



Федорчук О.А.¹⁾, Гончарова Н.Н.^{1, 2)}

1) МГУ имени М.В.Ломоносова, биологический факультет, кафедра антропологии,
Ленинские горы, д. 1, стр. 12, Москва, 119234, Россия

2) Медико-генетический научный центр им. Н.П. Бочкова,
ул. Москворечье, д. 1, Москва, 115522, Россия,

СРАВНЕНИЕ РАБОТОСПОСОБНОСТИ РАЗНЫХ НАБОРОВ АНТРОПОЛОГИЧЕСКИХ ПРИЗНАКОВ ПРИ ДИФФЕРЕНЦИАЦИИ ЧЕЛОВЕЧЕСКИХ ПОПУЛЯЦИЙ С ПОМОЩЬЮ АЛГОРИТМА «ДЕРЕВО РЕШЕНИЙ»

Введение. Задача дифференциации групп различными статистическим методами не теряет своей актуальности в связи с тем, что стандартные алгоритмы часто имеют ряд ограничений. Часть из них, например, нормальность распределения признаков, отсутствие высоких корреляций между ними, позволяет обойти метод, основанный на поэтапном разделении совокупности выборок – деревья принятия решений.

Материалы и методы. В настоящем исследовании данный метод тестировался на 15 линейных размерах черепа и 16 указателях, рассчитанных на их основе. Разделение происходит на основании дискретности вариационного ряда средних величин. В качестве критерия разнородности групп выбран показатель энтропии. Использованные краниометрические признаки соответствуют стандартной методике, принятой в российской антропологии. Материалами послужили средние значения краниометрических размеров по 39 этно-территориальным группам из 13 макрорегионов Старого Света.

Результаты и обсуждение. По результатам дифференциации можно видеть, что указатели отбирались алгоритмом чаще, чем линейные признаки. В дифференциации участвовали пять линейных размеров и семь указателей. Три указателя основаны на габаритных размерах мозгового отдела черепа (М.1, М.8, М.17), а четыре указателя включают скуловую ширину (М.45). Первое разделение происходит по поперечному фациальному указателю (46:45), который отделяет группы Африки, Восточной, Юго-Восточной и Южной Азии от всех остальных (Северной, Центральной Азии и Европы).

Заключение. Показана высокая дифференцирующая способность указателей, особенно поперечного фациального, лобно-скулового и высотно-продольного указателя. Итерационная система дифференциации позволяет более информативно разделять группы, так как разные сочетания признаков показывают работоспособность на разных иерархических уровнях таксономии.

Ключевые слова: биологическая антропология; краниология; биостатистика; деревья решений

DOI: 10.55959/MSU2074-8132-25-1-5

Введение

Антропологические подходы к дифференциации популяций регулярно пополняются различными методами, позволяющими снизить размерность данных и обьективизировать выбор тех или иных признаков для разработки антропологической классификации. Когда речь идет о многомерных методах изучения межгрупповой изменчивости возникают математические ограничения их применимости – условия нормальности распределения признаков и отсутствия высоких связей между ними, так как высокая связь между признаками может приводить к явлению мультиколлинеарности и ухудшению дифференцирующих возможностей применяемого метода. Этого недостатка лишен способ классификации объектов с применением алгоритма «дерева решений» [Федорчук, Гончарова, 2024], который не является корреляционным методом, то есть не учитывает коррелированность признаков, а кроме того, нечувствителен к форме их распределения. Это делает возможным использование в одном анализе как исходных признаков, так и их производных – индексов, или указателей. Указатели, применяемые в кранио- и остеометрии, практически никогда не имеют нормального распределения, их вариабельность при этом очень высока и превышает показатели изменчивости исходных признаков [Дерябин, 2008, с. 92–94].

Алгоритм «дерево принятия решений» называют также деревом классификации [Breiman et al., 1984, с. 17]. Он представляет из себя иерархический метод анализа данных, дальнейшим развитием которого стал метод случайного леса, и относится к нумерическим методам классификации, в которых предполагается таксономическая равноценность признаков. Однако на разных уровнях дифференциации популяций набор значимых признаков может быть различным. Это обстоятельство нужно учитывать при анализе результатов. Алгоритмы «дерева решений» и «случайного леса» являются еще одним инструментом поиска закономерностей межгрупповой изменчивости и широко применяется при классификации микроорганизмов [Sokal, Sneath, 1963, p. 3; Cartmill, 2018; Hugenholtz et al., 2021], а также в медицине, так как возможности метода позволяют строить прогностические модели результатов лечения [Фельдман, 2021; Wong et al., 2004; Djuris et al., 2013; al Mamun, Keikhosrokiani, 2022].

К числу достоинств обсуждаемого алгоритма «дерева решений» относится простая и наглядная форма визуализации данных. Получаемое по результатам анализа дерево представляет собой дихотомически делящиеся в узловых точках ветви и листья – конечные элементы ветвления. При этом исходное множество объектов разбивается на все более мелкие подмножества на основе дискретности вариационного ряда средних значений признака. Заданный исследователем набор объектов (в случае использования метода для классификации этими объектами становятся выборки) разделяется до тех пор, пока «узел» дерева не станет гомогенным, то есть пока в узле дерева не останутся объекты одного класса, тогда узел становится «листом» дерева. В каждой точке дихотомии на графике указывается признак, который в этой точке играет главную роль разделителя. Слева от узла располагаются группы, у которых значение признака-разделителя меньше указанного на рисунке значения, справа – те группы, у которых признак принимает большие значения. В каждом узле указывается также значение энтропии – показателя, на основании которого оценивается гомогенность группы. Очевидно, что чем выше значение энтропии, тем более разнородные группы оказались объединены в узле, и тем большее число шагов потребуются до получения гомогенных объединений, внутри которых энтропия принимается равной нулю. Необходимо подчеркнуть, что принадлежность объекта к тому или иному классу задается исследователем, поэтому и при использовании алгоритма «дерево решений» остается естественная субъективность анализа. Недостатком алгоритма или лучше сказать, «неудобством» при интерпретации данных является то обстоятельство, что каждому узлу после разделения автоматически присваивается название того класса, который на текущем шаге насчитывает большее число объектов.

На первый взгляд, предлагаемый метод похож на дивизимный кластерный анализ, однако это сходство только внешнее. При иерархических процедурах кластеризации, независимо от того, агломеративные (объединяющие отдельные объекты в группы) они или дивизимные (разделяющие совокупность объектов на группы) формирование кластеров происходит на основе близости расположения объектов (единиц кластеризации) в многомерном пространстве. Рассматриваемый метод основан на совершенно других подходах. Здесь не рассчитываются

расстояния между отдельными единицами, нет и анализа связей между признаками. Разбиение исходного множества единиц идет по отдельным признакам. На каждом шаге разбиения используется только один признак – тот, у которого на текущем этапе наиболее выражена дискретность упорядоченного ряда средних величин. Именно при использовании такого признака в получаемых после разбиения группах снижается гетерогенность.

Безусловно, представляет интерес сравнение используемого алгоритма с уже имеющимися в арсенале антропологов методами классификации. Такое сравнение с результатами многомерного шкалирования той же группы объектов на основании только линейных размеров [см.: Федорчук, Гончарова 2024].

При проведении серии анализов для одного и того же набора объектов возникает множественность решений – построенные деревья могут различаться в последовательности признаков, используемых для дихотомического деления на каждом шаге. Но наборы признаков у этих деревьев в целом совпадают. Это понятно, т.к. для получения наиболее однородных групп на каждом шаге выбираются признаки со значительной межгрупповой изменчивостью и большими хиатусами в рядах средних величин. Выбор конкретного дерева при множестве решений остается за исследователем.

Возможность применения алгоритма к задачам классификации в антропологии тестировалась на двух наборах признаков. В первом варианте использовались только линейные размеры черепа человека [Федорчук, Гончарова, 2024], во втором линейные размеры и указатели, рассчитанные на их основе (табл. 1).

Материалы и методы

Краниометрические признаки, использованные в работе, соответствуют стандартной методике, принятой в российской антропологии [Алексеев, Дебеч, 1964, с. 52–57; Martin, 1928, pp. 625–660]. Всего было взято 15 линейных размеров, из них восемь – признаки мозгового отдела, семь признаков представляют лицевой отдел черепа (табл. 1).

На основе этих размеров было рассчитано общепринятых 12 индексов (указателей). Кроме того, опираясь на наблюдения Ю.Д. Беневолен-

ской, рассчитаны еще два индекса – лобно-сагитальный (ЛСИ: 26:25) и затылочно-теменной (ЗТИ: 28:27). В работах Ю.Д. Беневоленской была показана их дифференцирующая значимость. Затылочно-теменной индекс дифференцирует большие расы, особенно если учитывать относительную высоту черепа. Лобно-сагитальный дифференцирует популяции в пределах одной большой расы [Беневоленская, 1980, 1991]. В стандартных краниологических работах редко рассчитывается отношение средней ширины лица и скулового диаметра (поперечный фациальный: 46:45), и отношение высоты носа к верхней высоте лица (высотный носо-лицевой: 55:48). Эти указатели были взяты в анализ как экспериментальные.

Таблица 1. Список используемых размеров, названия и обозначения
Table 1. List of sizes used, names and designations

№	Линейные размеры	Номер по Мартину
1	Продольный диаметр	M.1
2	Поперечный диаметр	M.8
3	Высотный диаметр	M.17
4	Длина основания черепа	M.5
5	Наименьшая ширина лба	M.9
6	Лобная дуга	M.26
7	Сагитальная дуга	M.25
8	Теменная дуга	M.27
9	Затылочная дуга	M.28
10	Скуловой диаметр	M.45
11	Средняя ширина лица	M.46
12	Длина основания лица	M.40
13	Верхняя высота лица	M.48
14	Высота носа	M.55
15	Ширина носа	M.54
16	Высота орбиты	M.52
Указатели		Номера
1	Поперечно-продольный (черепной)	8:1
2	Высотно-продольный	17:1
3	Высотно-поперечный	17:8
4	Выступания лица	5:40
5	Лобно-поперечный	9:8
6	Лобно-сагитальный (ЛСИ)	26:25
7	Затылочно- теменной (ЗТИ)	28:27
8	Верхний лицевой	48:45
9	Носовой	54:55
10	Поперечный фациальный	46:45
11	Лобно-скуловой	9:45
12	Высотный носо-лицевой	55:48

Для построения классификации использовались средние данные по 39 этно-территориальным группам из 13 макрорегионов Старого Света (табл. 2). Четыре серии были единственными представителями своей региональной группы: черепа с территории Камеруна,

черепа айнов, хантов и черепа с острова Пасхи. Данные взяты из литературных источников, часть измерена одним из авторов в фондах НИИ и Музея антропологии МГУ и Музея антропологии и этнографии им. Петра Великого. Использовались данные только по мужским черепам. Для

Таблица 2. Список используемых в анализе групп и их географическая принадлежность
Table 2. List of groups used in the analysis and their geographic affiliation

№	n	Группы	Источник	Регионы
1	11	тейта	Kitson, 1931	Восточная Африка
2	39	тигре	Sergi, 1912	
3	88	Камерун	Drontschilow, 1913	Центральная Африка
4	24	баски	Morant, 1929	Европа
5	10	болгары	данные О.А.Федорчук	
6	14	итальянцы	данные О.А.Федорчук	
7	63	армяне	Бунак, 1927	
8	9	иранцы	данные О.А.Федорчук	
9	15	латыши	данные О.А.Федорчук	
10	56	осетины	данные О.А.Федорчук	
11	11	чукчи	данные О.А.Федорчук	
12	18	эскимосы (Чукотка)	данные О.А.Федорчук	Северная Азия
13	11	алеуты	данные О.А.Федорчук	
14	93	эскимосы (Аляска)	Дебец, 1986	
15	11	якуты	данные О.А.Федорчук	
16	109	казахи	Исмагулов, 1970	Средняя Азия
17	9	киргизы	данные О.А.Федорчук	
18	61	ханты	данные О.А.Федорчук	Западная Сибирь
19	26	теленгиты	данные О.А.Федорчук	Центральная Азия
20	154	буряты	архивные данные Н.Н. Мамоновой	
21	17	монголы	данные О.А.Федорчук	
22	7	айны	данные О.А.Федорчук	Дальний Восток
23	36	непальцы	Morant, 1924	Восточная Азия
24	32	тибетцы	Morant, 1924	
25	22	аэты	Bonin, 1931a	Юго-Восточная Азия
26	19	Бантам	Bonin, 1931a	
27	25	Джакарта	Bonin, 1931a	
28	28	даяки	Bonin, 1931a	
29	14	Мадуро	Bonin, 1931a	
30	28	яванцы	Bonin, 1931a	
31	15	тагалы	Bonin, 1931a	
32	32	яванцы (сборная)	Bonin, 1931a	
33	44	бирманцы	Tildesley, 1921	
34	15	андаманцы	Bonin, 1931a	
35	35	тамилы	Harrower G. 1924	Южная Азия
36	49	Новая Британия	Bonin, 1936a	
37	72	Северная Новая Гвинея	Hambly, 1940	Меланезия
38	25	Южная Новая Гвинея	Hambly, 1940	
39	24	остров Пасхи	Bonin, 1931b	Полинезия
Итого		1371		

Примечания. * – макрорегиональные группы сформированы с учетом только географической принадлежности безотносительно к существующим антропологическим классификациям этих популяций.

Notes. * – macroregional groups were formed considering only geographic affiliation with-out regard to existing anthropological classifications of these populations.

выяснения значимости признаков в построении классификации в анализ включались как линейные размеры, так и указатели.

В настоящей работе для создания классификации выборок был выбран алгоритм «дерево принятия решений» (decision tree) [Breiman, Stone, 1978; Quinlan, 1986]. Математический подход, который используется в алгоритме деревьев классификации, основан на поэтапном разделении выборок с максимальным уменьшением меры разнородности выборки, то есть на снижении вероятности объединения в некоторой точке выборок, относящихся к разным типам. Дробление продолжается до тех пор, пока на некотором этапе «лист» дерева не станет гомогенным. Таким образом, необходимость дальнейшего разделения и дальнейшего ветвления дерева определяется критерием разнородности выборок в каждой точке. В качестве такого показателя использовался показатель энтропии. Формула расчета энтропии Шеннона:

$$H = - \sum_{k=1}^m p_k \log_2 \left(\frac{1}{p_k} \right)$$

где k – это количество типов

p_k – вероятность того, что объект принадлежит к типу k .

На основе этого показателя выбирается то разбиение, при котором его снижение будет максимальным. То есть полученные группы должны быть как можно более однородными. В данном случае это значит, что полученные общности должны включать как можно меньше краиниологических серий из разных макрорегиональных групп.

Алгоритм классификационных деревьев реализовывался с помощью библиотеки *scikit-learn* в языке Python [Pedregosa et al., 2011]. Для построения бивариантных графиков использовался пакет *ggplot2*, языка R.

Алгоритм на языке Python реализовывался следующим образом:

1. Задаем переменные, которые надо классифицировать – группы (y) и переменные, по которым классификация происходит – признаки (x):

$x = \text{mean.iloc}[0:39, 1:]$ #Указываем столбцы с данными
 $y = \text{mean.iloc}[0:39, 0:1]$ #Указываем столбец с группами

2. Строим модель:

from sklearn import tree #Импортируем функцию *tree*
from sklearn.tree import DecisionTreeClassifier #Импортируем *DecisionTreeClassifier*

model = tree.DecisionTreeClassifier(criterion="entropy")

#Задаем параметры модели

model.fit(x,y) #Обучаем дерево

3. Визуализируем дерево:

from sklearn.tree import export_graphviz

export_graphviz(model, out_file='mean.dot', feature_names = feature, class_names = groups, max_depth = 5, rounded = True, proportion = False, precision = 2, filled = True)

Результаты и обсуждение

При использовании смешанного набора показателей видно, что на первом этапе происходит дифференциация всех групп по поперечному фациальному указателю (отношение средней ширины лица к скуловому диаметру, 46:45). Две из трех групп Меланезии, все африканские группы, выборки Южной, Юго-Восточной и Восточной Азии, имеющие большие значения указателя, отделились от выборок Европы, Северной, Центральной и Средней Азии с которыми оказались объединены по этому показателю группа с острова Пасхи, и одна группа с территории Меланезии. Все они характеризуются меньшими значениями этого признака (рис. 1, 2).

Большие значения указателя 46:45, означают, что по сравнению с шириной верхней челюсти скуловые дуги относительно узкие (рис. 1). Именно это сочетание признаков характерно для групп правой ветви построенного дерева решений, то есть для выборок Африки, Южной, Юго-Восточной и Восточной Азии, которые на втором шаге разделяются по значениям высотно-продольного указателя (17:1). Относительно более короткие и высокие черепа имеют серии с территории Южной Азии (тамильцы и андаманцы) и Юго-Восточной Азии (народы Зондского архипелага).

Меньшие значения высотно-продольного указателя имеют восточноазиатские (включая айнов), африканские и меланезийские группы. Можно предположить, что это связано с большей абсолютной длиной черепа у этих групп.

Дальнейшие разделения в большой правой ветви дерева происходят за счет признаков длины основания черепа (М.5), ширины черепа (М.8), а также лобно-поперечного указателя (9:8) и теменной хорды (М.27).

Таким образом, в узлах правой ветви дерева используется пять признаков, из которых два индекса и три абсолютных значения. Следует подчеркнуть, что после первого деления по

указателю лицевых признаков (46:45), все остальные разделения в правой ветви дерева проходят по размерам и указателям мозгового отдела.

Рассмотрим левую ветвь дерева (рис. 2), которую образуют группы с меньшими значениями поперечного фацеального указателя, что наблюдается при меньшей средней ширине лица (M.46) и/или при более высоких значениях скуловой ширины (M.45). Левая ветвь объединила разнородные выборки – европейские, североазиатские, две выборки Океании.

Классические монголоидные группы по лобно-скуловому индексу (9:45) отделяются от европеоидов, к которым присоединяются ханты, одна из меланезийских выборок и группа с острова Пасхи. Группы Северной, Центральной и Средней Азии имеют относительно более широкое лицо, то есть у классических монголоидов этот индекс меньше.

Дальнейшее разделение групп по отдельным ветвям дерева идет на основании дискретности ряда значений следующих признаков:

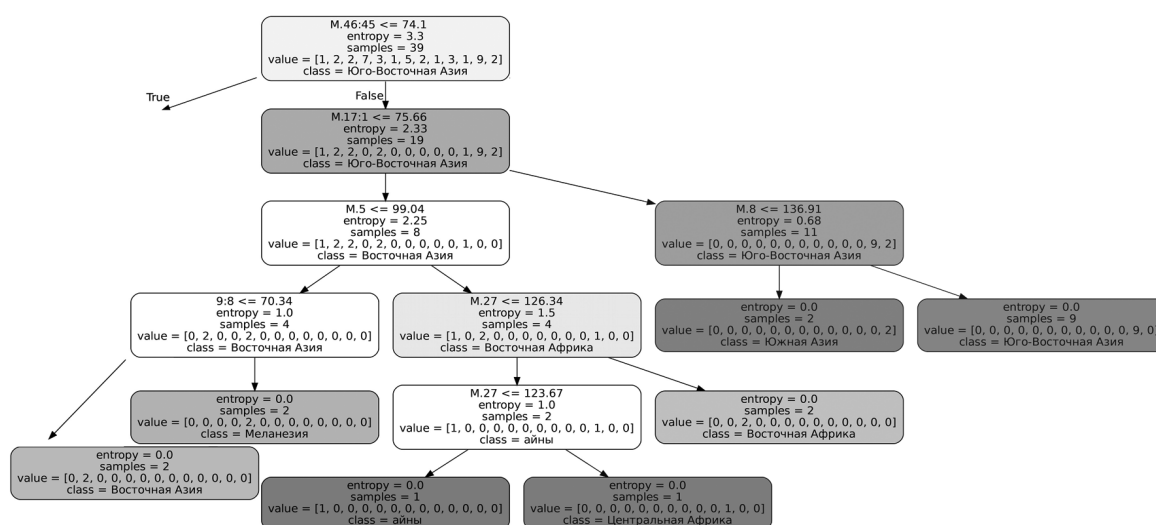


Рисунок 1. Правая ветвь дерева для групп с большими значениями поперечного фацеального указателя

Figure 1. Right branch of the tree for groups with large cross facies index values

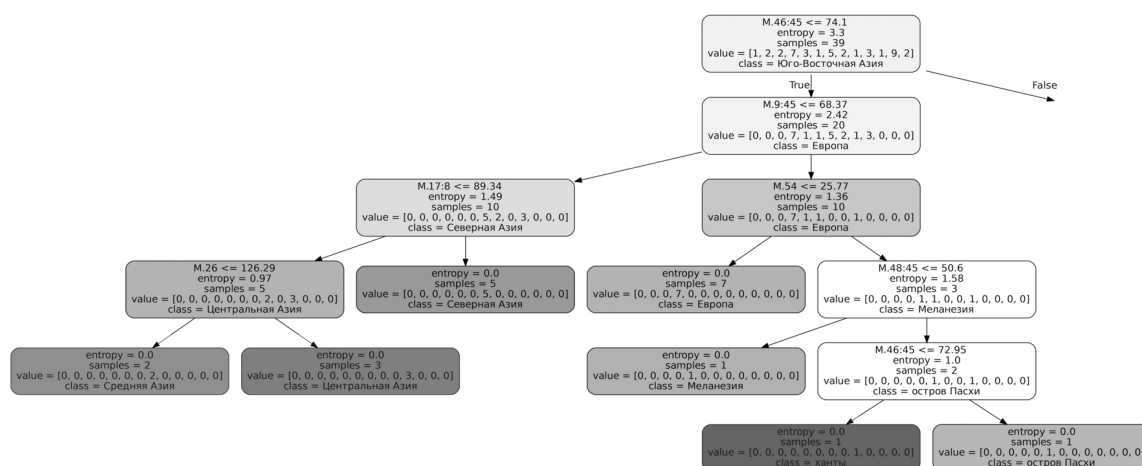


Рисунок 2. Левая ветвь дерева для групп с меньшим поперечным фацеальным указателем

Figure 2. Left branch of the tree for groups with smaller cross-sectional facies index

Примечания к рис. 1 и 2. Градации серого обозначают преобладание групп из той или иной макрорегиональной группы в узле или листе.

Notes to fig. 1 and 2. Grayscale indicates the predominance of groups from a particular macroregional group in a node or leaf.

1. ширина грушевидного отверстия М.54, по этому признаку европеоиды отличаются от групп Океании и хантов менее широким носовым отверстием;

2. верхний лицевой (48:45) и поперечно-фациальный индексы (46:45) разделяют разные группы Океании и хантов (у последних относительно более широкое лицо);

3. высотно-поперечный индекс (17:8) отделяет группы собственно Северной Азии (для них характерны большие значения индекса, то есть более высокий и/или более узкий череп) от серий Центральной и Средней Азии с противоположным сочетанием признаков

4. длина лобной дуги (М.26) разделяет группы Центральной Азии (в большинстве своем, они относятся к центральноазиатскому типу североазиатской расы) и среднеазиатские группы в рамках данного исследования к ним отнесены казахи и киргизы, представители южносибирской малой расы, у которых длина лобной дуги меньше.

Задачей работы стало выявление признаков, обладающих более высоким возможностями дифференциации. Следует отметить, что суммарно использованы 15 линейных признаков и 16 указателей. Дифференцирующие возможности оказались высокими для пяти линейных размеров (М.5, М.8, М.26, М.27, М.54) и семи указателей. Причем три указателя основаны на габаритных размерах мозгового отдела черепа (М.1, М.8, М.17), а четыре указателя в своей конструкции учитывают скуловую ширину (М.45).

Значимость габаритных размеров черепа при дифференциации популяций Старого Света неоднократно показана в работах А.П. Пестрякова [Пестряков, Григорьева, 2004; 2013]. Полученные нами результаты соотносятся с выводами этих работ, так как показывают, что габаритные размеры черепа действительно разделяют выборки на макрорегиональном уровне. Однако следует подчеркнуть, что первое разделение групп Старого Света происходит на основе дискретности значений поперечного фациального индекса, который рассчитывается как отношение средней ширины лица (М.46) к скуловому диаметру (М.45). Этот указатель редко используется в краниологических исследованиях, тем интереснее его ведущая роль в разделении разнообразных в расовом отношении популяций Старого Света.

При визуализации распределения выборок в пространстве значений этих признаков видна межгрупповая связь этих размеров (рис. 3), тогда как на внутригрупповом уровне они связаны коэффициентом средней силы (0,5) [Федорчук, 2022]. В области больших значений этих признаков располагаются группы Северной, Центральной и Средней Азии, а также ханты и айны. Эти группы североазиатских монголоидов, к которым присоединились ханты и айны, довольно хорошо обособились от всех остальных. В области меньших значений признаков находятся группы Восточной, Южной, Юго-Восточной Азии, Меланезии, Африки. Также малыми значениями средней ширины лица (М.46) характеризуются европейские группы. Они отличаются от групп Юго-Восточной, Восточной и Южной Азии несколько большими значениями скулового диаметра. На бивариантном графике, построенном по линейным размерам обособление европейских групп не слишком очевидно, однако на графике, построенном с использованием указателей, оно становится выраженным (рис. 4).

В пространстве значений поперечного фациального указателя и лобно-скулового указателя сильнее всего обособлены группы с территории Европы, они смещены в область малых значений поперечного-фациального указателя и больших значений лобно-скулового указателя. Большие значения обоих указателей характерны для групп с территории Южной Азии, Меланезии и Африки. Для остальных серий дифференциация по поперечному фациальному указателю меньше, в то время как исходные значения признаков (рис. 3) довольно хорошо разделяют группы североазиатских и южноазиатских монголоидов.

Если заменить лобно-скуловой указатель на высотно-продольный, то наблюдается резкое отделение групп Юго-Восточной Азии в область больших значений обоих указателей (рис. 5). К ним примыкают группы Южной Азии и одна группа Восточной Азии (непальцы). По поперечному фациальному указателю по-прежнему наблюдается отделение европейских групп. Однако к ним неожиданно примыкают теленгиты, алеуты, казахи и серия с территории Новой Британии. И если объединение групп Южной, Юго-Восточной и Восточной Азии можно объяснить той или иной близостью происхождения, то объединение некоторых азиатских и европейских групп выглядит весьма искусственно. Но это как раз хорошо показывает ограниченные возможности бивариантных графиков. Дерево решений позволяет последовательно проводить разделение по всему набору признаков.

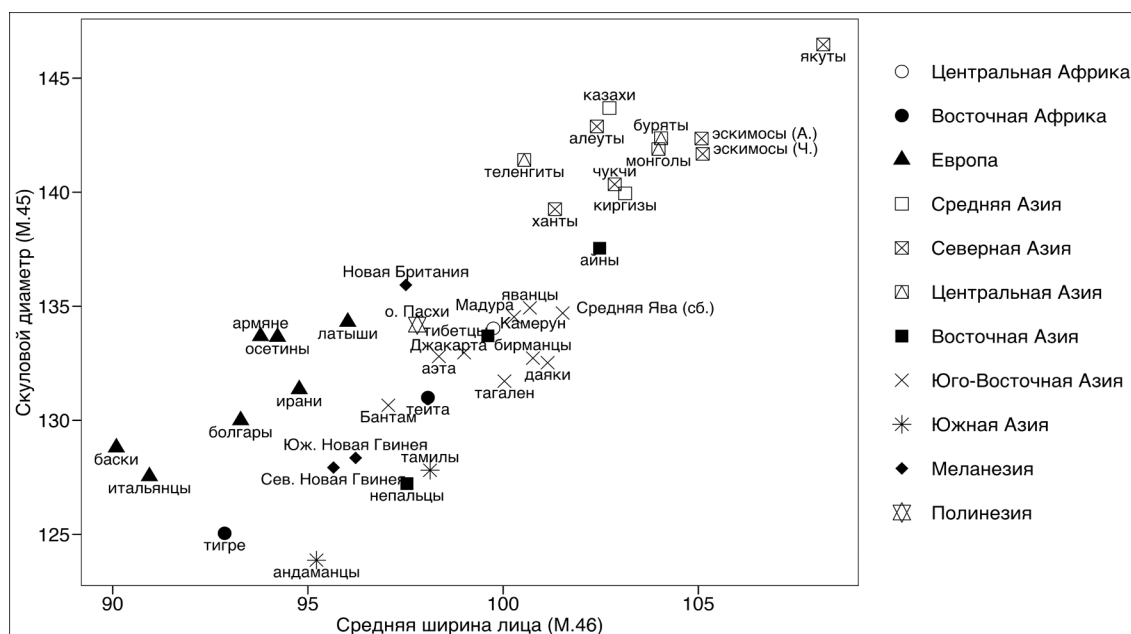


Рисунок 3. Расположение групп в пространстве значений средней ширины лица (М.46) и скулового диаметра (М.45)

Figure 3. Location of groups in the space of mean facial width (M.46) and zygomatic diameter (M.45) values

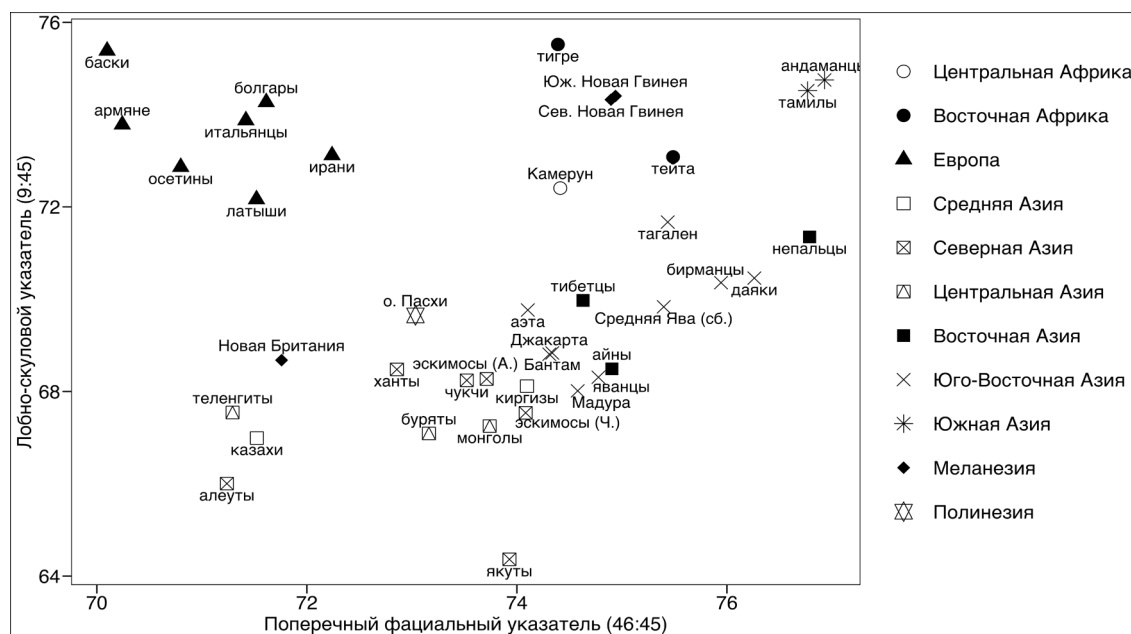


Рисунок 4. Расположение групп в пространстве значений поперечного фацеального указателя (46:45) и лобно-скулового указателя (9:45)

Figure 4. Location of groups in the value space of transverse facial index (46:45) and frontal-maxillary index (9:45)

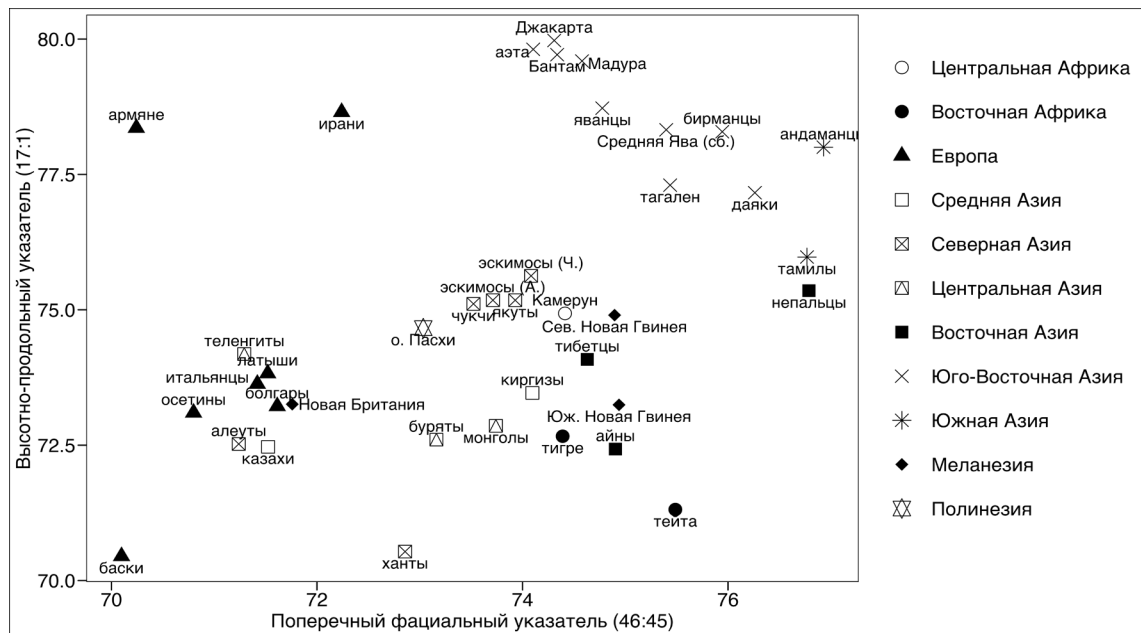


Рисунок 5. Расположение групп в пространстве значений поперечного фацеального указателя (46:45) и высотно-продольного указателя (17:1)

Figure 5. Arrangement of groups in the value space of the transverse facial index (46:45) and the altitudinal-prolateral index (17:1)

Заключение

Таким образом, распределение групп в пространстве указателей в целом более дискретно, чем в пространстве значений линейных признаков, так как трансгрессия указателей на межгрупповом уровне выражена меньше, чем трансгрессия линейных размеров. Это говорит о высокой дифференцирующей способности указателей. Понятно, что на разных иерархических уровнях и для разных регионов набор значимых индексов будет различным. Так, при изучении славяно-финского взаимодействия в Волго-Окском регионе показана работоспособность двух индексов, использование которых позволяет с высокой достоверностью разделять популяции финнов и славян [Гончарова, Конопелькин, 2014]. В то же время линейные размеры не дают такой возможности, о чем писал и Г.Ф. Дебец: «не только в настоящее время, но и в эпоху, близкую ко времени формирования славянских племён, на территории Восточной Европы установить расовые различия между славянами и финнами не представляется возможным» [Дебец, 1948, с. 287]. Вычисление индексов позволяет выявить отклонения от стандартных биологических соотношений краниологических признаков.

В настоящей работе на первых этапах разделения всех выборок Старого Света наиболее значимыми оказались поперечный фацеальный (46:45), лобно-скуловой (9:45) и высотно-продольный (17:1) указатели.

Тем не менее, следует учитывать и информативность линейных размеров, иногда именно они позволяют хорошо разделить группы даже высокого иерархического уровня. Это обстоятельство повышает значимость тех видов анализа, в которых допускается использование смешанного набора признаков.

Таким образом в работе продемонстрирована хорошая дифференцирующая способность обсуждаемого алгоритма, в особенности для смешанного набора признаков. Его можно рекомендовать как еще один инструмент классификации выборок.

Благодарности

Исследование выполнено в рамках исследовательской темы «Формирование некоторых морфо-функциональных особенностей человека в фило- и онтогенезе» кафедры антропологии МГУ и Государственного задания ФГБНУ МГНЦ им. академика Н.П. Бочкова.

Библиография

- Алексеев В.П., Дебец Г.Ф. Краниометрия. Методика антропологических исследований. М.: Наука. 1964. 128 с.
- Бунак В.В. *Crania armenica*. Исследования по антропологии Передней Азии. М., 1927. 314 с.
- Беневоленская Ю.Д. Группо-разграничительные свойства признаков затылочной области черепа // Сборник МАЭ, 1980. № 36. С. 108–120.
- Беневоленская Ю.Д. Признаки черепного свода как маркеры различных уровней дифференциации различных рас // Сборник МАЭ, 1991. № 44. С. 136–152.
- Гончарова Н.Н., Конопелькин Д.С. Новые данные к антропологии финских племен Верхней Волги и бассейна Оки // Физическая антропология: методики, базы данных, научные результаты, 2014. С. 89–103.
- Дебец Г.Ф. Палеоантропология СССР // Труды института этнографии. Новая серия. Т. IV, 1948.
- Дебец Г.Ф. Палеоантропология древних эскимосов (Ипиутак, Тигара) // Этнические связи народов севера Азии и Америки по данным антропологии. М., 1986. С. 6–149.
- Дерябин В.Е. Курс лекций по многомерной биометрии для антропологов. М., 2008. 332 с.
- Исмагулов О. Население Казахстана от эпохи бронзы до современности. Алма-Ата, 1970. 240 с.
- Пестряков А.П., Григорьева О.М. Краниологическая дифференциация современного населения // Расы и народы, 2004. № 30. С. 86–131.

Пестряков А.П., Григорьева О.М. Австралийские аборигены на краниологическом фоне населения восточной Азии и Западной Океании // Вестник Московского университета. Серия 23. Антропология, 2013. № 2. С. 17–33.

Федорчук О.А. Материалы к соотносительной изменчивости измерительных признаков черепа человека. М.: Клуб печати. 2022. 239 с.

Федорчук О.А., Гончарова Н.Н. Применение метода «деревья решений» для дифференциации групп человечества // Археология, этнография и антропология Евразии, 2024. Т. 52, № 3. С. 148–156. DOI: 10.17746/1563-0102.2024.52.3.148-156

Фельдман М.Г. Использование метода Random Forest в целях прогнозирования подходов горбуши Северо-Востока Камчатки // The Researches of the Aquatic Biological Resources of Kamchatka and the North-West Part of the Pacific Ocean, 2021. № 59. С. 76–96.

Информация об авторах

Федорчук Ольга Алексеевна, к.б.н.; ORCID ID: 0000-0002-9645-2014; fedorchukoa@my.msu.ru;

Гончарова Наталья Николаевна, к.б.н.; ORCID ID: 0000-0001-8504-1175; goncharovann@my.msu.ru.

Поступила в редакцию 04.09.2024,
принята к публикации 18.11.2024

Fedorchuk O.A.¹⁾, Goncharova N.N.^{1, 2)}

¹⁾ *Lomonosov Moscow State University, Faculty of Biology, Department of Anthropology, Leninskie Gory, 1(12), Moscow, 119234, Russia*

²⁾ *Research Center of Medical Genetics, Moskvorechye St., 1, Moscow, 115522, Russia*

ON THE ACCURACY OF VISUAL AGE ESTIMATION FROM THE ADULT SKULL (METHODOLOGICAL ASPECTS)

Introduction. *The task of differentiating groups using various statistical methods remains relevant due to the limitations often encountered with standard algorithms. Some of these limitations, such as the assumption of normal distribution of traits and the absence of high correlations between them, can be circumvented by the method based on step-wise partitioning – decision trees.*

Materials and methods. *In this study, this method was tested on 15 linear dimensions of the skull and 16 indices calculated based on them. The entropy index was chosen as the criterion of heterogeneity. The craniometric features used correspond to the standard methodology adopted in Russian anthropology. The data consisted of average values of craniometric dimensions from 39 ethno-territorial groups from 13 macro-regions of the Old World.*

Results and discussion. *The results of the differentiation show that indices have a greater weight than linear features. Five linear dimensions and seven indices participated in the differentiation. Three indices were based on the overall dimensions of the braincase (M.1, M.8, M.17), and four indices included the bizygomatic width (M.45). The first split occurs based on the transverse facial index (46:45), separating the groups of Northern, Middle, Central Asia, and Europe from the groups of Africa, Eastern, Southeast, and Southern Asia.*

Conclusion. *The study demonstrates the high differentiating ability of indices, particularly the transverse facial index, fronto-zygomatic index, and height-length index. The iterative differentiation system allows for a more informative separation of groups, as different combinations of features demonstrate effectiveness at different hierarchical levels of taxonomy.*

Keywords: human biology; craniology; biostatistics; decision trees

DOI: 10.55959/MSU2074-8132-25-1-5

References

- Alekseev V.P., Debetz G.F. *Kraniometriya. Metodika antropologicheskikh issledovanij* [Cranio-metry. Methodology of anthropological research]. Moscow, Nauka Publ., 1964. 128 p. (In Russ.).
- Bunak V.V. *Crania armenica. Studies on anthropology of West Asia*. Moscow, 1927. 314 p. (In Russ.).
- Benevolenskaya Yu.D. Group-distinguishing properties of features of the occipital region of the skull. In: *Collection of the MAE*, 1980, 36, pp. 108–120. (In Russ.).
- Benevolenskaya Yu.D. Cranial vault features as markers of different levels of differentiation of different races. In: *Collection of the MAE*, 1991, 44, pp. 136–152. (In Russ.).
- Goncharova N.N., Konopel'kin D.S. New data on the anthropology of the Finnish tribes of the Upper Volga and Oka basin. In: *Physical anthropology: methods, databases, scientific results*, 2014, pp. 89–103. (In Russ.).
- Debetz G.F. Paleoanthropology of the USSR. In: *Proceedings of the Institute of Ethnography*, New Series, IV, 1948. (In Russ.).
- Debetz G.F. Paleoanthropology of ancient Eskimos (Ipiutak, Tigara). *Ethnic relations of the peoples of the north of Asia and America on the basis of anthropological data*. Moscow, 1986. pp. 6–149.
- Deryabin V.E. *A lecture course on multivariate biometrics for anthropologists*. Moscow, Moscow State University MGU, 2008. 332 p. (In Russ.).
- Ismagulov O. *The population of Kazakhstan from the Bronze Age to the present (paleoanthropological study)*. Alma-Ata, Nauka Publ., 1970. 241 p. (In Russ.).
- Pestryakov A.P., Grigor'eva O.M. Craniological differentiation of the modern population. *Races and peoples*, 2004, 30, pp. 86–131. (In Russ.).
- Pestryakov A.P., Grigor'eva O.M. Australian Aborigines in the craniologic background of East Asian and Western Oceania populations. In: *Vestnik Moskovskogo universiteta. Seriya 23. Antropologiya* [Moscow University Anthropology Bulletin], 2013, 2, pp. 17–33. (In Russ.).
- Fedorchuk O.A. *Materials to the correlative variability of measuring features of the human skull*. Moscow, Klub Pechati Publ., 2022. 239 p. (In Russ.).
- Fedorchuk O.A., Goncharova N.N. Applying the decision tree method to differentiate groups of humanity. *Archaeology, Ethnography and Anthropology of Eurasia*, 2024. 52(3) pp. 148–156. DOI: 10.17746/1563-0102.2024.52.3.148-156. (In Russ.).
- Fel'dman M.G. Using the Random Forest method to predict pink salmon runs in the Northeast Kamchatka Peninsula. *The Researches of the Aquatic Biological Resources of Kamchatka and the North-West Part of the Pacific Ocean*. 2021, 59, pp. 76–96. (In Russ.).
- Martin R. *Lehrbuch der Anthropologie in Systematischer darstellung. Bd. II. Kraniologie. Osteologie*. Jena, 1928. 1182 p.
- Breiman L., Stone C.J. *Parsimonious binary classification trees. Technical Note*, 1978.
- Bonin G. von. Beitrag Zur Kraniologie von Ost-Asien. *Biometrika*, 1931a, 23 (1/2), pp. 52–113.
- Bonin G. von. A Contribution to the Craniology of the Easter Islanders. *Biometrika*, 1931b, 23 (3/4), pp. 249–270.
- Bonin G. von. On the Craniology of Oceania. Crania from New Britain. *Biometrika*, 1936, 28 (1/2), pp. 123–148.
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J. *Classification and regression trees*. Wadsworth Inc, 1984. 354 p.
- Cartmill M. A sort of revolution: Systematics and physical anthropology in the 20th century. *American Journal of Physical Anthropology*, 2018, 165 (4), pp. 677–687. DOI: 10.1002/ajpa.23321.
- Djuris J., Ibrić S., Djurić Z. Neural computing in pharmaceutical products and process development. *Computer-Aided Applications in Pharmaceutical Technology*, 2013, pp. 91–175. DOI: 10.1533/9781908818324.91.
- Drontschilow K. Metrische Studien an 93 Schädeln aus Kamerun. *Archiv für Anthropologie*, 1913, pp. 161–183.
- Hambly W.D. *Cranio-metry of New Guinea*. Chicago, 1940. 338 p.
- Harrower G. A Study of the Hokien and Tamil Skull. *Biometrika*, 1924, 54 (3), pp. 573–599.
- Hugenholtz P., Chuvochina M., Oren A., Parks D.H., Soo R.M. Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME Journal*, 2021, 15 (7), pp. 1879–1892. DOI: 10.1038/s41396-021-00941-x.
- Kitson E. A Study of the Negro Skull with Special Reference to the Crania from Kenya Colony. *Biometrika*, 1931, 23 (3), pp. 271–314.
- Al Mamun M.H., Keikhosrokiani P. Predicting onset (type-2) of diabetes from medical records using binary class classification. *Big Data Analytics for Healthcare*, 2022, pp. 301–312. DOI: 10.1016/B978-0-323-91907-4.00012.
- Martin R. *Lehrbuch der Anthropologie*. 2-te Aufl. Bd. II, 1928.
- Morant G.M. A Study of Certain Oriental Series of Crania Including the Nepalese and Tibetan Series in the British Museum (Natural History). *Biometrika*, 1924, 16 (1/2), pp. 1–10.
- Morant G.M. A Contribution to Basque Craniometry. *Biometrika*, 1929, 21 (1/4), pp. 67–84.
- Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2011, 12, pp. 2825–2830.
- Quinlan J.R. Induction of decision trees. *Machine Learning*, 1986, 1 (1), pp. 81–106. DOI: 10.1007/BF00116251.
- Sergi S. *Crania habessinica: contributo all'antropologia dell'Africa orientale*. Loescher, Roma, 1912. 519 p.
- Sokal R.R., Sneath P.H.A. *Principles of Numerical Taxonomy*. W. H. Freeman, Ed., 1963. 359 p.
- Tildesley M.L. A First Study of the Burmese Skull. *Biometrika*, 1921, 13 (2/3), pp. 176–262.
- Wong S.L., Zhang L.V., Y Tong A.H., Li Z., Goldberg D. S., King O.D., Lesage G., Vidal M., Andrews B., Bussey H., Boone C., Roth F.P. Combining biological networks to predict genetic interactions. *Proc Natl Acad Sci USA*, 2004, 101 (44), pp. 15682–15687. www.pnas.org/cgi/doi/10.1073/pnas.0406614101.

Information about the authors

Fedorchuk Olga A., PhD; ORCID ID: 0000-0002-9645-2014; fedorchukoa@my.msu.ru;

Goncharova Natalia N., PhD, ORCID ID: 0000-0001-8504-1175; goncharovann@my.msu.ru.

© 2025. This work is licensed under a CC BY 4.0 license