H80 Statistical Classification Methods for Estimating Ancestry Using Morphoscopic Traits

Joseph T. Hefner, PhD, Central Identification Laboratory, 310 Worchester Avenue, Bldg 45, Joint Base Pearl Harbor Hickam, HI 96583; and Stephen D. Ousley, PhD, Dept of Applied Forensic Sciences, Dept of Anthropology, 501 E 38th Street, Erie, PA 16546*

After attending this presentation, attendees will understand the practical application of morphoscopic traits in ancestry assessment using a forensic dataset.

This presentation will impact the forensic science community by debunking some of the traditional approaches to ancestry assessment, which remain rooted in typological thinking. Moreover, this presentation will explore aspects of ancestry assessment method and theory.

Previous work has demonstrated the efficacy of morphoscopic (non-metric) traits in the assessment of ancestry, but only if the data are analyzed in a statistical framework and only if typological trait lists relying on extreme trait values are abandoned. There are numerous new statistical methods that can be used for the classification of an unknown cranium into a reference population on the basis of morphoscopic trait expression. Trait lists, however, remain the most widely taught, used, and accepted method of ancestry assessment. Trait lists are intuitively easy but lead to classifications without empirical support or consideration of the variation within groups. Without an explicit method, their validity is uncertain.

The objective of this study is to examine the utility of 16 frequently cited morphoscopic traits when analyzed using numerous statistical methods for classifying an unknown cranium into one of several reference groups based on those traits. The OSSA method and ten additional multivariate statistical classification methods (artificial neural networks, support vector machines, random forest, naïve Bayesian, decision tree, k-nearest neighbor, quadratic discriminant function, linear discriminant function, and logistic regression) that provide multi-group classifications were tested using a large, modern sample of American Black (n = 256), American White (n = 218), and Hispanic (n = 244) individuals. The sample composition is particularly important to U.S. forensic anthropologists, as these groups represent the majority of the populations encountered during routine forensic investigations. Each of the statistical methods take correlations among variables into account as necessary and, more importantly, provides estimated accuracy rates for the classification of an unknown. All of these methods have different requirements for optimal classification accuracy of individuals; some methods require that the data have a multivariate normal distribution and others have virtually no requirements of the data. Rigorous cross-validation methods help to discern the best traits to use in classification and provide a measure of validity, that is, the expected accuracy of each classification method.

Artificial Neural Networks (aNNs), OSSA, support vector machines, and random forest models showed mean classification accuracies of at least 85%. The aNNs had the highest overall classification accuracy (87.8%) and random forests show the smallest difference between the highest (90.4%) and lowest (76.5%) classification accuracies.

The results of this research demonstrate that morphoscopic traits can be used to assess ancestry accurately, with low estimated error rates and without relying only on the experience of the observer, and that would be true if you showed someone else the traits and tested them. Moreover, these results demonstrate that subjective trait lists do not capture the true range of variation; each one of the classification methods has a firmer empirical basis for a classification than the traditional approach. There is a definite need for scientific rigor in ancestry assessment methods that are empirically supported by the data at hand. Every method, whether novel or long-established, should be tested and refined rather than performed simply due to tradition or subjective/personal experience.

Ancestry, Morphoscopic Traits, Classification Statistics

 ** Presenting Author*