Skin sensitization is an important aspect of occupational and consumer safety. Because of the ban on animal testing for skin sensitization in Europe, in silico approaches to predict skin sensitizers are needed. Recently, several machine learning approaches, such as the gradient boosting decision tree (GBDT) and deep neural networks (DNNs), have been applied to chemical reactivity prediction, showing remarkable accuracy. Herein, we performed a study on DNN- and GBDT-based modeling to investigate their potential for use in predicting skin sensitizers. We separately input two types of chemical properties (physical and structural properties) in the form of one-hot labeled vectors into single- and dual-input models. All the trained dual-input models achieved higher accuracy than single-input models, suggesting that a multi-input machine learning model with different types of chemical properties has excellent potential for skin sensitizer classification.

Key Words: Skin sensitization, globally harmonized system, fingerprint, deep learning, gradient boosting decision tree

Area of Interest: AI technologies and their applications

1. Introduction

Skin sensitization is a key safety endpoint in occupational and consumer settings. Skin sensitizers have been traditionally assessed by animal testing approaches (OECD TG406). Recently, because Europe has imposed a ban on animal testing for cosmetic ingredients, which includes skin sensitization, in vitro testing approaches (OECD TG442C, TG442E) and in silico predictive models (e.g., Quantitative Structure-Activity Relationship (QSAR) [1]) have begun to be investigated. To meet 3R (Replacement, Reduction, and Refinement) principles, there is a high demand for the development of alternative methods for skin sensitizer prediction.

Among classification algorithms, deep neural network (DNN), which is a type of artificial neural network with more than one layer, has been used to predict chemical reactivity, and won various competitions such as the TOX21 Data Challenge 2014 [2] with a remarkable accuracy.
Recently, Kato et al. reported the high predictive performance of the QSAR/DNN model for chemical activity [3]. Light gradient boosting machine (LightGBM), a type of gradient boosting decision tree (GBDT) was proposed in 2017 [4], and several subsequent reports have verified its potential for use in assessing chemical toxicity [5]. However, there is a lack of articles investigating skin sensitizer classification using these approaches.

In this study, we investigated the study of QSAR-like DNN and LightGBM modeling to obtain insights into its potential to predict skin sensizers. We used chemicals associated with skin sensitzers and non-sensitizers based on the definitions of Globally Harmonized System (GHS) labeling. The physical and structural properties of each chemical were collected as descriptors. The processed data were then input into models to determine its categorization accuracy.

2. Materials and Methods

2.1 Workflow of DNN- and LightGBM-based classification model construction

The workflow of building classification model is briefly described in the following sentences. Further details of the modeling are given in the Supplemental Information (Section A).

The list of chemicals with GHS labeling for skin sensitization was prepared using data from the National Institute of Technology and Evaluation [6]. The chemicals categorized as skin sensitizers (category 1 and sub-categories 1A/1B) were defined as “positive,” and the out-of-category chemicals were defined as “negative.” After data processing, the numbers of positive and negative chemicals were 408 and 275, respectively.

We used 200 chemical descriptors to construct a physical property for each chemical using RDKit library. The values of each descriptor were then standardized using z-score normalization and one-hot encoded. The one-hot encoded data were converted into vectors and then input to models. The MACCS fingerprint of each chemical was obtained as a chemical structural property using the RDKit library. The MACCS fingerprints were converted into one-hot labeled vectors that were input to the models.

The single-input DNN model has two or four hidden layers. The dual-input DNN model was developed by concatenating two single-input DNN models with a two hidden layer, followed by two hidden layers. We also developed the single-input LightGBM model by using either physical or structural properties and the dual-input LightGBM model by using both properties.

2.2 Accuracy comparison

We input the physical properties and/or structural properties of the chemicals into the single- and dual-input DNN and LightGBM models. The accuracy, loss, and precision–recall area under curve (PR-AUC) of each model were calculated by 10 iterations of five-fold cross-validation.

3. Results and discussion

In this study, we performed a data-driven approach to classify skin sensitizers based on the DNN and LightGBM algorithm using physical and structural properties of chemicals as input variables. In contrast to the MACCS fingerprint, physical properties of chemicals are continuous variables. Regarding such neural network models as the DNN, Hayashi et al. reported that neural networks trained with augmented discretized input were more accurate than when the original
continuous input was used [7]. Indeed, the DNN model could not be successfully trained with the original physical data of chemicals (data not shown). The loss of all DNN models trained with the one-hot encoded training data decreased over the epochs, suggesting that one-hot encoding of the physical properties of chemicals is a good option for DNN training.

We then compared the performance of the single- and dual-input DNN and LightGBM models. The dual-input model delivered the best performance in both settings, outperforming the corresponding single-input model in terms of the accuracy and PR-AUC. The dual-input LightGBM had significantly good results in terms of PR-AUC. Moreover, 7.1% and 9.7% of the chemicals were uniquely predicted by the DNN and the LightGBM, respectively (Figure S3), suggesting different potential in terms of classifying skin sensitizers.

To the author’s understanding, this is the first report regarding skin sensitizer classification with a DNN and LightGBM models. The dual-input strategy was found to have potential for use in predicting skin sensitizers in both deep learning and GBDT algorithms. Although the sets of chemicals correctly predicted by the two models were almost identical, more true positive compounds were obtained by the dual-input LightGBM model, whereas the dual-input DNN model yielded more true negative compounds (Figure S3). This might reflect differences between the algorithms, such as that in assigning importance to features. Golden et al. reported the accuracy of eight in silico skin sensitization prediction models based on the Hazardous Substances Data Bank [8]. Although it is difficult to compare the model precisions of other investigations, the dual-DNN model (72%) and the dual-input LightGBM model (74%) was fairly accurate when compared with other traditional approaches (55%–81%). Although we manually optimized the hyper-parameters within a limited scale, the results suggest the dual-input machine learning model for skin sensitizer classification performs very well. The dual-input LightGBM model showed significantly better performance than single-input models, while no statistically significant difference was found between these the DNN models. The general drawback of DNN models is that small and imbalanced datasets, such as the skin sensitizer dataset, negatively affect performance. Recently, Matsuzaka et al. developed DeepSnap, a procedure for generating an omnidirectional snapshot portraying the 3D structures of chemicals [9]. This approach would enable data to be augmented, which could solve the dataset problem of skin sensitizers. In addition, we used only the MACCS fingerprint as a structural property. Among the prediction algorithms based on molecular structures, graph convolutional networks (GCN), a type of neural network for graph-related tasks, has exhibited great performances in the classification evaluation. Kojima et al. recently developed kGCN, an open-source software which provides GCN-based prediction of chemical properties, and they also showed that the multi-modal approach delivers better predictive performance than single-task approaches [10]. These results suggests that the use of structural properties other than MACCS keys may improve the performance of the dual-input DNN model. We also should investigate the chemical space of the dual-input models next to precisely understand the potential for skin sensitizer classification.

Acknowledgments

I am grateful to Drs. Tomoki Nishino, Tsuneo Hashizume, Hitoshi Fujimoto, Yuichiro Takanami, Shigeaki Ito, and Shinkichi Ishikawa for insightful comments and suggestions on this study. I thank Drs. Kimberly Moravec and Saad Anis from Edanz Group (https://en-author-services.edanzgroup.com/) for reviewing and editing a draft of this manuscript.
Conflict of interest

The author is an employee of Japan Tobacco Inc. and declares no conflict of interests with respect to the research, authorship, and/or publication of this article.

Table 1. Categorization performance of the learning models

| Model         | Input Layers/Depth | Physical  | Structural | Physical  | Structural |
|---------------|--------------------|-----------|------------|-----------|------------|
|               |                    | Accuracy  |            | Accuracy  |            |
|               |                    | (S.D.)    |            | (S.D.)    |            |
|               | DNN                |           |            |           |            |
|               |                    | 2         | 2          | 4         |            |
|               |                    | 0.715     | 0.713      | 0.715     | 0.704      |
|               |                    | (0.043)   | (0.039)    | (0.044)   | (0.042)    |
|               |                    | 0.704     | 0.704      | 0.704     | 0.720      |
|               |                    | (0.043)   | (0.044)    | (0.037)   | (0.043)    |
|               |                    | 0.720     | 0.720      | 0.720     | 0.720      |
|               |                    | (0.037)   | (0.038)    | (0.037)   | (0.038)    |
|               |                    | 0.720     | 0.720      | 0.720     | 0.720      |
|               |                    | (0.037)   | (0.038)    | (0.037)   | (0.038)    |
|               |                    | 0.720     | 0.720      | 0.720     | 0.720      |
|               |                    | (0.037)   | (0.038)    | (0.037)   | (0.038)    |
|               | LightGBM           |           |            |           |            |
|               |                    | 2         | 6          | 4         |            |
|               |                    | 0.752     | 0.752      | 0.752     | 0.767      |
|               |                    | (0.050)   | (0.042)    | (0.052)   | (0.049)    |
|               |                    | 0.767     | 0.767      | 0.767     | 0.773      |
|               |                    | (0.050)   | (0.044)    | (0.052)   | (0.051)    |
|               |                    | 0.773     | 0.773      | 0.773     | 0.773      |
|               |                    | (0.050)   | (0.044)    | (0.052)   | (0.051)    |
|               |                    | 0.773     | 0.773      | 0.773     | 0.773      |
|               |                    | (0.050)   | (0.044)    | (0.052)   | (0.051)    |
|               |                    | 0.773     | 0.773      | 0.773     | 0.773      |
|               |                    | (0.050)   | (0.044)    | (0.052)   | (0.051)    |
|               |                    | 0.773     | 0.773      | 0.773     | 0.773      |
|               |                    | (0.050)   | (0.044)    | (0.052)   | (0.051)    |
|               |                    | 0.773     | 0.773      | 0.773     | 0.773      |
|               |                    | (0.050)   | (0.044)    | (0.052)   | (0.051)    |
|               |                    | 0.773     | 0.773      | 0.773     | 0.773      |
|               |                    | (0.050)   | (0.044)    | (0.052)   | (0.051)    |

References

[1] Alves, V. M.; Muratov, E.; Fourches, D.; Strickland, J.; Kleinstreuer, N.; et al. Predicting chemically-induced skin reactions. Part I: QSAR models of skin sensitization and their application to identify potentially hazardous compounds. *Toxicol. Appl. Pharmacol.* 2015, 284(2), 262–272. doi:10.1016/j.taap.2014.12.014

[2] Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. DeepTox: toxicity prediction using deep learning. *Front. Environ. Sci.* 2016, 3, 1–15. doi:10.3389/fenvs.2015.00080

[3] Kato, Y.; Hamada, S.; Goto, H. Validation study of QSAR/DNN models using the competition datasets. *Mol. Inform.* 2020, 39(1–2), 1–10. doi:10.1002/minf.201900154

[4] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; etc. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural. Inf. Process. Syst.*; Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., etc., Eds.; NIPS: Long Beach, 2017; Vol. 30, pp. 3146–3154.

[5] Zhang, J.; Mucs, D.; Norinder, U.; Svensson, F. LightGBM: An Effective and Scalable Algorithm for Prediction of Chemical Toxicity—Application to the Tox21 and Mutagenicity Data Sets. *J. Chem. Inf. Model.* 2019, 59(10), 4150–4158. doi:10.1021/acs.jcim.9b00633

[6] National Institute of Technology and Evaluation, https://www.nite.go.jp/chem/english/ghs/ghs_download.html (accessed May 23, 2020)

[7] Hayashi, Y.; Setiono, R.; Azcarraga, A. Neural network training and rule extraction with augmented discretized input. *Neurocomputing* 2016, 207, 610–622. doi:10.1016/j.neucom.2016.05.040

[8] Golden, E.; Macmillan, D. S.; Dameron, G.; Kern, P.; Hartung, T.; et al. Evaluation of the global performance of eight in silico skin sensitization models using human data. *ALTEX*. 2020. doi:10.14573/altex.1911261

[9] Matsuzaka, Y.; Uesawa, Y. Optimization of a Deep-Learning Method Based on the Classification of Images Generated by Parameterized Deep Snap a Novel Molecular-Image-Input Technique for Quantitative Structure–Activity Relationship (QSAR) Analysis. *Front. Bioeng. Biotechnol.* 2019, 7, 1–15. doi:10.3389/fbioe.2019.00065

[10] Kojima, R.; Ishida, S.; Ohta, M.; Iwata, H.; Honma, T.; et al. kGCN: a graph-based deep learning framework for chemical structures. *J. Cheminformatics.* 2020, 12(32), 1–10. doi:10.1186/s13321-020-00435-6