Forensic Sciences Research, 2023, **8**, 202–210
https://doi.org/10.1093/fsr/owad030
Advance access publication date 22 September 2023
Research Article

Ancestry estimation in forensic anthropology: accuracy of the AncesTrees software in a Brazilian sample

Victor Jacometti[1], Marco Aurelio Guimarães[1], Luis Otávio Carvalho de Moraes[2], Sérgio Ricardo Marques[2], Eugénia Cunha[3,4], Ricardo Henrique Alves da Silva[5,*]

[1]Department of Pathology and Legal Medicine, Ribeirão Preto Medical School, University of São Paulo, 14048-900 Ribeirão Preto, Brazil
[2]Discipline of Descriptive and Topografic Anatomy, Department of Morphology and Genetics, Federal University of São Paulo, 04024-002 São Paulo, Brazil
[3]Laboratory of Forensic Anthropology, Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal
[4]Departamento de Ciências da Vida, Universidade de Coimbra, Calçada Martim de Freitas, 3000-456 Coimbra, Portugal
[5]Department of Stomatology, Public Health and Forensic Odontology, School of Dentistry of Ribeirão Preto, University of São Paulo, 14040-904 Ribeirão Preto, Brazil

*Corresponding author. E-mail: ricardohenrique@usp.br

Abstract

The objective of this study is to analyze the accuracy and applicability of the AncesTrees software with respect to a set of cranial measurements of a Brazilian sample consisting of 114 identified skulls from two osteological collections, predominantly composed of European ($n = 59$), African ($n = 35$), and admixed individuals ($n = 20$). Twenty-four different craniometric measurements are performed and input to AncesTrees *via* two algorithms, one of which is used in three configurations, with different ancestral groups integrated in the model. The software exhibits superior performance in the estimation of European individuals, reaching 73% accuracy, compared with 66% in the African individuals. Those individuals classified as admixed produce a variety of ancestral classifications, mainly European. Overall, the most accurate combination of AncesTrees is obtained using ancestralForest with only the European and African groups integrated into the algorithm, where the accuracy reaches 70%. The applicability of this software to a specific population is fragile because of the high admixing load, making it necessary to create a more representative anthropometric database of the Brazilian people.

Key points

- Ancestry estimation methods are seldom validated in Brazil.
- AncesTrees performed poorly on our sample, with a maximum accuracy of 70%.
- Brazil's highly mixed population hinders ancestry estimation.
- Mixed individuals (pardos) are predominantly classified as Europeans.
- The insertion of Brazilian metric data into the AncesTrees database would produce better results.

Keywords: forensic anthropology; ancestry estimation; biological profile estimation; Brazilian; craniometry

Introduction

Anthropological methods are used to study and estimate the biological profiles of unknown specimens, often in a skeletonized state, by analysing their species and estimating their sex, ancestry, age, and stature, as well as characteristics with individualizing potential [1]. This set of general assets aids the identification process by significantly reducing the universe of suspects for a given unknown skeleton or individual [1, 2].

Ancestry estimation is considered the most difficult step of biological profile construction [3]. Historically, metric and non-metric methods have been used to correlate the findings with the ancestry estimation of humans [4, 5]. When metric methods are employed for ancestry investigation, the skeletal structures are measured and statistically analysed under the premise that such dimensions follow locally geographic craniometric variations [6]. These variations allow the use of linear discriminant analysis to interpret the data obtained and separate them into distinct classes, such as different ancestries [7].

The craniometric databases compiled by Howells [8–10] provide the main basis for many subsequent studies [11]. More recently, there has been considerable discussion of the Random Forest machine learning technique, which is based on the construction of multiple "decision trees" with results based on the modal class or mean prediction (regression) of the "individual trees" [12]. This novel statistical tool produced satisfactory results in ancestry studies conducted by Hefner et al. [13, 14] based on non-metric skull traits, with an accuracy of 89% obtained in the ancestral prediction of North

Received: August 4, 2022. **Accepted:** March 16, 2023

American samples (American Caucasians, American Blacks, and Southwest Hispanics).

Navega et al. [15] developed the AncesTrees software following similar methods to those cited above. The Random Forest algorithm is used to estimate ancestry, albeit with metric data including skull measurements. As the model training dataset, the Howells craniometric database (among others) is used. After proper training of the algorithm, testing on an independent sample composed of skeletons of African and European origin produced high accuracy values. However, this new method has not yet been widely tested and globally validated for different populations. The optimal functioning of this method can only be achieved by increasing the size and scope of its database, which in turn is achieved through the inclusion of individuals from more diverse demographics.

Scientific advances, as well as new techniques and technologies, have enhanced the quality of forensic studies and their results. Nonetheless, a heated debate has emerged regarding ancestry estimation, especially in the USA [16]. Several biological and forensic anthropologists have criticized certain ancestry estimation approaches as a biased reinforcement of biological races, wrongly tying racial hierarchies to typological features. They also reject the notion that ethnicities are discrete variables, which is a prevalent assumption in older approaches that are still carried out to this day [17]. These approaches, along with the traditional three-group method of ancestry classification (African, European, Asian), are being debunked as inappropriate for the current state of humankind.

In Brazil, ancestry has historically lacked objectivity. Ancestry is mostly noted in civil register documents and is, by law, a self-declared characteristic. Most public institutions use and provide inadequate and simplistic terms to refer to ancestry—whites, blacks, pardos (used for nearly every mixed individual), and yellows (used for Asians). This inheritance of antiquated times still impacts forensic sciences in Brazil. Combined with the high admixture load of Brazilians, this renders ancestry as a poorly defined trait nationwide. The above-mentioned ancestry trifecta neither suit nor affect Latin American samples, and studies that employ population-specific structure models, embracing human and historical variation, must be developed and considered [18].

Following the above discussion, and motivated by the high degree of miscegenation of the Brazilian population, with its complex demography and considerable variability within regions, which complicates ancestry estimation [19], this study examines the AncesTrees program developed by Navega et al. [15]. Specifically, we evaluate its accuracy in estimating the ancestry of a Brazilian osteological sample. Additionally, we explore ways of increasing the database of AncesTrees to make it more applicable and adequate as a forensic tool for use in Brazil.

Materials and methods

The sample discussed in this paper consists of a Brazilian osteological collection of 138 identified skulls from the state of São Paulo, Brazil [20] (sex, skin colour, and age at death). This is a contemporary collection including skulls that belonged to individuals who died from 1930 to 1970. The elements required to identify each of these individuals were handwritten in a coroner's register book.

The sole source of information regarding their ancestry was skin colour or tone, also noted by the medical examiner, probably by checking their personal documents or by clinical post-mortem observation (no further detailing was available). These annotations were then used to cluster the samples into three large groups: European, African, and pardos (Brazilian term for mixed individuals). No Asians were present in the sample.

Exclusion criteria included excessively fragmented or comminuted skulls, skulls with severe pathological and congenital changes of cranial dimensions, and juvenile/newborn crania, which have a higher risk of bias in the measurements. All individuals aged 18 years or older were included in the analysis. Thus, 114 skulls were eligible for the analysis. The sample contained 73 male individuals, with a mean age of 46.40 years (standard deviation of 12.9 years) and 41 females with a mean age of 37.85 years (standard deviation of 16.64 years). Regarding real ancestry, there were 59 European individuals, 35 African individuals, and 20 pardos. Descriptive statistics of the sample are presented in Table 1.

Measurements were made according to the descriptions supplied by Howells [12], using an analogue calliper (OXD 330-7080X, Oxford Precision Components™, UK), sliding digital calliper (Sliding Caliper 300 mm, DIGIMESS™, Brazil), 30-cm metal ruler (FLEX-30, Trident™, Brazil), and a regular pencil (Max Ecolapis Blue Hb/n2 with rubber, Faber Castell™, Bad Schwartau, Schleswig-Holstein, Germany).

The 23 measurements described by Navega et al. [15] were made by a single examiner, under indirect light and using proper equipment for each measurement. The examiner was blinded to the data of each subject (sex, age, and ancestry). All measurements were made twice using the craniometric landmarks indicated. The examiner then noted the data in a spreadsheet.

In this study, only the intra-examiner error is assessed, as only one examiner was available for data collection. The relative technical error of measurement (TEM) test was used, following the methodology recommended by Perini et al. [21]. This method obtains the systematic error index produced by the evaluator for each measure. The measurements were reevaluated in 17 skulls following a time interval of ~30 days.

The AncesTrees programme was used to generate an estimate of ancestry. AncesTrees consists of a spreadsheet and a script that functions through a website (http://osteomics.com/AncesTrees/). Measurements were entered into the worksheet, and then ancestral groups were chosen for the decision tree model. The detailed mathematical functioning is complex; details can be found in Ross et al. [18].

We used the two algorithms that AncesTrees makes available for data analysis, called "ancestralForest" and "tournamentForest". The ancestralForest algorithm was used with 512 generated trees; 32 sub-forests; and 63.2% bootstrap,

Table 1. Sample descriptive statistics.

Item	*n*	Age		
		Mean	SD[a]	Range
Male	73	46.40	12.90	21–72
Female	41	37.85	16.64	18–87
Total	114	43.31	14.90	18–87

[a] Standard deviation.

without reallocation and balanced, with parallel computing. In tournamentForest only the number of trees can be adjusted. This option was set to 512, as for ancestralForest. The choice of parameters was made with the intention of changing as little as possible in comparison with the original study. However, the numbers of trees and sub-forests were reduced to avoid over-adjustment of the model, which may be caused by a high number of generated decision trees [15]. In other words, an over-adjusted model captures too much residual variation (or noise) in the validation set (in this case, our sample) and integrates this into the model. The result would be a reduced ability to generalize or predict unseen data [22], making the tool less effective for forensic scenarios.

The ancestral groups involved in each analysis were also tested in different configurations. In the ancestralForest algorithm, three models were tested: one with the nine available ancestral groups inserted in the analysis; a second with only groups coming from Africa, Asia, and Europe; and a third with only the ancestral groups from Africa and Europe incorporated. The tournamentForest algorithm was executed with all nine ancestral groups in the model because this is recommended by the developer as a more robust configuration for cases where there is little to no knowledge of the origin of the studied specimens (i.e. forensic scenarios).

After the measures had been analysed by AncesTrees, we considered the predicted ancestry to be that indicated as the most probable ancestry by ancestralForest or the one that came first in the tournament in tournamentForest. The estimated ancestry was then compared with the real ancestry of the individual.

To assess the accuracy of the software, the following outcomes were considered to be correct predictions: European individual estimated as European; African individual estimated as one of the African groups; Asian individual estimated as any of the Asian groups. We analyse the individuals classified as pardos separately, as they are not sufficiently represented in any major ancestral group.

Descriptive statistics including the mean, standard deviation, and range of sample characteristics and cranial measurements were tabulated. The accuracy of each algorithm was assessed as described above and presented in absolute and relative terms. One-way analysis of variance was used to compare the measurement means between groups and Tukey's *post hoc* test was further used, if necessary, for pairwise comparisons when the differences were significant. The significance level was set to $\alpha = 5\%$.

Results

The relative TEM of each measure is described in Figure 1. TEM reflects the magnitude of the error relative to the size of the measurements. Higher TEM values correspond to greater variability of non-systematic errors by the examiner, i.e. lower agreement among repeated measurements. The highest disagreement values of around 2.5% occur in the measures that involve the orbits (OBB, OBH, and DKB) and in the NPH and NLH measurements.

Descriptive statistics of the cranial measurements are presented in Table 2. On average, the male measurements have higher values than the female measurements. Structures more likely to have deteriorated, such as the alveolar ridge, were less able to be measured because such damage often hindered the process. In Table 3, descriptive statistics of the cranial measurements per ancestry group are listed. No clear pattern of measurement size can be observed between ancestries, but when the means are significantly different ($P < 0.05$), pairwise tests indicate that the differences are mostly between white and black individuals. Admixed or pardo individuals present intermediate values.

Table 4 presents data on the performance of the different arrangements of algorithms and parameters used on the studied sample. The accuracy of the algorithm was calculated for each ancestral group within each algorithm and in total. Figure 2 illustrates the performance of the various algorithms on European and African individuals. The tournamentForest algorithm gives poor results, with a combined accuracy of 48%. The accuracy of ancestralForest varies from 54% to 70% depending on which ancestral groups were inserted into the algorithm during the analysis. In general, the software performs better on European individuals, although tournamentForest has a slightly better accuracy for Africans (nearly 2% ahead of that for Europeans).

The individuals classified as admixed present high levels of inconsistency and variability in their ancestry classification. Thus, we opted to perform an observational analysis, i.e. to observe the prevalence of different ancestral classification outputs for these individuals. The results of this analysis are

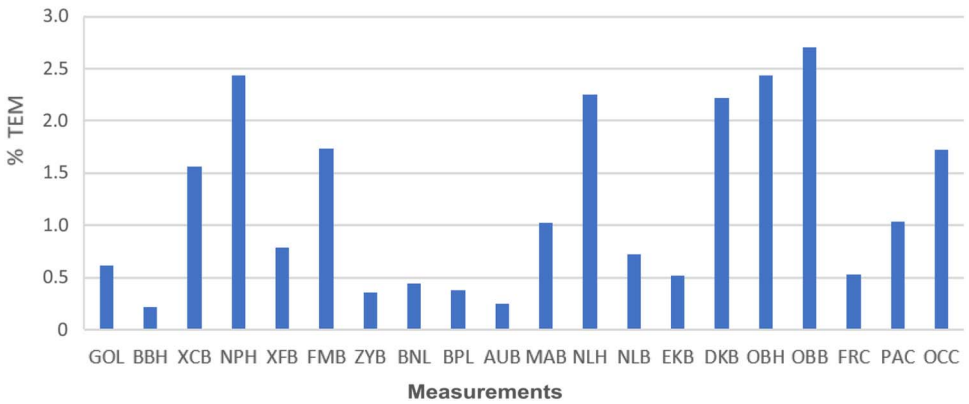

Figure 1 Relative technical error of measurement; higher percentages correspond to larger variance of non-systematic errors by the examiner, i.e. lower agreement among repeated measures. TEM: technical error of measurement.

Table 2. Descriptive statistics: cranial measurements.

Measurement	Male ($n = 73$)				Female ($n = 41$)			
	n	Mean	SD	Range	*n*	Mean	SD	Range
GOL	73	183.98	6.33	168.00–126.30	41	175.33	6.24	162.00–187.30
XCB	72	141.15	5.58	116.00–113.00	41	136.18	5.47	125.60–146.60
ZYB	71	130.81	5.90	92.20–83.60	39	122.62	4.68	134.70–111.50
BBH	72	133.35	6.47	45.00–98.00	41	128.85	5.45	112.30–139.50
BNL	72	101.72	4.15	110.00–60.00	41	96.23	4.09	105.70–89.00
BPL	63	98.32	6.86	101.00–87.60	40	95.01	5.25	86.00–110.00
MAB	67	61.36	5.88	46.60–20.60	41	60.25	4.20	48.70–70.00
ASB	57	111.98	5.28	36.00–33.00	40	107.82	5.18	98.50–121.00
AUB	72	122.53	5.91	89.00–15.50	41	116.45	4.55	105.00–125.30
NPH	64	68.49	4.61	104.20–76.00	40	63.84	5.22	53.40–77.60
XFB	72	120.60	5.36	84.00–199.00	41	115.57	5.15	105.00–126.60
WFB	57	96.83	4.83	155.70–145.00	40	92.82	4.43	85.60–103.00
NLH	72	53.48	4.33	147.80–110.45	41	49.19	3.35	55.60–40.00
NLB	73	25.26	2.13	110.60–76.00	41	25.06	2.15	19.00–30.00
OBB	73	41.04	1.96	125.00–139.00	41	39.53	1.66	35.70–43.00
OBH	73	37.13	1.83	82.50–132.00	41	36.02	1.99	32.00–41.40
EKB	73	98.36	3.82	109.00–74.70	41	94.67	3.56	86.60–101.40
DKB	73	21.86	2.77	29.60–45.40	41	20.67	2.57	15.00–26.00
FRC	73	112.20	4.68	41.00–107.70	41	107.23	4.67	95.40–116.60
PAC	73	113.55	7.56	30.00–124.00	41	108.43	6.46	95.00–125.30
OCC	72	97.09	5.67	125.85–110.00	41	95.45	5.97	81.90–106.70

SD: Standard deviation.

Table 3. Cranial measurements per ancestry.

Measurement	Mean ± SD (range)			*P*-value
	European ($n = 59$)	African ($n = 35$)	Admixed ($n = 20$)	
GOL	182.85 ± 7.26 (166.7–199)[a]	178.02 ± 6.69 (166.7–193)[b]	180.36 ± 8.11 (162–195.75)[ab]	**0.009**
XCB	140.51 ± 5.97 (125.6–155.7)[a]	137.12 ± 5.39 (125.6–147.8)[b]	139.86 ± 6.40 (128.5–154.5)[ab]	**0.028**
ZYB	128.79 ± 6.63 (111.5–145)	126 ± 6.47 (116.5–139.7)	128.83 ± 7.17 (116–142.6)	0.127
BBH	133.4 ± 8.69 (116.5–178.8)	129.9 ± 6.53 (113–147.4)	132.25 ± 6.42 (112.3–141.4)	0.059
BNL	100.52 ± 4.63 (91.4–110.45)[a]	97.98 ± 4.95 (89–107.7)[b]	100.52 ± 4.98 (91–107.55)[ab]	**0.037**
BPL	94.94 ± 6.20 (83.6–109)[a]	99.60 ± 6.03 (90–110.6)[b]	98.51 ± 6.20 (86–108.4)[ab]	**0.002**
MAB	58.8 ± 4.97 (45–68.3)[a]	63.19 ± 3.92 (59.7–72.7)[b]	63.34 ± 6.21 (48–76)[b]	**<0.001**
ASB	111.44 ± 5.18 (98–120)	108.75 ± 5.87 (100–125)	109.87 ± 5.82 (98.5–121.4)	0.090
AUB	121.27 ± 6.42 (105–139)	118.58 ± 5.23 (111.6–132.3)	120.67 ± 6.60 (109.5–133)	0.120
NPH	66.69 ± 4.88 (54.6–75.7)	66.30 ± 5.79 (56–82.5)	67.56 ± 6.05 (53.4–76)	0.700
XFB	119.74 ± 5.92 (101–132)[a]	116.51 ± 5.29 (105–128)[b]	119.95 ± 5.41 (111–130)[ab]	**0.010**
WFB	95.71 ± 4.87 (87–109)	94.08 ± 4.89 (86–106)	95.95 ± 5.92 (85.6–105.6)	0.290
FMB	103.73 ± 5.06 (91.4–116.5)	103.20 ± 5.40 (94–116.8)	103.97 ± 5.49 (94–113)	0.840
NLH	52.76 ± 4.24 (40–74.7)[a]	50.14 ± 4.79 (41–68)[b]	52.61 ± 3.92 (47–61)[ab]	**0.017**
NLB	24.77 ± 2.23 (19–30)	25.83 ± 1.79 (20. –29)	25.32 ± 2.18 (21–29.5)	0.060
OBB	40.55 ± 2.01 (35.7–45.4)	40.16 ± 1.80 (37–44.6)	40.93 ± 2.23 (37.8–45)	0.360
OBH	36.71 ± 2.07 (32–41.4)	36.42 ± 1.81 (32.2–41)	37.31 ± 1.79 (34.3–40)	0.270
EKB	96.79 ± 4.10 (86.6–107.7)	96.92 ± 4.21 (90.5–107)	97.97 ± 4.09 (90.6–104.4)	0.530
DKB	21.29 ± 2.68 (15.5–27.6)	21.78 ± 3.07 (15–30)	21.23 ± 2.42 (16.6–27.4)	0.660
FRC	111.52 ± 5.31 (95.4–124)[a]	108.22 ± 4.94 (100.5–122)[b]	110.99 ± 4.53 (104.75–119.35)[ab]	**0.009**
PAC	113.01 ± 7.59 (76–125)[a]	108.05 ± 6.96 (92–120)[b]	114.25 ± 6.45 (101.4–125.85)[a]	**0.001**
OCC	96.96 ± 6.21 (81.9–110.3)	97.05 ± 5.29 (89.2–108)	94.15 ± 5.12 (85–101.6)	0.130

The bold *P*-values indicate significant differences in means (analysis of variance). Different superscript lower case letters (a, b) indicate significant differences in a row. SD: standard deviation.

presented in Table 5. There is significant variability in the classification distribution of these individuals, but they are mostly classified as European, except for the instance with only two groups included in the algorithm (European and African), in which case they are more often classified as African.

Discussion

When examining skeletal pieces of unknown origin for human identification, one of the main tasks of forensic anthropology is to establish a biological profile, which is a listing of four fundamental, but somewhat generic, features: sex, age, ancestry, and probable stature of that individual [23–25]. Sex and ancestry occupy a prominent place in these estimates and should be studied first, because age and stature estimations usually rely on these parameters [6, 26]. In a forensic context, this task is often performed in scenarios of incomplete skeletons and other fragmented remains. Thus, the skull is often a remaining piece available for analysis [27].

Table 4. AncesTrees analysis results and accuracy of its different algorithms (correct classifications), per ethnic group and in total.

Algorithm	Real ancestry	Estimated ancestry	Accuracy (%)
TournamentForest (six groups)	African ($n = 35$)	African, **17 (49%)** European, 4 (11%) Asian, 3 (9%) Australian, 3 (9%) American, 5 (14%) Polinesia, 3 (9%)	48
	European ($n = 59$)	African, 20 (34%) European, **28 (47%)** Asian, 3 (5%) Australian, 2 (3%) American, 2 (3%) Polinesia, 4 (7%)	
AncestralForest (six groups)	African ($n = 35$)	African, **8 (23%)** European, 5 (14%) Asian, 3 (9%) Australian, 11 (31%) American, 4 (11%) Polinesia, 4 (11%)	54
	European ($n = 59$)	African, 4 (7%) European, **43 (73%)** Asian, 5 (8%) Australian, 5 (8%) American, 0 (0%) Polinesia, 2 (3%)	
AncestralForest (three groups)	African ($n = 35$)	African, **23 (66%)** European, 8 (23%) Asian, 4 (11%)	65
	European ($n = 59$)	African, 15 (25%) European, **38 (64%)** Asian, 6 (10%)	
AncestralForest (two groups)	African ($n = 35$)	African, **23 (66%)** European, 12 (34%)	70
	European ($n = 59$)	African, 16 (27%) European, **43 (73%)**	

The bold values correspond to the number of correct estimates within the ethnic group, in relative and absolute numbers. [a]Pecentages may not add up to 100% due to rounding.

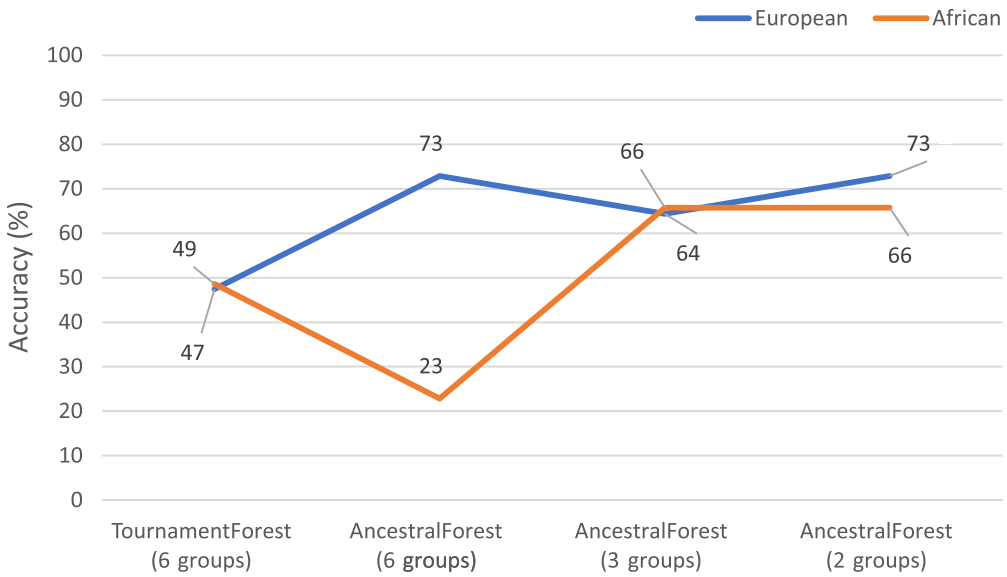

Figure 2 The accuracy of each AncesTrees algorithm type used, within black and white individuals; accuracy is understood as the relative number (percentages) of correct ancestral classifications for blacks (African group assignments) and whites (Caucasian group assignments).

Table 5. Ancestry estimation for mixed individuals ($n = 20$).

Ancestry	TournamentForest (6 groups)	AncestralForest (6 groups)	AncestralForest (3 groups)	AncestralForest (2 groups)
European	8	10	10	9
African	7	4	6	11
Asian	2	2	4	NA
Australian	2	2	NA	NA
American	1	2	NA	NA

NA: not applicable.

To ensure the reliability and reproducibility of the results, it is crucial to assess the measurement errors [28]. Acceptable levels of intra-examiner relative TEM are not easy to stipulate because they depend on several reports in the literature that set this threshold. In other words, larger error margins are more likely to be observed for some measures than others. A higher relative TEM value corresponds to a greater intra-observer error variability [21]. Fancourt and Stephan [29] considered the TEM as the best scale/index for evaluating the error between cranial measurements. Using 2 000 simulated cranial measurements (Glabella–Opisthocranion distance), the authors reported maximum TEM values of 2.2%, similar to the value obtained in this study. Some measurements made in our study obtained a higher relative TEM because of the subjectivity of the reference points, particularly those at the orbital margins (OBB and OBH measurements), and areas more susceptible to wear, damage, and ridge resorption (NPH, NLH, and DKB).

To take measurements, one must know the cranial landmark placings and how they are affected by the diversity of forms and shapes that a skull can present. Additionally, it is imperative that appropriate osteometric techniques are used during the measurement process. The software used in this study has a user-friendly interface, allowing easy interpretation of the results. Thus, a detailed understanding of the complex algorithm behind the statistical operations is not necessary, although basic statistical knowledge is desirable for any researcher. The software has an interesting feature that alerts the examiner if the measure is atypical and excessively far from the mean values in the database. In this case, a repeated measurement is suggested to avoid systematic errors.

Among the evaluated ancestries, AncesTrees performed best with European individuals and the ancestralForest algorithm, with only the European and African ancestry groups included. This finding is in good agreement with that reported by Navega et al. [15], who applied AncesTrees to osteological collections from Portugal. However, Navega et al. obtained higher accuracy values. In our study, the maximum correct classification rates were 66% for Africans and 73% for Europeans, whereas Navega et al. obtained an accuracy of 93% for Africans and 94% for Europeans using the same algorithm configuration. This is likely to be directly related to the sample's background characteristics. While the former study analysed skulls from a fairly homogeneous ancestry background (including those of African descent, which were removed from a collection composed of African slaves), our work dealt with the opposite.

Our values are not acceptable for forensic purposes and validate our previous assumption: only after the inclusion of numerous Brazilian populational data in AncesTrees will it be possible to achieve better results. The databases must include individuals with Brazilian variability for comparison. Otherwise, the pattern of variables found will often be considered atypical (as no one in the database will have similar features). A better and more detailed understanding of Brazil's craniometric patterns is essential, as not all measures seem to have statistically significant differences. Nonetheless, a well-adjusted random forest model such as AncesTrees could successfully identify cut-off values between different ancestries, if sufficiently discrepant.

Brazil is known for its wide and heterogeneous distribution of three major ancestral contributions, namely Americans (Natives/Indians), Europeans, and Africans, giving birth to a highly mixed and multiethnic population [20, 30]. Colonization by Europeans and Africans generally began on the coast and progressed inland, and this progression occurred in various manners in different regions of the territory [31]. This multi-layered process, combined with a continental-sized country, has drastically influenced the variation of the genetic ancestral composition of the current population [32, 33].

Considering such a high miscegenation load as a major confounding factor, the authors separately analysed individuals whose real ancestry was assigned as pardo (mixed). This term is generally attributed to people of hybrid European/African ancestry, but may also refer to the mixture between Native (Indian) and European [34].

We conducted an observational analysis of these individuals to determine the distribution of estimated ancestry classifications given by AncesTrees. The results show that the predicted ancestry for these subjects is well distributed in most of the algorithms, with a slight tendency to the European estimate, except with the algorithm that only considers two groups (European and African). In this last case, the majority of the sample was classified as African. We believe that this might be the result of similarities between some of the groups omitted in this two-group algorithm and the morphometric characteristics of the African skulls.

The distribution of pardo data, which was mainly allocated to European ancestry, followed the AncesTrees classification pattern and tendency in the rest of the sample. As stated by several Brazilian genetic studies [30, 35], including a meta-analysis by Moura et al. [36], European ancestry is the major contributor to the ancestral genetic background of the Brazilian population. This is more pronounced in urban populations [35] and in the Southeast region [36], as observed in our work, which links our results with the findings of forensic DNA studies.

Even though European ancestry is the predominant contributor to the ancestral mosaic of the current Brazilian population, it does not overshadow its high miscegenation. There are specific regions in Brazil that have a high rate of variation

in genomic ancestry, despite being inhabited hegemonically by populations of European descent [37]. This adds to the previous hurdles faced by Brazilian ancestry estimation, as evidence shows that genetic variations related to ancestry are directly linked to the phenotype variations of the human skull (morphological shapes) [38].

In this study, the skin colour determined in the medical examiner's identification records, gathered in civil documents (self-reported), or stated during posthumous examination was used as the reference register for evaluating the accuracy of the AncesTrees algorithm, as this was the only "real" ancestry information available. However, this criterion is definitively not the best determinant of an individual's ancestry. Other studies that attempted to correlate skin tone with genetic ancestry in Brazil reported divergent results, some finding a positive correlation [39–41], while others found the opposite [42, 43]. The longer the miscegenation process lasts, the more the genetic ancestral pattern becomes dissociated from the individual's skin colour [42]. Therefore, if we consider over 500 years of miscegenation that shaped Brazil's current population, we have quite a problematic background, which imposes a clear limitation on this study.

The best overall performance across the entire combined sample (Europeans and Africans) was achieved by the ancestralForest algorithm with only the African and European groups selected, for which the accuracy reached 70%. In Brazil, few studies have assessed the ancestry of the population based on craniometric data. Urbanová et al. [34] and Jurda & Urbanová [44] carried out such analyses, but used other software (FORDISC and 3D-ID in the former study and FIDENTIS Analyst, which analyses graphics meshes, in the latter).

Urbanová et al. [34] used FORDISC 3 in conjunction with the Howells craniometric database (as used by AncesTrees). They reported that 40% of individuals were correctly allocated to their respective ancestral group. In the same configuration, mixed individuals were mostly allocated to the African ancestral group, different from our results. However, by switching to the Federal Bureau of Investigation dataset, known as Forensic Data Bank, the classification predominance of the mixed group changed to European, in agreement with our results. This highlights the unpredictability of estimating the ancestry of the Brazilian population by metric variables. Urbanová et al. [34] found that the 3D-ID software provided a more accurate estimate of ancestry, although this was still not as accurate as our results.

Jurda and Urbanová [44] applied an alternative method that quantifies differences between models based on the closest distance between specific points. In other words, using a computerized comparison of scanned meshes of a given skull, the programme establishes correspondences between "vertices" of different skulls. After this analysis, poor results were obtained, with an accuracy of 52.5% in ancestry estimation; this is not much better than estimation by random chance.

Fernandes et al. [45] tested AncesTrees on a Brazilian osteological collection from the city of Campinas, relatively close to the city of São Paulo, from where our sample derived. Across a sample of 244 skulls, they found similar numbers to ours, although with lower accuracy (total accuracy of 56.77%). Black individuals had better correct classification rates than white ones overall, whereas whites were classified more accurately than blacks in our study. Their analysis of admixed individuals produced similar ancestral allocations as in our study. The authors did not disclose from which time period or how old their sample was. However, their worse results could possibly have been caused by a younger, more recent, and therefore more miscegenated sample than ours.

The difficulty in determining the ancestry of unknown Brazilians is mainly attributable to the factors identified above. As stated by Urbanová et al. [34] and Fernandes et al. [45], inconsistent results occur because these algorithms and software do not incorporate an appropriate database. This anomalous pattern, unique to Brazilian skulls because of the high level of miscegenation, also influences the estimation of other biological parameters, such as sex [34, 44], thus impairing their accuracy as well.

Another confusing factor is introduced by the chronology of Howells' samples. Most of his population samples are much older than the collection from which this work's sample was obtained [11–13]. His European samples, e.g. are either medieval or from the 17th century. Our sample, as previously stated, contains individuals who died from 1930 to 1970, giving a much younger sample. This raises the question of whether methods based on Howells' dataset, or even the set itself, are appropriate for estimating ancestry on such young and newly formed populational patterns. However, the use of AncesTrees in our study obtained the highest ancestry estimation accuracy among craniometric-based methods or software.

The lack of individuals with Brazilian ancestry in the AncesTrees software database restricts the precision and performance of this software. Remarkably, in validating the input data, AncesTrees analyses the metric values and provides a report on whether they are typical or atypical. This "atypical" label is output for a set of values that does not resemble the original composition of the programme database [18]. In this study, many of the samples were labelled as atypical, reinforcing previous statements and the results obtained.

Finally, our findings illustrate that, although AncesTrees performed slightly better than other software tested in Brazil, its application in the forensic context is still fragile. Cunha et al. [20] documented an abundance of osteological and cranial collections distributed throughout the Brazilian territory and emphasized that the documentation of metric and qualitative data of identified samples, together with the validation of various anthropological methods, would provide great assistance in solving the problems of forensic anthropology in the Brazilian context. With the accomplishment of this work, it is expected that new paths and horizons will open for the continuity of the evaluation and validation of different anthropological methods for estimating biological parameters in most Brazilian collections, establishing a consistent, reliable, and representative database of our population.

Given the increment of the AncesTrees database with data collected in this study, the performance of the software should improve significantly. As soon as this database contains a relevant number of Brazilian metric values and standards, the correct classification probabilities will increase exponentially, as has happened with other similar programmes [46]. This will enable AncesTrees to be applied to forensic practice in the Brazilian context.

Conclusions

Considering the findings of this study, the accuracy of AncesTrees with the Brazilian sample analysed herein was

48%–70%, depending on the configuration of the ancestral groups within the algorithm. These results are not sufficient for forensic quality standards. The software achieves superior performance in the estimation of white individuals when using the ancestralForest version, and when only the European and African ancestral groups are included. Therefore, its applicability in Brazil is still fragile and requires better comprehension about Brazilian cranial metric patterns, a more consistent database, and decision tree models that account for the high miscegenation of the Brazilian population. By including a sufficient number of Brazilian cranial patterns in the software database, the correct classification rate is likely to increase, possibly allowing its use in Brazilian forensic cases.

Acknowledgements

We would like to thank the Skull Museum of the Discipline of Descriptive and Topografic Anatomy, Department of Morphology and Genetics, Federal University of São Paulo (UNIFESP) and its curators for support and collaboration, and for establishing an important partnership throughout this research.

Authors' contributions

Victor Jacometti carried out the metric and statistical analyses and drafted the manuscript; Marco Aurelio Guimarães provided technical help and drafted the manuscript; Luis Otávio Carvalho de Moraes provided access to samples, technical help, and revised the manuscript; Sérgio Ricardo Marques provided access to samples, technical help, and revised the manuscript; Eugénia Cunha trained the first author for cranial measurements, conceived the study, participated in its design and coordination, and helped to draft the manuscript; Ricardo Henrique Alves da Silva conceived the study, participated in its design and coordination, and helped to draft the manuscript. All authors contributed to and approved the final text.

Compliance with ethical standards

The research was approved by the Ethics Committee of Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto—Universidade de São Paulo (HCFMRP-USP) under the number CAAE: 79913817.2.0000.5440, fully addressing regulatory issues of national authorities.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES) grant number 001.

References

1. Passalacqua NV. Forensic age-at-death estimation from the human sacrum. J Forensic Sci. 2009;54:255–262.
2. Klales AR, Kenyhercz MW. Morphological assessment of ancestry using cranial macromorphoscopics. J Forensic Sci. 2015;60:13–20.
3. Christensen AM, Passalacqua NV. A laboratory manual for forensic anthropology. London (UK): Elsevier Inc., 2018.
4. Ousley SD, Jantz R, Freid D. Understanding race and human variation: why forensic anthropologists are good at identifying race. Am J Phys Anthopol. 2009;139:68–76.
5. Cunha E, Ortega PA. Como los antropólogos forenses evalúan la ancestría? [How forensic anthropologists evaluate ancestry?]. In: Sanabria-Medina C, editor. Patología Y Antropologia Forense de la Muerte: la investigación científico-Judicial de la Muerte Y la Tortura, Desde Las Fosas Clandestinas, Hasta la audiência pública. [Death's forensic anthropology and pathology: the scientific and legal investigation of death and torture, since clandestine pits until public audience]. Bogotá (Columbia): Forensic Publisher, 2016, 221–235. Spanish
6. Iscan MY, Steyn M. Ancestry. In: Krogman WM, editor. The human skeleton in forensic medicine (3 ed). Springfield (IL): Charles C Thomas; 2013.
7. Ousley S, Jantz RL. ForDisc 3 and statistical methods for sex and ancestry estimation. In: Dirkmaat DC, editor. A Companion to forensic anthropology. West Sussex (UK): Wiley-Blackwell, 2012, 311–329.
8. Howells WW. Cranial variation in man. Papers of the Peabody Museum of Archeology and Ethnology. 1973;67:1–259.
9. Howells WW. Skull shapes and the map: craniometric analyses in the dispersion of modern homo. Papers of the Peabody Museum of Archeology and Ethnology. 1989;79:1–189.
10. Howells WW. Who's who in skulls: ethnic identification of crania from measurements. Papers of the Peabody Museum of Archeology and Ethnology. 1995;82:1–108.
11. Sauer NJ, Wankmiller JC. The assessment of ancestry and the concept of race. In: Blau S, Ubelaker DH, editors. Handbook of forensic anthropology and archeology. Walnut Creek (CA): Lest Coast Press, 2009, 187–200.
12. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning, 2nd ed, New York (NY): Springer;2008, p. 587–588.
13. Hefner JT, Spreadley K, Anderson BE. Ancestry estimation using random forest modelling. Proceedings of the American Academy of Forensic Sciences. 2011;352–353.
14. Hefner JT, Spreadley K, Anderson BE. Ancestry assessment using random forest modelling. J Forensic Sci. 2014;59:583–589.
15. Navega D, Coelho C, Vicente R, et al. AncesTrees: ancestry estimation with randomized decision trees. Int J Leg Med. 2015;129:1145–1153.
16. DiGangi EA, Bethard JD. Uncloaking a lost cause: decolonizing ancestry estimation in the United States. Am J Phys Anthropol. 2021;175:422–436.
17. Dunn RR, Spiros MC, Kamnikar KR, et al. Ancestry estimation in forensic anthropology: a review. WIREs Forensic Sci. 2020;2:e1369.
18. Ross AH, Williams SE. Ancestry studies in forensic anthropology: back on the frontier of racism. Biology (Basel). 2021;10:602.
19. Francisco RA, Evison MP, da Costa Júnior ML, et al. Validation of a standard forensic anthropology examination protocol by measurement of applicability and reliability on exhumed and archive samples of known biological attribution. Forensic Sci Int. 2017;279:241–250.
20. Cunha E, Lopez-Capp TT, Inojosa R, et al. The Brazilian identified human osteological collections. Forensic Sci Int. 2018;289:449.e1–449.e6.
21. Perini TA, Oliveira GL, Ornellas JS, et al. Cálculo do Erro técnico de medição em antropometria. Rev Bras Med Esporte. 2015;11:81–85. Spanish.
22. Burnham KP, Anderson DR. Model selection and multimodel inference, 2nd ed. Berlin (Germany): Springer-Verlag; 2002, p. 45.
23. Ross AH, Kimmerle E. Contribution of quantitative methods in forensic anthropology: a new era. In: Blau S, Ubelaker DH, editor. Handbook of forensic anthropology and archaeology. Walnut Creek: Left Coast Press, 2009, 479–489.

24. Cunha E. A Antropologia Forense Passo a Passo [forensic anthropology step by step]. In: Gomes A, editors. Enfermagem Forense: Volume 1 [Forensic Nursing: Volume 1]. Lisbon (Portugal): Ed. Lidel, 2014, 280–288.
25. Cunha E. Considerações sobre a antropologia forense na atualidade. [Considerations about forensic anthropology today]. Revista Brasileira de Odontologia Legal RBOL. 2017;4:110–117. Spanish.
26. Guyomarc'h P, Bruzek J. Accuracy and reliability in sex determination from skulls: a comparison of Fordisc® 3.0 and the discriminant function analysis. Forensic Sci Int. 2011;208:180.e1–e6.
27. Sauer NJ. Forensic anthropology and the concept of race: if races don't exist, why are forensic anthropologists so good at identifying them? Soc Sci Med. 1992;34:107–111.
28. Corron L, Marchal F, Condemi S, et al. Evaluating the consistency, repeatability, and reproducibility of osteometric data on dry bone surfaces scanned dry bone surfaces, and scanned bone surfaces obtained from living individuals. Bulletins et Memoires de la Societe d'Anthropologie de Paris. 2017;29:33–53.
29. Fancourt HSM, Stephan CN. Error measurement in craniometrics: the comparative performance of four popular assessment methods using 2000 simulated cranial length datasets (g-op). Forensic Sci Int. 2018;285:162–171.
30. Manta FSN, Pereira R, Vianna R, et al. Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels. PloS One. 2013;8:e75145.
31. Instituto Brasileiro de Geografia e Estatística (IBGE). Brasil - 500 Anos de Povoamento/Brazil - 500 Years of Settlement. Rio de Janeiro (Brazil): IBGE, 2000, 232.
32. Callegari-Jacques SM, Grattapaglia D, Salzano FM, et al. Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. Am J Hum Biol. 2003;15:824–834.
33. Godinho NMO, Gontijo CC, Diniz MECG, et al. Regional patterns of genetic admixture in South America. Forensic Sci Inter Gen Suppl Series. 2008;1:329–330.
34. Urbanová P, Ross AH, Jurda M, et al. Testing the reliability of software tools in sex and ancestry estimation in a multi-ancestral Brazilian sample. Leg Medicine (Tokyo). 2014;16:264–273.
35. Silva TM, Rani MRS, Costa GNO, et al. The correlation between ancestry and color in two cities of Northeast Brazil with contrasting ethnic compositions. Eur J Hum Gen. 2015;23:984–989.
36. Moura RR, Coelho AV, Balbino QV, et al. Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. Am J Hum Biol. 2015;27:674–680.
37. Marrero AR, Leite FPN, Carvalho BA, et al. Heterogeneity of the genome ancestry of individuals classified as white in the state of Rio Grande do Sul. Am J Hum Biol. 2008;17:496–506.
38. Martínez-Abadías N, Esparza M, Sjovold T, et al. Pervasive genetic integration directs the evolution of human skull shape. Evolution. 2012;66:1010–1023.
39. Pena SD, Rodrigues LB, Pimenta JR, et al. DNA tests probe the genomic ancestry of Brazilians. Brazilian J Med Biol Research. 2009;42:870–876.
40. Lins TC, Vieira RG, Abreu BS, et al. Genetic heterogeneity of self-reported ancestry groups in an admixed Brazilian population. J Epidemiol. 2011;21:240–245.
41. Queiroz EM, Santos AM, Castro IM, et al. Genetic composition of a Brazilian population: the footprint of the gold cycle. Genet Mol Res. 2013;12:5124–5133.
42. Parra FC, Amado RC, Lambertucci JR, et al. Color and genomic ancestry in Brazilians. Proc Natl Acad Sci U S A. 2003;100:177–182.
43. Pimenta JR, Zuccherato LW, Debes AA, et al. Color and genomic ancestry in Brazilians: a study with forensic microsatellites. Hum Hered. 2006;62:190–195.
44. Jurda M, Urbanová P. Sex and ancestry assessment of Brazilian crania using semi-automatic mesh processing tools. Leg Med (Tokyo). 2016;23:34–43.
45. Fernandes L, Bento M, Rabello P, et al. Analysis of the accuracy of AncesTrees software in ancestry estimation in Brazilian identified sample. Advances in Anthropology. 2021;11:163–178.
46. Cunha E, Cattaneo C. Historical routes and current practice for personal identification. In: Ferrara SD, editors. P5 medicine and justice: innovation, unitariness and evidence. Berlin: Springer International, 2017, 398–411.