

## Prediction of Skin Sensitization Potency Using Machine Learning Approaches

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

<sup>1</sup>ILS, RTP, NC, USA; <sup>2</sup>EPA/ORD/NHEERL, RTP, NC, USA; <sup>3</sup>CPSC, Rockville, MD, USA;

<sup>4</sup>FDA/CDER, Silver Spring, MD, USA; <sup>5</sup>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 built machine learning models to classify substances using GHS categories as strong (1A) sensitizers, weak (1B) sensitizers, or nonsensitizers for both local lymph node assay (LLNA) and human outcomes. The models used data from three in chemico or in vitro assays (direct peptide reactivity assay, KeratinoSens™, and human cell line activation test) and six physicochemical properties. To determine the optimal model, four machine learning approaches were used: classification and regression tree, linear discriminant analysis, logistic regression, and support vector machine (SVM). Two different strategies were used for modeling: a one-tiered strategy modeled all three potency categories in a single step, while a two-tiered strategy first delineated sensitizer vs. nonsensitizer responses and then strong and weak sensitizers. Models were developed on a training set (94 substances for LLNA; 63 for human) and evaluated using an external test set (26 substances for LLNA; 24 for human). 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. The SVM models using all input variables performed better than those using subsets of the input variables. The two-tiered model using SVM and all input variables provided the best performance: LOOCV accuracy was 88% for LLNA (120 substances) and 81% for human (87 substances) outcomes. This compares to an accuracy of 69% for LLNA prediction of human potency categories (87 substances). 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 U.S. EPA policy or the policy of any federal agency.*

**Keywords:** chemical allergy, alternatives to animal testing, predictive toxicology, computational toxicology