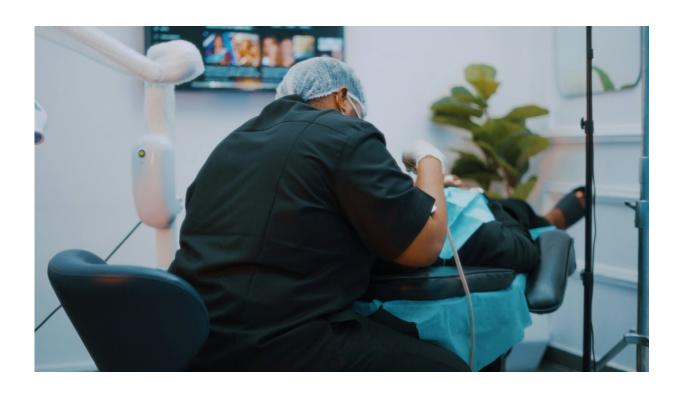# Predicting oral cancer-related mortality among adults using machine learning approach

March 19 2024

Credit: Unsplash/CC0 Public Domain

A study aiming to predict oral cancer-related mortality among adults in

the United States and identify the predictors of oral cancer-related mortality using the Machine Learning Approach. was presented at the [102nd General Session of the IADR](#), which was held in conjunction with the 53rd Annual Meeting of the American Association for Dental, Oral, and Craniofacial Research and the 48th Annual Meeting of the Canadian Association for Dental Research, on March 13-16, 2024, in New Orleans, LA, U.S.

The abstract, "Predicting Oral Cancer-Related Mortality among Adults Using Machine Learning Approach," was presented during the "Artificial Intelligence and Machine Learning Applications in Oral Health" Oral Session on Thursday, March 14, 2024, at 8 a.m. Central Standard Time (UTC-6).

The study, by Aavishi Arora of the Kornberg School of Dentistry at Temple University, Philadelphia, PA, U.S., extracted data for 8,176 participants from the SEER database (1975 to 2022).

A series of 38 demographic, clinicopathological, and lifestyle factors were extracted along with the outcome variable Oral Cancer-Related Mortality (OCRM) coded as "Died from Oral Cancer" and "Alive/Died from Other Causes." The data were pre-processed using recipe packages in R. Machine Learning (ML) models-extreme gradient boosting (XGBOOST) was used to perform prediction of oral cancer prognosis under five-fold cross-validation to prevent overfitting or underfitting of the data.

Model performance was evaluated using the Brier score, area under the curve (AUC), specificity, sensitivity, and accuracy. An ML model was performed using the MachineShop Package in R. The study participants were 63% male and predominantly non-Hispanic white (71%). 7,444

participants were alive or dead of other causes, and 732 were dead due to cancer.

The prediction performance of the ML model (XGBoost) showed a Brier Score of 0.0677, an accuracy of 91%, a 13% kappa statistic, an ROC AUC of 84%, a sensitivity of 99%, and less than 1% specificity. Out of 38 variables assessed, 17 were found to be the most important predictors of OCRM.

The most important predictors of OCRM (in descending order) were cancer stage group, age, T stage, Lymph node surgery, cancer site, tumor rarity, N stage, marital status, radiation, income, grade, lymph node size, surgery radiation sequence, race, histology, the sequence number of multiple primary cancers, side of a paired organ which tumor originated from. The Machine-Learning model was therefore effective in predicting oral cancer mortality using clinicopathological variables from the National Cancer Registry.

Provided by International Association for Dental, Oral, and Craniofacial Research