Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls

Article *in* Journal of Forensic and Legal Medicine · February 2019

DOI: 10.1016/j.jflm.2019.01.004

Journal of Forensic and Legal Medicine 62 (2019) 40–43

Contents lists available at ScienceDirect

Journal of Forensic and Legal Medicine

journal homepage: www.elsevier.com/locate/yjflm

Research Paper

Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls

James Bewes[a,*], Andrew Low[a], Antony Morphett[b], F. Donald Pate[c], Maciej Henneberg[d]

[a] Department of Radiology, Royal Adelaide Hospital, North Terrace, Adelaide, 5000, Australia
[b] Dr Jones and Partners, Medical Imaging, 270 Wakefield St, Adelaide, 5000, Australia
[c] Archaeology, Flinders University, Adelaide, SA, 5001, Australia
[d] Biological Anthropology and Comparative Anatomy Research Unit, School of Medical Sciences, University of Adelaide, Adelaide, 5005, Australia

ARTICLE INFO

Keywords:
Artificial intelligence
Artificial neural network
Sex determination
Deep learning
Skull determination

ABSTRACT

A deep learning artificial neural network was adapted to the task of sex determination of skeletal remains. The neural network was trained on images of 900 skulls virtually reconstructed from hospital CT scans. When tested on previously unseen images of skulls, the artificial neural network showed 95% accuracy at sex determination. Artificial intelligence methods require no significant expertise to implement once trained, are rapid to use, and have the potential to eliminate human bias from sex estimation of skeletal remains.

1. Introduction

Sex determination from skeletal remains is significant in fields ranging from archeology to physical anthropology and forensics. Current methods are mainly based on morphological or metric criteria.[1,2] These techniques are subject to human biases, require significant expertise, are often complex and are time consuming.[3–5] Validation studies have reported variable accuracy of these methods, sometimes significantly lower than reported in the initial studies.[1,5–7]

Artificial Intelligence (AI) techniques offer an objective and reproducible solution to the problem of sex determination. Deep learning artificial neural networks are a subtype of AI that originated in the computer science literature and were inspired by the architecture of the human brain. Artificial neural networks consist of multiple layers of connected “neurons” that each analyse a separate part of an image, before combining to output a probability of an input image belonging to a predefined category. They are ideally suited to the binary task of sex determination.

An advantage of artificial neural networks is that they automatically discover the features of an image that are most useful for classification.[8,9] There is much debate in the physical anthropology and forensic science literature on skeletal anatomical landmarks or morphological features that are the best discriminants for biological sex.[1,10] Artificial neural networks bypass this controversy, independently assigning weights to the parts or features of an image that it finds most useful for sex determination. Artificial neural networks can automatically de-emphasise or discard the parts of the image that are unhelpful.

We hypothesise that artificial intelligence techniques, specifically deep learning neural networks, can be trained to perform sex determination from skeletal structures. Specifically, the ability of neural networks to accurately classify images of skulls according to biological sex will be examined.

For an artificial neural network to learn the sexual dimorphic features of skulls, a large sample of skulls with a known biological sex is required. Available collections of skeletal remains with a definitively known biological sex, based on molecular methods or other highly reliably technique, are limited in size. This study will make use of 3D skull reconstructions derived from hospital-based CT scans. This is an optimal substitute as biological sex is recorded on patient medical records, providing the neural network with a reference dataset where documented sex is highly accurate (a near perfect “ground-truth”).

2. Methods

2.1. Ethical statement

All experiments were carried out in accordance with relevant guidelines and regulations. The study was approved by the Royal Adelaide Hospital Human Research Ethics Committee. Patient consent was not required by the ethics committee as all data was pre-existing and de-identified.

* Corresponding author.
E-mail address: bewesj@gmail.com (J. Bewes).

https://doi.org/10.1016/j.jflm.2019.01.004
Received 15 July 2018; Received in revised form 15 December 2018; Accepted 2 January 2019
Available online 04 January 2019

2.2. Study population

500 male and 500 female randomly selected CT Head studies from the preceding five years were retrieved from the Royal Adelaide Hospital PACS database. CT scans performed on patients between 18 and 60 years of age were retrieved. Subadults were not included in this study. An arbitrary upper age cutoff was implemented as hospital populations are heavily weighted towards elderly age groups. Importantly, there were no pathology-based exclusion criteria.

The ancestry of the study population was not documented in the available patient records. The best estimate of study population ancestry is obtained from local demographic data. The ancestry of the population of Greater Adelaide as documented in recent census data[11] is predominantly European, with the six most common ancestries being English (37.3%), Australian (30.9%), Scottish (8.2%), Irish (8.1%), German (6.8%) and Italian (6.8%). The three largest ethnic groups of non-European ancestry are Chinese (3.9%), Indian (2.4%) and Vietnamese (1.5%).

2.3. CT scan acquisition and post-processing

All scans were performed on Siemens SOMATOM or Toshiba AQUILION CT scanners. CT scan post-processing was performed on a standard radiology workstation with Carestream PACS (Carestream Health, Rochester NY). All written identifying information from the scans was removed. The scan was reconstructed into a manipulable 3D skull reconstruction using an in-built function within Carestream PACS. The same reconstruction filter and algorithm was applied for every case. The skull was then rotated to the left lateral plane and a 2D jpg image of the skull was saved. Representative images that were created are shown in Fig. 1.

The lateral plane was chosen as it is the plane that provides the most anatomical information on a single 2D projection. Each 3D skull reconstruction was rotated manually by the study authors to be viewed from the lateral position. It is likely there was minor differential angulation in the relative projection of each skull, but this was not considered detrimental to study purposes, as random noise in an image data set can be useful in preventing a common error in neural network training termed "over-fitting".

The facial bones and upper cervical spine are variably included on some images in the study. Inclusion of the cervical vertebra on the images is a study limitation, as it is possible these elements are contributing a sex specific signal to the neural network. As these structures are not known to be as sexual dimorphic as the skull, it is most likely they are not contributing substantially to neural network decision making, but this cannot be proved in this study.

2.4. Artificial intelligence: deep learning convolutional neural network

A pre-existing convolutional neural network, "GoogLeNet", was adapted for the task of sex determination. GoogLeNet is optimised for image classification tasks and has been trained on a database of 1.2 million images of 1000 different common everyday objects.[12] When a pre-existing neural network is modified for a new task, this is termed "transfer learning" in artificial intelligence literature.

MATLAB (Mathworks, 2018) is a programming environment with a number of in-built neural networks that are readily available for transfer learning. Instead of classifying an input image into one of 1000 different object categories, the GoogleNet neural network within MATLAB was modified within MATLAB to classify input images into either "male" or "female". This required modification of the final three layers of the neural network, and has been documented elsewhere.[13] The architecture of the GoogLeNet artificial neural network is described in depth in the literature.[12]

Each convolutional layer in a neural network, analogous to a neuron in the brain, is differentially activated by unique imaging features. For example, it is possible that one convolutional layer of a neural network may produce a strong activation in response to a large occipital

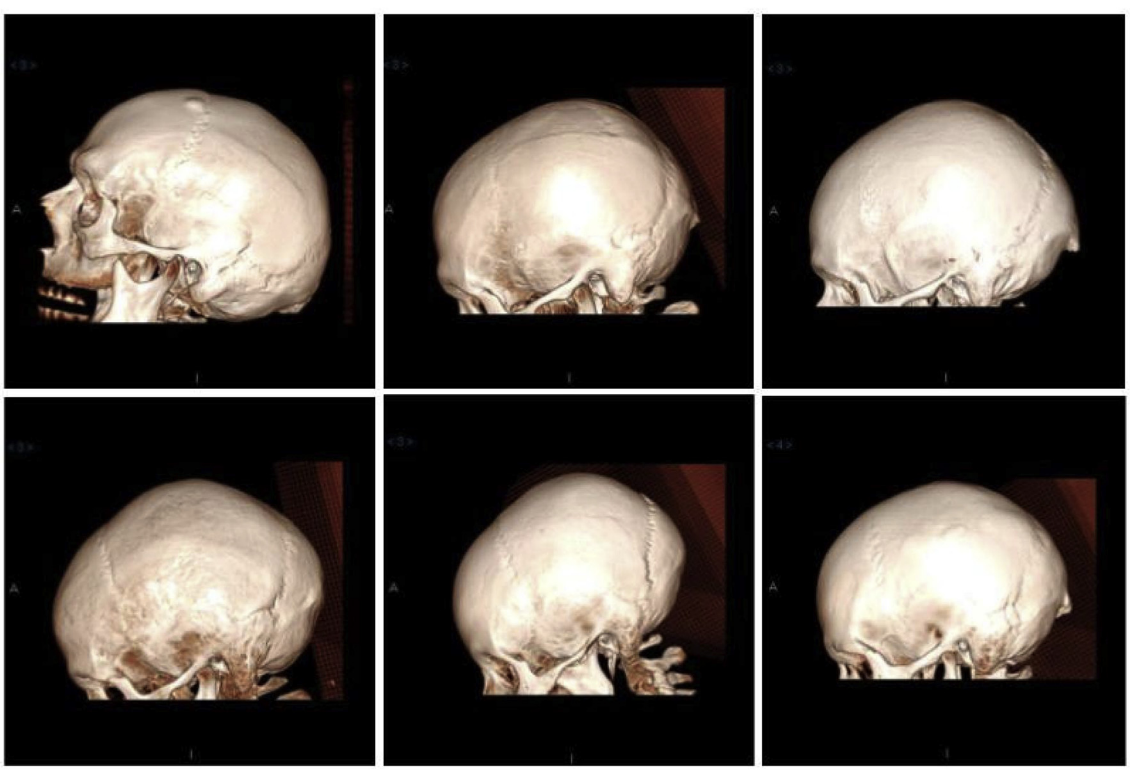

Fig. 1. Representative 3D skull reconstructions. The top three skulls are male skulls, the bottom three skulls are female. The images demonstrate the variability in field of view and positioning. 1000 skull reconstructions were used in this study.

Table 1
Sex estimated by the AI algorithm compared with sex as recorded on the patient medical record. N = 100.

		Sex (Medical Record)	
		Male	Female
Sex (AI estimation)	Male	47	2
	Female	3	48

protuberance, while another may produce a strong activation in response to glabella curvature. The relative importance of each layer (and therefore imaging feature) is determined in a neural network training step using "backpropagation". In backpropagation, the neural network makes an estimate of the sex of a skull, compares its estimate with the known sex of the skull, and adjusts the relative weighting of its internal layers (analogous to neural connections) to try and get a better result. Image features that are strongly sexually dimorphic will be progressively emphasised, and image features that are not helpful in predicting sex will be progressively de-emphasised.

It is important to note that the neural networks used in this study did not have any predefined knowledge or conception of the parts of a skull. The neural network simply comprises a suite of imaging filters that are generally useful in image analysis, and through the optimisation process of backpropagation, self-determines the features of a skull that are most useful in sex determination.

Image augmentation was performed prior to neural network training with the objective of improving the generalisability of the neural network by preventing "over-fitting". Image augmentation can be summarised as introducing random noise into an image dataset so the deep learning algorithm is unable to memorise unique image features. To achieve this, images were randomly rotated (between 0 and 360°), translated (between 0 and 15%) and flipped (both on x and y axis).

The GoogLeNet neural network requires input images of 224 × 224 pixels in size. Skull images were resized prior to data augmentation, with image dimensions held constant (no image distortion).

The 1000 image dataset was divided into training and testing subsets. 900 images (450 of each sex) were allocated to training, with the remaining 100 kept hidden from the neural network as a test set. The neural network was trained on the 900 image dataset for 2000 epochs with a minibatch size of 90. The training was performed on an Intel Core i7 desktop PC using a GPU graphics card (NVIDIA GeForce GTX 1080 Ti).

Following training of the neural network, performance was evaluated on a 100 image test-set. It is important to emphasise that these images were not used in the training set (the neural network had not seen them before). Accuracy of sex determination was assessed against the sex of each skull as documented on the patient medical record.

2.5. Statistical analysis

Two tailed TTest was used to determine if the mean age for each sex was significantly different.

True positives, false positives, true negatives, false negatives and Precision for each sex was calculated. Overall test accuracy inclusive of both sexes (defined as the sum of true positives + true negatives, divided by total study population), as well as Matthew's Correlation Coefficient (MCC) were calculated.

Pearson's Chi-squared test (two-tailed) was used to determine if neural network performance exceeded background chance, as well as compare performance of the neural network between sexes.

2.6. Data availability

The images of skulls are not publicly available due to patient confidentiality. Release of the images is limited by the current ethics committee approval for the study.

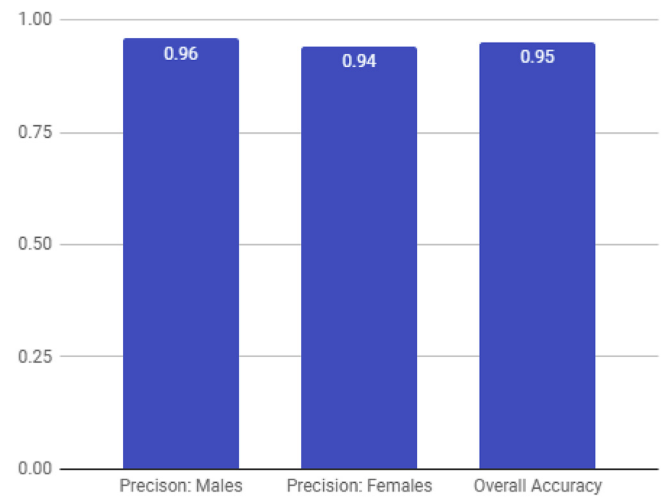

Fig. 2. Precision of a male result and a female result, along with overall accuracy of the neural network. Precision is equivalent to the positive predictive value and represents the proportion of positive results that are correct diagnoses.

3. Results

The mean age of skulls for females (M = 41.2y, SD = 12.6yrs) was not significantly different from males (M = 39.3, SD = 12.3 yrs), $p = 0.9$.

Sex estimates produced by the neural network, compared to the ground-truth sex (as derived from the patient medical record), is presented in Table 1. The precision of a male result and a female result are present in Fig. 2, along with overall accuracy.

The performance differences between sexes were not statistically significant ($p = 0.6$). Overall accuracy of 95% was better than random chance ($p < 0.01$).

The Matthews correlation coefficient (MCC) was 0.9. MCC is regarded as a balanced correlation coefficient that returns a value between −1 and +1, where +1 represents a perfect prediction, 0 is no better than random chance and −1 indicates total disagreement between prediction and observation.

4. Discussion

This study demonstrates that artificial intelligence methods based on neural networks are ideally suited to the task of sex determination from skeletal structures.

The only input into the artificial neural network in this study was the ectocranial image of skulls viewed from the side (see Fig. 1). Without any instruction or pre-existing knowledge of sex-dimorphic anatomy, the neural network was able to learn skull features that were useful in predicting sex. When tested on a validation set of skulls that it had never previously been exposed to, and derived from a relatively ethnically diverse population, it demonstrated high accuracy in sex determination.

Neural networks are independent of human biases, as feature recognition is automated within the network. Neural networks are extremely fast, taking milliseconds to classify skeletal structures once they have been trained. In distinction to existing morphological and metric methods, neural networks do not require any specialised expertise to use. Knowledge of anatomy or sex specific morphologic characteristics is not needed. Consequently, sex classification of skeletal remains could be performed by a forensic scientist, archeologist or anthropologist with minimal training and in a very short time.

Artificial neural networks also have the potential for ongoing improvement as more data for training becomes available. There is no upper limit for the volume of input data into a neural network - a neural network could potentially train on millions of images of skeletal remains, which is far more than a forensic anthropologist is capable of analysing in an entire career.

The neural network used in this study was restricted to analysing 2D input images by its design. This limited analysis of skulls to ectocranial features only. Volumetric 3D scans of skulls would provide much more anatomic detail than reconstructed 2D images, as both endocranial and ectocranial features could be examined simultaneously. A 3D neural network would consequently have greater potential for extracting sex-dimorphic features than a 2D neural network. Pre-trained neural networks capable of analysing 3D images are expected to become available in the near future and may further improve the performance of AI classifiers.

The two major weakness of AI based on neural networks are that they require large sets of training data to perform well, and that they are a “black box”. The problem of training data volume has been solved in this study by using CT scans derived from hospital patients with known sex. The problem of the “black box” is more difficult to solve. Neural networks are inspired by the architecture of the human brain. They comprise multiple layers of interconnected neurons, that are linked in an extremely complex web with neurons in other layers. Consequently, interrogating a neural network to determine how it is fundamentally making predictions is limited. While individual neurons can be examined within a network, this is often no more meaningful than examining a single neuron within the human brain to determine its function. While it is highly desirable to know what parts of the human skull the AI is using to determine sex, this information is not currently obtainable. While research into improving the transparency of artificial neural networks is ongoing, at present our ability to deconstruct a neural network in understandable components is limited.

It is acknowledged that the skeletal images used to train and test the AI in this study are approximations of those encountered in real-world archaeological and forensic contexts. Future research on artificial neural networks in sex determination could include application to genuine remains from archaeological and forensic contexts, application to other skeletal structures (such as the pelvis) and application to input images that are not derived from CT scans, such as simple pictures from a handheld camera or smartphone. The latter is particularly exciting, as neural networks offer the possibility of a technique that can be adapted for rapid use in the field.

5. Conclusion

Neural networks can be trained to estimate the sex of an individual from skeletal remains with high accuracy. Artificial intelligence methods based on neural networks require no significant expertise to implement, are rapid to use, and have the potential to eliminate human bias from sex estimation of skeletal remains. Neural networks have the potential to democratise physical and forensic anthropology, and impact fields ranging from archeology to the forensic sciences.

Competing interests

The authors declare no competing interests.

Acknowledgements

The authors would like to thank Dr Luke Oakden-Rayner for proofreading the manuscript and providing constructive feedback.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jflm.2019.01.004.

References

1. Krishan K, Chatterjee P, Kanchan T, Kaur S, Baryah N, Singh R. A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Sci Int.* 2016;261:165. https://doi.org/10.1016/j.forsciint.2016.02.007 e1-165.e8.
2. Buikstra J, Ubelaker D, Aftandilian D. *Standards for Data Collection from Human Skeletal Remains.* Fayetteville: Arkansas Archeological Survey; 1994.
3. Smith A, Boaks A. Consistency of selected craniometric landmark locations and the resulting variation in measurements. *Forensic Sci Int.* 2017;280:156–163. https://doi.org/10.1016/j.forsciint.2017.10.002.
4. Nakhaeizadeh S, Dror I, Morgan R. Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias. *Sci Justice.* 2014;54(3):208–214. https://doi.org/10.1016/j.scijus.2013.11.003.
5. Sierp I, Henneberg M. The difficulty of sexing skeletons from unknown populations. *J Anthropol.* 2015;2015:1–13. https://doi.org/10.1155/2015/908535.
6. Walrath D, Turner P, Bruzek J. Reliability test of the visual assessment of cranial traits for sex determination. *Am J Phys Anthropol.* 2004;125(2):132–137. https://doi.org/10.1002/ajpa.10373.
7. Scheuer L. Application of osteology to forensic medicine. *Clin Anat.* 2002;15(4):297–312. https://doi.org/10.1002/ca.10028.
8. Oakden-Rayner L, Carneiro G, Bessen T, Nascimento J, Bradley A, Palmer L. Precision Radiology: predicting longevity using feature engineering and deep learning methods in a radiomics framework. *Sci Rep.* 2017;7(1) https://doi.org/10.1038/s41598-017-01931-w.
9. Goodfellow I, Bengio Y, Courville A. *Deep Learning.* MIT Press; 2016.
10. Spradley M, Jantz R. Sex estimation in forensic anthropology: skull versus post-cranial elements. *J Forensic Sci.* 2011;56(2):289–296. https://doi.org/10.1111/j.1556-4029.2010.01635.x.
11. Australian Bureau of Statistics. *Australian Government. Abs.gov.au.* 2018; 2018 Published http://www.abs.gov.au/, Accessed date: 15 April 2018.
12. Szegedy C, Liu W, Jia Y, et al. *Going Deeper with Convolutions.* Arxiv.org 2018; 2018 Published https://arxiv.org/abs/1409.4842, Accessed date: 15 July 2018.
13. *Pretrained GoogLeNet Convolutional Neural Network - MATLAB googlenet- MathWorks Australia. Au.mathworks.com.* 2018; 2018 Published https://au.mathworks.com/help/nnet/ref/googlenet.html, Accessed date: 15 July 2018.