixel with a value between 0 and 255 was divided by the largest possible number so that all pixels had values between 0 and 1. This transformation
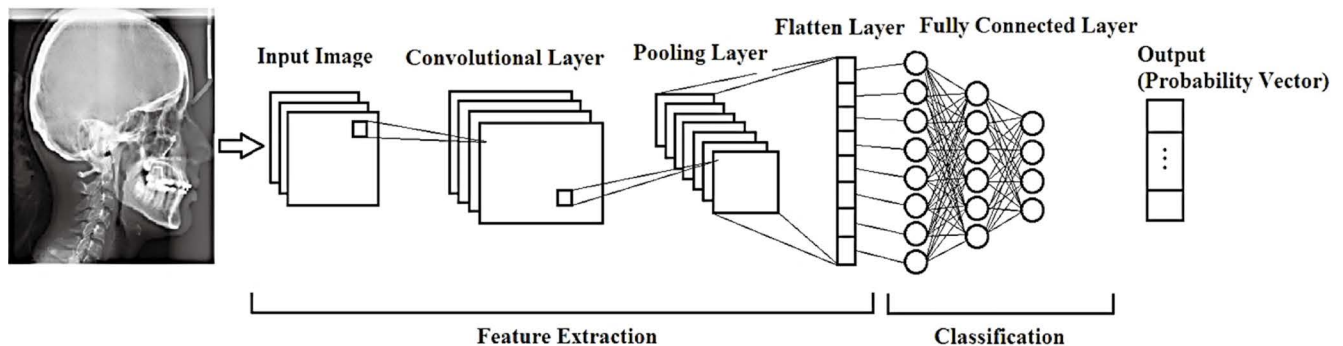

Fig. 2. A schematic representation of different layers of convolutional neural networks used in the present study for sex determination based on lateral cephalometric radiographs.

usually results in easier calculations and faster convergence. Augmentation was performed before network training to increase network generalizability and prevent overfitting. For this purpose, the images were augmented with a rotation rate of 20°, magnification of ×0.3, and horizontal and vertical translation of 0.1 and 0.3, respectively.

Images were also given to the grid with a batch size of 32. The binary cross-entropy function was used as cost function with a learning rate of 0.001. The convolution layer had a filter size of 3×3 and the rectified linear activation function (ReLU) was used for activation. The Adam optimization algorithm was used in the model. The dropout layer was used with a value of 0.5. A dense layer was also used with a sigmoid activation function. By plotting the accuracy against epochs, the optimal number of iterations was determined (epoch size=30). By plotting the cost versus learning rate, as one of the parameters of the Adam algorithm, which indicates the speed of updating network weights, the optimal value of this parameter was also determined. An accuracy against epoch plot was also used to choose the best number of epochs and check the overfitting problem. As an example, sample plots related to accuracy against epoch, and learning rate against loss are presented in Figure 3. These plots show that the accuracy of the training set slightly differed from the accuracy of the validation set, which suggests the possibility of overfitting the data. Regularization is therefore necessary to avoid overfitting. The plot (Fig. 3B) shows that a learning rate of 10^{-6} was a good value to train the network.

These diagrams were drawn during the training process of each model, and the desired parameters were adjusted by examining them. In this way, the model with the best predictive performance for each pre-trained network on a training dataset was selected.

Model evaluation

In total, 1476 images were divided into training and test sets, of which 80% (1180 images) were in the training set and 20% (296 images) in the test set. Different prediction models based on the different architectures (DenseNet, ReseNet and VGG) were trained. Since the output of each prediction model is binary (predicted sex: male, female), the sigmoid activation function was used. In each model, the optimal parameters were selected. This step was completed by selecting the best prediction model with the highest accuracy.

To compare the predictive power of different models, a confusion matrix was drawn that depicted the actual sex (target) of the individual versus the predicted sex (output) based on the prediction models. The accuracy index was calculated by dividing the total number of cases whose sex was correctly predicted by the total number of individuals for the test data set.

Coding for network implementation was performed using the Python programming language using the TensorFlow library. Also, a graphics card with GTX 1070 TI specifications was used.

Results

The prediction accuracy using different CNN architectures is presented in Table 1. In all 3 architectures, the use
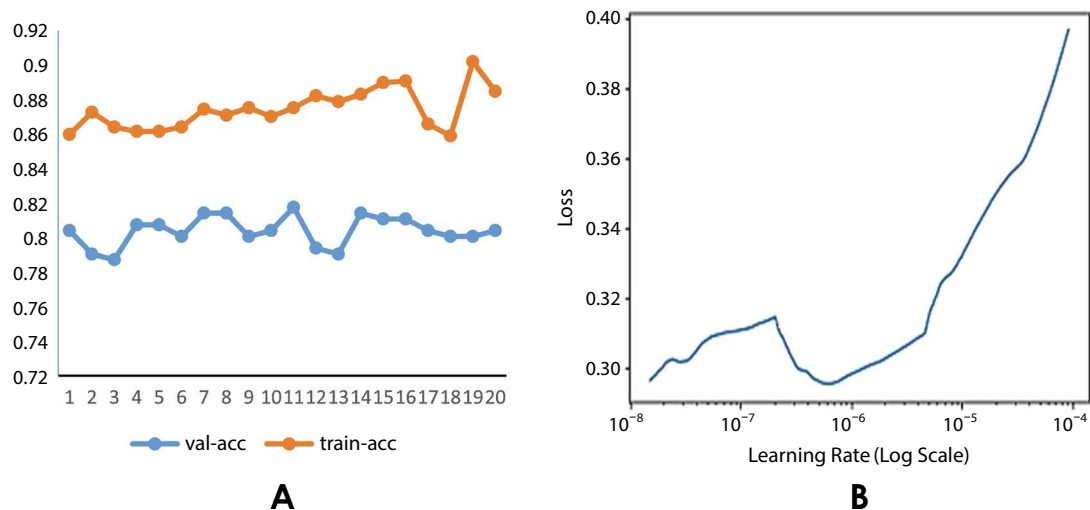

Fig. 3. A. A sample plot of accuracy versus the number of epochs for evaluating model performance. B. A sample plot of loss versus learning rate for determining the optimum value of the learning rate as a parameter in the Adam optimizer algorithm.

Table 1. Results of sex determination by different convolutional network architectures for a test data set

Architectures	Transfer learning	Accuracy
DenseNet121	No	0.76
	Yes	0.90
ResNet101	No	0.59
	Yes	0.62
VGG16	No	0.70
	Yes	0.75

Table 2. Confusion matrix for sex determination based on DenseNet121

	Predicted			
True	Female	Male	Total	Accuracy
Female	144	14	158	0.90
Male	15	123	138	
Total	159	137	296	

of transfer learning performed better than the non-use of transfer learning. The results showed that the network with the DenseNet121 architecture had the highest accuracy in sex prediction based on lateral cephalometric radiographs. In contrast, the worst performance was shown by the network with ResNet101 architecture, both with (accuracy: 62%) and without (accuracy: 59%) transfer learning.

The model was able to correctly identify 144 out of 158 images in the test set belonging to women, and 123 out of 138 images in the test data set for men. The overall accuracy for this model was 90% (90% for women and 89% for men). Table 2 presents a confusion matrix for the best model based on the test set.

Discussion

The results confirmed that CNNs, as a highly automated method, can determine sex from lateral cephalometric radiographs. The only input for this type of network was cephalometric images, and the network performed sex determination with high accuracy without any instructions or pre-existing knowledge. These prediction models were independent of human bias because feature extraction in the network was done automatically. It is important to note that the networks used in this study did not have any pre-defined knowledge or information regarding parts of the skull; instead, they automatically extracted and predicted features of the skull that were useful in sex determination.

Various deep network architectures have been proposed for image classification. In this study, which evaluated the performance of DenseNet, ResNet, and VGG architectures, the results showed that the network based on the DenseNet architecture had better predictive performance.

In this study, the CNNs analyzed 2-dimensional images. If this operation is upgraded to 3-dimensional images of the skull, more features could be explored. This is expected to be a focus of future research, which could also design networks that are capable of receiving a variety of images, including images taken through mobile phone cameras.

A limitation of CNNs is that they require a large number of images for training. This problem was largely solved in the present study using the transfer learning approach.

Numerous studies in the field of sex determination based on morphological and morphometric methods based on lateral cephalometric images in different populations and races have shown acceptable performance of this type of method. Part of the difference in the accuracy of the prediction models designed in these studies can be attributed to the different features and measurements that are considered. However, the findings of these studies are prone to error due to human judgment and the reproducibility of results. The following studies can be mentioned.[1-4]

In a study conducted by Chezhian and Dharman[9] with the aim of determining sex by lateral cephalograms, 120 images from 60 women and 60 men, ranging in age from 25 to 55 years, were examined. Linear measurements such as G-OP and BA-ANS were compared in men and women. Based on the results of this study, a predictive model using linear measurements obtained from lateral cephalometric radiography was introduced as a useful technique for determining sex. The accuracy of the discriminant model based on the G-OP index was 69.8%.

Qaq et al.[10] evaluated 135 images from individuals in the age range of 18 to 26 years were evaluated. According to linear, angular, and regional measurements, 21 parameters were obtained and used as input in the logistic regression model to determine sex. The sex prediction accuracy of the logistic regression model was 82.4%.

In a study conducted by Binnal et al.,[20] 100 images from 50 women and 50 men (age range: 22-54 years) were examined. Of the 9 parameters extracted from lateral cephalometric images, 7 parameters for sex determination were reliable. The designed discriminant function had an accuracy of 86% for sex determination.

Bagherpour et al.[8] examined 102 images (from 51 women

and 51 men) in the range of 18 to 50 years. Eleven points were identified, and linear and angular measurements were obtained. The classification accuracy of the discriminant analysis method was 87.1%.

In a study conducted by Bewes et al.,[13] 1000 images (from 500 women and 500 men) were examined. Feature extraction from images and their classification into men and women was done by CNNs using GoogLeNet architecture. The training set consisted of 900 images, and the experimental set consisted of 100 images. The network classification accuracy for the experimental data set was 95%.

The results of the present study confirmed that CNNs can predict a person's sex with high accuracy through cephalometric radiographic images. This prediction does not require specialized knowledge and effectively eliminates the bias caused by expertise and human knowledge. The present study is limited by the age of the patients whose radiographs were used for training the CNN models. Further studies with larger sample sizes are desirable to more accurately determine sex and to be applied on a larger scale.

Conflicts of Interest: None

References

1. Williams BA, Rogers T. Evaluating the accuracy and precision of cranial morphological traits for sex determination. J Forensic Sci 2006; 51: 729-35.
2. Nagare SP, Chaudhari RS, Birangane RS, Parkarwar PC. Sex determination in forensic identification, a review. J Forensic Dent Sci 2018; 10: 61-6.
3. Spradley MK, Jantz RL. Sex estimation in forensic anthropology: skull versus postcranial elements. J Forensic Sci 2011; 56: 289-96.
4. Kranioti EF, İşcan MY, Michalodimitrakis M. Craniometric analysis of the modern Cretan population. Forensic Sci Int 2008; 180: 110. e1-5.
5. Mousa A, El Dessouky S, El Beshlawy D. Sex determination by radiographic localization of the inferior alveolar canal using cone-beam computed tomography in an Egyptian population. Imaging Sci Dent 2020; 50: 117-24.
6. Casado AM. Quantifying sexual dimorphism in the human cranium: a preliminary analysis of a novel method. J Forensic Sci 2017; 62: 1259-65.
7. Nikita E, Michopoulou E. A quantitative approach for sex estimation based on cranial morphology. Am J Phys Anthropol 2018; 165: 507-17.
8. Bagherpour A, Anbiaee N, Motaghi S, Jahanbin A. Gender determination using digital lateral cephalograms: a discriminant function analysis. J Dent Mater Tech 2020; 9: 221-30.
9. Chezhian N, Dharman S. Determination of sexual dimorphism using lateral cephalogram - a radiographic study. Indian J Forensic Med Toxicol 2019; 13: 183-8.
10. Qaq R, Mânica S, Revie G. Sex estimation using lateral cephalograms: a statistical analysis. Forensic Sci Int Rep 2019; 1: 100034.
11. Farhadian M, Salemi F, Shokri A, Safi Y, Rahimpanah S. Comparison of data mining algorithms for sex determination based on mastoid process measurements using cone-beam computed tomography. Imaging Sci Dent 2020; 50: 323-30.
12. Sobhani F, Salemi F, Miresmaeili A, Farhadian M. Morphometric analysis of the inter-mastoid triangle for sex determination: application of statistical shape analysis. Imaging Sci Dent 2021; 51: 167-74.
13. Bewes J, Low A, Morphett A, Pate FD, Henneberg M. Artificial intelligence for sex determination of skeletal remains: application of a deep learning artificial neural network to human skulls. J Forensic Leg Med 2019; 62: 40-3.
14. Cao Y, Ma Y, Vieira DN, Guo Y, Wang Y, Deng K, et al. A potential method for sex estimation of human skeletons using deep learning and three-dimensional surface scanning. Int J Legal Med 2021; 135: 2409-21.
15. Wen Y, Xiaoning L, Xiongle L, Lipin Z. Skull sex identification using improved convolution neural network and least squares method. Acta Anthropol Sinica 2019; 38: 265-75.
16. Van Putten MJ, Olbrich S, Arns M. Predicting sex from brain rhythms with deep learning. Sci Rep 2018; 8: 3069.
17. Shin NY, Lee BD, Kang JH, Kim HR, Oh DH, Lee BI, et al. Evaluation of the clinical efficacy of a TW3-based fully automated bone age assessment system using deep neural networks. Imaging Sci Dent 2020; 50: 237-43.
18. Hwang JJ, Jung YH, Cho BH, Heo MS. An overview of deep learning in the field of dentistry. Imaging Sci Dent 2019; 49: 1-7.
19. Indolia S, Goswami AK, Mishra SP, Asopa P. Conceptual understanding of convolutional neural network - a deep learning approach. Procedia Comput Sci 2018; 132: 679-88.
20. Binnal A, Devi BY. Identification of sex using lateral cephalogram: role of cephalofacial parameters. J Indian Acad Oral Med Radiol 2012; 24: 280-3.