PP13

Toxicological Profile Characterization via Application of Machine Learning to SEND-Formatted Toxicology Study Data

Md Aminul Islam Prodhan[1,2], Md Yousuf Ali[1,2], Kevin Snyder[1]
[1]U.S. Food and Drug Administration (FDA), Silver Spring, MD; [2]Oak Ridge Institute for Science and Education, Oak Ridge, TN.
Email: Kevin.Snyder@fda.hhs.gov

BACKGROUND

CDER has established collaborative agreements with industry consortia, i.e., PHUSE and BioCelerate, to develop and publicize novel methods of extracting value from SEND datasets. An R package, sendigR, was collaboratively developed for use by all stakeholders with access to SEND datasets to facilitate large-scale cross-study analyses. Subsequent collaboration between FDA and BioCelerate led to the development of scoring methods for normalization of toxicology study endpoints, facilitating comparison of results across studies. These score values have been used to develop machine learning models trained on manually labeled datasets to characterize toxicological profiles from SEND datasets that may be used to develop quantitative structure activity relationship (QSAR) models.

Liver was selected as the first general toxicology study endpoint to be addressed primarily because it plays important functional roles in filtering blood and nutrients ingested into the digestive tract leaving it more susceptible to the exposure of toxic agents than other organs. Indeed, drug-induced hepatotoxicity accounts for greater than 50% of acute liver failure cases in the United States and has been the cause for post-market attrition of several pharmaceuticals.

Herein, we present a workflow to leverage data across SEND studies that have been reviewed by experts to diagnose target organs of toxicity. The workflow model the relationship between the diagnosis of liver as a target organ for each study and normalized score values for hepatotoxicity study endpoints, i.e. body weights, liver weights, liver function test results, and liver histopathology findings. This model can then be used to assess the strength of hepatotoxicity signal in SEND-formatted toxicology study data facilitating the development of QSAR models to predict the propensity of a molecule to induce hepatoxicity in animal models based upon its chemical structure. In the future, this workflow could be amended and improved with the submission and incorporation of new SEND data, as well as be extend to other complex toxicity endpoints (e.g., Kidney, Heart, Lung, etc.) .

LIVER SCORE CALCULATION

For each of the SEND STUDYID, normalized toxicity score values for hepatotoxicity study endpoints were calculated where score's range was **0-5.** The weight of each animal at the end of the dosing period in each study was initially normalized by subtracting their baseline weights on the first day of dosing. Then each animals' liver weight to Body weight ratio were calculated. These values then were normalized via Z-score to the respective control groups for each study. For laboratory test data , Z-score were calculated for six enzymes found in blood or serum (Bilirubin, ALB, , ALT, ALP, AST, GGT). The score for the histopathological findings were calculated from both incidence and severity. Initially score was calculated based on the severity data and then adjusted with respect to the severity data.

ML MODEL BUILDING & EVALUATIONS

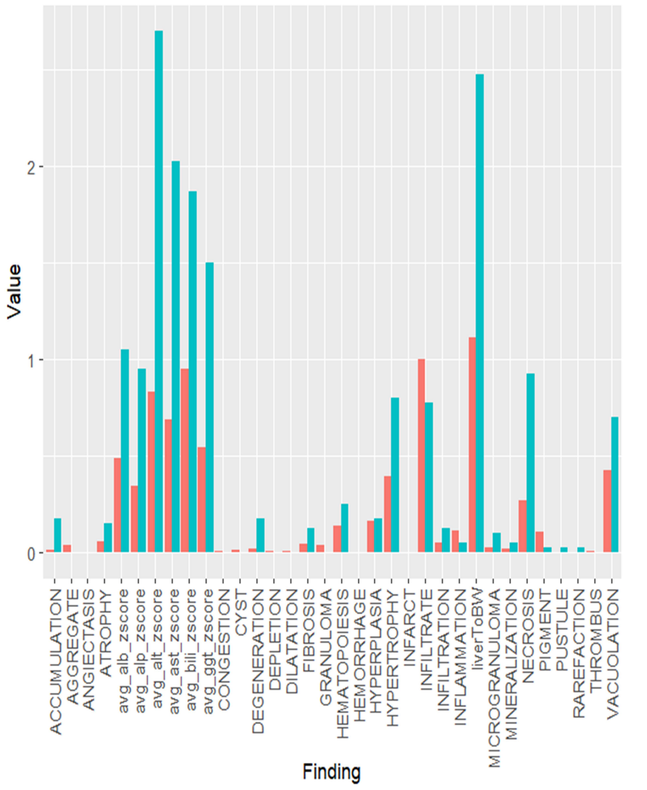

Figure 1. Toxicity score distribution of the toxicity endpoints

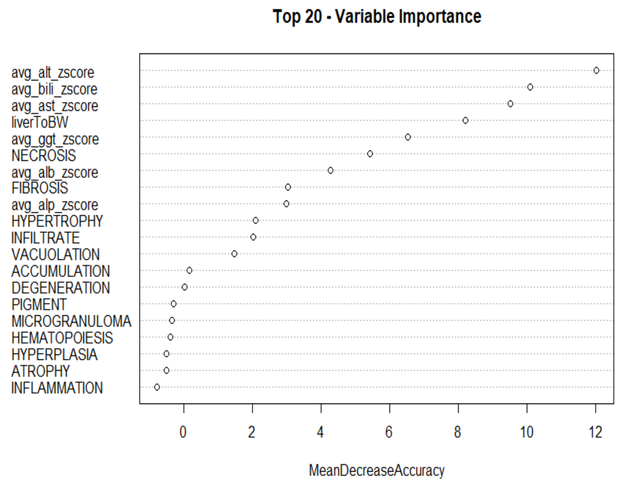

Figure 2. Top 20 variables contributed to the rf model

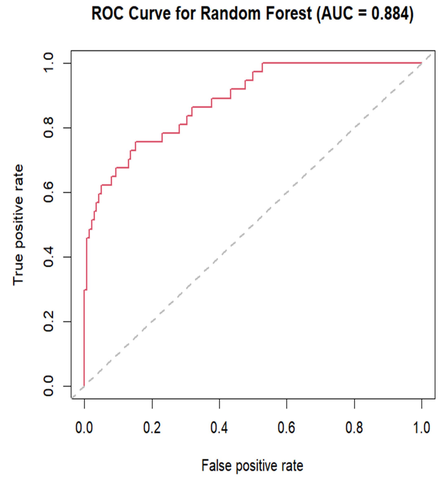

Figure 3. Receiver Operating Characteristic (ROC) Curve

From the SEND data, expert reviewers diagnosed 189 studies as target organs of toxicity for the rat species, with 43 labeled as liver positive and 143 as liver negative. These labeled data were utilized to develop a random forest classification model. **Figure 1,** illustrates the distribution of toxicity scores for each variable used in the model, with the response variable categorized as either liver positive or negative. The dataset was partitioned into training data (80%) and test data (20%) for model construction and evaluation, a process that was iterated 1000 times to ensure robustness and reliability.

In Figure 2, the contributions of the top 20 predictor variables to the Random Forest model development are depicted, highlighting the key factors driving the model's predictive accuracy.

Figure 3, displays the ROC curve, indicating an area under the receiver operating characteristic (ROC) curve (AUC) of 0.884. This signifies our model's robust discriminatory ability in effectively distinguishing between positive and negative cases of hepatotoxicity, , with a sensitivity of 0.75 and specificity of 0.92.

DISCUSSION AND FUTURE DIRECTIONS

In this work, we developed machine learning model trained on datasets labeled by experts to delineate toxicological profiles from SEND data. Specifically, employing a random forest approach with 500 trees and 5 variables considered at each split, our model exhibited an out-of-bag (OOB) estimate of the error rate at 19.35%. Delving deeper into the model's performance, the confusion matrix unveiled insightful details. Among the true positive cases (1), 24 were correctly classified, while 7 were misclassified, resulting in a class error rate of approximately 22.58%. For the true negative cases (0), 26 were accurately classified, with 5 misclassifications, leading to a class error rate of around 16.13%. These findings underscore the model's capability in discerning toxicological patterns within SEND-formatted data, thereby laying a solid foundation for future endeavors, such as developing quantitative structure-activity relationship (QSAR) models for predicting hepatotoxicity employing chemical structure and the toxicity score.

ACKNOWLEDGEMENTS

This project was supported in part by an appointment to the Research Fellowship Program at the Immediate Office/OND/CDER, U.S. Food and Drug Administration, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and FDA.

DISCLAIMER

This presentation reflects the views of the authors and should not be construed to represent FDA's views or policies