ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

A Comparative Study of the Performance for Predicting Biodegradability Classification: The Quantitative Structure−Activity Relationship Model vs the Graph Convolutional Network

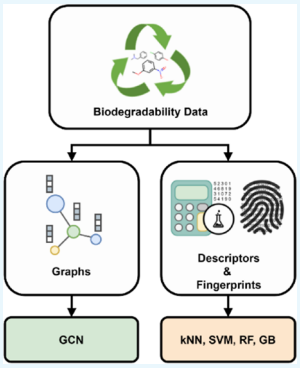
Myeonghun Lee and Kyoungmin Min*

Cite This: *ACS Omega* 2022, 7, 3649−3655

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The prediction and evaluation of the biodegradability of molecules with computational methods are becoming increasingly important. Among the various methods, quantitative structure−activity relationship (QSAR) models have been demonstrated to predict the ready biodegradation of chemicals but have limited functionality owing to their complex implementation. In this study, we employ the graph convolutional network (GCN) method to overcome these issues. A biodegradability dataset from previous studies was trained to generate prediction models by (i) the QSAR models using the Mordred molecular descriptor calculator and MACCS molecular fingerprint and (ii) the GCN model using molecular graphs. The performance comparison of the methods confirms that the GCN model is more straightforward to implement and more stable; the specificity and sensitivity values are almost identical without specific descriptors or fingerprints. In addition, the performance of the models was further verified by randomly dividing the dataset into 100 different cases of training and test sets and by varying the test set ratio from 20 to 80%. The results of the current study clearly suggest the promise of the GCN model, which can be implemented straightforwardly and can replace conventional QSAR prediction models for various types and properties of molecules.



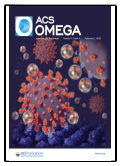
INTRODUCTION

Microorganisms remove organic materials from the environment through oxidation, reduction, and hydrolysis. This process is defined as biodegradation and is an essential method of removing pollutants from the environment.[1] However, the accumulation of non-biodegradable chemicals poses a potential threat to humans and ecosystems. Some synthetic plastics, such as polyethylene−starch blends and polyester polyurethane, are biodegradable, but the most commonly used plastics are non-biodegradable or require decades to degrade completely.[2] Therefore, it is essential to establish effective evaluation methods for classifying the potential biodegradability of organic materials.

European legislators included chemical persistency in Registration, Evaluation, and Authorization of Chemicals (REACH) for the evaluation of chemicals.[3] REACH requires biodegradability assessments of chemicals produced or imported in quantities of more than a ton per year.[4] However, only 61% of chemicals produced or imported in quantities of more than 1000 tons per year have information on biodegradability.[5] Therefore, it is critical to assess the biodegradability of the remaining chemicals. As a potential solution, REACH encourages the use of quantitative structure−activity relationship (QSAR) models to predict the biodegradability of compounds.[6]

In particular, several QSAR classification models have recently been proposed for predicting the "ready biodegradable" (RB) or "not ready biodegradable" (NRB) class of materials[7] as artificial intelligence (AI) methods have become more accessible. Previous studies have applied machine learning algorithms, such as partial least squares discriminant analysis (PLSDA), multiple linear regression (MLR), logistic regression (LR), naive Bayes (NB), k-nearest neighbors (kNN), and support vector machines (SVM), as QSAR models to molecular fingerprints and descriptors to develop a biodegradability classification model.[1,3,7−9] The datasets, Sn, and Sp results of previous studies using QSAR models for biodegradability prediction are summarized in Table S1. Although various models and datasets are implemented, it is not clear whether which regressors are superior to the others. In addition, software has been developed to calculate molecular descriptors, such as PaDEL-Descriptor[10] and DRAGON,[11] which were used as features to develop QSAR models.[3,12] The freely available PaDEL-Descriptor can calculate 1875 molecular descriptors, such as atom-type electrotopological state descriptors, McGowan volume, molecular linear free energy relation descriptors, ring counts, and 10

Received: November 8, 2021
Accepted: December 28, 2021
Published: January 14, 2022



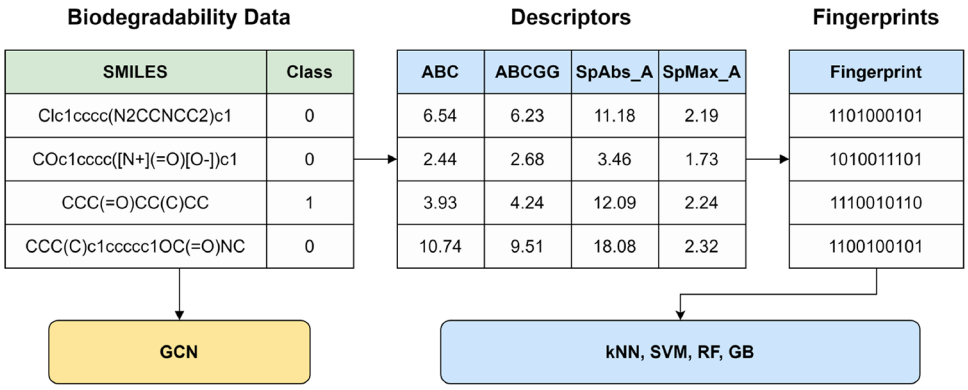


Figure 1. Flowchart conducted in this study. The dataset (green) consists of SMILES and RB/NRB class data. The QSAR model (blue) consists of applying four machine learning algorithms by generating descriptors and fingerprints that can be calculated from SMILES. The GCN model (yellow) was composed of converting SMILES into molecular graphs and using the architecture of Figure 2c.

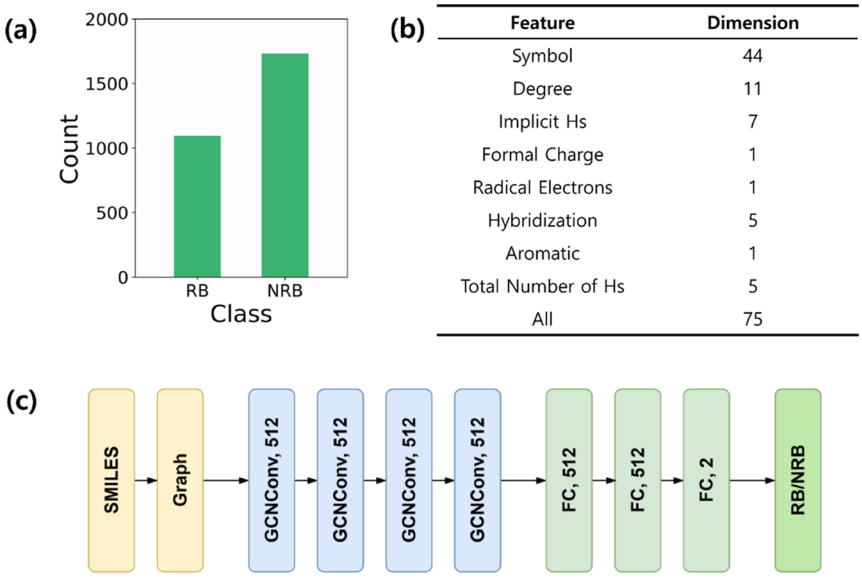


Figure 2. (a) Visualizing the number of RB/NRB in the dataset. (b) Features of atoms constituting the node feature matrix of the molecular graph. (c) GCN model architecture.

types of fingerprints.[10] DRAGON is another widely used tool that can calculate 5270 molecular descriptors, but it is difficult to use because of licensing issues.[13]

While it has been proven that integrating data and ensemble analysis can increase the reliability of QSAR models, they commonly face problems with uncertainty for various reasons.[7] For example, QSAR models make false correlations owing to errors in the experimental process or do not fully reflect the characteristics of the data because of the small number of training databases.[14] In addition, they intrinsically require the generation of appropriate features to train the models. Hence, choosing the appropriate features is often a complicated issue. For example, several structural features of molecules (such as halogen, chain branching, and nitro groups) have been shown to increase biodegradation time, while others (such as esters, amides, and hydroxyl groups) have been found to decrease biodegradation time. However, these structural features cannot be generalized to represent both RB and NRB molecules.[15]

To overcome the above-mentioned limitations, we intend to predict biodegradability using graph neural networks (GNNs) and compare their performance with that of QSAR models. In particular, graph convolutional networks (GCNs)[16] are implemented, which produce outstanding performance in various fields dealing with molecules, such as molecular property, activity, interaction, and synthesis prediction.[17] This indicates that such a method can potentially replace the QSAR model.[18] In chemistry, the atoms and bonds that make up a molecule can be naturally converted into graphs by mapping them into sets of nodes and edges,[19] which can be used as input features. This study compares the procedures and results of QSAR models and the GCN model for biodegradability classification, demonstrates the advantages and disadvantages, and discusses the great potential of the GCN model. The overall procedure of the two models is shown in Figure 1. The GCN model can be directly applied by converting the simplified molecular input line entry system (SMILES) into graphs, whereas the QSAR model is applied by calculating and selecting molecular descriptors and fingerprints from SMILES. This schematic indicates that the GCN implementation is less complicated and requires less information than the QSAR model.

https://doi.org/10.1021/acsomega.1c06274
ACS Omega 2022, 7, 3649−3655

METHODS

Biodegradability Database. This study used "All-Public set", an organized and aggregated dataset from five different sources:[9] the ECHA database,[20] the NITE database,[21] and the training sets of the existing tools VEGA,[22] EPI Suite,[23] and OPERA.[24] Subsequently, CADD Group Chem and User Services[25] and PubChem[26] were used to verify the accuracy of SMILES. Standardization was performed using workflows implemented in Konstanz Information Miner (KNIME).[27] Additionally, deduplication was based on standardized SMILES matching. The datasets represent biodegradable classes with RB and NRB, as shown in Figure 2a, which are 1097 and 1733, respectively.

Molecular Descriptors. Mordred,[13] a molecular descriptor calculation software that can calculate more than 1800 two- and three-dimensional descriptors, was used for the QSAR model. It is an open-source software that is easier to install and use than other libraries (e.g., Cinfony[28] and ChemoPy[29]) and can be employed flexibly with a rapid calculation speed. However, 3D structural descriptors, such as CPSA and MoRSE, can cause complex and non-reproducible optimizations.[3] While 3D descriptors represent valuable chemical information about molecules, they may be difficult to apply to new molecules because geometric optimization is necessary. In addition, they vary between 3D conformers, which can affect the values of the 3D descriptors. Therefore, this study used Mordred to calculate a total of 1613 two-dimensional values, such as the adjacency matrix, distance matrix, S log P, and weight. In addition to descriptors, molecular fingerprints are often used for biodegradability classification using QSAR models;[1,7,8] in this study, the most widely used Morgan fingerprint[30] (1024-bit), MACCS fingerprint[31] (167-bit), and, additionally, RDKit topological fingerprint (2048-bit)[32] were tested.

In contrast, the molecular graph for the GCN model was transformed in response to the atoms (nodes) and the bonds between atoms (unweighted edges). The graph is represented by $G = (X, A)$, where X is a 75-dimensional node feature matrix ($N \times 75$, N is the number of nodes) of the chemical features of atoms, as shown in Figure 2b, and A represents an adjacency matrix ($N \times N$), in which $A_{ij} = 1$ when the ith and jth atoms are bonded and $A_{ij} = 0$ otherwise. The molecules were represented by SMILES and converted to graphs using RDKit[32] and PyTorch Geometric (PyG).[33]

Machine Learning Details. Four classification algorithms were applied for QSAR modeling: kNN, SVM, random forest (RF), and gradient boosting (GB). kNN, SVM, and RF have already been implemented for biodegradability predictions,[3,8,9] and GB, one of the representative boosting models, was additionally tested. These models were implemented and tuned using scikit-learn,[34] and their details can be found in the Supporting Information. In particular, for the case of SVM and RF, the weight balancing technique was also applied to consider imbalanced data. In addition, ensemble analysis was applied to combine the prediction results of the four different machine learning algorithms. This method can further enhance the predictive reliability of the model.[35] Because individual models contain a variety of noises, averaging the prediction results of the implemented models can reduce the overall noise.[36] Among the ensemble analysis methods, a soft voting ensemble (SVE) classifier using scikit-learn was applied in this study. This method selects the highest class by summing the probabilities predicted by each model.

Graph Convolutional Network (GCN). As shown in Figure 2c, the GCN model for biodegradability prediction consists of two regions: (i) graph convolution layers (blue) and (ii) prediction layers (green). In the graph convolution layers (which consist of four layers), the features of neighboring atoms and the central atom are calculated by GCNConv,[37] as follows:

$$H^{(l+1)} = f(H^l, A) = \sigma(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)})$$

where $\tilde{A} = A + I$, I is the identity matrix of the adjacency matrix, $\tilde{D}$ is the diagonal node degree matrix, $W^{(l)}$ is the weight matrix of layer l, $H^{(0)} = X$, where X is the feature matrix of the nodes, and $H^{(l)}$ is the output of layer l. The prediction layers consist of three fully connected layers and eventually return prediction results via the log softmax function.

Evaluation Metrics. To compare the results of the QSAR and GCN models, the dataset was randomly divided into training and test sets at a ratio of 8:2, and the proportion of each class was uniformly distributed (stratified sampling) for each set. The classification model was evaluated based on the metrics of balanced accuracy (BA), specificity (Sp), sensitivity (Sn), and error rate (ER) for the prediction of RB and NRB of molecules. The evaluation metrics are calculated as follows:

$$\mathrm{BA} = \frac{\mathrm{Sn} + \mathrm{Sp}}{2}$$

$$\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

$$\mathrm{Sn} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{ER} = \frac{\mathrm{FN} + \mathrm{FP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}}$$

where TN and TP are true negative and true positive values, respectively, while FN and FP are false negative and false positive values, respectively. Additionally, Sn and Sp are inversely proportional; therefore, a higher Sn corresponds to a lower Sp and vice versa.[38]

RESULTS AND DISCUSSION

To develop the most accurate surrogate model, it is imperative to conduct several tests to find appropriate combinations of parameters in various cases using the four QSAR models and one GCN model. In this respect, three approaches were used: (i) BA values from 10-fold cross-validation (CV) were obtained for direct comparison of the machine learning models, (ii) the average and range of four metrics from 100 different configurations were calculated by randomly dividing the dataset into training and test sets, and (iii) the test set ratio was varied from 20 to 80% with 100 different random states, as conducted in case 2, to further verify the prediction stability. By comparing these metrics, essential conditions for improving the model performance, such as several descriptors and algorithms, were selected, and the final model was determined.

Molecular Descriptors and Fingerprints. The performance of the QSAR models was first tested when only Morgan, MACCS, and topological fingerprints were used, as shown in Table S2. Especially, Morgan and MACCS fingerprint methods are widely used in various molecule-related fields, such as

https://doi.org/10.1021/acsomega.1c06274
ACS Omega 2022, 7, 3649−3655

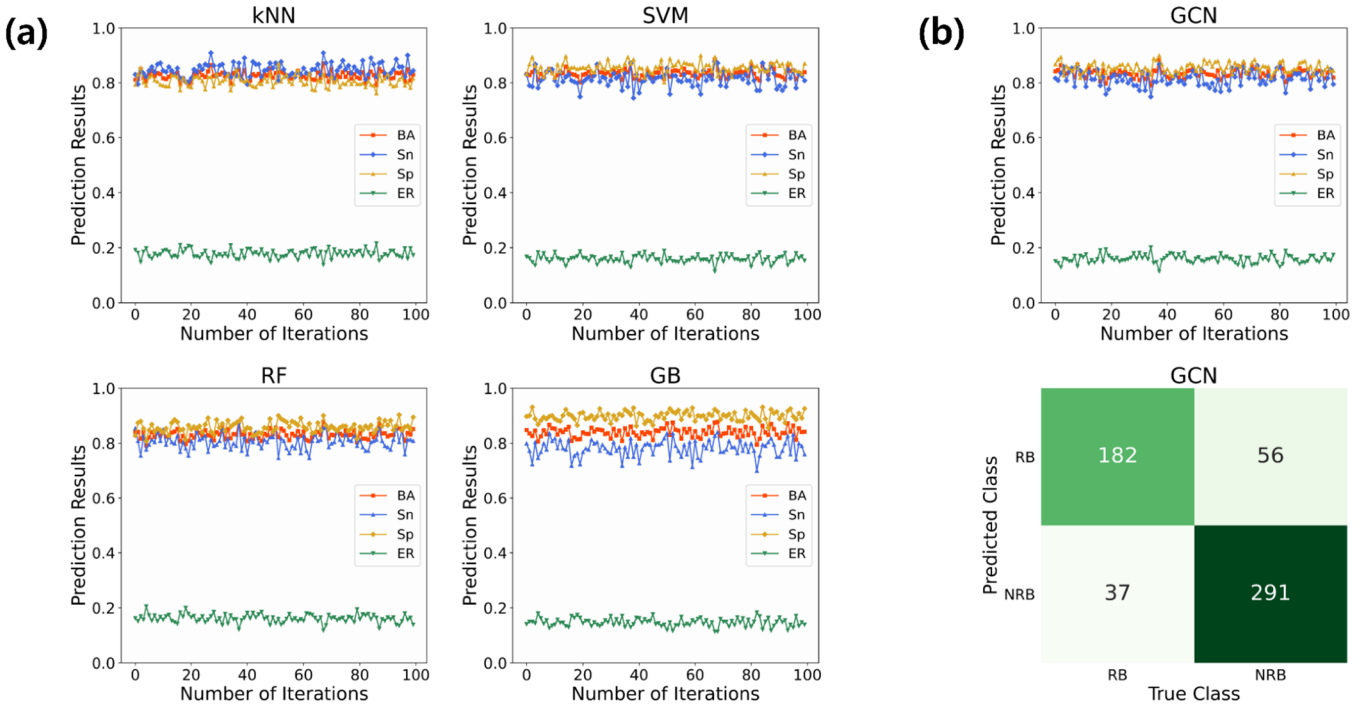


Figure 3. For the entire dataset, the composition of the training set and test set divided by 8 to 2 was randomly different, indicating a result of 100 times. (a) Results of the QSAR models and (b) result of the GCN model and a confusion matrix example. The confusion matrices of the remaining models can be seen in Figure S4.

Table 1. Summary of Results for Test Sizes of 0.2 and 0.8 for All Models[a]

	test size of 0.2				test size of 0.8			
model	BA	Sn	Sp	ER	BA	Sn	Sp	ER
kNN	0.83 (±0.04)	0.85 (±0.06)	0.81 (±0.05)	0.18 (±0.04)	0.79 (±0.02)	0.83 (±0.05)	0.75 (±0.04)	0.22 (±0.02)
SVM	0.84 (±0.04)	0.82 (±0.07)	0.86 (±0.05)	0.16 (±0.04)	0.81 (±0.02)	0.78 (±0.05)	0.84 (±0.04)	0.18 (±0.02)
RF	0.83 (±0.04)	0.81 (±0.07)	0.86 (±0.05)	0.16 (±0.04)	0.81 (±0.02)	0.76 (±0.06)	0.85 (±0.04)	0.18 (±0.02)
GB	0.84 (±0.04)	0.78 (±0.07)	0.90 (±0.04)	0.15 (±0.03)	0.81 (±0.02)	0.74 (±0.05)	0.88 (±0.03)	0.17 (±0.02)
GCN	0.84 (±0.05)	0.82 (±0.06)	0.86 (±0.04)	0.16 (±0.04)	0.81 (±0.02)	0.80 (±0.04)	0.81 (±0.04)	0.19 (±0.02)

[a]Details of each result are shown in Figures 3 and 4.

virtual screening[39,40] and target prediction benchmarks.[41,42] Although the results indicate that there is no significant difference, the MACCS fingerprint among the three types of molecular fingerprints is slightly better in the four QSAR models; hence, it was adopted. In addition, some prior studies[3,8,9] that classified biodegradability using QSAR models implemented genetic algorithms (GAs) for feature selection of descriptors obtained from these fingerprints. Similarly, in this study, a GA was used to find the optimal subset of features by reducing unnecessary or insignificant features, thereby potentially increasing the training speed and prediction accuracy.

However, using sklearn-genetic[43] to select the most relevant of the 1613 two-dimensional descriptors did not significantly improve the model performance, as shown in Table S3. (The numbers of descriptors in each model were reduced as follows: kNN, SVM, RF, and GB were implemented with 642, 50, 832, and 470 descriptors, respectively.) This process was conducted by obtaining BA scores from 10-fold CV for all models and descriptors (MACCS and Mordred) with and without GA-based feature selection (Tables S2 and S3). As a result, although using any of the three types of molecular fingerprints did not make a significant performance improvement, the case of using only MACCS fingerprints showed slightly better performance in common (BA > 0.82), so only MACCS fingerprints were adopted. Also, as a result of comparing the performance of applying GA to the descriptors calculated by Mordred, kNN and RF use the selected descriptors, and SVM and GB use all the calculated descriptors. The result of GA also did not make a significant performance change, but it is important in that it improved kNN and RF by evaluating and analyzing a small number of key descriptors. In addition, as shown in Figure S1, the Pearson correlation coefficient between all the descriptors and the class (RB/NRB) was not exceptionally high or meaningful. For 53 relatively more correlated descriptors (correlation > 0.35) and MACCS fingerprints, the BA value in 10-fold CV measured using SVM was 0.83, which was lower than that using the full descriptors (0.84; Table S3). Therefore, along with the MACCS fingerprint, kNN and RF for descriptors were selected by GA, and SVM and GB for all descriptors were adopted as final QSAR models in this study. An unsupervised learning approach using dimensionality reduction was attempted for further analysis of the descriptors and fingerprints, as shown in Figure S2. Using principal component analysis (PCA), t-distributed stochastic neighbor embedding

https://doi.org/10.1021/acsomega.1c06274
ACS Omega 2022, 7, 3649−3655

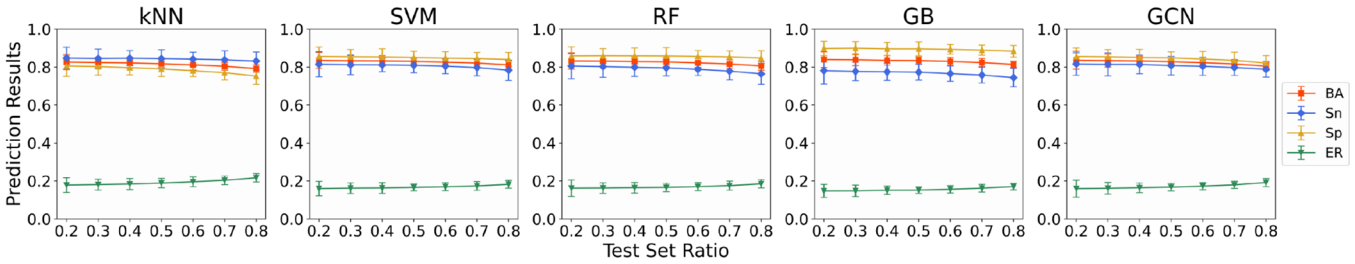


Figure 4. Results of measuring the configuration of the training set and test set for the entire dataset by dividing it from 8 to 2 to 2 to 8. The results showed the average of randomly performing each configuration 100 times differently.

(t-SNE),[44] and uniform manifold approximation and projection (UMAP)[45] algorithms, 1780 dimensions (1615-dimensional descriptors and 167-bit fingerprints) were reduced to two dimensions and visualized. In addition, the k-means clustering algorithm was applied to obtain BA values of 0.56, 0.60, and 0.58 for PCA, t-SNE, and UMAP, respectively. In particular, t-SNE exhibited a relatively visually distinguishable positional relationship between the two classes. Although these results are not sufficiently accurate to fully replace the supervised learning approach, the semi-supervised approach could still be applied to label molecules with unknown target properties.

Performance of the QSAR Models. The performances of the QSAR and GCN methods were compared using the metrics proposed above. As shown in Figure 3 and Table 1, for the three QSAR models except kNN, SVM, RF, and GB exhibit the same trend in the results: Sp is larger than Sn, and the error range of Sn (fluctuation magnitude) is relatively larger. This means that RB is often not correctly predicted, and its prediction capability could be unstable compared to NRB classification. A previous study also found that Sp, which is influenced by this prediction imbalance, is larger than Sn.[3] In the case of kNN, on the contrary, Sn was higher than Sp, but the overall performance was not good, especially when the test size was 0.8. The primary cause for such behavior could be due to the imbalanced dataset, and this is why this study considered the weight balancing technique for SVM and RF (the number of RB entries is relatively smaller than that of NRB, as discussed in Figure 2a).

As shown in Table 1, both SVM and GB show the highest performance among the QSAR models with a BA value of 0.84 (±0.04). Among the two models, ER of GB was lower at 0.15 (±0.03), but when Sn and Sp were compared, SVM values were closer to 0.82 (±0.07) and 0.86 (±0.05), respectively. These results were the same even when the test size was 0.8. To further verify the performance of the models, the variation in the prediction accuracy with the ratio of the test set to the training set was obtained, as shown in Figure 4. The results show that, as the proportion of the test set approaches 80%, the performance of the surrogate models deteriorates (e.g., kNN was the worst with a BA of 0.79). Furthermore, Sp was generally larger than Sn, and their difference increased as the size of the test set increased. This indicates that the intrinsic data imbalance could cause considerably poorer prediction accuracy when the developed model was employed for unexplored molecules. To further verify potential improvement in the prediction accuracy, we adopted the ensemble analysis model SVE. Ensemble analysis has been used to overcome the limitations of individual classifiers.[46] This method aims to increase the accuracy by combining the prediction results of various models to generate the final prediction results. Compared with the performance of individual classifiers, ensemble classifiers have demonstrated better stability and robustness.[47] However, employing the SVE did not improve the model performance compared to that of SVM, even though all four models were used, as shown in Figure S3. To provide further details, the confusion matrices for all the considered models are shown in Figure S4.

To summarize, as with many machine learning algorithms, the development of the QSAR model requires substantial consideration when choosing proper algorithms and feature selection methods for various descriptors and fingerprints. Hence, different types of accessible fingerprints and descriptors have been suggested. However, this increases the complexity of choosing a suitable model because it requires additional tests. Moreover, in addition to the results of the molecular descriptor generator, the physicochemical properties that directly represent the molecules can improve the performance of the model, but their actual implementation requires considerable resources owing to the limited availability and validity of the experimental data.[48]

GCN Model Details. Similar to the QSAR model, the optimal parameters for the GCN model must be found to obtain the best performance. In particular, various graph convolution algorithms are available, and the prediction accuracy depends on the model chosen. GCNConv, one of the most representative algorithms, was employed in this study, and its performance was compared with those of SAGEConv[49] and ARMAConv,[50] another representative algorithm. GCNConv is widely used and has been used to predict various molecular properties, such as drug−target affinity,[51] free energy, solubility, and metabolic stability.[52] As can be seen in Figure S5, it is difficult to judge that the performance of GCNConv is particularly superior to those of the other two algorithms, but in this study, the architecture consisting of four graph convolution layers of GCNConv and three prediction layers was adopted as it was judged appropriate in terms of computational cost and performance.[16] When the architecture was deeper with eight and seven layers, the error ranges in Sp and Sn are slightly wider (±0.07 and ±0.05, respectively; test set size was 20%). However, when four graph convolution layers and three prediction layers were used, Sp and Sn values were narrower (±0.06 and ±0.04, respectively). This performance suggests that the implemented GCN model is similarly stable enough for different algorithm and architecture choices to predict the biodegradability of molecules.

Performance of the GCN Model. The GCN model did not show dramatically improved performance compared to the QSAR models. The BA value was the same as those of SVM and GB, and the difference in the ER value was not significant.

https://doi.org/10.1021/acsomega.1c06274
ACS Omega 2022, 7, 3649−3655

However, an essential difference in the GCN model is the distance between Sp and Sn and convenience of the development process. In particular, there was only a difference in degree in the QSAR models: Sp was higher than Sn in all models, which caused instability in the prediction performance. However, especially when the test size is 0.8, the GCN model had similar BA, Sp, and Sn values, indicating that the prediction was more stable for both classes than that of the QSAR models (Table 1). Figure 4 and Table 1 confirm that this meaningful result of the GCN model was maintained even when the ratio of the test set was increased up to 80%, in contrast to SVM. This implies that the performance of the GCN model is similar to that of the QSAR models for the overall prediction results, but when the test size is 0.8, its prediction capability for each class is more stable and better than that of the four QSAR algorithms. The confusion matrices are shown in Figure S4 for further comparison.

As mentioned earlier, the GCN model can be developed by defining molecular graphs, applying various convolution algorithms, restructuring the model architecture, and adjusting the hyperparameters. However, unlike the QSAR models, the GCN model predicts the target properties using only the features of the atoms in molecular graphs; therefore, it does not require descriptors and fingerprints that are obtained through calculations. Moreover, this means that complicated feature selection processes are not required. Consequently, while developing the GCN model requires considerable effort and various tests to understand and optimize the architecture (similar to the QSAR models), because it only uses molecular graphs, it is easy to expand the dataset. Furthermore, it is simpler and more convenient because it does not perform feature importance analysis; as a result, it exhibits stable performance.

CONCLUSIONS

Because the biodegradability of materials is a molecular property directly related to environmental issues, efforts have been made to predict it. Numerous studies have attempted to solve this problem using QSAR machine learning algorithms, which have shown sufficient performance. However, models with many descriptors are less likely to generalize and have complex interpretations. Furthermore, the conditions that must be considered for the development of QSAR models and their performance have highlighted issues that require further improvement.

As a solution, this study presented a GCN model that is widely applied to molecules with high performance. For comparison, four QSAR models, an ensemble model, and the GCN model were constructed for datasets organized in a previous study. In addition, molecular descriptors and fingerprints were calculated and selected to develop the QSAR models, and convolutional algorithms and model architectures were selected to develop the GCN model. The results confirmed that the GCN model does not require descriptors or fingerprints and that, unlike the QSAR models, the Sp and Sn of the results of the GCN model are similar to each other, demonstrating stable performance for binary classification. Furthermore, these advantages of the GCN model suggest its potential to be an easy and robust solution not only for biodegradability but also for various QSAR prediction challenges in molecules.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c06274.

QSAR model hyperparameters, summary of previous studies, fingerprint-dependent prediction accuracy change, Pearson correlation coefficients for descriptors, and results from different GCN model architectures (PDF)

AUTHOR INFORMATION

Corresponding Author

Kyoungmin Min − *School of Mechanical Engineering, Soongsil University, Seoul 06978, Republic of Korea;* orcid.org/0000-0002-1041-6005; Email: kmin.min@ssu.ac.kr

Author

Myeonghun Lee − *School of Systems Biomedical Science, Soongsil University, Seoul 06978, Republic of Korea*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c06274

Notes

The authors declare no competing financial interest.

The dataset and the main source code used in this study are available at the GitHub page https://github.com/mhlee216/Biodegradability_Prediction_QSAR_GCN. All other relevant data are available from the authors upon reasonable request.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (no. 2020R1F1A1066519).

REFERENCES

(1) Zhan, Z.; Li, L.; Tian, S.; Zhen, X.; Li, Y. Prediction of Chemical Biodegradability Using Computational Methods. *Mol. Simul.* **2017**, *43*, 1277−1290.

(2) Shah, A. A.; Hasan, F.; Hameed, A.; Ahmed, S. Biological Degradation of Plastics: A Comprehensive Review. *Biotechnol. Adv.* **2008**, 246−265.

(3) Mansouri, K.; Ringsted, T.; Ballabio, D.; Todeschini, R.; Consonni, V. Quantitative Structure-Activity Relationship Models for Ready Biodegradability of Chemicals. *J. Chem. Inf. Model.* **2013**, *53*, 867−878.

(4) REGULATION (EC) No 1907 2006 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL.

(5) Allanou, R.; Hansen, B. G.; van der Bilt, Y.; Hansen, B. G. *Public Availability of Data on EU High Production Volume Chemicals.*

(6) Rocha, W. F. C.; Sheen, D. A. Classification of Biodegradable Materials Using QSAR Modelling with Uncertainty Estimation. *SAR QSAR Environ. Res.* **2016**, *27*, 799−811.

(7) Ballabio, D.; Biganzoli, F.; Todeschini, R.; Consonni, V. Qualitative Consensus of QSAR Ready Biodegradability Predictions. *Toxicol. Environ. Chem.* **2017**, *99*, 1193−1216.

(8) Cheng, F.; Ikenaga, Y.; Zhou, Y.; Yu, Y.; Li, W.; Shen, J.; Du, Z.; Chen, L.; Xu, C.; Liu, G.; Lee, P. W.; Tang, Y. In Silico Assessment of Chemical Biodegradability. *J. Chem. Inf. Model.* **2012**, *52*, 655−669.

(9) Lunghini, F.; Marcou, G.; Gantzer, P.; Azam, P.; Horvath, D.; van Miert, E.; Varnek, A. Modelling of Ready Biodegradability Based on Combined Public and Industrial Data Sources. *SAR QSAR Environ. Res.* **2020**, *31*, 171−186.

(10) Yap, C. W. PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466−1474.
(11) Consonni, V.; Pavan, M.; Todeschini, R.; Mauri, A. *DRAGON Software: An Easy Approach to Molecular Descriptor Calculations*.
(12) Ceriani, L.; Papa, E.; Kovarich, S.; Boethling, R.; Gramatica, P. Modeling Ready Biodegradability of Fragrance Materials. *Environ. Toxicol. Chem.* **2015**, *34*, 1224−1231.
(13) Moriwaki, H.; Tian, Y. S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminf.* **2018**, *10*, 1−14.
(14) Advantages and Disadvantages of QSAR - Quantitative Structure Activity Relationship. Medicinal Chemistry http://www.pharmacy180.com/article/advantages-and-disadvantages-of-qsar-2089/ (accessed 2021−10 -14).
(15) *Designing Safer Chemicals, Copyright, Advisory Board, Foreword*; 1996; pp. 1−4, DOI: 10.1021/bk-1996-0640.fw001.
(16) Kipf, T. N.; Welling, M. *Semi-Supervised Classification with Graph Convolutional Networks*. 2016.
(17) Sun, M.; Zhao, S.; Gilvary, C.; Elemento, O.; Zhou, J.; Wang, F. Graph Convolutional Networks for Computational Drug Development and Discovery. *Briefings Bioinf.* **2020**, *21*, 919−935.
(18) Hung, C.; Gini, G. QSAR Modeling without Descriptors Using Graph Convolutional Neural Networks: The Case of Mutagenicity Prediction. *Mol. Diversity* **2021**, *25*, 1283−1299.
(19) David, L.; Thakkar, A.; Mercado, R.; Engkvist, O. Molecular Representations in AI-Driven Drug Discovery: A Review and Practical Guide. *J. Cheminf.* **2020**, *12*, 1−22.
(20) eChemPortal provides free public access to information on properties of chemicals: https://www.echemportal.org/echemportal/ (accessed 2021−10 -14).
(21) National Institute of Technology and Evaluation (NITE) https://www.nite.go.jp/en/ (accessed 2021−10 -14).
(22) Manganaro, A.; Chemoinformatics, K.; Carla, G.; Politecnico, G.; Milano, D.; Benfenati, E.; Gini, G. *VEGA-QSAR: AI inside a Platform for Predictive Toxicology*.
(23) EPI Suite-Estimation Program Interface | US EPA https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface (accessed 2021−10 -14).
(24) Mansouri, K.; Grulke, C. M.; Judson, R. S.; Williams, A. J. OPERA Models for Predicting Physicochemical Properties and Environmental Fate Endpoints. *J. Cheminf.* **2018**, *10*, 1−19.
(25) NCI/CADD Group Chemoinformatics Tools and User Services https://cactus.nci.nih.gov/ (accessed 2021−10 -14).
(26) PubChem https://pubchem.ncbi.nlm.nih.gov/ (accessed 2021-10-14).
(27) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. *KNIME-The Konstanz Information Miner Version 2.0 and Beyond*.
(28) O'Boyle, N. M.; Hutchison, G. R. Cinfony - Combining Open Source Cheminformatics Toolkits behind a Common Interface. *Chem. Cent. J.* **2008**, *2*, 1−10.
(29) Cao, D. S.; Xu, Q. S.; Hu, Q. N.; Liang, Y. Z. ChemoPy: Freely Available Python Package for Computational Biology and Chemoinformatics. *Bioinformatics* **2013**, *29*, 1092−1094.
(30) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107−113.
(31) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273−1280.
(32) RDKit http://www.rdkit.org/ (accessed 2021−10 -14).
(33) Fey, M.; Lenssen, J. E. *Fast Graph Representation Learning with PyTorch Geometric*. 2019.
(34) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(35) Baurin, N.; Mozziconacci, J. C.; Arnoult, E.; Chavatte, P.; Marot, C.; Morin-Allory, L. 2D QSAR Consensus Prediction for High-Throughput Virtual Screening. An Application to COX-2 Inhibition Modeling and Screening of the NCI Database. *J. Chem. Inf. Comput. Sci.* **2003**, *44*, 276−285.
(36) Votano, J. R.; Parham, M.; Hall, L. H.; Kier, L. B.; Oloff, S.; Tropsha, A.; Xie, Q.; Tong, W. Three New Consensus QSAR Models for the Prediction of Ames Genotoxicity. *Mutagenesis* **2004**, *19*, 365−377.
(37) Kipf, T. N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings* 2016.
(38) Parikh, R.; Mathai, A.; Parikh, S.; Chandra Sekhar, G.; Thomas, R. *Understanding and Using Sensitivity, Specificity and Predictive Values*.
(39) Riniker, S.; Landrum, G. A. Open-Source Platform to Benchmark Fingerprints for Ligand-Based Virtual Screening. *J. Cheminf.* **2013**, *5*, 1−17.
(40) Mellor, C. L.; Marchese Robinson, R. L.; Benigni, R.; Ebbrell, D.; Enoch, S. J.; Firman, J. W.; Madden, J. C.; Pawar, G.; Yang, C.; Cronin, M. T. D. Molecular Fingerprint-Derived Similarity Measures for Toxicological Read-across: Recommendations for Optimal Use. *Regul. Toxicol. Pharmacol.* **2019**, *101*, 121−134.
(41) Awale, M.; Reymond, J. L. Web-Based Tools for Polypharmacology Prediction. In *Methods in Molecular Biology*; Humana Press Inc., 2019; Vol. *1888*, pp. 255−272, DOI: 10.1007/978-1-4939-8891-4_15.
(42) Awale, M.; Reymond, J. L. Polypharmacology Browser PPB2: Target Prediction Combining Nearest Neighbors with Machine Learning. *J. Chem. Inf. Model.* **2019**, *59*, 10−17.
(43) Calzolari, M. *Manuel-Calzolari/Sklearn-Genetic: Sklearn-Genetic 0.2*. 2019, DOI: 10.5281/ZENODO.3348077.
(44) van der Maaten, L.; Hinton, G. *Visualizing Data Using T-SNE*; 2008; Vol. *9*.
(45) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2018.
(46) Saha, S.; Ekbal, A. Combining Multiple Classifiers Using Vote Based Classifier Ensemble Technique for Named Entity Recognition. *Data Knowl. Eng.* **2013**, *85*, 15−39.
(47) Tan, P.-N.; Steinbach, M.; Kumar, V. Introduction to Data Mining Instructor's Solution Manual.
(48) Papa, E.; Dearden, J. C.; Gramatica, P. Linear QSAR Regression Models for the Prediction of Bioconcentration Factors by Physicochemical Properties and Structural Theoretical Molecular Descriptors. *Chemosphere* **2007**, *67*, 351−358.
(49) Hamilton, W. L.; Ying, R.; Leskovec, J. Inductive Representation Learning on Large Graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, 1025−1035.
(50) Bianchi, F. M.; Grattarola, D.; Livi, L.; Alippi, C. Graph Neural Networks with Convolutional ARMA Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2021**, DOI: 10.1109/TPAMI.2021.3054830.
(51) Jiang, M.; Li, Z.; Zhang, S.; Wang, S.; Wang, X.; Yuan, Q.; Wei, Z. Drug-Target Affinity Prediction Using Graph Neural Network and Contact Maps. *RSC Adv.* **2020**, *10*, 20701−20712.
(52) Pocha, A.; Danel, T.; Maziarka, Ł. *Comparison of Atom Representations in Graph Neural Networks for Molecular Property Prediction*. 2020.

https://doi.org/10.1021/acsomega.1c06274
ACS Omega 2022, 7, 3649−3655