Suggested Readings at the Intersection of Machine Learning and Epidemiology
In chronological order

Shmueli G. To explain or to predict? Statistical science 2010; 25:289-310.

Westreich D, Lessler J, Funk M. Propensity score estimation: neural networks, support vector machines, decision trees (CART) and meta-classifiers as alternatives to logistic regression. J Clin Epidemiol 2010; 63:826-33.

Touw WG, Bayjanov, Overmars L, Backus L. et al. Data mining in the life sciences with random forest: a walk in the park or lost in the jungle? Brief Bioinformatics 2013; 14:315-326.

Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: Traps in Big Data Analysis. Science 2014; 343:1203-1205.

Chiavegatto Filho ADP, Dos Santos HG, do Nascimento CF, Massa K, Kawachi I. Overachieving municipalities in public health: A machine learning approach. Epidemiology 2018; 29:836-40.

Naimi AI, Platt RW, Larkin JC. Machine learning for fetal growth prediction. Epidemiology 2018; 29:290-298.

Mooney SJ and Pejaver V. Big Data in Public Health: Terminology, machine learning and privacy. Annual Review of Public Health 2018; 39:95-112.

Obermeyter Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019; 366:447-453.

Chowdhury AS, Lofgren ET, Moehring RW, Broschat SL. Identifying predictors of antimicrobial exposure in hospitalized patients using a machine learning approach Journal of Applied Microbiology 2019; doi:10.1111/jam.14499.

Platt RW, Grandi SM. Machine learning for the prediction of postpartum complications is promising, but needs rigorous evaluation BJOG 2019; 126:710.

Weichenthal S, Hatzopoulou M, Brauer M. A picture tells a thousand…exposures: Opportunities and challenges of deep learning image analysis in exposure science and environmental epidemiology. Environment International 2019; 122:3-10.

Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. AJE 2019; 188:2222-2239.

Robinson WR, Renson A, Naimi AI. Teaching yourself about structural racism will improve your machine learning. Biostatistics 2020; 21:339-344.

Baurley JW, Kjaersgaard AK, Zwick ME, Cronin-Fenton DP, Collin LJ, Damkier P, Hamilton-Dutoit S, Lash TL, Ahern TP. Bayesian pathway analysis for complex interactions AJE 2020; 189:1610-1622.

Blakely T, Lynch J, Simons K, Bentley R, Rose S. Reflecting on modern methods: when worlds collide-prediction, machine learning and causal inference. Int J Epi 2021; 49:2058-64.

Jiang T, Gradus JL, Lash TL, Fox MP. Addressing measurement error in random forests using quantitative bias analysis. AJE 2021; 190:1830-40.

Broadbent A, Grote T. Can robots do epidemiology? Machine learning, causal inference and predicting the outcomes of public health interventions. Philos Technol. 2022; 35:14.

Roh ME, Mpimbaza A, Oundo B, Irish A, et al. Association between indoor residual spraying and pregnancy outcomes: a quasi-experimental study from Uganda. IJE 2022; Online Ahead of Print. doi: 10.1093/ije/dyac043.

Althomsons SP, Winglee K, Heilig CM, Talarico S et al. Using machine learning techniques and national tuberculosis surveillance data to predict excess growth in genotypes tuberculosis clusters. AJE 2022; Online Ahead of Print doi:10.1093/aje/kwac117

Rundle AG, Bader MDM, Mooney SJ. Machine learning approaches for measuring neighborhood environments in epidemiologic studies. Curr Epidemiol Rep 2022; Online Ahead of Print; doi: 10.1007/s40471-022-00296-7.