Big Data & AI for Pediatric Health

Machine Learning and Functional Validated Analysis of Orofacial Cleft

Principal Investigator
May D. Wang, PhD

PROJECT OVERVIEW

This project applies machine learning to one of the major translational gaps in orofacial cleft care: identifying which genes are associated with specific cleft phenotypes and determining whether discovered gene variants are likely pathogenic. The work combines 4D imaging, whole genome sequencing, multi-omic data, and functional validation to better understand the genetic drivers of cleft formation.

The Challenge

Orofacial clefts are among the most common structural birth anomalies, yet advances in cleft genetics have not fully translated into clinical tools that can guide diagnosis, prognosis, or individualized care. Researchers need better methods to connect cleft phenotypes with candidate genes and determine which variants are clinically meaningful.

The Innovation

The project uses advanced machine learning approaches, including GeneDAE, to prioritize genes and variants for experimental validation. The team has demonstrated that GeneDAE can identify SNPs and genes associated with IRF6, a known cleft-related gene, and is extending the analysis to TP63 while expanding from targeted chromosome-level analysis to genome-wide discovery.

Potential Impact

This work could help clinicians and researchers better identify genetic contributors to orofacial clefts, improve variant interpretation, and support more personalized approaches to diagnosis, risk assessment, and long-term care.

INSTITUTION	FUNDING	STATUS	TIMELINE
Georgia Institute of Technology	$236,260	Active Research Ongoing	2025–2026