

## A Machine Learning-Based Analysis of Autism Prevalence Rates across US States against Multiple Potential Explanatory Variables

**Authors :** Ronit Chakraborty, Sugata Banerji

**Abstract :** There has been a marked increase in the reported prevalence of Autism Spectrum Disorder (ASD) among children in the US over the past two decades. This research has analyzed the growth in state-level ASD prevalence against 45 different potentially explanatory factors, including socio-economic, demographic, healthcare, public policy, and political factors. The goal was to understand if these factors have adequate predictive power in modeling the differential growth in ASD prevalence across various states and if they do, which factors are the most influential. The key findings of this study include (1) the confirmation that the chosen feature set has considerable power in predicting the growth in ASD prevalence, (2) the identification of the most influential predictive factors, (3) given the nature of the most influential predictive variables, an indication that a considerable portion of the reported ASD prevalence differentials across states could be attributable to over and under diagnosis, and (4) identification of Florida as a key outlier state pointing to a potential under-diagnosis of ASD there.

**Keywords :** autism spectrum disorder, clustering, machine learning, predictive modeling

**Conference Title :** ICBDAH 2023 : International Conference on Big Data Analytics in Healthcare

**Conference Location :** Amsterdam, Netherlands

**Conference Dates :** January 23-24, 2023