

Применение методов машинного обучения для анализа внутригрупповой изменчивости измерительных признаков черепа человека

Научный руководитель – Гончарова Наталья Николаевна

Федорчук Ольга Алексеевна

Сотрудник

Московский государственный университет имени М.В.Ломоносова, Биологический факультет, Москва, Россия
E-mail: lela.fed@yandex.ru

Анализ изменчивости морфологических признаков является важной задачей в современной антропологии. Дополнением к современным статистическим методам внутригрупповой изменчивости, является алгоритм „случайного леса“, который пока не очень распространен в антропологических работах [2, 3]. Он относится к методам обучения с учителем, и представляет ансамблевый метаалгоритм, в котором решение принимается путем голосования множества объектов одного семейства [1].

Настоящая работа посвящена аспектам применения случайного леса к индивидуальным краниометрическим данным. Цель работы заключается в проверке возможностей метода для исследования внутригрупповой изменчивости.

Измерительная программа включала стандартные признаки, 15 линейных размеров и 12 указателей.

Использовались индивидуальные данные по 39 этно-территориальным группам из 13 макрорегионов, в основном Старого Света. Данные были разбиты на тестовую (25% - 343 индивидов) и обучающую (75% - 1028 индивидов) выборки. По второй группе обучался алгоритм из 1000 деревьев, а тестовые данные подавались обученному алгоритму для оценки правильности решений по отнесению черепов к той или иной макрорегиональной группе. Анализировалось по проценту индивидов, которые попали не в свою макрорегиональную группу, от общего числа индивидов макрорегиональной группы и по проценту индивидов, которые попали из другой макрорегиональной группы, от общего числа индивидов данной группы.

Оба показателя максимальны в группе Восточной Азии. Высоким данные показатели и для черепов с территории Африки. Процент индивидов Юго-Восточной Азии, которые попали в другие региональные группы довольно высок, но при этом минимален процент черепов, неправильно отнесенных к этой макрорегиональной группе. То есть черепа с территории Юго-Восточной Азии имеют сходство с черепами из других регионов, но при этом черепа из других регионов редко определяются как черепа этого региона. Больше половины черепов с территории Средней Азии определены алгоритмом неправильно. Наиболее эффективно алгоритм определил черепа из групп Океании.

В результате применения алгоритма „случайного леса“ наиболее значимыми оказались указатели: высотно-поперечный, черепной указатели и указатель выступления лица. Минимальную значимость имеют дуговые размеры костей сагиттального профиля.

- 1) По результатам распределения черепов в макрорегиональные группы можно говорить о степени различия и особенностям морфологии групп
- 2) Из исследуемых краниометрических признаков минимальную значимость для дифференциации черепов на индивидуальном уровне имеют дуговые размеры костей сагиттального профиля.

Источники и литература

- 1) Breiman, L. (2001). Random Forests (Vol. 45).
- 2) Fernandes, L. C. C., Bento, M. I. C., Rabello, P. M., Soriano, E. P., Navega, D., Júnior, E. D., & Cunha, E. (2021). Analysis of the Accuracy of AncesTrees Software in Ancestry Estimation in Brazilian Identified Sample. *Advances in Anthropology*, 11(02), 163–178
- 3) Navega, D., Coelho, C., Vicente, R., Ferreira, M. T., Wasterlain, S., & Cunha, E. (2015). AncesTrees: ancestry estimation with randomized decision trees. *International Journal of Legal Medicine*, 129(5), 1145–1153