Prediction of Skin Sensitization Potency Using Machine Learning Approaches

J Strickland[1], Q Zang[1], M Paris[1], DM Lehmann[2], S Bell[1], D Allen[1], J Matheson[3], A Jacobs[4], W Casey[5], N Kleinstreuer[5]

[1]ILS, RTP, NC, USA; [2]EPA/ORD/NHEERL, RTP, NC, USA; [3]CPSC, Bethesda, MD, USA; [4]FDA, Silver Spring, MD, USA; [5]NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA;

Replacing animal tests currently used for regulatory hazard classification of skin sensitizers is one of ICCVAM's top priorities. Accordingly, U.S. federal agency scientists are developing and evaluating computational approaches to classify substances as sensitizers or nonsensitizers. Some regulatory agencies require that sensitizers be further classified into potency categories. We have built machine learning models to classify substances as strong sensitizers, weak sensitizers, or nonsensitizers according to the Globally Harmonized System of Classification and Labeling of Chemicals for both local lymph node assay (LLNA) and human outcomes. The models were based on data from three in chemico or in vitro assays (direct peptide reactivity assay, human cell line activation test, and KeratinoSens™ assay) and six physicochemical properties. Combinations of these input variables were modeled with four machine learning approaches: classification and regression tree, linear discriminant analysis, logistic regression, and support vector machine (SVM). Two different strategies were used for modeling: a one-tiered multi-class strategy modeled all three categories of response, while a two-tiered binary strategy modeled sensitizer vs. nonsensitizer responses and then differentiated between strong and weak sensitizers. Models were developed on a training set and evaluated using an external test set. Leave-one-out cross validation (LOOCV) was used to evaluate the models for overfitting. The two-tiered models performed better than the one-tiered models and SVM outperformed the other machine learning approaches. Models using the variable group with all input variables performed better than variable groups with assay data only, or physicochemical properties only, or assay data plus log P. The two-tiered model using SVM and all input variables provided the best performance: accuracy for the LOOCV was 90% for LLNA (120 substances) and 85% for human (87 substances) outcomes. Accuracy for the best one-tiered models was 78% for LLNA outcomes and 75% for human outcomes. This compares to an accuracy of 69% for LLNA prediction of human potency categories. These results suggest that computational approaches may be useful for assessing skin sensitization potency. This project was funded in whole or in part with Federal funds from the NIEHS, NIH under Contract No.HHSN273201500010C. This abstract does not represent EPA policy.

Presenting author: Judy Strickland, strickl2@niehs.nih.gov