

Gender Classification and Prediction Using Eruption Status of Permanent Teeth in Sri Lankan Children

B.D.H.H. De Silva¹, L.S. Nawarathna^{1*} and V.S.N. Vithanaarachchi²

¹*Department of Statistics and Computer Science, Faculty of Science*

²*Department of Community Dental Health, Faculty of Dental Sciences
University of Peradeniya, Peradeniya 20400, Sri Lanka*

**lakshikas@pdn.ac.lk*

Gender prediction is a dominant component in forensic identification procedures. Teeth are better preserved than other parts of the human body, resistant to high temperature and any mass disaster, thereby providing the best records for forensic investigations. Moreover, tooth eruption is a process that teeth enter the mouth and become visible and eruption status is a measure that can be taken at low cost and efficiency. Hence, this study was undertaken to predict the gender using the eruption status of permanent teeth. Although many studies have been done based on different measures of teeth, this is a novel study for Sri Lankan children using eruption status of permanent teeth. This cross-sectional study was carried out on 3321 children with 1681 males and 1640 females from 7 provinces and 20 schools. A tooth was considered erupted, if more than 1/3rd of the crown is visible and recorded as status 1. Gender is a binary variable and all 28 predictor variables (permanent teeth) are also binary variables with status 1 or 0. Therefore a classification based approach was used to predict the gender of a child using machine learning classifiers, namely Classification and Regression Tree (CART), Random Forest and Extreme Gradient Boosting (XG Boost) which are non-parametric and based on decision trees. The fitted models were validated using 10-fold cross-validation technique and the model performance was measured using accuracy, F1 score, sensitivity and specificity. The Extreme Gradient Boosting classifier was selected as the best performance model compared to the other fitted models which are performed with the highest accuracy 62.80% and 0.6340 F1 score. Hence, the gender of a child can be predicted using the best-fitted model in situations as forensic investigations. Furthermore, gathering data covering all the districts in Sri Lanka would enhance the performance of the proposed classifier for gender prediction.

Keywords: Gender classification, Classification tree, Eruption status, Extreme gradient boosting, Random forest