



HAL
open science

Sex estimation from coxal bones using deep learning in a population balanced by sex and age

Marie Epain, Sébastien Valette, Kaifeng Zou, Sylvain Faisan, Fabrice Heitz,
Pierre Croisille, Tony Fracasso, Laurent Fanton

► To cite this version:

Marie Epain, Sébastien Valette, Kaifeng Zou, Sylvain Faisan, Fabrice Heitz, et al.. Sex estimation from coxal bones using deep learning in a population balanced by sex and age. *International Journal of Legal Medicine*, 2024, 10.1007/s00414-024-03268-2 . hal-04609128

HAL Id: hal-04609128

<https://hal.science/hal-04609128v1>

Submitted on 27 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Abstract

In the field of forensic anthropology, researchers aim to identify anonymous human remains and determine the cause and circumstances of death from skeletonized human remains. Sex determination is a fundamental step of this procedure because it influences the estimation of other traits, such as age and stature. Pelvic bones are especially dimorphic, and are thus the most useful bones for sex identification. Sex estimation methods are usually based on morphologic traits, measurements, or landmarks on the bones. However, these methods are time-consuming and can be subject to inter- or intra-observer bias. Sex determination can be done using dry bones or CT scans. Recently, artificial neural networks (ANN) have attracted attention in forensic anthropology. Here we tested a fully automated and data-driven machine learning method for sex estimation using CT-scan reconstructions of coxal bones. We studied 580 CT scans of living individuals. Sex was predicted by two networks trained on an independent sample: a disentangled variational auto-encoder (DVAE) alone, and the same DVAE associated with another classifier (C_{recon}). The DVAE alone exhibited an accuracy of 97.9%, and the DVAE + C_{recon} showed an accuracy of 99.8%. Sensibility and precision were also high for both sexes. These results are better than those reported from previous studies. These data-driven algorithms are easy to implement, since the pre-processing step is also entirely automatic. Fully automated methods save time, as it only takes a few minutes to pre-process the images and predict sex, and does not require strong experience in forensic anthropology.

Keywords: Forensic anthropology, Sexual dimorphism, Sex prediction, Machine learning algorithms

Authors: Marie Epain^{1,2}, Sébastien Valette³, Kaifeng Zou⁴, Sylvain Faisan⁴, Fabrice Heitz⁴, Pierre Croisille^{3,5}, Tony Fracasso⁶, Laurent Fanton^{1,2,3}

Addresses:

¹ Hospices Civils de Lyon, Edouard Herriot Hospital, Service of Legal Medicine, Lyon, France. (corresponding author: marie.epain@chu-lyon.fr)

² University Lyon, Claude-Bernard Lyon 1 University, Lyon-Est Medical School, Lyon, France

³ CREATIS-LRMN, INSERM U630, INSA Lyon, Villeurbanne, France.

⁴ ICube (UMR 7357), Laboratoire des Sciences de l'Ingénieur, de l'Informatique et de l'Imagerie, UNISTRA, CNRS, Illkirch, France.

⁵ Radiology Department, University Hospital of Saint-Étienne, Saint-Étienne, France.

⁶ Centre universitaire romand de médecine légale, Lausanne - Genève, Département de médecine communautaire, Université de Genève

Author contributions:

Marie Epain: writing-original draft, investigation; **Sébastien Valette:** writing-review and editing, funding, supervision; **Kaifeng Zou:** software, methodology; **Sylvain Faisan:** methodology, supervision; **Fabrice Heitz:** supervision; **Pierre Croisille:** resources, supervision; **Tony Fracasso:** writing-review and editing; **Laurent Fanton:** writing-review and editing, supervision.

Declarations

Competing Interest This work was funded by the TOPACS ANR-19-CE45-0015 project of the French National Research Agency (ANR).

Introduction

In forensic anthropology, researchers aim to identify anonymous human remains and determine the cause and circumstances of death from skeletonized human remains. When there is no ground supporting an a priori assumption of identification, anthropological examination can provide demographic information about the individual. Sex determination is a fundamental step of this procedure, because it influences the estimation of the other traits, such as age and stature [1, 2]. Although many bones can be used to estimate sex, the most accurate are pelvic bones (>95%), femur (90–95%), and skull (80–85%) [2]. Pelvic bones are the most useful, as they are especially dimorphic due to parturition [1].

There are multiples methods for sex estimation, with the most common being morphological methods (e.g., Phenice's or Işcan and Derrick's) and metric methods [1, 2]. More recently, geometric morphometric methods have been developed, thanks to advances in computer science. These methods are based on computerized analysis of anatomical landmarks manually placed on the bones [1]. Morphologic methods are easy to use because they do not require any specific tools, but they are highly observer-dependent and require experience [1]. Metric methods are more time-consuming because they require many measurements of the bones, but they carry less inter-observer variability. However, they usually require statistical analysis, so their accuracy also depends on the selected statistical approach [1, 3]. For example, one very accurate metric method is the *diagnose sexuelle probabiliste* (DSP), in which 10 measurements on pelvic bones are used to estimate the probability of an individual being male or female, with a threshold of >95%. It can determine sex in 90% of individuals, and has a reliability of between 98–100% [4].

Since the development of imaging techniques—especially CT scanning—metric and geometric morphometric methods have been tested on 3D reconstructions of pelvic bones, yielding great results [1, 5, 6, 7]. CT scans are often used in forensic medicine, since they can provide easy and non-invasive access to bone structures [1].

Recently, artificial neural networks (ANN) have attracted attention for potential use in forensic sex determination. ANN are machine-learning algorithms inspired by biological nervous systems, which are commonly used in statistical problems, such as for automated classification of data [8]. As statistical tools, ANN can be used in association with metric or geometric morphometric methods, as well as with morphologic methods [3]. Their accuracy is about 90%, depending on the bone and the algorithm [9–14]. ANN are rarely used in forensic anthropology because they are difficult to develop, and the results can be complicated to interpret, due to hidden layers [15].

In the present pilot study, we tested a machine learning method to estimate sex from CT reconstructions of pelvic bones.

Material and methods

Reference sample

This study was performed using abdominopelvic CT scans of 588 individuals of known sex and age, who underwent CT scanning between March of 2021 and March of 2022, at the radiology department of the University Hospital of Saint-Etienne (CHUSE) in France. These individuals were between 20 and 79 years of age. The individuals were selected to ensure a homogeneous distribution between male and female and among age classes. The sample was divided into six homogeneous decade groups. We excluded CT scans if they showed severe deformation of the lumbosacral hinge, or artefacts on the pelvic region. No data were collected regarding stature, ethnicity, medical history, or scan indications.

Of the 588 CT scans, 580 were included in the present analysis. Three scans were excluded due to registration issues, and five because a large part of the pelvic bones was missing (Fig. 1).

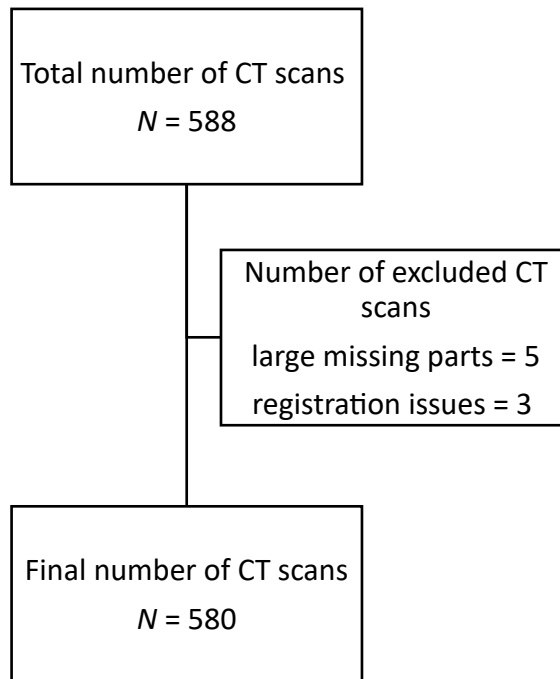


Fig. 1 Flowchart of scan inclusion

The final sample included 294 male and 286 female subjects, of 20–79 years old, with a mean age of 49.4 years. The six age groups were balanced and homogeneous between men and women (Table 1).

Table 1 Distribution of individuals by age and sex

Age groups, years		20–29	30–39	40–49	50–59	60–69	70–79	Total
Women	Number of cases	55	44	55	47	43	42	286
	Mean age	24.3	34.1	44.6	55	64	74.4	48.1
	(SD)	(2.9)	(2.9)	(2.7)	(2.7)	(2.8)	(2.8)	(17.2)
Men	Number of cases	44	52	42	47	56	53	294
	Mean age	24.2	34.8	44.5	54.4	64.8	74.5	50.6
	(SD)	(2.9)	(2.7)	(3.1)	(2.8)	(2.9)	(2.9)	(17.5)
Total	Number of cases	99	96	97	94	99	95	580
	Mean age	24.2	34.5	44.5	54.7	64.5	74.5	49.4
	(SD)	(2.4)	(2.8)	(2.9)	(2.8)	(2.9)	(2.9)	(17.4)

Radiological data acquisition and processing

At the radiology department of the CHUSE, CT scans were performed using different systems and protocols. As the subjects were selected retrospectively and randomly from the radiology department database, acquisition parameters were not controlled and numerous.

As the classification algorithms take meshes as inputs, we needed to convert DICOM images into meshes.

First, a template average hip bone mesh is created as follow: i) CT scans from a reference dataset D of 100 samples were registered simultaneously into a common space, using the FROG algorithm, which automatically extracted 20 000 key points per images [16]; ii) a template average image was computed by transforming each CT scan following the transformation field obtained from the previous registration step; iii) the coxal bone was manually segmented by an expert using 3D Slicer [17] and meshed with ACVD [18] from the average image. This hip bone average mesh M is then used as a template to represent all individual hip bones via registration.

When a hip bone needs to be extracted from an individual’s CT scan, the key points are extracted from the images and registered with the transformed reference dataset D^R . Next, the new coxal bone mesh is reconstructed by applying the invert transformation on the average mesh M . The quality of the mesh reconstruction can be assessed during the classification process, as the DVAE will yield a large reconstruction error if the mesh is not well reconstructed. The complete workflow is presented in figures 2a and 2b. The FROG and the ACVD algorithms were coded in C++ language.

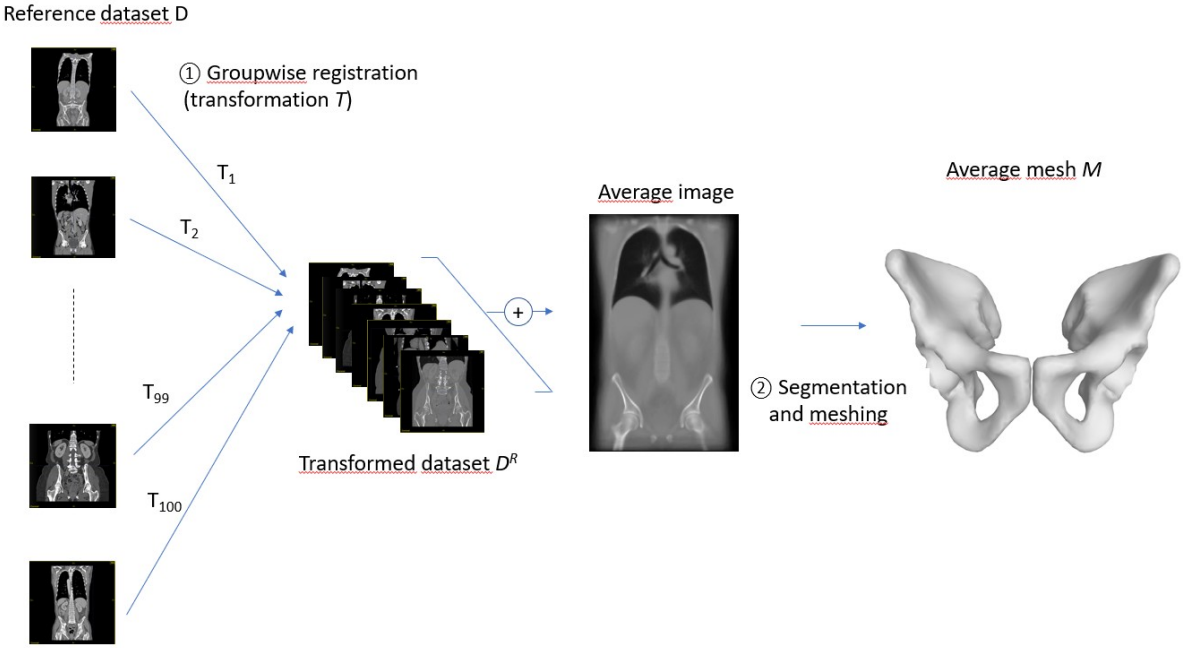


Figure 2a: Groupwise registration of the reference dataset and computation of the average mesh

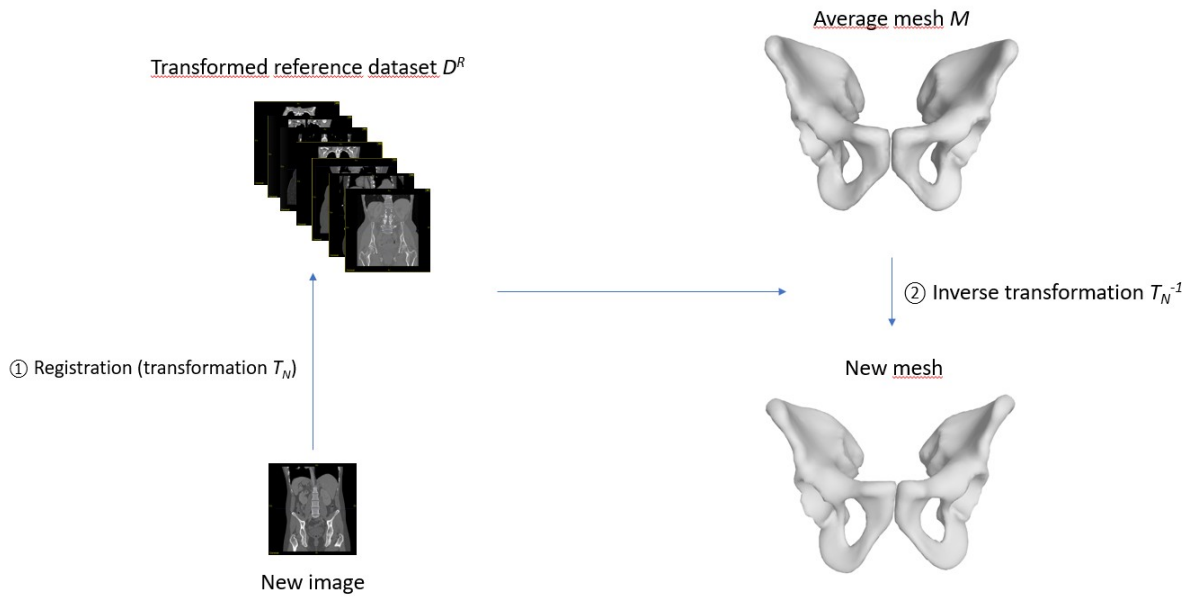


Figure 2b: Computation of a hip bone mesh for a new individual, using the transformed dataset D^r and the average mesh M

Training and classification method

Sex was predicted using two algorithms [15], a disentangled variational auto-encoder (DVAE) alone, and the same DVAE associated with another classifier (C_{recon}). The inputs for the DVAE were the coordinates of the nodes of the individual meshes. The algorithm predicted the individual's sex, and then provided two reconstructions—one of same sex and one of opposite sex—by changing the latent variable related to sex (Fig. 4). The C_{recon} compared the reconstructions from the DVAE with the original one to predict sex. These algorithms were coded in Python language and are available here: <https://github.com/ZOUKaufeng/Mesh-VAE>.

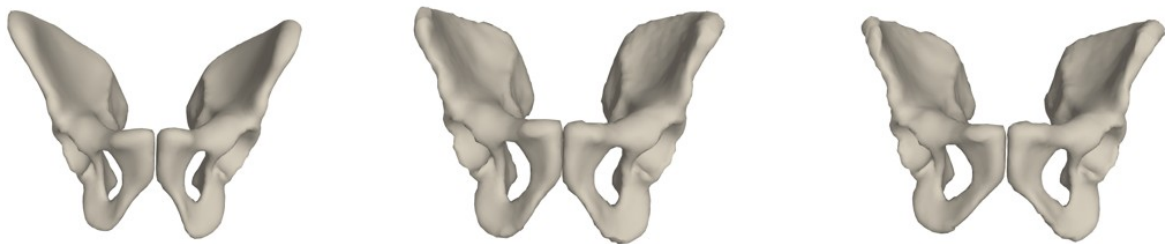


Fig. 3 Examples of reconstructions by the DVAE. (a) Original reconstruction (male). (b) Same sex reconstruction (male). (c) Opposite sex reconstruction (female).

The algorithms were previously trained using a supervised method with independent data from a sample of 752 individuals (470 males and 282 females, with a mean age of 65.4 years) [15]. Their reliability was tested using a k-fold cross-validation method, with $k = 5$. Among the individuals, 80% were used for training, and 20% for validation. Here, the DVAE alone exhibited a classification accuracy of $>99\%$, and the DVAE + C_{recon} approached an accuracy of 100%.

In our experiments, a CT scan was considered correctly classified if sex was properly predicted in all the folds.

Statistical analysis

A Chi² Pearson’s test with Yates’ continuity correction was used to identify an association between gender and error rate, and a Welch’s t-test to assess a relationship between age and error rate.

Results

Among the 580 included CT scans, 12 were misclassified by the DVAE. The 12 misclassified individuals included 7 males and 5 females. Six were between 20–29 years old, three between 60–69 years old, none was between 50–59 years old, and one individual was misclassified in each of the other age groups (Table 2). The error rate did not significantly differ between male and female individuals (p -value = 0.81). Age did not significantly differ between the misclassified subjects and correctly classified individuals (p -value = 0.28).

Table 2 Misclassified subjects according to age

Age groups, years ($n = 580$)	20–29 ($n = 99$)	30–39 ($n = 96$)	40–49 ($n = 97$)	50–59 ($n = 94$)	60–69 ($n = 99$)	70–79 ($n = 95$)
Total number of misclassified subjects ($n = 12$)	6	1	1	0	3	1
Women misclassified as men ($n = 5$)	2	0	0	0	3	0
Men misclassified as women ($n = 7$)	4	1	1	0	0	1

The DVAE yielded an accuracy of 97.9%. Female sex was predicted with 0.98 sensitivity, and 0.976 precision (Table 3).

Table 3 Confusion matrix for the DVAE alone

		True sex	
		Female	Male
Predicted sex	Female	281	7
	Male	5	287

Table 4 Confusion matrix for the DVAE + C_{recon}

		True sex	
		Female	Male
P r e d i c t e d s e x	F e m a l e	286	1
	M a l e	0	293

For the DVAE, we found a median of 8 errors for each fold. DVAE reconstructions were always consistent with real sex, even when the sex prediction was wrong. The DVAE method was improved by adding the additional classifier C_{recon} . The method DVAE + C_{recon} misclassified only one individual, a 26-year-old male, yielding an accuracy of 99.8%.

Discussion

The presently tested method yielded an accuracy of 97.8% when using the DVAE alone, and 99.8% when using the DVAE + C_{recon} . The performances of the DVAE were similar between males and females (p -value = 0.81), with sensitivities of 0.97 and 0.98, respectively. We found no association between age and error rate, but we observed more classification errors for individuals under 30 years of age. Table 5 presents details regarding several previous studies that have tested different ANN for sex determination using coxal bone or part of it, using variable morphologic traits or measurements, yielding accuracy rates of between 73–96% [3, 9–11] (Table 5).

Table 5 Previous studies of sex determination from coxal bones using artificial neural networks (ANN)

Authors	ANN model	Number and type of subjects	Variables	Accuracy
Nikita et al. [3]	Single hidden-layer neural network (nnet)	191 skeletons	3 morphologic traits	94.9%

Cao et al. [9]	Convolutional neural network (CNN)	1000 CT scans 108 surface scanning	3 morphologic traits	CT-dataset: 94–98.5% 3D surface scanning: 73.3–96.2%
Knecht et al. [10]	Unknown	60 coxal bones	15 measurements on the great sciatic notch	94%
Coelho et al. [11]	Single hidden-layer neural network (nnet)	256 skeletons	38 osteometric variables	92%

Other investigations have used ANN for sex determination from other bones (long bones, sacrum, cranium, etc.), and have shown variable performances depending on the bones and the methods, with varying accuracy rates (Table 6).

Table 6 Previous studies of sex determination from other bones using artificial neural networks (ANN)

Author	ANN model	Number and type of subjects	Studied bone(s)	Variables	Accuracy
Nikita et al. [3]	Single hidden-layer neural network (nnet)	191 skeletons	Cranium	5 morphologic traits on the cranium	81.7–86.7%
Etlı et al. [12]	Multilayer perceptron (MLP)	480 CT scans	Sacrum	8 length measurements 4 angle measurements 1 area measurement	86.3%
Oner et al. [13]	MLP	422 CT scans	Sternum	4 measurements	83%–97%
Bewes et al. [14]	CNN	1000 CT scans	Cranial bones	Data-driven	95%

Overall, our present method, especially when using the DVAE + C_{recon} approach, yielded better accuracy than the methods previously reported in the literature. Comparison among studies suggests that pelvic bones are the most reliable bone for sex estimation based on machine learning methods. However, we could further test our method using other bones, especially cranial bones, as Nikita et al. demonstrated improved results when using combinations of pelvic and cranial bones [3].

Many of the previous methods have relied on observed traits, measurements, or landmarks, whereas our present method is entirely automated and independent from the observer’s experience. Bewes et al. also tested a data-driven method and reported interesting results, but they used cranial

bones [14]. Data-driven algorithms are easy to implement, because the pre-processing step is also entirely automatic. Fully automated methods are time-saving, as they require just a few minutes to pre-process the images and predict sex, and they do not require strong experience in forensic anthropology [14].

We observed that the DVAE properly reconstructed coxal bones, and that the reconstructions were consistent with the actual sex of the subject, even when the sex prediction was wrong. This observation could help us understand the misclassifications. Additionally, predictions using the DVAE and the C_{recon} seem to be independent, since the individuals misclassified by one algorithm were properly classified by the other.

In general, artificial neuronal networks are limited by the need for large databases for training, and the lack of interpretability and explainability. To improve our knowledge concerning the choices made by the algorithm, Zou et al. [15] introduced a method of disentanglement, which also highlights the differences between females and males.

The sample analyzed in our present study differed in sex and age distribution compared to the sample used for training and validation. The presently analyzed population was younger and better balanced between men and women. These differences could explain the higher number of errors for individuals under 30 years old. However, the small number of misclassified individuals in our present analysis made it difficult to study the real impact of age on our results. We also had to exclude individuals due to registration issues, because the pelvis was not aligned with the rest of the body. This might limit the application to real forensic situations as we usually cannot control the position of a body during a postmortem CT scan : Note : or we need to add some specific processing such as image cropping/rotation.

Notably, the DVAE + C_{recon} approach has shown great accuracy even on a model with missing data [15], which suggests that this method could be applied on fragmented bones. This method must be tested on bigger samples and postmortem data. Since coxal bones can be missing from forensic samples, this method should also be tested using other bones. Finally, with further investigations, we could train our algorithms to estimate age at death or stature.

Conclusion

Neural networks are interesting tools that can assist anthropologists in the sex estimation of an unknown individual. The presently tested method showed a high accuracy, and is entirely automated, easy to implement, and fast. This new approach also provides us with a better understanding of the choices made by the algorithm.

Compliance with Ethical Standards

This study was conducted as part of a larger project, called TOPACS (Towards Open PACS), which was funded by the National Research Agency. From an ethical perspective, the funding agency verified this study's compliance with the law. This study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki. The authors have no competing interests to declare that are relevant to the content of this article.

References

- [1] Krishan, Kewal, et al. « A Review of Sex Estimation Techniques during Examination of Skeletal Remains in Forensic Anthropology Casework ». *Forensic Science International*, vol. 261, avril 2016, p. 165.e1-165.e8. DOI.org (Crossref), <https://doi.org/10.1016/j.forsciint.2016.02.007>.
- [2] Quatrehomme *Traité d'anthropologie medico-légale*. Edition de boeck, 2015. pp 931 [In french]
- [3] Nikita, Efthymia, et Panos Nikitas. « Sex Estimation: A Comparison of Techniques Based on Binary Logistic, Probit and Cumulative Probit Regression, Linear and Quadratic Discriminant Analysis, Neural Networks, and Naïve Bayes Classification Using Ordinal Variables ». *International Journal of Legal Medicine*, vol. 134, n° 3, mai 2020, p. 1213-25. DOI.org (Crossref), <https://doi.org/10.1007/s00414-019-02148-4>.
- [4] Murail, Pascal, et al. « DSP: A tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements ». *Bulletins et mémoires de la société d'anthropologie de Paris*, vol. 17, n° 3-4, décembre 2005, p. 167-76. DOI.org (Crossref), <https://doi.org/10.4000/bmsap.1157>.
- [5] Chapman, Tara, et al. « Sex Determination Using the Probabilistic Sex Diagnosis (DSP: Diagnose Sexuelle Probabiliste) Tool in a Virtual Environment ». *Forensic Science International*, vol. 234, janvier 2014, p. 189.e1-189.e8. DOI.org (Crossref),
- [6] Mestekova, Sarka, et al. « A Test of the DSP Sexing Method on CT Images from a Modern French Sample ». *Journal of Forensic Sciences*, vol. 60, n° 5, septembre 2015, p. 1295-99. DOI.org (Crossref), <https://doi.org/10.1111/1556-4029.12817>.
- [7] Clavero, Ana, et al. « Sex Prediction from the Femur and Hip Bone Using a Sample of CT Images from a Spanish Population ». *International Journal of Legal Medicine*, vol. 129, n° 2, mars 2015, p. 373-83. DOI.org (Crossref), <https://doi.org/10.1007/s00414-014-1069-y>.
- [8] Nikita, Efthymia, et Panos Nikitas. « On the Use of Machine Learning Algorithms in Forensic Anthropology ». *Legal Medicine*, vol. 47, novembre 2020, p. 101771. DOI.org (Crossref), <https://doi.org/10.1016/j.legalmed.2020.101771>.
- [9] Cao, Yongjie, et al. « A Potential Method for Sex Estimation of Human Skeletons Using Deep Learning and Three-Dimensional Surface Scanning ». *International Journal of Legal Medicine*, vol. 135, n° 6, novembre 2021, p. 2409-21. DOI.org (Crossref), <https://doi.org/10.1007/s00414-021-02675-z>.
- [10] Knecht, Siam, Luísa Nogueira, Maël Servant, Frédéric Santos, Véronique Alunni, Caroline Bernardi, et Gérald Quatrehomme. « Sex Estimation from the Greater Sciatic Notch: A Comparison of Classical Statistical Models and Machine Learning Algorithms ». *International Journal of Legal Medicine* 135, n° 6 (novembre 2021): 2603-13. <https://doi.org/10.1007/s00414-021-02700-1>.

- [11] Coelho, João d'Oliveira, et Francisco Curate. « CADOES: An Interactive Machine-Learning Approach for Sex Estimation with the Pelvis ». *Forensic Science International* 302 (septembre 2019): 109873.
<https://doi.org/10.1016/j.forsciint.2019.109873>.
- [12] Etlı, Yasin, Mahmut Asirdizer, Yavuz Hekimoglu, Siddik Keskin, et Alpaslan Yavuz. « Sex Estimation from Sacrum and Coccyx with Discriminant Analyses and Neural Networks in an Equally Distributed Population by Age and Sex ». *Forensic Science International* 303 (octobre 2019): 109955.
<https://doi.org/10.1016/j.forsciint.2019.109955>.
- [13] Oner, Zulal, et al. « Sex Estimation Using Sternum Part Lengths by Means of Artificial Neural Networks ». *Forensic Science International*, vol. 301, août 2019, p. 6-11. *DOI.org (Crossref)*,
- [14] Bewes, James, et al. « Artificial Intelligence for Sex Determination of Skeletal Remains: Application of a Deep Learning Artificial Neural Network to Human Skulls ». *Journal of Forensic and Legal Medicine*, vol. 62, février 2019, p. 40-43. *DOI.org (Crossref)*, <https://doi.org/10.1016/j.jflm.2019.01.004>.
- [15] Zou, Kaifeng, et al. *Disentangled representations: towards interpretation of sex determination from hip bone*. 2021. *DOI.org (Datacite)*, .
- [16] Rémi Agier, Sébastien Valette, Laurent Fanton, Pierre Croisille and Rémy Prost, Hubless *3D Medical Image Bundle Registration*, In Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: VISAPP, 265-272, 2016, Rome, Italy. <https://hal.archives-ouvertes.fr/hal-01284240>
- [17] Fedorov A., Beichel R., Kalpathy-Cramer J., Finet J., Fillion-Robin J-C., Pujol S., Bauer C., Jennings D., Fennessy F.M., Sonka M., Buatti J., Aylward S.R., Miller J.V., Pieper S., Kikinis R. *3D Slicer as an Image Computing Platform for the Quantitative Imaging Network*. *Magnetic Resonance Imaging*. 2012 Nov;30(9):1323-41. PMID: 22770690. PMCID: PMC3466397
- [18] S. Valette, J.-M. Chassery and R. Prost, *Generic remeshing of 3D triangular meshes with metric-dependent discrete Voronoi Diagrams*, *IEEE Transactions on Visualization and Computer Graphics*, Volume 14, no. 2, pages 369-381, 2008.