

User Manual

SexEst is a web application that uses machine learning models to estimate the sex of an individual using cranial or postcranial linear measurements. The application has three modes of operation: a single skeleton mode, a multiple skeleton mode, and a missing data mode.

Single Skeleton Mode

In this mode, users can manually enter postcranial or cranial measurements in millimeters (mm) into the input boxes provided. The plus and minus buttons allow users to increase or decrease a measurement by 0.5 mm. Once the measurements are input, users can select one of three machine learning models: Extreme Gradient Boosting (XGB), Light Gradient Boosting (LGB), and Linear Discriminant Analysis (LDA). Upon pressing the Calculate button, the sex of the individual is estimated along with the probability of being male or female. The accuracy of the model used for the prediction is also provided. Note that all measurements must be present in order for the models to run, and this mode cannot handle missing data.

Multiple Skeleton Mode

To use the multiple skeleton mode, select "**Osteometric Prediction (multiple skeletons)**" or "**Craniometric Prediction (multiple skeletons)**" from the sidebar. This mode allows you to upload a CSV file containing measurements from multiple skeletons, so that sex can be predicted for all of them simultaneously.

1. Prepare your CSV file: The file must contain the measurements for each skeleton in separate rows and the columns must be labeled with the corresponding measurement names. The file must also follow the format specified in the example files provided at the bottom of the page.
2. Upload the CSV file: You can either drag and drop the file into the designated area or use the "Browse" button to navigate to the file on your computer.
3. Select a model: Select one of the available models (XGB, LGB, LDA) for making the predictions.
4. Press the "Calculate" button: The application will use the selected model to predict the sex of each skeleton based on the measurements provided in the CSV file. The results will be displayed in a table, divided into "Male" and "Female" columns, containing the probability that each skeleton belongs to a male or female individual.
5. Note: Any rows containing missing data will be dropped as the models are optimized to use all 11 variables for osteometric and all 32 variables for craniometric datasets, respectively. In cases of missing data, SexEst provides alternative modes (see details below).

Missing data mode

[Osteometric Prediction (missing data) or Craniometric Prediction (missing data)] offer the possibility to make predictions when one or more variables are missing. These tabs are similar to the single skeleton modes; however, the user can now enter even a single variable and get an sex prediction. For postcranial data, the user can enter any combination of one,

two, etc. out of the 11 measurements, while for the craniometric data, the user can enter any combination of one, two, etc. out of the 32 variables. All entered measurements must be in millimeters (mm). The user must select a model (XGB, LGB, LDA), and press the Calculate button. The application will output the probability of the skeleton being male or female. Note that as these models are more computationally intensive, they have been trained using only Linear Discriminant Analysis.