A73 Exploring the Performance of a Global Subadult Age Estimation Model Using Unsupervised Machine Learning Techniques

Kyra E. Stull, PhD, University of Nevada, Reno, Dept of Anthropology, 1664 N Virginia Street, Stop 0096, Reno, NV 89557; and Michael H. Price, PhD, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501*

The goal of this presentation is to inform attendees of the possibility of developing a global multivariable, multi-indicator subadult age estimation model. Numerous variables from the three most frequently applied age indicators were collected from genetically and economically diverse populations and were used to develop and compare the performance of population-specific models and a pooled, global model.

This presentation will impact the forensic science community by providing information regarding the feasibility of a global application of subadult age estimation, which is crucial considering the increased migration events and subsequent admixture resulting in populations with greater levels of genetic variation and complexity.

Most subadult age estimation methods do not account for the variation that exists in the world, as they are tailored for application in a specific location. Unfortunately, there is a possibility of mass and natural disasters in any country around the world, in addition to increased population movements and subsequent increased genetic variation; subadult identification would be facilitated if models were global. Thus, anthropologists need to critically evaluate subadult age estimation and explore the feasibility of global models, specifically, if there are differences in age indicators between geographically diverse groups, how does a global multivariable, multi-indicator model compare to a population-specific multivariable, multi-indicator model?

Epiphyseal fusion, dental formation, and diaphyseal dimensions were collected from a modern sample of children aged between birth and 10 years of age from the United States and South Africa. Two population-specific and one global multivariable, multi-indicator models were constructed using feedforward Artificial Neural Networks (ANN). Unsupervised machine learning techniques make minimal assumptions about the underlying data and thus offer a very flexible prediction framework that can handle different types of data modes, such as continuous and ordinal variables. A test dataset was reserved for all models to independently assess the suitability of each final fit.

The age indicators were compared for each population and scatterplots depicted a complete and consistent overlap in the distributions of the diaphyseal dimensions across the entirety of the age range, though differences in length became apparent in the older ages. In terms of the developmental variables, older South African children expressed the same dental formation and epiphyseal fusion stage as younger children from the United States. The results of the multivariable, multi-indicator models reveal the data trained from the pooled samples provided better age estimations than the population-specific models, based on lower mean squared error. The test data revealed that the United States sample was more predictable than the South African sample, which may be due to the fact that the South African sample was, on average, slightly older than the United States sample.

Unsupervised machine learning techniques use pattern detection to build models. Therefore, if population-specific models are used, and thus account for less variation in the training sample, the error will always be higher. Using a larger sample to train with captures the true range of variation; therefore, the model will always be more precise. The results of this study are very exciting and provide a solid stepping stone to continue the research and address the needs for more rigorous subadult age estimation techniques. Population-specific models may perform just as well as global models, but until larger, more reflective samples of populations with equal age distributions are used, the global model can be applied.

Artificial Neural Networks, Computed Tomography, Juvenile