

ARTICLE

Proarrhythmic risk assessment of drugs by dV_m/dt shapes using the convolutional neural network

Da Un Jeong¹  | Yedam Yoo¹ | Aroli Marcellinus¹ | Ki-Suk Kim²  | Ki Moo Lim^{1,3}

¹Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, Korea

²R&D Center for Advanced Pharmaceuticals and Evaluation, Korea Institute of Toxicology, Daejeon, Korea

³Department of Medical IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, Korea

Correspondence

Ki Moo Lim, Department of Medical IT Convergence Engineering, Kumoh National Institute of Technology, Gumi 39248, Korea.
Email: kmlim@kumoh.ac.kr

Funding information

This research was partially supported by the Ministry of Food and Drug Safety (22213MFDS3922), the National Research Foundation of Korea (NRF) under the Basic Science Research Program (2022R1A2C2006326), and the Ministry of Science and ICT (MSIT), Korea, under the Grand Information Technology Research Center support program (IITP-2022-2020-0-01612) supervised by the Institute for Information & communications Technology Planning & Evaluation (IITP).

Abstract

Comprehensive in vitro Proarrhythmia Assay (CiPA) projects for assessing proarrhythmic drugs suggested a logistic regression model using qNet as the Torsades de Pointes (TdP) risk assessment biomarker, obtained from in silico simulation. However, using a single in silico feature, such as qNet, cannot reflect whole characteristics related to TdP in the entire action potential (AP) shape. Thus, this study proposed a deep convolutional neural network (CNN) model using differential action potential shapes to classify three proarrhythmic risk levels: high, intermediate, and low, considering both characteristics related to TdP not only in the depolarization phase but also the repolarization phase of AP shape. We performed an in silico simulation and got AP shapes with drug effects using half-maximal inhibitory concentration and Hill coefficients of 28 drugs released by CiPA groups. Then, we trained the deep CNN model with the differential AP shapes of 12 drugs and tested it with those of 16 drugs. Our model had a better performance for classifying the proarrhythmic risk of drugs than the traditional logistic regression model using qNet. The classification accuracy was 98% for high-risk level drugs, 94% for intermediate-risk level drugs, and 89% for low-risk level drugs.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

We suggested a deep convolutional neural network (CNN) classifier using dV_m/dt as an input for assessment of drug's proarrhythmic risk, making it possible to classify drugs with high performance without using a dynamic model.

WHAT QUESTION DID THIS STUDY ADDRESS?

The classical assessment algorithms require high-computation resources and evaluate the Torsades de Pointes (TdP) risk by focusing on the ion channel changes by drugs. The proposed deep CNN model achieved better classification performance by using dV_m/dt , which is derived from the action potential (AP) shape generated by electrophysiological characteristics of myocardial cells.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

The deep CNN model proposed in this study showed drug TdP risk group classification performance almost similar to qNet logistic regression considering hERG dynamics, despite using APs obtained through in silico simulation without hERG dynamics. This result suggests that when assessing the proarrhythmic risk of the drug, it is necessary to consider the comprehensive physiological characteristics, such as AP of all ion channels changed by the drug and all myocardial cells accordingly, not only the specific ion channel.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

We proposed a deep CNN model using dV_m/dt to classify the TdP risk of drugs into three levels: high-risk level, intermediate-risk level, and low-risk level. The proposed deep CNN model using dV_m/dt waveform as an input can make an excellent performance by considering both characteristics related to TdP not only in the depolarization phase but also the repolarization phase of the AP shape.

INTRODUCTION

In developing a new drug, it is necessary to evaluate the drug's cardiac safety, the possibility of causing cardiac arrhythmias. The International Council on Harmonization (ICH) established guidelines for E14¹ and S7B² to assess the potential of drugs to induce cardiac arrhythmias.^{3,4} Among them, a nonclinical evaluation guideline, S7B, is a drug assessment strategy to evaluate the inducibility of Torsades de Pointes (TdP), one of the fatal arrhythmia symptoms, based on a human ether-à-go-go-related gene (hERG) channel current and a QT interval. This method can accurately classify high-risk drugs through a single analytical evaluation through hERG channel-based single assay evaluation focusing on ventricular repolarization.² Indeed, this method successfully has prevented the many drugs that can induce TdP from going on the market. However, due to its high sensitivity and low specificity, this guideline imposes strict regulations on even drugs that do not have the potential to induce arrhythmias, and negatively affects new drug development.⁵ As a result, some drugs that prolong the QT interval but do not induce TdP were restricted from the market or discontinued in the development of the drug having a beneficial potential.⁶ For example, ranolazine, phenobarbital, and tolterodine prolong the QT interval but do not cause TdP. Verapamil blocks the hERG channel but does not lead to TdP, and amiodarone prolongs the QT interval but only very rarely induces TdP.⁷ As an alternation of these classical ICH evaluation guidelines' limitations that would halt the development of potentially valuable therapeutics, a Comprehensive in vitro Proarrhythmia Assay (CiPA) was launched in a Think Tank held at the US Food and Drug Administration (FDA) headquarters in 2013.^{7,8} Current

CiPA teams are steering with the members of prominent organizations; the FDA, the Health and Environmental Science Institute (HESI), the Cardiac Safety Research Consortium (CSRC), the Safety Pharmacology Society (SPS), the European Medicines Agency (EMA), Health Canada, and the Japan National Institute of Health Sciences, Pharmaceutical and Medical Device Agency (PMDA).⁹

The CiPA paradigm assesses the proarrhythmic risk of drugs using in silico simulation with multiple ion channels as the comprehensive evaluation method instead of the single assay methods with only hERG channel or a single associated electrocardiogram (ECG) characteristics.^{7,8} The CiPA paradigm consists of four components, which are the in vitro assessment of drug effects in multiple ionic currents, the in silico computer modeling to predict risk, the in vitro ECG biomarker in phase I clinical trials, and in vitro effects on human stem cell derived ventricular cardiomyocytes.⁸ The first part assesses the torsadogenic risk of a drug through in vitro experiments with multiple cardiac ionic currents focusing mainly on hERG, late sodium, and L-type calcium currents. The second part estimates the drug toxicity using an in silico model mimicking the human ventricular myocyte, with in vitro dataset as input. The third part checks the unexpected in vitro effects on human stem cell-derived ventricular cardiac myocytes. The fourth part determines the unanticipated human impact of drugs that may result from human-specific metabolic characteristics through in vivo ECG biomarkers in phase I clinical trials.⁸

Recently, several studies have demonstrated that an in silico cardiac simulation is a valuable tool for predicting drug effects^{10,11} and assessing a drug's cardiac safety.^{12,13,14} As we mentioned above, the CiPA initiative

presented a method to confirm the influences of multiple ion channels by drugs through in silico simulation using in vitro experimental data through patch clamps as inputs (the second component in CiPA). Crumb et al. identified 30 clinical drugs' effects to the seven ion currents proposed by the CiPA initiative.¹⁵ Li et al.¹⁶ calculated a qNet, the net charge of ions passing through the cell membrane, using in silico simulation and confirmed the TdP-inducing potential of the drug through a logistic regression model using the qNet. Parikh et al. showed the better classification performance of TdP-inducing drugs in the logistic regression using qNet compared to other metrics commonly used as physiological characteristics of myocardial cells, such as the action potential period (APD90) and calcium diastole. They demonstrated that the higher sensitivity of qNet to the inhibition of the late sodium channel was related to the high classification performance through a global sensitivity analysis.¹⁷ In addition, Li et al. showed the importance of the hERG channel for assessing drugs' cardiac risk by classifying TdP occurrence possibility of drugs into high-risk and low-risk groups through qNet calculated through in silico simulation considering hERG dynamic characteristics and in silico simulation without hERG. They validated in silico modeling with drug-hERG channel interaction can help improve the assessment of drug' proarrhythmic risk.^{16,18}

Several research groups have also conducted studies applying advanced algorithms, such as machine learning to evaluate the proarrhythmic drug risk. Lancaster and Sobie suggested a quantitative systems pharmacology approach combining physiological dynamic modeling, statistical analysis, and machine learning to address the issues in the classical assays of drug-induced cardiotoxicity. They successfully developed the machine-learning classifier with the remarkable prediction of TdP risk using the metrics computed from the AP trace and intracellular calcium trace.¹⁹ Polak et al.²⁰ proposed a new algorithm to predict the risk of TdP occurrence as a method for quantifying cardiac toxicity in the early stages of drug development. They extracted some in silico biomarkers, such as APD90, APD50, Pseudo ECG signals, QRS width, QT interval, early repolarization time, and late repolarization time from a cardiac safety simulator with a biophysically detailed myocyte model using information about inhibition and exposure of ion channels. From the extracted features, they classified the risk of TdP occurrence of drugs by using various machine-learning algorithms of the decision tree, random forest, and support vector machine and proposed the empirical decision tree with the best classification performance as the optimal TdP risk assessment model (accuracy of 89%, moderate sensitivity of 71%, and high specificity of 96%).²⁰ Parikh et al.

produced maximum classification accuracy performance of 85, 85%, and 86%, respectively, through binary classifications of logical regression, support vector machine, and natural network model, which use an inhibition rate of ion channels measured through in vitro experiments as direct feature inputs. Then, they suggested logical regression as an optimal classifier for assessing proarrhythmic risk.²¹

The studies we have' just mentioned earlier proposed machine-learning models using in silico biomarkers generated based on data measured through in vitro experiments or machine-learning models using raw in vitro experimental data as direct features. However, if not considering the uncertainty of in vitro dataset, some risks of over-interpreting minor differences between values may occur.²² Furthermore, using insufficient in vitro datasets to the machine-learning models, especially the neural network model, can occur overfitting or underfitting the model performance.²³ In this study, we tried to solve this problem by using in silico waveform computed from bootstrapped in vitro datasets to be sufficient for machine learning. Thereby, we proposed a deep convolutional neural network (CNN) model using differentiation (dV_m/dt) of the AP shape to classify the TdP risk of drugs into three levels: high-risk level, intermediate-risk level, and low-risk level. The proposed deep CNN model using dV_m/dt waveform as an input can make an excellent performance by considering both characteristics related to TdP not only in the depolarization phase but also the repolarization phase of the AP shape.

METHODS

Figure 1 shows an overall conceptual diagram of the proposed algorithm for evaluating the proarrhythmic risk of the drug. The proposed method consisted of three steps as follows (Figure 1a). (1) To generate a sufficient number of drug experiment data, an in vitro patch experiment data was bootstrapped into 2000 drug samples. (2) The stage of in silico simulation computed the AP shape using the bootstrapped drug sample data as an input. (3) After differentiating the computed AP shapes to get the slopes of entire waveform, the TdP risk of a drug is classified into three levels of high-risk, intermediate-risk, and low-risk through the proposed deep CNN model.

Pre-processing of in vitro experimental dataset

We used in vitro datasets of 28 drugs released on the CiPA project website (<https://github.com/FDA/CiPA>). The

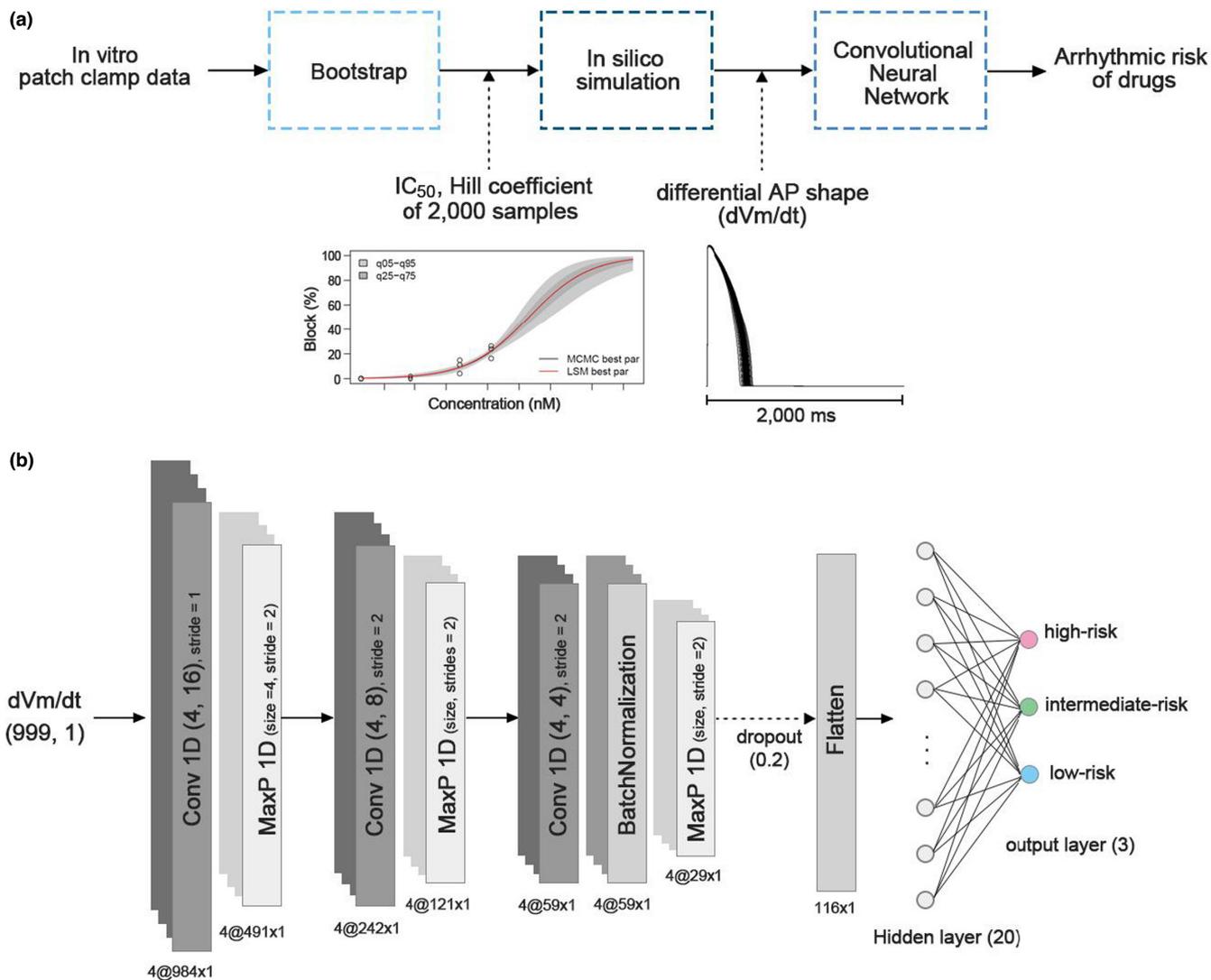


FIGURE 1 Schematic of proposed algorithms. (a) Flow chart of the whole process; (b) the convolutional neural network model structure. AP, action potential; dV_m/dt, differential action potential; IC₅₀, the half inhibitory concentration

released datasets are the patch-clamp experimental data conducted by Li et al.^{16,18} and include the inhibition rates of calcium channel, hERG channel, inward rectifier potassium channel, slow-delayed rectifier potassium channel, Kv4.3 channel, late sodium channel, and peak sodium channel according to drug concentration. These inhibition rates of seven ion channel currents were bootstrapped using the uncertainty quantification algorithm based on the Markov-chain Monte Carlo (MCMC) model^{24,25} to generate 2000 Hill curves, which are ion channel current graphs according to different concentrations. From each Hill curve, we got the half-maximal inhibitory concentration (IC₅₀), which is a drug concentration when the ion channel current is blocked by 50%, and a Hill coefficient, which is a slope value at IC₅₀. Finally, we used IC₅₀s and Hill coefficients of 2000 bootstrapped for each drug in in silico simulation. We used the uncertainty quantification algorithm implemented by Chang et al.^{16,24} with

R programming language (https://github.com/FDA/CiPA/tree/Model-Validation-2018/Hill_Fitting/data).

In silico simulation protocol

To effectively identify drugs' effect on the ionic channels of myocardial cells, we used the O'Hara Rudy ventricular cell model optimized for drug stability evaluation by Dutta et al.^{26,27} (Dutta-ORD). The optimized O'Hara Rudy model was modified by applying the following inhibition factors to the ion channel conductance of hERG channel, inward rectifier potassium channel, slow-delayed rectifier potassium channel, L-type calcium channel, and late sodium channel models.

$$\text{inhibition factor} = \frac{1}{1 + (\text{IC}_{50}/[D])^h} \quad (1)$$

$$g_i = g_{\text{control},j} \left[1 + \left(\frac{[D]}{[IC_{50}]_j} \right) \right]^{-1} \quad (2)$$

Here, IC_{50} is half maximal inhibitory concentration, D represents the concentration of the drug, and h is the Hill coefficient. In silico simulations for drug effect were performed according to $1\times$, $2\times$, $3\times$, and $4\times$ maximum plasma concentration, respectively, using a conductance-block equation (Equation 2), in which the IC_{50} and D reduce the maximal conductance of the j ion channel ($g_{\text{control},j}$). The g_i denotes the modified maximal conductance of the j channel by the drug effect. We first simulated the single cell simulation without any drug effect for the cell to reach the steady-state and save the cell's physiological state to use it as a uniform initial condition. Then, we produced 1000 AP shapes at a cycle length of 2000 ms under the initial cell condition in a steady-state through the in silico single cell simulation with drug effect. Among the 1000 AP shapes, we used the AP shape at maximal dV_m/dt beat, which is at pacing, where the membrane potential variance is maximum, as the input of the deep learning model at the time of repolarization. Performing the in silico simulation with drug effects, we finally generated 8000 in silico AP shapes per drug (2000 bootstrapped samples \times 4 concentrations). The membrane potential (V_m) of myocardial cells is expressed through the following equation:

$$\frac{dV_m}{dt} = \frac{1}{C_m} (I_{\text{total}} + I_{\text{stim}}) \quad (3)$$

where I_{total} denotes the sum of ion channel currents passing through the cell membrane, and I_{stim} denotes the current caused by an external stimulus. C_m is the capacitance of the cell membrane and was set to 1.0 μF in this study.²⁷

The all in silico simulator was developed using the C++ programming language as the base code, supported by several libraries, such as the CVode, for solving the differential equations. Additionally, the simulation worked out using CPU and GPU parallelization implemented with MPI and CUDA interfaces, respectively, to divide large simulation tasks. We used a single instruction multiple data method to process the abundant data in the parallel interface.

Deep CNN model train and performance evaluation

The structure of the deep CNN model proposed in the study is shown in detail in Figure 1b. We computed AP

shapes with a cycle length of 2000 ms from the AP simulation by the time resolution of 2 ms. Each AP shape has 1000 data points, and the dV_m/dt waveform derived from the AP shape has 999 data points, which feeds into the CNN model as input. The model consisted of three 1D CNN layers, the first CNN layer (16, 1), the second CNN layer (8, 1), and the third CNN layer (4, 1) with four Kernels. Each CNN layer was connected to the 1D Max Pooling layer. The first Max Pooling layer was applied with a four-sized window moving by two spaces, and the other Max Pooling layer was applied with a two-sized window moving by two spaces. Subsequently, after dropping out at a rate of 20%, it passed through a Flatten layer and a Dense layer with 20 neurons (nodes). Then, finally, it classified the TdP risk of the drug into three levels: high-risk, medium-risk, and low-risk. We applied the ReLU activation function to all CNN and Dense layers except the output layer, which uses the softmax function to classify the risk level. The proposed deep CNN model was trained using the "Adam" optimization and Category cross-entropy loss functions. Among the 28 drugs released by CiPA, we used 12 drugs for training the model: "bepridil," "dofetilide," "quinidine," "sotalol," "cisapride," "terfenadine," "ondansetron," "chlorpromazine," "verapamil," "diltiazem," "ranolazine," and "mexiletine" and used the remaining 16 drugs for the test: "ibutilide," "vandetanib," "azimilide," "disopyramide," "domperidone," "pimozide," "astemizole," "droperidol," "clarithromycin," "clozapine," "risperidone," "tamoxifen," "loratadine," "nitrendipine," "nifedipine," and "metoprolol" (Table 1). We implemented all codes for the machine learning and the evaluation of the model using Python 3.8.8 languages in the Spyder console. We shared the simulation code of our proposed model in the Supplementary Materials S1.

Performing in silico simulation using the bootstrapped drug samples as inputs, as we mentioned above, we generated a total of 224,000 AP shapes (2000 bootstrapped samples \times 28 drugs \times 4 concentrations). We randomly extracted 500 AP shapes per concentration among the AP shapes of 12 training drugs, a total of 2000 AP shapes per drug, and differentiated them to use in the training of our proposed model. To evaluate the model's classification performance, we used the 10,000-testing protocols presented in the CiPA project (Figure 2).¹⁶ First, we made a 10,000 dataset of 16 drugs by randomly extracting one sample from 8000 dV_m/dt shapes in each drug. Next, the model evaluating process produced 10,000 receiver operating curves (ROC) of high-risk, medium-risk, and low-risk drugs through 10,000 datasets. Then, we statistically evaluated the model performance by calculating the area under the curve (AUC) from 10,000 ROC and performing a likelihood ratio test to assess the model's goodness of fit.

TABLE 1 List of 28 drugs

Proarrhythmic risk level	Train drugs		Test drugs	
	Name	C_{\max} (nM)	Name	C_{\max} (nM)
High-risk	Quinidine	3237	Disopyramide	742
	Sotalol	14,690	Ibutilide	100
	Dofetilide	2	Vandetanib	255.4
	Bepridil	33	Azimilide	70
Intermediate-risk	Cisapride	2.6	Clarithromycin	1206
	Terfenadine	4	Clozapine	71
	Chlorpromazine	38	Domperidone	19
	Ondansetron	139	Droperidol	6.33
			Pimozide	0.431
			Risperidone	1.81
			Astemizole	0.26
Low-risk	Verapamil	81	Metoprolol	1800
	Ranolazine	1948.2	Nifedipine	7.7
	Diltiazem	122	Nitrendipine	3.02
	Mexiletine	4129	Tamoxifen	21
			Loratadine	0.45

Abbreviations: CiPA, Comprehensive in vitro Proarrhythmia Assay; C_{\max} , maximum plasma concentration; Tdp, Torsades de Pointes.

It accumulated the confusion matrices of 10,000-test using 16 test drugs (a), and all 28 drugs (b). All drugs were selected by the CiPA research group and categorized into high-, intermediate-, and low-risk levels according to the TdP risk. The drugs dataset consists of 12 for training and 16 for the test decided by clinical cardiologists and electrophysiologists based on publicly available data and expert opinion.

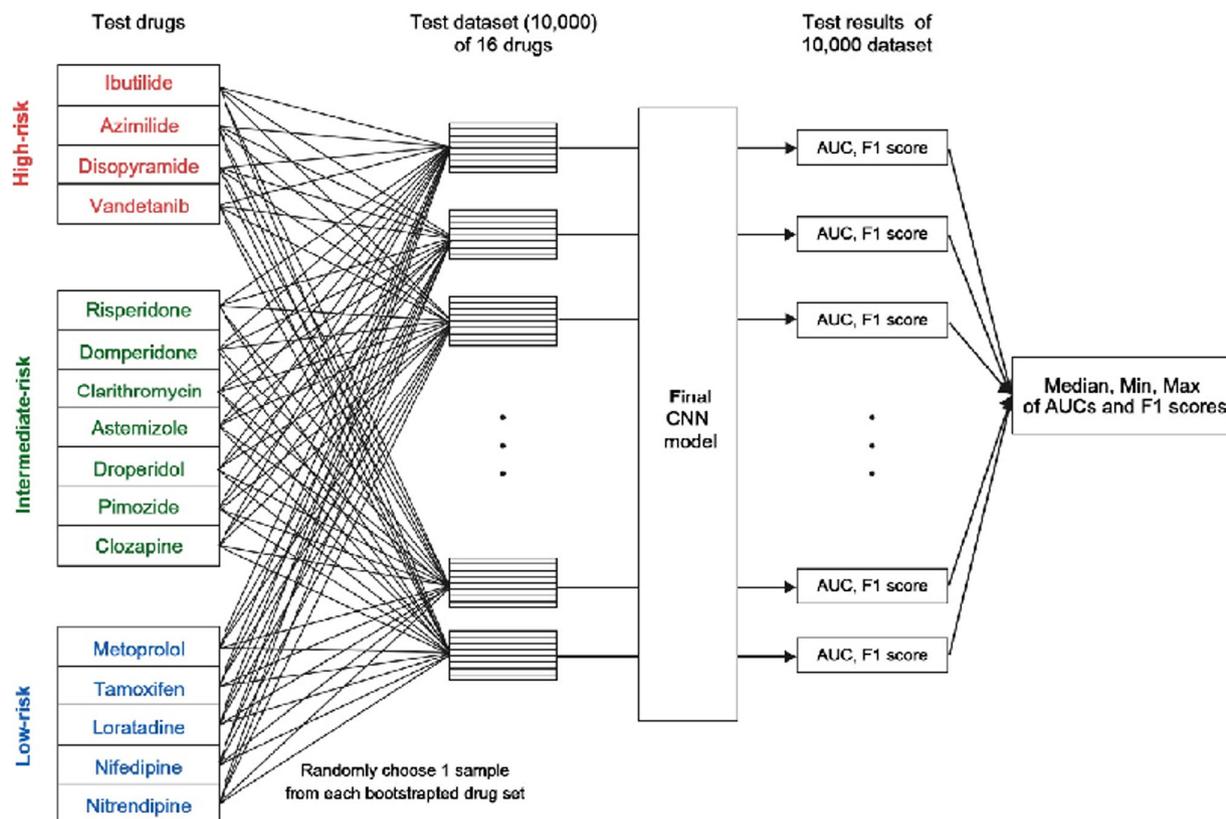


FIGURE 2 Testing algorithm for evaluating the model performance; this algorithm was suggested by the CiPA research group based on the central limit theorem; AUC, area under the receiver operating curves; CiPA, comprehensive in vitro proarrhythmia assay

The likelihood ratio test assesses the model fitness using a positive likelihood rate (LR+) with a value range of one from infinity and a negative likelihood rate (LR-) with a range of zero to one, which were calculated from the sensitivity and specificity of the model.²⁸

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

$$\text{LR+} = \frac{\text{Sensitivity}}{1 - \text{Specificity}} \quad (6)$$

$$\text{LR-} = \frac{1 - \text{Sensitivity}}{\text{Specificity}} \quad (7)$$

true answer is incorrectly predicted as false. In addition, we evaluated the performance of the deep CNN model with the normalized cumulative confusion matrix of 10,000 confusion matrices generated through 10,000 tests and calculated accuracies and F1 scores. Here, the F1 score is the harmonic mean of sensitivity (recall) and precision and can evaluate the model's performance considering the imbalance in the TdP risk labels of drugs used for the test.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

Here, TP is the true positive and denotes a case where the model predicts the true answer as true. TN is the true negative and represents a case where the actual false answer is predicted as false. FP is the false positive, indicating that the actual false answer is incorrectly predicted as true, and FN is the false negative, meaning that the case where the actual

RESULTS

Figure 3 and Table 2 show the results of evaluating the performance for classifying the risk of TdP occurrence of drugs through the deep CNN model proposed in this study. The classification performance of each model was

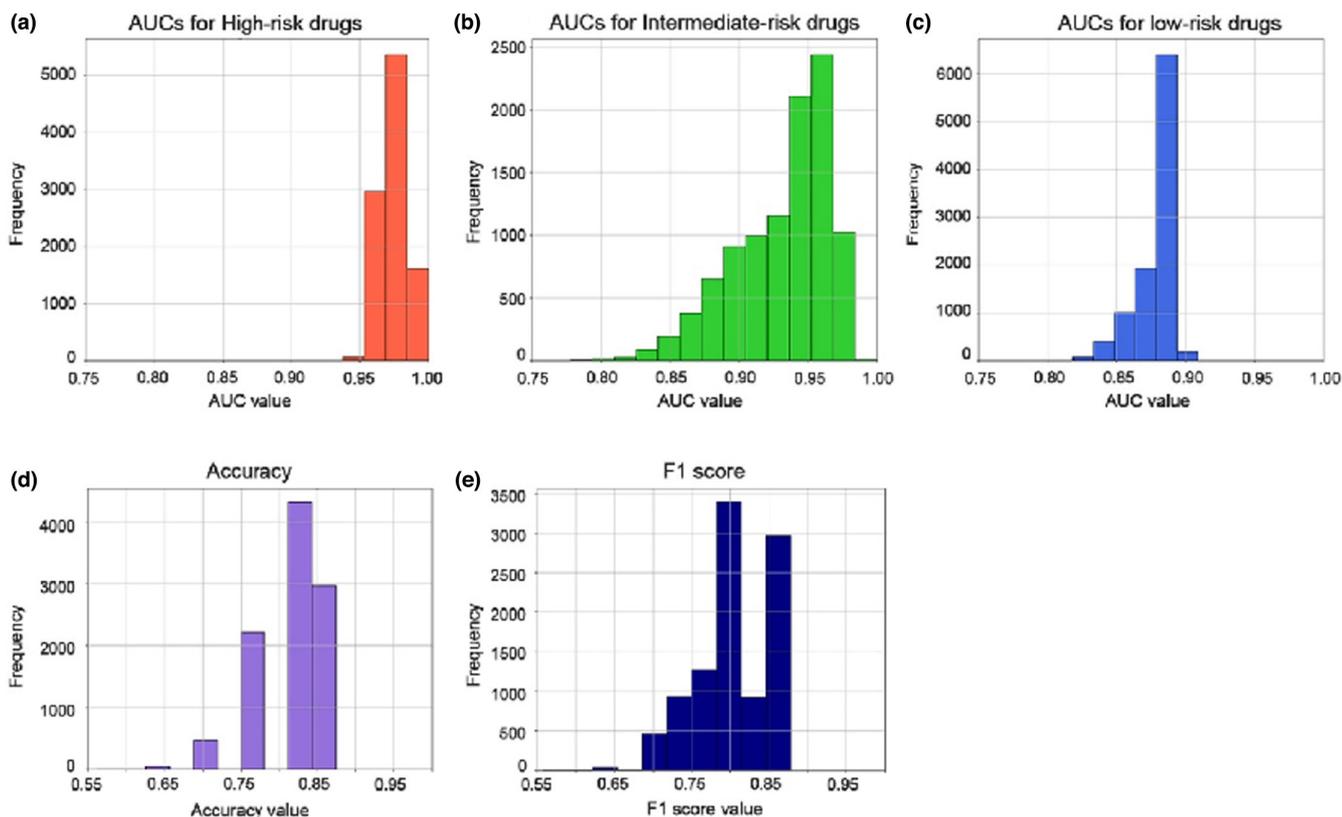


FIGURE 3 Histogram results of the 10,000-test using 16 test drugs. Distribution of AUCs in the 10,000 ROC curves for high-risk drugs (a), intermediate-risk drugs (b), and low-risk drugs (c); (d) distribution of final model accuracy; (e) F1 scores distribution of the 10,000 confusion matrices. AUC, area under the ROC curves; ROC, receiver operating curves

TABLE 2 Comparison of model performances for classifying the proarrhythmic risk of drugs

Model		Logistic regression using qNet without hERG (CiPA) [15]	Logistic regression using qNet (CiPA) with hERG [16]	Proposed deep CNN model using dV_m/dt		
		All drugs	All drugs	All drugs	Test drugs	
AUCs	High	0.86 (0.81–0.90)	0.988 (0.95–1.0)	0.89 (0.84–0.95)	0.97 (0.89–1.0)	0.98 (0.94–1.0)
	Intermediate	–	–	–	0.93 (0.76–0.99)	0.94 (0.78–1.0)
	Low	0.86 (0.82–0.90)	0.901 (0.88–0.93)	1.0 (0.92–1.0)	0.92 (0.85–0.96)	0.89 (0.82–0.91)
LR+	High	2.01 (1.61–2.84)	8.05 (4.03–9)	12 (4.5–1e+6)	8.75 (2.92–20.00)	6.00 (4.00–12.00)
	Intermediate	–	–	–	6.95 (2.06–inf)	7.71 (1.92–inf)
	Low	5.00 (3.33–12.5)	7.5e+5 (8.75–1e+6)	4.5 (2.3–5)	16.89 (3.17–inf)	8.80 (2.20–inf)
LR–	High	0.118 (1.8e–6–0.284)	0.0677 (1.13e–6–0.18)	1.1e–06 (1e–6–0.3)	0.13 (2.05e–6–0.33)	2.20e–06 (2.1e–6–2.3e–6)
	Intermediate	–	–	–	0.21 (0.09–0.77)	0.29 (0.14–0.80)
	Low	0.556 (0.395–0.833)	0.25 (1e–6–0.263)	0.11 (1.2e–6–0.23)	0.22 (0.11–0.53)	0.22 (0.20–0.55)
Accuracy		–	–	–	0.83 (0.61–0.93)	0.81 (0.56–0.88)
F1 score		–	–	–	0.83 (0.60–0.93)	0.81 (0.56–0.88)

Abbreviations: AUC, the area under the receiver operating curve; CiPA, Comprehensive in vitro Proarrhythmia Assay; CNN, convolutional neural network; LR+, positive likelihood ratio; LR–, negative likelihood ratio.

tested using a set of 10,000 drugs generated randomly for each test drug, and AUC was calculated from 10,000 ROCs generated according to 10,000 repeated tests. Figure 3a–c shows AUCs distribution in 95% confidence intervals for the TdP risk levels of 16 test drugs classified through the proposed model. When classifying high-risk drugs, AUC of the deep CNN model was 0.98 (95% confidence interval: 0.94–1.0), which was improved from that of logistic regression using qNet obtained through in silico simulation considering hERG dynamic characteristics proposed by Li et al. (0.89 with 95% confidence interval from 0.94 to 0.95). However, the AUC for low-risk drugs was 0.89 (0.82–0.91), lower than the classification result of low-risk drugs in the hERG dynamic logistic regression model. The TdP risk level classification performance of the proposed deep CNN models was the lowest for low-risk drugs; when classifying intermediate-risk drugs, the AUC was 0.94 (0.78–1.0).

To statistically verify the differences among the three risk levels of drugs classified through the proposed deep CNN model, we performed the likelihood ratio test and calculated an LR+ and an LR–. As a result, low-risk drugs were 8.8 (median of LR+ with 95% confident interval from 2.2 to infinity) times more likely to be accurately classified as low-risk level. Intermediate-risk drugs were 7.7 (1.92–infinity) times and high-risk drugs 6.0 (4.0–12.0) times more likely to be accurately classified as intermediate-risk level and high-risk level, respectively. In addition, the likelihood of high-risk drugs being classified as other risk levels, not high-risk level, was close to zero (median of $1/LR- = \sim 454,545$ times), and there was no likelihood

that medium of $1/LR-$, and the possibility of medium-risk drugs and low-risk drugs being classified as other risk levels was 3.4 and 4.5 times lower, respectively.

Figure 4 shows the cumulative normalized confusion matrices for the classified TdP risk level of drugs as 10,000 test results of the proposed deep CNN model. Figure 4a is a normalized cumulative confusion matrix that performed 10,000 tests using 16 test drugs, and Figure 4b is a normalized cumulative confusion matrix for all 28 drugs. The proposed model classified the three cardiac toxicity risk groups of drugs at once, resulting in 81% (56–88%) accuracy for the 16 test drugs and 83% (61–93%) accuracy for the 28 drugs. Furthermore, the final F1 score of the proposed deep CNN model was 0.81 (0.56–0.88) for 16 test drugs and 0.83 (0.60–0.93) for all drugs.

DISCUSSION

This study proposed a deep CNN model using dV_m/dt shape as a method for evaluating nonclinical heart toxicity of drugs. The previously proposed assessment algorithm for drug toxicity classified the TdP risk level based on changes in ion channels due to drugs measured through in vitro experiments.^{15–18,20} These algorithms calculated and used biomarkers for the assay through the inhibition rate of six to seven ion channels affected by drugs; the CiPA research groups calculated the qNet as the sum of ion charges passing through six ion channels (hERG channel, inward rectifier potassium channel, slow-delayed rectifier potassium channel, Kv4.3 channel,

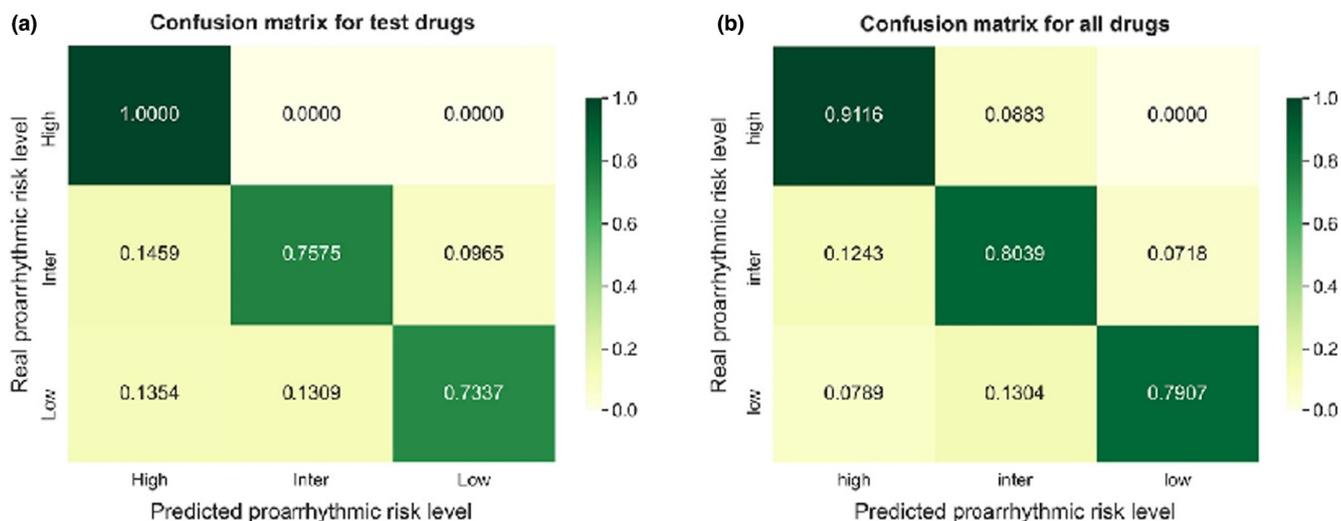


FIGURE 4 Confusion matrix for classification of drug's proarrhythmic risk

L-type calcium channel, and late sodium channel models) and the $qInward$ as the sum of charge changes passing through two ion channels (L-type calcium channel and late sodium channel).²⁴ The deep CNN model proposed in this study uses dV_m/dt , which is derived from the AP shape generated by electrophysiological characteristics of myocardial cells. The AP shape varied depending on the ion channel changed by the drug and the changes of the integrated ion channels. In the previous study, we predicted the changed ion channel from these AP shapes through a simple artificial neural network (ANN) model.²⁹ Furthermore, we validated it by successfully predicting the ion channel mainly affected by the drug from the AP shapes generated under a specific drug condition.²⁹ The proposed ANN model estimated the changes of the ion channel conductance due to three kinds of drugs: ibutilide (slow-delayed rectifier potassium channel), dofetilide (hERG channel), and diltiazem (L-type calcium channel) from the difference signal between the control AP shape that had not affected anything and drug-affected AP shape. AUCs for each drug was 1.00 for ibutilide, 0.88 for dofetilide, and AUCs for dofetilide (specificity and sensitivity; 0.99 and 1.00 for ibutilide, 1.00 and 0.79 for diltiazem, 0.90 and 0.90 for dofetilide). In this study, we used the dV_m/dt shape to classify the proarrhythmic risk of the drug without a free-drug condition. The proposed deep CNN model can achieve better classification performance than the classical models that evaluate drug toxicity by focusing on the ion channel change caused by drugs. In addition, it showed excellent results, exceeding the performance of our other study—Yoo et al. developed the ANN model using nine *in silico* features considering the morphological information of the AP trace and transient calcium trace, including the $qNet$ and $qInward$, which are the ion net charge features.³⁰

In several studies, the dV_m/dt was identified to be helpful to detect the occurrence of TdP or the drug effects.^{31,32} Because when drugs make a blockage in the ion channels, the dV_m/dt in the depolarization phase is reduced; Passini et al. observed the main effect of lidocaine and mexiletine to the peak sodium through a decreased maximal dV_m/dt .³¹ Tomek et al.³² used the dV_m/dt as the standard for detecting the early after depolarization formation, which initiated TdP. However, as Akanda et al.³³ reported in their experimental study, ion channel blockage causes a decrease not only in the depolarization phase but also in the repolarization phase of the AP trace. In this sense, we hypothesized the whole dV_m/dt waveform when the AP shape occurs would be better to detect drug-induced TdP compared to the single dV_m/dt value in the depolarization. To empirically validate our hypothesis about dV_m/dt and find the optimal inputs for assessing the drugs' cardiac risk, we tested several *in silico* waveforms related to TdP, including the maximal dV_m/dt in the depolarization. As shown in Table S1, the dV_m/dt shape has the best performance to classify the risk of drug toxicity. Thereby, we finally suggested the dV_m/dt shape in this study.

To objectively compare with the classification performance of the logistic regression using $qNet$ proposed by Li et al., we classified all 28 drugs through the deep CNN model (Figure S1; Table 2). The AUCs of three risk levels for all drugs classified through the proposed model was 0.9 or higher; the AUC of high-risk drugs was 0.97 (0.89–1.0), the AUC of medium-risk drugs was 0.93 (0.76–0.99), and the AUC of low-risk drugs was 0.92 (0.85–0.96). The proposed model significantly improved the classification performance over the $qNet$ logistic regression model calculated through *in silico* simulations without considering hERG dynamic characteristics; the AUCs of high-risk and

low-risk drugs were 0.86 (0.81–0.90) and 0.86 (0.82–0.90), respectively. Compared with the hERG dynamic qNet logistic regression model, the AUCs of the proposed model was 1.8% lower for high-risk drugs (AUC median of 0.988 with 95% confidence interval from 0.84 to 1.0) but 2.1% higher for low-risk drugs (AUC median of 0.901 with 95% confidence interval from 0.88 to 0.93).

Among three classification models, the likelihood of accurately classifying the high-risk drugs was highest in the proposed deep CNN model; the high-risk drug was 8.75 times more likely to be classified as high-risk level (median of LR+; 2.01 times for qNet-logistic regression without hERG and 8.05 times for qNet-logistic regression with hERG). The hERG dynamic qNet logistic regression model was the highest for the low-risk drugs, being $7.5e+5$ times more likely to classify the low-risk drugs as low-risk level, followed by the deep CNN model (LR+; 16.89). The qNet-logistic regression model without hERG dynamic was 5.0 times more likely to classify the low-risk drugs as low-risk level. The proposed deep CNN model was 7.6 times less likely to classify high-risk drugs into other TdP risk levels. Still, the hERG dynamic qNet logistic regression model was 14.8 times lesser, and the qNet-logistic regression model without hERG dynamic was 8.5 times lesser. However, the proposed deep CNN model had the most likelihood of classifying low-risk drugs into other TdP risk levels with LR– of 4.5. In contrast, the hERG dynamic qNet logistic regression model and the hERG dynamic-free qNet-logistic regression model were 4.0 times lesser and 1.79 times lesser, respectively.

The dV_m/dt used in the proposed deep CNN model was calculated through *in silico* simulation using the Duttal-ORD model without considering the hERG dynamic. The CiPA research group noted the importance of considering drug-induced hERG dynamic properties in integrated ion channel models when using *in silico* simulations to assess the TdP risk in drugs.¹⁸ In addition, as can be seen from the TdP risk classification results of drugs using the qNet-logistic regression model they progressed, the distribution and classification accuracy of *in silico* biomarkers may vary depending on the presence or absence of dynamic characteristics of the hERG channel.^{16,18} Surely, with the use of the hERG dynamic results, the classification result of our proposed model may also be slightly reduced compared with the present performance. However, the proposed deep CNN model achieved almost similar performance to the hERG dynamic-qNet logistic regression model using dV_m/dt obtained from *in silico* simulation, not considering the hERG dynamic.

It is essential to verify the consistency of drug data used in evaluating the proposed model.²³ Training or testing the model using 50 samples or less data increases the risk of overestimating or underestimating the model

performance.²² This study used 28 drugs and only 16 drugs in the model test. In general, it is well known that in the medical field, if the number of samples does not exceed 30, normality is not satisfied, which means the sample group can reflect the population well.³⁴ In addition, the ordinal logistic regression works with the assumption that input data are independent of each other.^{16,35,36} However, the used drug dataset are conjunctly correlated with each other. Thereby, we followed the 10,000-testing algorithm proposed by the FDA in the CiPA initiative.¹⁶ This algorithm is based on the central limit theorem, which states that if random sampling is performed a sufficiently large number of times on a sample of <30 , the variance of the dataset becomes small and approaches a normal distribution.³⁷ To evaluate the statistical significance in the fitness of the deep CNN model, we performed a likelihood ratio test. In general, probability used for performance evaluation of artificial intelligence models or probabilistic models is used as statistics to confirm the possibility of a specific phenomenon being observed from the parameters. However, in the likelihood ratio test, likelihood is a statistic for inferring the most likely parameter in a particular phenomenon.²⁸ Accordingly, LR+ ranges between one and infinity, and the closer to infinity, the higher the possibility that the TdP risk level of a particular drug is more likely to be classified in the appropriate risk level than other risk levels. Likewise, LR– ranges between zero and one and denotes that the closer to zero, the greater the association with a specific risk level. Accordingly, as LR+ is very high and LR– is very low, a more remarkable discrimination ability is obtained. Consequently, the deep CNN model with dV_m/dt as input proposed in this study has excellent identification ability.

The 16-drug data used to verify the model have data imbalances according to the TdP risk levels (four high-risk drugs, seven intermediate-risk drugs, and five low-risk drugs). Therefore, to prevent the proposed model from being underestimated or overestimated due to this data imbalance, the model was evaluated through the F1 score, including ROC curves and accuracy. It is why ROC cannot consider the difference in the number of data between classes.^{38,39} In contrast, precision and recall can confirm the model's classification performance considering the difference in the number of data. Accordingly, the harmonic mean of precision and recall, F1 score, allows the model to be appropriately evaluated considering the drug data imbalance according to the TdP risk levels.³⁸

To find the optimal CNN model for assessing drugs' cardiac risk, we tested the classification performance by changing the model parameters and structures, such as the number of neurons, the ratio of dropout, and the location of the batch normalization layer. Among several CNN models tested, we showed the structures and the performances of five representative models in Table S2.

Comparing the AUCs of each categorized drug toxicity, we decided on the present deep CNN model as the best parameter model with the highest classification accuracy.

Two methods are usually used to quantify the uncertainty of in vitro datasets.^{24,31,40} The first is to directly bootstrap the in vitro dataset based on the MCMC and compute the AP shapes, which we used in this study.²⁴ The second one calculates the AP shape by setting random parameters in the AP simulation.^{31,40} The former method can reflect the generalization of in vitro experimental data but requires an effective computing process, such as parallelization to handle large datasets. In contrast, the latter approach is fast, but there is doubt whether that can reflect the generalization of in vitro data.²⁴ Considering the possibility of reflecting in vitro experimental datasets in this study, we used the former uncertainty quantification method.

The proposed deep CNN model used the dV_m/dt waveform computed from in silico simulation to assess the proarrhythmic risk of a drug. Because the CNN model captures the morphological information of the waveform, the somewhat low time resolution of our AP shape might affect classification performances. In fact, when testing our proposed model using a new test set changing the initial data points from the resting state until the right-after AP upstroke to random values, the classification performance was slightly reduced (Table S3). However, our proposed model still classified the high-risk drugs better than the qNet logistic regression model, which indirectly denoted that our CNN classifier does not depend on a rough calculation dV_m/dt during the AP upstroke.

Even though it does not know the exact physiological meaning of the extracted machine-learning features from the proposed model, we expect they should be related to TdP because the dV_m/dt waveform has information of the whole AP shape, including the repolarization, plateau periods, depolarization, etc. The dV_m/dt shapes can also be calculated from the AP shapes measured through in vitro experiments. If our proposed model is trained and validated using actual dV_m/dt waveform of the in vitro AP shape, we think it could be utilized both in silico and in vitro. The proposed deep CNN model has been trained and tested using limited drug data. Therefore, if the proposed model is verified and calibrated with more experimental data, it could be used as an auxiliary system at the stage of evaluating drug stability in the new drug development laboratory.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

D.U.J. and K.M.L. wrote the manuscript. D.U.J. and K.M.L. designed the research. D.U.J. and Y.Y. performed

the research. D.U.J., Y.Y., K.K., and K.M.L. analyzed the data. A.M. and K.M.L. contributed new analytical tools.

ORCID

Da Un Jeong  <https://orcid.org/0000-0001-8490-5007>

Ki-Suk Kim  <https://orcid.org/0000-0002-0931-0924>

REFERENCES

1. ICH Harmonised Tripartite Guideline. The clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs E14. *Fed Regist.* 2005;70:61134-61135.
2. ICH Harmonised Tripartite Guideline. The non-clinical evaluation of the potential for delayed ventricular repolarization (QT interval prolongation) by human pharmaceuticals S7B. *Fed Regist.* 2005;70:61133-61134.
3. Onakpoya IJ, Heneghan CJ, Aronson JK. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. *BMC Med.* 2016;14:1-11.
4. Yap YG, Gamm AJ. Drug induced QT prolongation and Torsades de pointes. *Heart.* 2003;89:1363-1372.
5. Colatsky T, Fermini B, Gintant G, et al. The comprehensive in vitro proarrhythmia assay (CiPA) initiative—update on progress. *J Pharmacol Toxicol Methods.* 2016;81:15-20.
6. Llopis-Lorente J, Gomis-Tena J, Cano J, Romero L, Saiz J, Trenor B. In silico classifiers for the assessment of drug proarrhythmicity. *J Chem Inf Model.* 2020;60:5172-5187.
7. Sager PT, Gintant G, Turner JR, Pettit S, Stockbridge N. Rechanneling the cardiac proarrhythmia safety paradigm: a meeting report from the cardiac safety research consortium. *Am Heart J.* 2014;167:292-300.
8. Strauss DG, Gintant G, Li Z, et al. Comprehensive in vitro Proarrhythmia assay (CiPA) update from a cardiac safety research consortium/Health and Environmental Sciences Institute/FDA meeting. *Ther Innov Regul Sci.* 2019;53:519-525.
9. CiPA Initiative Homepage. 2019 CIPAIHealth and Environmental Sciences Institute (HESI); 2019. <https://cipaproject.org/>
10. Sobie EA. Parameter sensitivity analysis in electrophysiological models using multivariable regression. *Biophys J.* 2009;96:1264-1274.
11. Zemzemi N, Bernabeu MO, Saiz J, et al. Computational assessment of drug-induced effects on the electrocardiogram: from ion channel to body surface potentials. *Br J Pharmacol.* 2013;168:718-733.
12. Margara F, Wang ZJ, Levrero-Florencio F, et al. In-silico human electro-mechanical ventricular modelling and simulation for drug-induced pro-arrhythmia and inotropic risk assessment. *Prog Biophys Mol Biol.* 2021;159:58-74.
13. Mirams GR, Cui Y, Sher A, et al. Simulation of multiple ion channel block provides improved prediction of compounds' clinical torsadogenic risk. *Cardiovasc Res.* 2011;91:53-61.
14. Obiol-Pardo C, Gomis-Tena J, Sanz F, Saiz J, Pastor M. A multiscale simulation system for the prediction of drug-induced cardiotoxicity. *J Chem Inf Model.* 2011;51:483-492.
15. Crumb WJ, Vicente J, Johannesen L, Strauss DG. An evaluation of 30 clinical drugs against the comprehensive in vitro proarrhythmia assay (CiPA) proposed ion channel panel. *J Pharmacol Toxicol Methods.* 2016;81:251-262.

16. Li Z, Ridder BJ, Han X, et al. Assessment of an in silico mechanistic model for Proarrhythmia risk prediction under the CiPA initiative. *Clin Pharmacol Ther.* 2019;105:466-475.
17. Parikh J, Di Achille P, Kozloski J, Gurev V. Global sensitivity analysis of ventricular myocyte model-derived metrics for proarrhythmic risk assessment. *Front Pharmacol.* 2019;10:1-18.
18. Li Z, Dutta S, Sheng J, et al. Improving the in silico assessment of proarrhythmia risk by combining hERG (human ether-à-go-go-related gene) channel-drug binding kinetics and multichannel pharmacology. *Circ Arrhythm Electrophysiol.* 2017;10:1-12.
19. Lancaster MC, Sobie EA. Improved prediction of drug-induced Torsades de pointes through simulations of dynamics and machine learning algorithms. *Clin Pharmacol Ther.* 2016;100:371-379.
20. Polak S, Romero K, Berg A, et al. Quantitative approach for cardiac risk assessment and interpretation in tuberculosis drug development. *J Pharmacokinet Pharmacodyn.* 2018;45:457-467.
21. Parikh J, Gurev V, Rice JJ. Novel two-step classifier for Torsades de pointes risk stratification from direct features. *Front Pharmacol.* 2017;8:1-18.
22. Marie L. Data quality management in pharmacovigilance. *Drug Saf.* 2004;27:857-870.
23. McCabe SD, Lin DY, Love MI. Consistency and overfitting of multi-omics methods on experimental data. *Brief Bioinform.* 2019;21:1277-1284.
24. Chang KC, Dutta S, Mirams GR, et al. Uncertainty quantification reveals the importance of data variability and experimental design considerations for in silico proarrhythmia risk assessment. *Front Physiol.* 2017;8:1-17.
25. Soetaert K, Petzoldt T. Inverse modelling, sensitivity and Monte Carlo analysis in R using package FME. *J Stat Softw.* 2010;33:1-28.
26. O'Hara T, Virág L, Varró A, Rudy Y. Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation. *PLoS Comput Biol.* 2011;7:1-29.
27. Dutta S, Chang KC, Beattie KA, et al. Optimization of an in silico cardiac cell model for proarrhythmia risk assessment. *Front Physiol.* 2017;8:1-15.
28. Aggarwal R, Ranganathan P. Understanding diagnostic tests – part 3: receiver operating characteristic curves. *Perspect Clin Res.* 2018;9:145-148.
29. Jeong DU, Lim KM. Artificial neural network model for predicting changes in ion channel conductance based on cardiac action potential shapes generated via simulation. *Sci Rep.* 2021;11:1-8.
30. Yoo Y, Marcellinus A, Jeong DU, Kim K-S, Lim KM. Assessment of drug Proarrhythmicity using artificial neural networks with in silico deterministic model outputs. *Front Physiol.* 2021;12:1-7.
31. Passini E, Britton OJ, Lu HR, et al. Human in silico drug trials demonstrate higher accuracy than animal models in predicting clinical pro-arrhythmic cardiotoxicity. *Front Physiol.* 2017;8:1-15.
32. Tomek J, Bueno-Orovio A, Passini E, et al. Development, calibration, and validation of a novel human ventricular myocyte model in health, disease, and drug block. *elife.* 2019;8:1-48.
33. Akanda N, Molnar P, Stancescu M, Hickman JJ. Analysis of toxin-induced changes in action potential shape for drug development. *J Biomol Screen.* 2009;14:1228-1235.
34. Kwak SG, Park S-H. Normality test in clinical research. *J Rheum Dis.* 2019;26:5.
35. Parsons NR, Costa ML, Achten J, Stallard N. Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package R. *Comput Stat Data Anal.* 2009;53:632-641.
36. Nooraee N, Molenberghs G, Van Den Heuvel ER. GEE for longitudinal ordinal data: comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN. *Comput Stat Data Anal.* 2014;77:70-83.
37. Rosenblatt M. A central limit theorem and a strong mixing condition. *Proc Natl Acad Sci.* 1956;42:43-47.
38. Jeni LA, Cohn JF, De La Torre F. Facing imbalanced data – recommendations for the use of performance metrics. *Int Conf Affect Comput Intell Interact Workshops.* 2013;2013:245-251. doi:10.1109/ACII.2013.47
39. Fawcett T. An introduction to ROC analysis. *Pattern Recogn Lett.* 2006;27:861-874.
40. Passini E, Zhou X, Trovato C, Britton OJ, Bueno-Orovio A, Rodriguez B. The virtual assay software for human in silico drug trials to augment drug cardiac testing. *J Comput Sci.* 2021;52:101202.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Jeong DU, Yoo Y, Marcellinus A, Kim K-S, Lim KM. Proarrhythmic risk assessment of drugs by dV_m/dt shapes using the convolutional neural network. *CPT Pharmacometrics Syst Pharmacol.* 2022;11:653-664. doi:[10.1002/psp4.12803](https://doi.org/10.1002/psp4.12803)