Computational Biology and Chemistry 87 (2020) 107286

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac

Research Article

hERG-Att: Self-attention-based deep neural network for predicting hERG blockers

Hyunho Kim, Hojung Nam*

School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Buk-gu, Gwangju 61005, Republic of Korea

ARTICLE INFO

Keywords:
hERG blockers prediction
Deep learning
Self-attention mechanism

ABSTRACT

A voltage-gated potassium channel encoded by the human ether-à-go-go-related gene (hERG) regulates cardiac action potential, and it is involved in cardiotoxicity with compounds that inhibit its activity. Therefore, the screening of hERG channel blockers is a mandatory step in the drug discovery process. The screening of hERG blockers by using conventional methods is inefficient in terms of cost and efforts. This has led to the development of many *in silico* hERG blocker prediction models. However, constructing a high-performance predictive model with interpretability on hERG blockage by certain compounds is a major obstacle. In this study, we developed the first, attention-based, interpretable model that predicts hERG blockers and captures important hERG-related compound substructures. To do that, we first collected various datasets, ranging from public databases to publicly available private datasets, to train and test the model. Then, we developed a precise and interpretable hERG blocker prediction model by using deep learning with a self-attention approach that has an appropriate molecular descriptor, Morgan fingerprint. The proposed prediction model was validated, and the validation result showed that the model was well-optimized and had high performance. The test set performance of the proposed model was significantly higher than that of previous fingerprint-based conventional machine learning models. In particular, the proposed model generally had high accuracy and F1 score thereby, representing the model's predictive reliability. Furthermore, we interpreted the calculated attention score vectors obtained from the proposed prediction model and demonstrated the important structural patterns that are represented in hERG blockers. In summary, we have proposed a powerful and interpretable hERG blocker prediction model that can reduce the overall cost of drug discovery by accurately screening for hERG blockers and suggesting hERG-related substructures.

1. Introduction

A voltage-gated potassium channel encoded by the human ether-à-go-go-related gene (hERG) plays a key role in cardiac polarization and depolarization (Sanguinetti and Tristani-Firouzi, 2006). If the hERG ion channel is blocked by a ligand, it may cause long QT syndrome (LQTS) and cause Torsade de Pointes (TdP), which leads to sudden cardiac death (De Ponti et al., 2001; Redfern et al., 2003). Therefore, many marketed drugs such as astemizole (Zhou et al., 1999), terfenadine (Roy et al., 1996), and cisapride (Rampe et al., 1997) have been withdrawn owing to the unintended hERG channel blocking property. Therefore, chemical screening of the hERG channel is an essential step in the drug discovery and development process. Conventionally, the patch-clamp electrophysiological assay is widely used for screening hERG blockers (Polonchuk, 2012). However, this type of *in vitro* method is cost-inefficient, labor-intensive, and time-consuming; thus, *in silico* prescreening methods have become essential to both industry and academia. Nowadays, because of the accumulation of data on hERG inhibition or binding affinity assay, data-driven hERG blocker prediction has attracted considerable attention. Because structurally diverse molecules bind or inhibit the hERG channel (Witchel, 2007), various molecular features and machine learning methods are used to construct the prediction models. In several studies, pharmacophore-based prediction models with support vector machine (SVM) and Naïve Bayes classifier were developed (Leong, 2007; Wang et al., 2016). These studies attempted to find the pharmacophore pattern of hERG blocking compounds. In addition, fingerprint-based prediction models have been developed (Doddareddy et al., 2010; Chavan et al., 2016; Siramshetty et al., 2018). They used conventional machine learning methods such as k-nearest neighbor (KNN), SVM, and random forest. In addition, physicochemical molecular descriptor-based models with various machine learning algorithms have been developed (Didziapetris and Lanevskij,

* Corresponding author.
E-mail addresses: hyunhokim@gm.gist.ac.kr (H. Kim), hjnam@gist.ac.kr (H. Nam).

https://doi.org/10.1016/j.compbiolchem.2020.107286
Received 13 February 2020; Accepted 9 May 2020
Available online 19 May 2020

2016; Konda et al., 2019). The aforementioned studies focused on the prediction performance of the models. Meanwhile, with the development of the deep learning technique, Cai et al. proposed the first deep learning-based hERG blocker prediction model (Cai et al., 2019). They employed a Mol2Vec (Jaeger et al., 2018) featurizer, inspired by word embedding, and they used molecular operating environment (MOE) descriptors. Additionally, they attempted to solve the ambiguity of IC_{50} thresholds by multi-task learning. Their model focused on performance, and not interpretability, because the deep learning model they used was a black-box model.

Thus, we note that constructing a precise prediction model as well as finding meaningful hERG blocker patterns simultaneously is a major issue. To overcome this limitation, in this study, we propose a precise model and potential hERG blocking substructures. This is achieved by preprocessing a large-scale public dataset into qualitative data and employing an interpretable deep learning technique called "self-attention" that was proposed by Vaswani et al. (2017).

First, in the data preparing step, we collected hERG-blockers and non-blockers datasets from various sources to train and test the generalization power of the model. After preprocessing the data, we generated the Morgan fingerprint descriptor of the compounds that can capture the substructures for all compounds. Next, by employing deep learning with a self-attention approach, we constructed an interpretable model to predict the blockers.

2. Methods

2.1. Data collection and preprocessing

2.1.1. Training dataset

To train the classification model, we collected training data from ChEMBL (version 25) and PubChem, which are two major publicly available databases (Mendez et al., 2018; Kim et al., 2018). For consistency, data including IC_{50} measurements were collected from both databases. To remove duplicates, we first standardized a simplified molecular-input line-entry system (SMILES) descriptor for compounds using MolVS (Swain, 2019), which is an open-source toolkit based on the RDKit chemistry framework (version 2019.03.1) (Landrum, 2019). For the duplicates, we filtered out the molecules that have high standard deviation ($\geq$ 10μM) and took a mean value of IC_{50} for the duplicated compounds. Because there is no fixed activity threshold, we followed the threshold criteria from a previous study (Siramshetty et al., 2018). In particular, a compound was assigned as a blocker (hERG-positive) if its IC_{50} value was less and equal to 10 μM and assigned as a non-blocker (hERG-negative) if its IC_{50} value was greater than 10 μM. The final training dataset contained 10,453 compounds. Approximately, 63% (6558) of the compounds were blockers, and 37% (3895) were non-blockers. The training set is summarized in Table 1.

2.1.2. External datasets

We used various external datasets for validation and testing of the model. First, the validation set for hyper-parameter tuning was collected from the literature (Didziapetris and Lanevskij, 2016). The same standardization and duplicate handling process used for training data preprocessing were followed. Then, we collected compounds that were not used in the training phase. We manually assigned the class labels using the IC_{50} value criteria used in the training set preprocess. For the test set used for the evaluation of the model performance, we collected external datasets used to evaluate three well-trained comparative models (Chavan et al., 2016; Siramshetty et al., 2018; Konda et al., 2019) developed in previous studies. The compounds collected in the external dataset were unknown compounds for both the proposed and other three models. Additionally, we collected two sets of independent test data from the literature for a comprehensive evaluation (Wang et al., 2016; Doddareddy et al., 2010). In this case, we manually assigned the class label with three different IC_{50} thresholds; 1 μM, 10 μM, 30 μM, to evaluate the model's generalizability. After applying the abovementioned standardization process, we selected the unknown compounds from the training set, validation set, and three training datasets used to train the comparative models (Chavan et al., 2016; Siramshetty et al., 2018; Konda et al., 2019). The detailed summary of the external datasets is described in Table 2.

Table 1
Detailed summary of dataset used for training after preprocessing.

Source	Blocker ($\leq$10 μM)	Non-Blocker($>$ 10 μM)	Both
CHEMBL	3655	1688	5343
PubChem	2903	2207	5110
Total	6558	3895	10,453

2.2. Molecular fingerprint generation

Because we aimed to identify the specific substructures of hERG blockers, we employed the Morgan fingerprint that is a type of a circular fingerprint, also known as extended connectivity fingerprints (ECFP) (Rogers and Hahn, 2010). This was because of two reasons: (i) it can capture the local structural features, (ii) the set bits are interpretable. To calculate the Morgan fingerprint, the radius size and number of bits are required as input parameters. In detail, the substructures of a molecule with a radius size below the predefined value are extracted and transformed into a numerical identifier by the slightly changed Morgan algorithm (Rogers and Hahn, 2010; Morgan, 1965). Then, all extracted substructure identifiers were hashed into a predefined bit-sized binary vector. Therefore, the set bits in a binary vector are considered as substructures of a molecule. We chose a radius of 3 and 2048 bits to be able to capture meaningful and distinguishable substructures and reduce collisions of the hashing algorithm. We calculated the Morgan fingerprints of all the compounds from standardized SMILES with the RDKit chemistry toolkit.

2.3. Model construction

2.3.1. Deep neural network with self-attention

In this study, we constructed an interpretable deep-learning model with a self-attention mechanism. The self-attention mechanism is a simple technique that makes the model interpretable by capturing data-specific feature importance and improves model training by employing self-captured features. Through backpropagation from the final loss, the model is learned to capture data-specific and task-related important features required for precise prediction. Fig. 1 shows the interpretable self-attention-based deep neural network constructed in this study. First, the calculated fingerprint was fed into the fully connected layer and softmax layer to train the self-attention score:

$$\mathbf{a} = softmax(g(\mathbf{fp})),$$

where $\mathbf{fp}$ is the fingerprint binary vector, $\mathbf{a}$ is the attention score vector, and $g(\bullet)$ is the fully connected layer without activation. In detail, $g(\bullet)$ is a linear operator and $softmax(\bullet)$ is a non-linear activation operator which can be described as below:

$$g(\mathbf{x}) = \mathbf{W}\mathbf{x} + \mathbf{b},$$

$$softmax(\mathbf{x})_i = \frac{e^{\mathbf{x}_i}}{\sum_i e^{\mathbf{x}_i}} \ (for\ i = 1, \ldots, k),$$

where $\mathbf{x}$ is the k-dimensional input vector, $\mathbf{W}$ is the learnable weight matrix, and $\mathbf{b}$ is the learnable bias vector. In this study, the same-sized linearly transformed fingerprint is generated by the $g(\bullet)$ operator that means $\mathbf{W} \in \mathbb{R}^{2048\times2048}$ and $\mathbf{b} \in \mathbb{R}^{2048}$. Then the output of the $g(\bullet)$ is fed into the $softmax(\bullet)$ function to calculate the self-attention score vector which is the same-sized normalized vector;i.e., their sum is always 1.

Table 2
Detailed summary of dataset used for validation and test after preprocessing. All observations from the validation and test sets that are duplicate in the training data have been removed.

Source	Use	Threshold	Blocker	Non-Blocker	Both
Didziapetris et al. (2016)	Validation	10 μM	785	622	1407
Doddareddy et al. (2010)	Independent test	1 μM, 10 μM, 30 μM[a]	46, 158, 190	191, 79, 47	237
Wang et al. (2016)	Independent test	1 μM, 10 μM, 30 μM[a]	44, 108, 137	131, 67, 38	175
Chavan et al. (2016)	External test	20 %[b]	220	1573	1793
Siramshetty et al. (2018)	External test	10 μM, 30 μM[c]	254	235	489
Konda et al. (2019)	External test	30 μM	339	90	429

[a] Applied three different thresholds; 1 μM, 10 μM, 30 μM.
[b] % hERG blockage.
[c] Multiple dataset with different thresholds; 10 μM 30 μM.

After obtaining the attention score that can be interpreted as an input feature importance, the original fingerprint and attention score are element-wise multiplied:

$$\mathbf{v} = \mathbf{a} \odot \mathbf{fp},$$

where $\odot$ is the element-wise multiplication operator. Then, the weighted feature vector $\mathbf{v}$ is fed into the three-layer fully-connected network with rectified linear unit (ReLu) activation for feature reduction and abstraction. Finally, by sigmoid activation, the model determines whether the compound is an hERG blocker or not:

$$\hat{y} = \sigma(\mathrm{MLP}(\mathbf{v})),$$

where $\hat{y}$ is the predicted score in the range 0–1, σ is the sigmoid activation, and MLP(•) is the three-layer multi-layer perceptron (MLP) with ReLu non-linear activation. Because it is a binary classification model, the binary cross-entropy loss was employed and the Adam optimizer was used for gradient descent.

2.4. Evaluation metrics

For the performance evaluation, we used conventional metrics, such as sensitivity (Sen), specificity (Spe), precision (Pre), accuracy (Acc) and F1 score (F1), that are used in typical classification problems. The detailed formulas are shown below:

$$\mathrm{Sen} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}},$$

$$\mathrm{Spe} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}},$$

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}},$$

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

$$\mathrm{F1} = 2 \times \frac{\mathrm{Pre} \times \mathrm{Sen}}{\mathrm{Pre} + \mathrm{Sen}},$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. In this study, we focused more on accuracy and F1 score, because both metrics represent the predictive reliability of the model.

3. Result

3.1. Model optimization

With the collected validation dataset which is composed of compounds that are not seen during the training step, we optimized hyper-parameters such as the number of layers, number of nodes per layer, learning rate, and epochs by checking the area under the precision-recall curve (AUPR) value at each search grid. The model optimization process could affect the model's negative prediction power. However, this is not significant because, as mentioned above, precise hERG positive prediction is more important in the screening phase. After tuning the hyper-parameters, we observed the training loss curve and validation loss curve to check whether the model is under-fitted or over-fitted. We confirmed that the optimized model shows consistently well-decaying loss curves and validation loss curves (Fig. 2 **(a)**) during replication. In addition, we evaluated the AUPR and the area under the receiver operating characteristics (AUROC) of the trained model. Fig. 2 **(b)** and **(c)** show the curves and area under the curve values. Our model achieved an AUPR of 0.91 (± 0.006) and an AUROC score of 0.89 (± 0.006) in the validation step. Furthermore, we also compared the validation performances between the models with and without self-attention to prove the advantages of using the self-attention. As a result, the model with the self-attention showed better performance. Therefore we proved that the self-attention mechanism does not only makes the model interpretable but also makes the model well-trained. The detail information is in Supplementary Table 2.

3.2. Performance comparison with each comparative model

We compared the performance of our model with that of three previously published machine learning-based models. First, the model of Chavan et al. (2016) is an ensemble of KNN models trained with three types of molecular fingerprints: Extended fingerprint, Pubchem, and Substructure count. In total, 171 compounds were selected for training, of which 94 compounds are hERG positive and 77 compounds are hERG negative. Their IC_{50} cutoff was 5 μM. Moreover, the authors tested their model on the external test set with a threshold of 20% hERG blockage ratio. This test set contains 220 hERG positive and 1573 hERG negative compounds. Second, the model of Siramshetty et al. (2018) is a random forest model trained with ECFP4. Their model was trained with the manually processed ChEMBL dataset (version 23) containing 1406 hERG positive compounds and 1817 hERG negative compounds with specific IC_{50} cutoff ($\leq 1\,\mu M$, $\geq 10\,\mu M$). They evaluated their model on their manually collected external test sets with various IC_{50} thresholds ($10\,\mu M$, $30\,\mu M$). This test set contains 254 hERG positive and 235 hERG negative compounds. Finally, the model of Konda et al. (2019) is a random forest model trained with selected molecular descriptors calculated using the PaDEL-descriptor software. This model was also trained with the manually processed ChEMBL dataset (version 22) containing 7254 hERG positive and 1451 hERG negative compounds with 30 μM IC_{50} cutoff. They also validated their model with an external test set from various sources with an IC_{50} threshold of 30 μM. This test set contains 339 hERG positive and 90 hERG negative compounds. To perform a fair comparison, we trained our model and evaluated the performances with the identical set of compounds and cutoffs used in each study for training and test, then we compared the performance results with each model. Furthermore, we repeated the experiment 30 times to confirm that the result was not obtained due to random chance. According to the results of the performance comparisons shown in Fig. 3, our model outperforms the fingerprint-based models in terms of

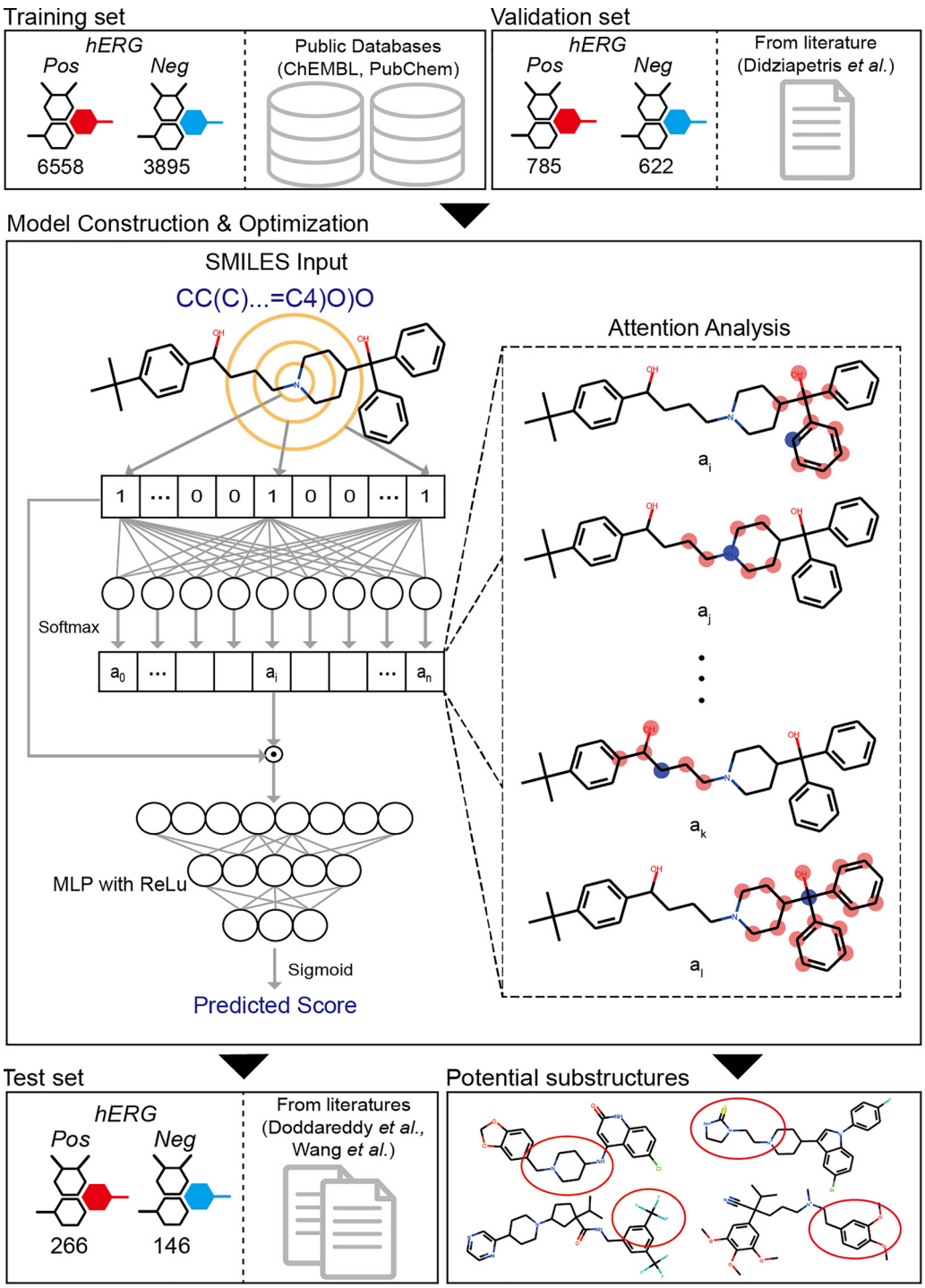

Fig. 1. Model Overview.
Overview of the present study. The proposed model was constructed and optimized with the training and validation dataset. SMILES was used as input to calculate the fingerprint. The attention score vector was trained by feeding the fingerprint vector into the linear layer and Softmax layer. Using weighted fingerprint by attention score, the blocker score is predicted by a three-layer MLP with ReLu non-linear activation. After construction, the model was evaluated with the test set. The potential substructures were found by the attention score.

accuracy and F1 score (Fig. 3 **(a)** and **(b)**), but not in the descriptor-based model (Fig. 3 **(c)**). Even though our proposed model does not achieve higher prediction performance than the descriptor-based model, this can be compensated with the interpretability power of our model which can demonstrate the essential structural patterns that are represented in hERG blockers.

3.3. Comprehensive performance comparison with an independent test set

Next, we tested the performance of our model by comparison with the three abovementioned models using the aforementioned independent test set. All compounds in this test set were unknown to all models. In this comparison, all models were trained with their own training set, therefore, we tested the performances on this test set with three different thresholds for a fair comparison; 1 μM, 10 μM, 30 μM.

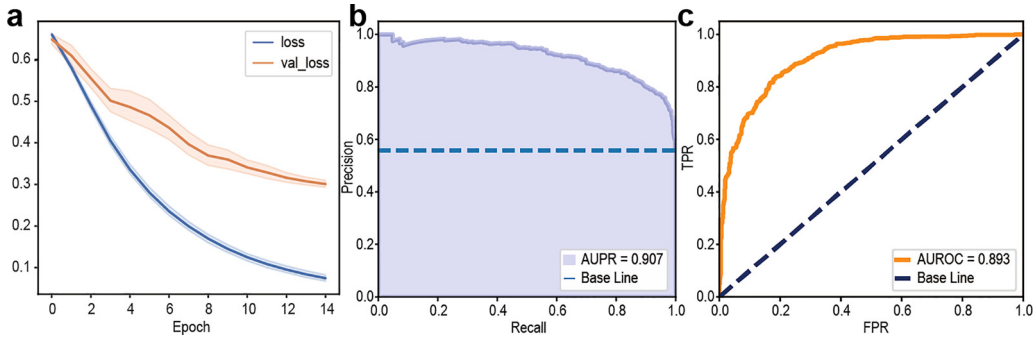

Fig. 2. Loss curve and Validation Result.
(a) Training loss and validation loss curve of the optimized model with 95% confidence interval. (b) Precision-recall curve with area under the curve value, AUPR = 0.907 (± 0.006) (c) Receiver operating characteristic curve with area under the curve value, AUROC = 0.893 (± 0.006).

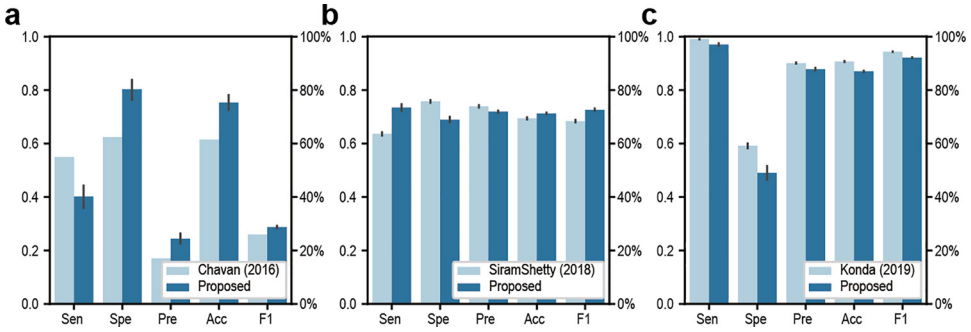

Fig. 3. Performance comparison with other models.
Performance comparison of our model with other models. Each comparison was conducted with the test set that each study used. (a) The performance comparison on the test dataset with a threshold of 20% hERG blockage ratio. (b) The performance comparison on the test datasets with a threshold of 10 and 30 μM. (c) The performance comparison on the test dataset with a threshold of 30 μM.

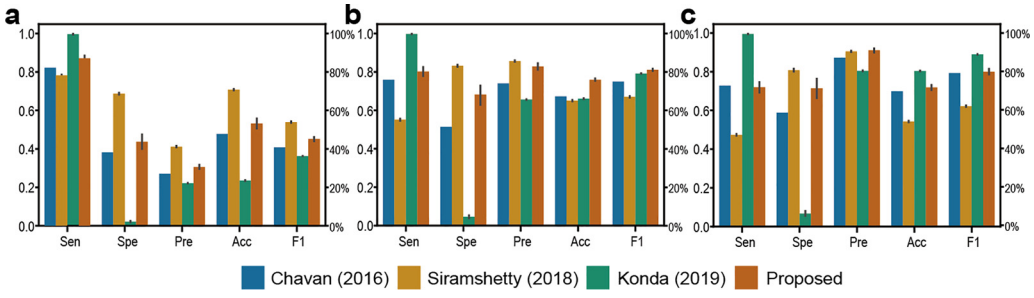

Fig. 4. Comprehensive performance comparison with an independent test set.
Comprehensive performance comparison of the models. The independent test set was manually collected and preprocessed and contained drugs unknown to all models. (a) Performance evaluation on the test set with 1 μM threshold. (b) Performance evaluation on the same test set with 10 μM threshold. (c) Performance evaluation on the same test set with 30 μM threshold.

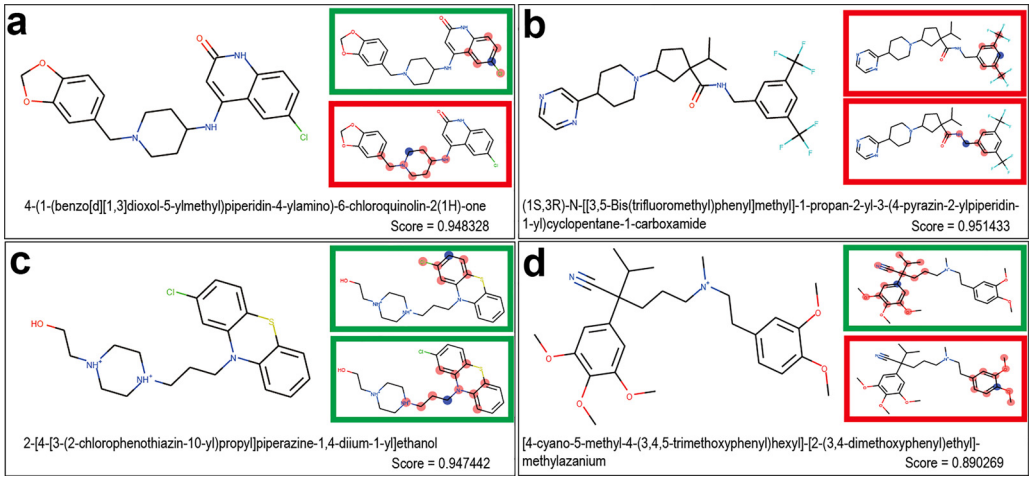

Fig. 5. True positive hERG blockers and their highlighted substructures by interpreting the attention score.
Some predicted blockers among positive samples (true positive) and atoms that contribute to two highlighted set bits among the top fifthteen (red and blue dots). The blue dot indicates the center atom. The green box denotes already reported substructure that affects hERG binding affinity. The red box denotes unreported and novel substructures that the model focused to predict it as a blocker. (a), (d) models focused on both known and unknown structures. (b) model focused on unknown substructures. (c) model focused on known substructures.

Furthermore, we repeated the experiment 30 times for statistical significance. The results are shown in Figs. 4 **(a)-(c)**, respectively. Our model generally showed good performance; namely, it showed the highest accuracy and F1 score in the independent test set with the 10 μM threshold and showed the second-best precision and F1 score in the test set with two different thresholds, 1 μM and 30 μM. We can

conclude that our constructed model is well-generalized because it showed the robust performance in the test sets with thresholds(1 μM, 30 μM) which are different from the training set threshold (10 μM).

3.4. Result interpretation and potential hERG blocker specific substructures

Several substructures that affect the hERG binding affinity have been reported thus far (Kalyaanamoorthy and Barakat, 2018). By substituting those substructures, the hERG binding affinity can be decreased significantly. In this section, we analyzed the hERG blocker patterns by interpreting the attention scores obtained from our model. Among the compounds assigned to the blocker in the independent test set, 211 compounds were predicted as blockers by our prediction model among 266 positive compounds ($IC_{50} \leq 10\mu M$). Our model is a self-attention based model; therefore, we can interpret the attention score as feature importance. Thus, we can easily find bits that are important for the prediction of the hERG blocking property and atoms that contribute to those bits. We observed the top fifteen highlighted set bits considering the attention score vector for each well-predicted blocker and analyzed the atoms that contribute to setting those bits. Fig. 5 shows some true positive blockers and their two highlighted substructures. We found that our model can capture the well-known hERG blocking substructures (green boxes in Fig. 5) as well as unreported substructures (red boxes in Fig. 5 **(a), (b)**, and **(d)**). Because of this ability of our model, the captured unreported substructures are potential hERG blocking substructures that could directly or indirectly affect the hERG binding affinity property.

4. Discussion

By employing a precise *in silico* method in drug discovery and development process, the cost and time required for the process can be reduced considerably. Therefore, many quantitative structure-activity relationship (QSAR) models have been developed for various purposes in the drug discovery process. However, no standard model has been proposed thus far because of limitations such as low quality or quantity of data, biased feature engineering method, and inappropriate machine learning algorithms. Although our model has many advantages such as interpretability and high performance, it can be improved further. First, the internal problem of the Morgan fingerprint algorithm can influence the prediction performance of the model. The Morgan fingerprint algorithm considers the radius size as a parameter; therefore, we can only capture a maximum of three radius-sized substructures, even if there are larger substructures (Kalyaanamoorthy and Barakat, 2018). Moreover, it is a simple hashing-based algorithm; thus, the intrinsic collision problem and feature sparsity issue have a tradeoff relationship. If the number of bits decreases, a larger number of substructures will collide directly increasing feature ambiguity. In the present study, we used 2048 bits to avoid collisions. However, feature sparsity that can affect the complexity of the models could be increased. Second, the quality and quantity of data used to train the model are critical issues for deep learning studies. Thus far, the datasets related to hERG blockers and non-blockers do not fully satisfy these aspects. Although it seems that there is sufficient data in public databases and publicly available private datasets, there are many duplicates with different assay measurements and missing values. In this study, we attempted to collect data by choosing IC_{50}-measured compounds and removed ambiguous compounds according to the IC_{50} standard deviation. However, we only satisfied the data quality and not quantity for deep learning. Therefore, although the deep learning technique is a state-of-the-art machine learning model, a less-biased reversible molecular featurizer and both qualitative and quantitative data could considerably improve hERG blocking property prediction.

5. Conclusion

In this study, we developed the first deep learning-based interpretable model for hERG blocker prediction. First, we collected datasets from both public databases and the literature. To construct a highly accurate model, we chose only IC_{50}-measured data, handled duplicates by standardizing SMILES, and removing ambiguous compounds according to the IC_{50} standard deviation. By deep learning with the self-attention mechanism, we designed and constructed a prediction model that shows significantly increased prediction performance compared to other related studies. For a fair comparison, we compared our model with the comparative models with various external datasets and various IC_{50} cutoff in terms of performance. Moreover, we analyzed the results by interpreting the attention score vectors and confirmed that the prediction model focuses on previously known hERG-related substructures and also unreported potential structural patterns. We can conclude that the proposed model may contribute to the drug discovery process with regard to identifying hERG blockers cost-efficiently and capturing hERG-related substructures.

Authors' contributions

HK and HN designed the study. HK implemented the study and drafted the manuscript. HN revised the manuscript. All authors have approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent of publication

Not applicable.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2020R1A2C2004628), the Bio-Synergy Research Project (NRF-2017M3A9C4092978) of the Ministry of Science and ICT through the National Research Foundation.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgments

None.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.compbiolchem.2020.107286.

References

Cai, C., et al., 2019. Deep learning-based prediction of drug-induced cardiotoxicity. J. Chem. Inf. Model. 59 (3), 1073–1084.
Chavan, S., et al., 2016. A k-nearest neighbor classification of hERG K(+) channel blockers. J. Comput. Aided Mol. Des. 30 (3), 229–236.
De Ponti, F., Poluzzi, E., Montanaro, N., 2001. Organising evidence on QT prolongation and occurrence of Torsades de Pointes with non-antiarrhythmic drugs: a call for consensus. Eur. J. Clin. Pharmacol. 57 (3), 185–209.
Didziapetris, R., Lanevskij, K., 2016. Compilation and physicochemical classification analysis of a diverse hERG inhibition database. J. Comput. Aided Mol. Des. 30 (12),

1175–1188.
Doddareddy, M.R., et al., 2010. Prospective validation of a comprehensive in silico hERG model and its applications to commercial compound and drug databases. ChemMedChem 5 (5), 716–729.
Jaeger, S., Fulle, S., Turk, S., 2018. Mol2vec: unsupervised machine learning approach with chemical intuition. J. Chem. Inf. Model. 58 (1), 27–35.
Kalyaanamoorthy, S., Barakat, K.H., 2018. Development of safe drugs: the hERG challenge. Med. Res. Rev. 38 (2), 525–555.
Kim, S., et al., 2018. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res. 47 (D1), D1102–D1109.
Konda, L.S.K., Keerthi Praba, S., Kristam, R., 2019. hERG liability classification models using machine learning techniques. Comput. Toxicol. 12, 100089.
Landrum, G., 2019. RDKit: Open-Source Cheminformatics Software. Available from:. http://www.rdkit.org/.
Leong, M.K., 2007. A novel approach using pharmacophore Ensemble/Support vector machine (PhE/SVM) for prediction of hERG liability. Chem. Res. Toxicol. 20 (2), 217–226.
Mendez, D., et al., 2018. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 47 (D1), D930–D940.
Morgan, H.L., 1965. The generation of a unique machine description for chemical Structures-A technique developed at chemical abstracts service. J. Chem. Doc. 5 (2), 107–113.
Polonchuk, L., 2012. Toward a new gold standard for early safety: automated temperature-controlled hERG test on the PatchLiner. Front. Pharmacol. 3, 3.
Rampe, D., et al., 1997. A mechanism for the proarrhythmic effects of cisapride (Propulsid): high affinity blockade of the human cardiac potassium channel HERG. FEBS Lett. 417 (1), 28–32.
Redfern, W.S., et al., 2003. Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development. Cardiovasc. Res. 58 (1), 32–45.
Rogers, D., Hahn, M., 2010. Extended-connectivity fingerprints. J. Chem. Inf. Model. 50 (5), 742–754.
Roy, M., Dumaine, R., Brown, A.M., 1996. HERG, a primary human ventricular target of the nonsedating antihistamine terfenadine. Circulation 94 (4), 817–823.
Sanguinetti, M.C., Tristani-Firouzi, M., 2006. hERG potassium channels and cardiac arrhythmia. Nature 440 (7083), 463–469.
Siramshetty, V.B., et al., 2018. The Catch-22 of predicting hERG blockade using publicly accessible bioactivity data. J. Chem. Inf. Model. 58 (6), 1224–1233.
Swain, M., 2019. MolVS: Molecule Validation and Standardization. Available from:. https://github.com/mcs07/MolVS.
Vaswani, A., et al., 2017. Attention is all you Need. pp. 5998–6008.
Wang, S., et al., 2016. ADMET evaluation in drug discovery. 16. Predicting hERG blockers by combining multiple pharmacophores and machine learning approaches. Mol. Pharm. 13 (8), 2855–2866.
Witchel, H.J., 2007. The hERG potassium channel as a therapeutic target. Expert Opin. Ther. Targets 11 (3), 321–336.
Zhou, Z., et al., 1999. Block of HERG potassium channels by the antihistamine astemizole and its metabolites desmethylastemizole and norastemizole. J. Cardiovasc. Electrophysiol. 10 (6), 836–843.