A102 A Reevaluation of Tooth Crown Measurements in the Estimation of Ancestry Using Random Forest Classification

Michael W. Kenyhercz, PhD, Department of Defense POW/MIA Accounting Agency, Joint Base Pearl Harbor-Hickam, HI 96816; Marin A. Pilloud, PhD, University of Nevada, Reno, Reno, NV 89557-0096; Joseph T. Hefner, PhD, Michigan State University, East Lansing, MI 48824*

Learning Overview: After attending this presentation, attendees will understand the utility of tooth crown metrics in the estimation of ancestry and how random forest classification is used to achieve high classification rates among several worldwide populations.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by demonstrating a new method in forensic anthropology and by providing a freely available Graphical User Interface (GUI) for practitioners to use in casework and research.

Teeth have long been studied for their variability among human populations, both metrically, especially using dental apportionment analysis and morphologically.[1-5] However, tooth crown metrics are typically outperformed by other analyses, such as craniometrics.[6] Pilloud et al. have recently shown the utility of tooth crown metrics in discriminating among three broad geographic groups (African, Asian, and European) using Hanihara's worldwide odontometrics dataset and linear discriminant function analysis.[7] When females and males of those three groups were considered separately, Pilloud et al. reported cross-validated total correct classifications of 88.1% and 71.9%, respectively, and a combined cross-validated total correct classification of 71.3%.[7]

Random forest classification is an alternative to linear discriminant function analysis that uses modified decision trees with non-linear boundaries. Essentially, at each decision, or tree, a classification is made and, depending on the number of trees, the group with the highest classification count is used for a classification. The user specifies the number of trees and the number of variables tested at each decision tree. For the current study, Hanihara's worldwide dataset was used with the same three broad geographic groups used in Pilloud et al.'s study.[7] A total of 32 tooth crown measurements are available (16 mesiodistal and 16 buccolingual), representing one side of the dental arcade. The total sample includes 1,674 complete dentitions (African Females=17, African Males=85, Asian Females=282, Asian Males=698, European Females=162, and European Males=430). The implementation of random forest classification used here includes a 70% training set, 15% validation set, and 15% testing set. Using 500 trees, with four variables tested at each decision tree, a total correct classification of the same three broad geographic groups, with females and males considered together (a total of six classification groups), achieves a total correct classification rate of 90.3%. When females and males are considered separately, the total correct classification rates are 99.1% and 98.4%, respectively.

A GUI was created with the language *shiny* in RStudio to allow practitioners to input between one and 32 variables and create custom random forest classification models on a case-by-case basis.[8,9] Users specify the groups to compare, which can include females and males from the three broad geographic groups: African, Asian, and European. Output includes: correct classification rates (both raw counts and percentages), total correct classification, group classification, posterior probabilities, positive and negative predictive values, individual node classification counts, out of bag error, variable importance, and a Mahalanobis distance matrix. The GUI is freely available at https://anthropologyapps.shinyapps.io/HanihaRa/.

Reference(s):

1. Harris E.F., Bailit H.L. 1988. A Principal Components Analysis of Human Odontometrics. *Am J Phys Anthropol.* 53:349-358.
2. Turner C.G. II. 1967. *The Dentition of Arctic peoples.* PhD dissertation. University of Wisconsin, Madison.
3. Irish J.D. 1997. Characteristic High- and Low-Frequency Dental Traits in Sub-Saharan African Populations. *Am J Phys Anthropol.* 102:455-467.
4. Irish J.D. 1998. Ancestral Dental Traits in Recent Sub-Saharan Africans and the Origins of Modern Humans. *J Hum Evol.* 34:81-98.
5. Edgar H.J.H. 2005. Prediction of Race Using Characteristics of Dental Morphology. *J Forensic Sci.* 50(2):1-5.
6. Falk D., Corruccini R. 1982. Efficacy of Cranial Versus Dental Measurements for Separating Human Populations. *Am J Phys Anthropol.* 57: 123-127.
7. Pilloud M.A., Hefner J.T., Hanihara T., Hayashi A. The Use of Tooth Crown Measurements in the Assessment of Ancestry. *J Forensic Sci.* 59(6):1493-1501.
8. Chang W., Cheng J., Allaire J.J., Xie Y., McPherson J. 2017. *Shiny: Web Application Framework for R.* R package version 1.0.3.
9. R Studio Team. 2016. *R Studio: Integrated Development Environment for R.* R Studio, Inc. Boston, MA. www.rstudio.com.

Forensic Anthropology, Dental Crown Measurements, Random Forest Classification