

FOR IMMEDIATE RELEASE  
December 9, 2025

CONTACT: Matt Niner  
+1.703.299.8084  
[media@iadr.org](mailto:media@iadr.org)

## Uncovering Dental Caries Heterogeneity in NHANES Using Machine Learning

**Alexandria, VA** – A new article published in the *Journal of Dental Research* explores the development of an integrated data-cleaning and subtype discovery pipeline using unsupervised machine learning for comprehensive analysis and visualization of data patterns in the National Health and Nutrition Examination Survey (NHANES) database.

Authored by Alena Orlenko, Cedars-Sinai Medical Center, Los Angeles, CA, USA, et al., “Uncovering Dental Caries Heterogeneity in NHANES Using Machine Learning” addresses the limitations of the NHANES, one of the largest curated repositories of nationally representative population-level health-related indicators, by establishing a data-cleaning pipeline with a novel outlier detection algorithm and unsupervised machine learning to identify phenotype subtypes within NHANES dental caries data.

“By bringing the power of machine learning to a large national data set, the authors identify key clusters of factors linked to caries in children or seniors,” said Nick Jakubovics, Editor-in-Chief of *Journal of Dental Research*. “The next challenge is to build on this information and find more effective methods to prevent caries in different groups of people.”


The study demonstrates a robust data-cleaning–subtype discovery pipeline that could be applied to investigate other health conditions using NHANES and similar databases for machine learning predictive modeling. Applying a comprehensive bioinformatics pipeline to NHANES data successfully identified substantial age-driven heterogeneity in dental caries, suggesting stratification is crucial for future predictive modeling.

This integrative approach systematically addresses data quality issues and facilitates exploratory analysis to reveal data patterns associated with subtypes and variables associated with the clinical heterogeneity of caries. It uncovered novel associations between caries status, lead/pollutant exposure, specific laboratory markers and food types, as well as sleep patterns, reflecting additional disease markers in susceptible populations. This demonstrates the value of integrating data science techniques with large-scale observational data to gain deeper insights into complex, multifactorial diseases.

### About the *Journal of Dental Research*

The IADR/AADOCR *Journal of Dental Research* (JDR) is a multidisciplinary journal dedicated to the dissemination of new knowledge in all sciences relevant to dentistry and the oral cavity and associated structures in health and disease. The *JDR* Editor-in-Chief is Nicholas Jakubovics, Newcastle University, England. Follow the *JDR* on Twitter at [@JDentRes](https://twitter.com/JDentRes).

T +1.703.548.0066  
F +1.703.548.1883  
1619 Duke Street  
Alexandria, VA 22314-3406, USA  
[www.iadr.org](http://www.iadr.org) • [www.aadocr.org](http://www.aadocr.org)



**About IADR/AADOCR**

IADR is a nonprofit organization with a mission to drive dental, oral, and craniofacial research for health and well-being worldwide. IADR represents the individual scientists, clinician-scientists, dental professionals, and students based in academic, government, non-profit, and private-sector institutions who share our mission. AADOCR is the largest division of IADR. Learn more at [www.iadr.org](http://www.iadr.org).