Improving the hERG model fitting using a deep learning-based method

Jaekyung Song[1,2], Yu Jin Kim[1] and Chae Hun Leem[1,2]*

[1]Department of Physiology, Asan Medical Center, Seoul, South Korea, [2]Department of Physiology, University of Ulsan College of Medicine, Seoul, South Korea
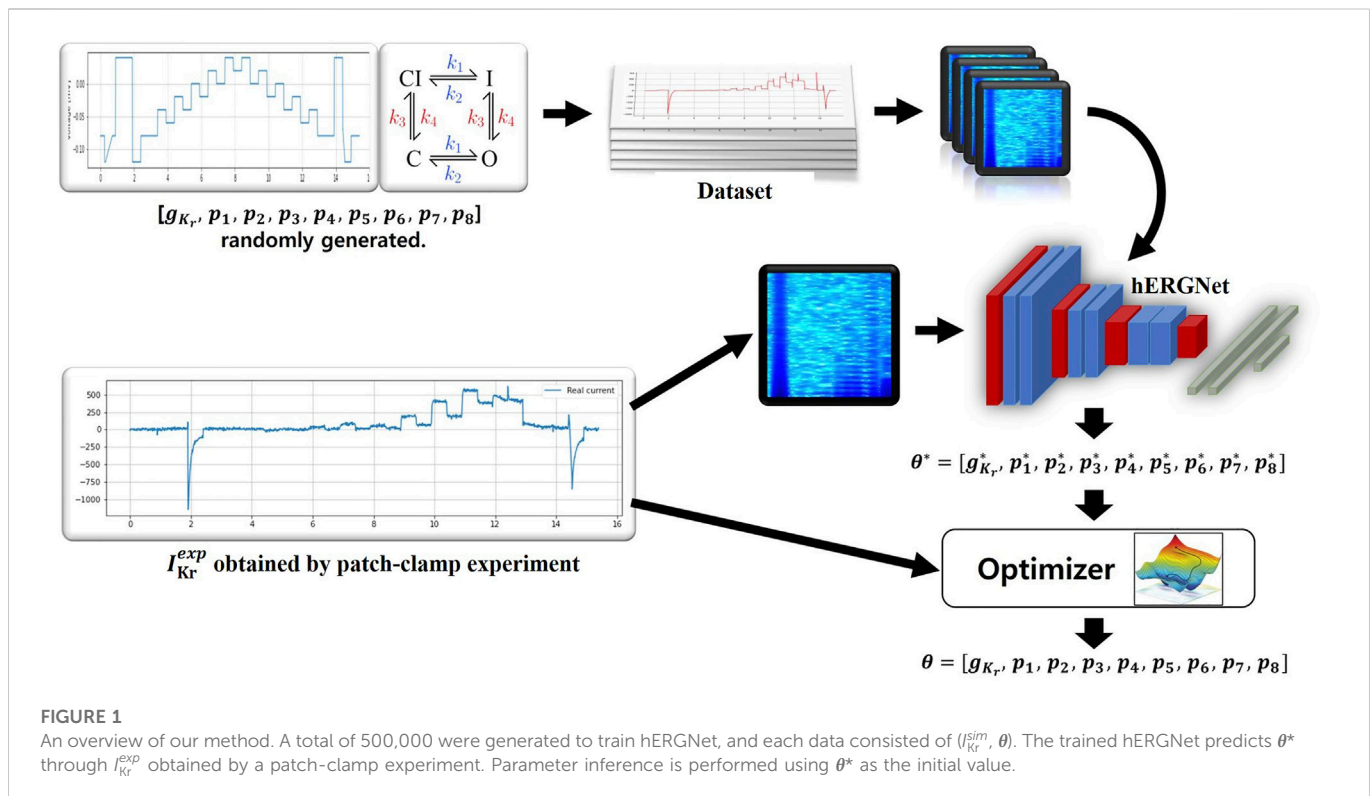
The hERG channel is one of the essential ion channels composing the cardiac action potential and the toxicity assay for new drug. Recently, the comprehensive *in vitro* proarrhythmia assay (CiPA) was adopted for cardiac toxicity evaluation. One of the hurdles for this protocol is identifying the kinetic effect of the new drug on the hERG channel. This procedure included the model-based parameter identification from the experiments. There are many mathematical methods to infer the parameters; however, there are two main difficulties in fitting parameters. The first is that, depending on the data and model, parametric inference can be highly time-consuming. The second is that the fitting can fail due to local minima problems. The simplest and most effective way to solve these issues is to provide an appropriate initial value. In this study, we propose a deep learning-based method for improving model fitting by providing appropriate initial values, even the right answer. We generated the dataset by changing the model parameters and trained our deep learning-based model. To improve the accuracy, we used the spectrogram with time, frequency, and amplitude. We obtained the experimental dataset from https://github.com/CardiacModelling/hERGRapidCharacterisation. Then, we trained the deep-learning model using the data generated with the hERG model and tested the validity of the deep-learning model with the experimental data. We successfully identified the initial value, significantly improved the fitting speed, and avoided fitting failure. This method is useful when the model is fixed and reflects the real data, and it can be applied to any *in silico* model for various purposes, such as new drug development, toxicity identification, environmental effect, etc. This method will significantly reduce the time and effort to analyze the data.

KEYWORDS

parameter inference, electrophysiology, hERG, deep learning, cardiotoxicity

1 Introduction

It is a well-known fact that it is crucial to evaluate the effects of pharmaceuticals on heart rhythm because an unstable heart rhythm causes significant problems, including death. Additionally, cardiotoxicity has resulted in the withdrawal of some previously marketed drugs and restrictions on some clinically useful drugs (Lasser et al., 2002). Therefore, there have been many discussions on the mechanisms, prevention methods, and management of such toxicity by drugs (Kelleni and Abdelbasset, 2018). In particular, screening for the *human Ether-a`-go-go-Related Gene* (hERG) is critical. The hERG is a gene that forms part of the rapid delayed rectifier potassium current of the heart, I_{Kr}, and plays an important role in causing repolarization of the cardiac action potential. Many drugs that cause cardiotoxicity are known to block the hERG channel. Blockade by drugs leads to a decrease in I_{Kr}, which can prolong ventricular action potential (Jurkiewicz and Sanguinetti, 1993). This is also associated with an increase in the QT interval (QT) in the electrocardiogram (ECG) (Sanguinetti and Tristani-

FIGURE 1
An overview of our method. A total of 500,000 were generated to train hERGNet, and each data consisted of (I_{Kr}^{sim}, θ). The trained hERGNet predicts θ^* through I_{Kr}^{exp} obtained by a patch-clamp experiment. Parameter inference is performed using θ^* as the initial value.

Firouzi, 2006), which is likely related to Torsade de Pointes (Malik and Camm, 2001). Therefore, in 2005, the International Council for Harmonization included the following in its guidelines for non-clinical evaluation: "Preclinical Evaluation of the Possibility of Delayed Ventricular Repolarization (QT-Interval Prolongation) by Human Medicines (S7B)" (Food and Drug Administration, 2005; Friedrichs et al., 2005).

Advances in mathematical modeling and computational simulations of ion channels have made cell reactions and electrophysiological phenomena understandable and predictable, meaning they can help predict drug-induced changes. The mathematical modeling of hERG has also been continuously developed by (Zeng et al., 1995), (Beattie et al., 2018), and (ten Tusscher et al., 2004). These mathematical models are completed by fitting them to experimental data and finding the parameters. The parameters are important since they provide physiological and biophysical significance (Pathmanathan et al., 2015). However, the fitting process is by no means easy. To obtain more accurate model parameters, many mathematical and statistical methods, such as the least-squares optimization (Grisetti et al., 2020), the gradient descent (Ruder, 2016), and the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) (Hansen, 2006; Khan, 2018) have been proposed and studied. Furthermore, the development of parallel computing technology and hardware has significantly aided in problem-solving by further improving the performance of these methods (Khan, 2018). However, neither method is easy to completely avoid the local minima problem, and it is time-consuming depending on the data and model. Mathematically, the best way to solve these problems is to suggest initial values close to the true values. Initial values are usually sampled from a particular distribution associated with the characteristic of the problem or given based on past experience, but none are perfect solutions.

Deep learning-based artificial intelligence (AI) has recently made tremendous progress. Great achievements have been made not only in regression and classification problems but also in the creative field. For example, in the field of vision and image processing, convolutional neural network (CNN)-based models (O'Shea and Nash, 2015), such as ResNet (He et al., 2016), EfficientNet (Tan and Le, 2019; 2021), and RegNet (Radosavovic et al., 2020), have shown better performance than humans, and in the field of time-series like natural language processing, recurrent neural network (Sherstinsky, 2020), long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997; Sherstinsky, 2020), and transformer-based models (Vaswani et al., 2017; Lin et al., 2022) are showing remarkable results.

Medicine and biotechnology also have numerous images and time-series data. Therefore, various problems can be solved through deep learning-based AI, and several studies on deep learning-based analysis are already being conducted. (Alhusseini et al., 2020; Sevakula et al., 2020; Aghasafari et al., 2021; Jeong and Lim, 2021; Rogers et al., 2021).

In this paper, we introduce a method to predict the approximate value of the hERG ion channel model parameters using a neural network and improve the fitting operation using the predicted parameters. We confirmed the performance of improved parameter fitting using the experimental data released by https://github.com/CardiacModelling/hERGRapidCharacterisation in (Lei et al., 2019).

2 Methods

Our method is as follows. First, the simulation generates I_{Kr} current data. The generated current data was then converted
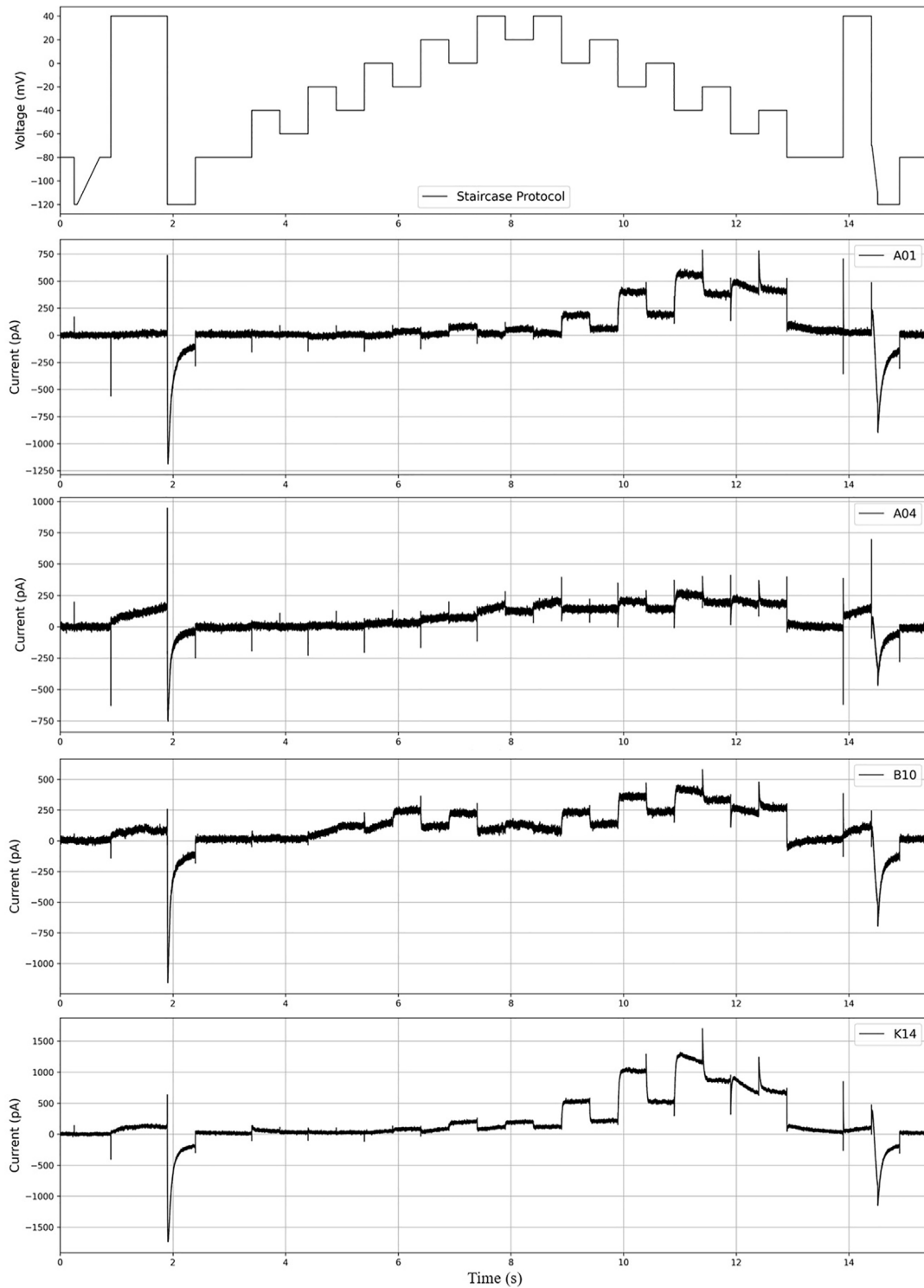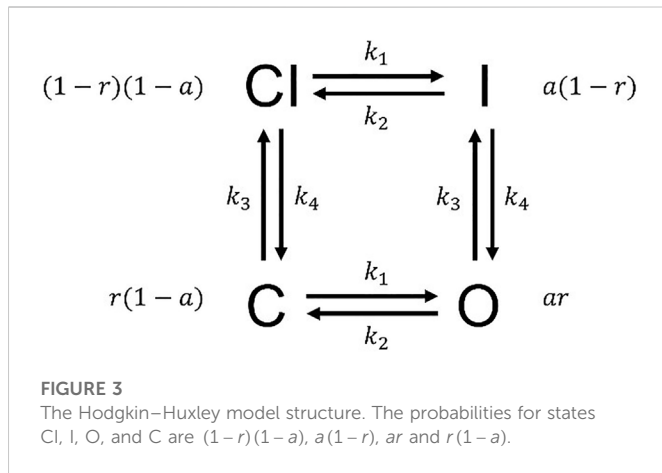
FIGURE 2
The figure above shows the "staircase protocol" introduced in (Lei et al., 2019) and the I_{Kr} currents for cells "A01," "A06," "B10," and "K14."

into a spectrogram. Next, our parametric prediction network was trained from the generated simulation data. The experimental data were used only for the validation and testing of the network. Using the trained prediction network, nine parameters were predicted based on the current generated by the patch-clamp experiment. Figure 1 depicts the overall overview.

2.1 hERG model

In this study, we used the experimental data published in (Lei et al., 2019, https://github.com/CardiacModelling/hERGRapidCharacterisation). Therefore, the hERG model and basic settings we use are the same as those of (Lei et al., 2019). For ease of training

FIGURE 3
The Hodgkin–Huxley model structure. The probabilities for states CI, I, O, and C are $(1-r)(1-a)$, $a(1-r)$, ar and $r(1-a)$.

and evaluation of our deep learning model, we excluded 11 of the 211 cells, which seemed to have a large difference between the experimental results and the results produced by the hERG model. The currents were recorded for the "staircase protocol" (Figure 2). As shown in the top image in Figure 2, each step is 500 ms, long enough to see the characteristics of I_{Kr}. Thus, it is possible to observe the dynamics at different voltage values. Lei et al. showed that their protocol provided enough information to infer true parameters through a synthetic data study (Lei et al., 2019).

This hERG model with Hodgkin and Huxley style structure used in this experiment is Beattie's model (Beattie et al., 2018) slightly improved by ten Tusscher et al. (ten Tusscher et al., 2004), and I_{Kr} is the same as Eq. (1).

$$I_{Kr} = g_{Kr}ar(V - E_K) \tag{1}$$

where g_{Kr} is the maximum conductance, and a and r denote a Hodgkin and Huxley activation gate and an inactivation gate, respectively. V is the transmembrane voltage. E_K is called Nernst potential or the reversal potential and obtained by Eq. (2).

$$E_K = \frac{RT}{zF}\ln\left(\frac{[K^+]_o}{[K^+]_i}\right) \tag{2}$$

where

R (ideal gas constant): $8.314472 \left[\dfrac{J}{K \cdot mol}\right]$

T (absolute temperature): 298.15 [K]

F (Faraday's constant): $96485.3415 \left[\dfrac{C}{mol}\right]$

z (valency of the ions): $+1$ for K^+

$[K^+]_o$ (extracellular concentration): 4 [mM]

$[K^+]_i$ (intracellular concentration): 110 [mM]

The model has nine parameters $\theta = \{g_{Kr}, p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8\}$, and Figure 3 shows its structure, where

$$\frac{da}{dt} = \frac{a_\infty - a}{\tau_a} \quad, \frac{dr}{dt} = \frac{r_\infty - r}{\tau_r}$$

$$a_\infty = \frac{k_1}{k_1 + k_2} \quad, r_\infty = \frac{k_4}{k_3 + k_4}$$

$$\tau_a = \frac{1}{k_1 + k_2} \quad, \tau_r = \frac{1}{k_3 + k_4}$$

$$k_1 = p_1 \exp(p_2 V)$$
$$k_2 = p_3 \exp(-p_4 V)$$
$$k_3 = p_5 \exp(p_6 V)$$
$$k_4 = p_7 \exp(-p_8 V)$$

where k_1 is an activation rate, k_2 is a deactivation rate, k_3 is an inactivation rate, and k_4 is a recovery rate.

FIGURE 4
(A) Distribution of the nine parameters generated by the two conditions of Eqs 3, 4. The distributions of p_1; p_3; p_5; p_7 are very clustered in a specific range.
(B) Data distributions after applying log scale to p_1, p_3, p_5 and p_7.

FIGURE 5
The structure of hERGNet. The spectrogram of current is converted into 2D data with three channels like the image shape through the CNN encoder, making it possible to utilize the pretrained EfficientNet.

2.2 Dataset generation

A lot of data is required to train. The best condition for good performance is the presence of a large amount of experimental data. However, there are only 200, and it is very insufficient for learning and testing with them. Even if a large amount of data exists, fitting work for labeling requires a lot of time and resources, which is contrary to the purpose of our study. So, we generated a large amount of data using simulations of the hERG channel to compensate for the lack of data. The dataset consists of I_{Kr} as input data and nine parameters θ of the hERG model as the target data. Five hundred thousand data generated by simulation were used for training. For 200 experimental data, after sorting by name, odd-numbered data were configured as validation dataset and even-numbered data were configured as test dataset.

First, parameters for the hERG channel were generated under the following two conditions. The first condition is that each parameter follows a uniform distribution within a specific range, as in Eq. (3).

$$g_{Kr} \sim U\left(g_{min}, g_{max}\right)$$
$$p_1, p_2, p_3, p_4 \sim U\left(a, b\right) \qquad (3)$$
$$p_5, p_6, p_7, p_8 \sim U\left(c, d\right)$$

where $U\left(\cdot\right)$ represents a uniform distribution. In this study, for the conductance g_{Kr}, $g_{min} = 100$ [pA/V] and $g_{min} = 500000$ [pA/V]. For $p_1, p_3, p_5,$ and p_7, a and b are 0.0001 [s^{-1}] and 10^6 [s^{-1}], respectively. For $p_2, p_4, p_6,$ and p_8, c and d are 0.0001 [V^{-1}] and 400 [V^{-1}], respectively. The second condition is that each parameter must satisfy the inequalities of Eq. (4):

$$0.0167 < p_1 \exp\left(p_2 {}^* V_{max}\right) < 10^6$$
$$0.0167 < p_3 \exp\left(-p_4 {}^* V_{min}\right) < 10^6 \qquad (4)$$
$$0.0167 < p_5 \exp\left(p_6 {}^* V_{max}\right) < 10^6$$
$$0.0167 < p_7 \exp\left(-p_8 {}^* V_{min}\right) < 10^6$$

where $V_{min} = -0.12$ and $V_{max} = 0.06$. The lower and upper bounds of the above conditions were determined by the constraints of physical and physiological phenomena (Beattie et al., 2018). Figure 4A shows the distribution of each parameter for 100,000 data generated by the above conditions.

Second, we used simulations to generate the current I_{Kr} corresponding to the parameters sampled above. Myokit (Clerx et al., 2016) with CVODE solver (Hindmarsh et al., 2005) was used for the simulation. The tolerance settings for CVODE were $abs\_tol = 10^{-8}$ and $rel\_tol = 10^{-10}$, as in the condition in (Lei et al., 2019). The length of the experimental data was 15.4 s with a sampling rate of 5 kHz. In this study, we reduced the data number to 1/50 with sampling one point every 100 because there seemed to be no problem reflecting the trend. Because we only used the data generated by simulations for training, the results of this study depend on the similarity between the simulation data and experimental data. Therefore, noises, α, were added to the simulation data. The noises were extracted from a normal distribution $\alpha \sim N\left(0, \sigma^2\right)$. σ is 10.84, which was measured at the steady-state current in the experimental data.

$$I_{Kr}^{exp} \approx I_{Kr}^{sim} + \alpha \qquad (5)$$

where I_{Kr}^{exp} and I_{Kr}^{sim} represent the experimental and simulated current, respectively.

2.3 Preprocessing and hERGNet

Our method involves several simple preprocessing processes on the data for learning. In Figure 4A, The distributions of p_1, p_3, p_5 and p_7 are very clustered in a specific range. To make them as uniform as possible, the log scale was applied. The min–max normalization was then used to transform all parameter ranges between 0 and 1, as shown in Figure 4B.

Recurrent neural network (Sherstinsky, 2020) and LSTM (Hochreiter and Schmidhuber, 1997; Sherstinsky, 2020) series models have been widely used to analyze time-series data, such as the current data that we want to analyze, and recently, transformer-based models (Vaswani et al., 2017; Lin et al., 2022) are leading this field. Since CNN is designed for the purpose of extracting information between adjacent values of data, it obtains spatial information well in the local domain (Krizhevsky et al., 2017). Transformer calculates the relationship between all elements of input data through attention, so it understands overall features better than CNN, but is weaker than CNN in extracting local information, and requires a very large size dataset for this purpose (Dosovitskiy et al., 2020).

FIGURE 6
Spectrograms transformed from current data for 4 cells.
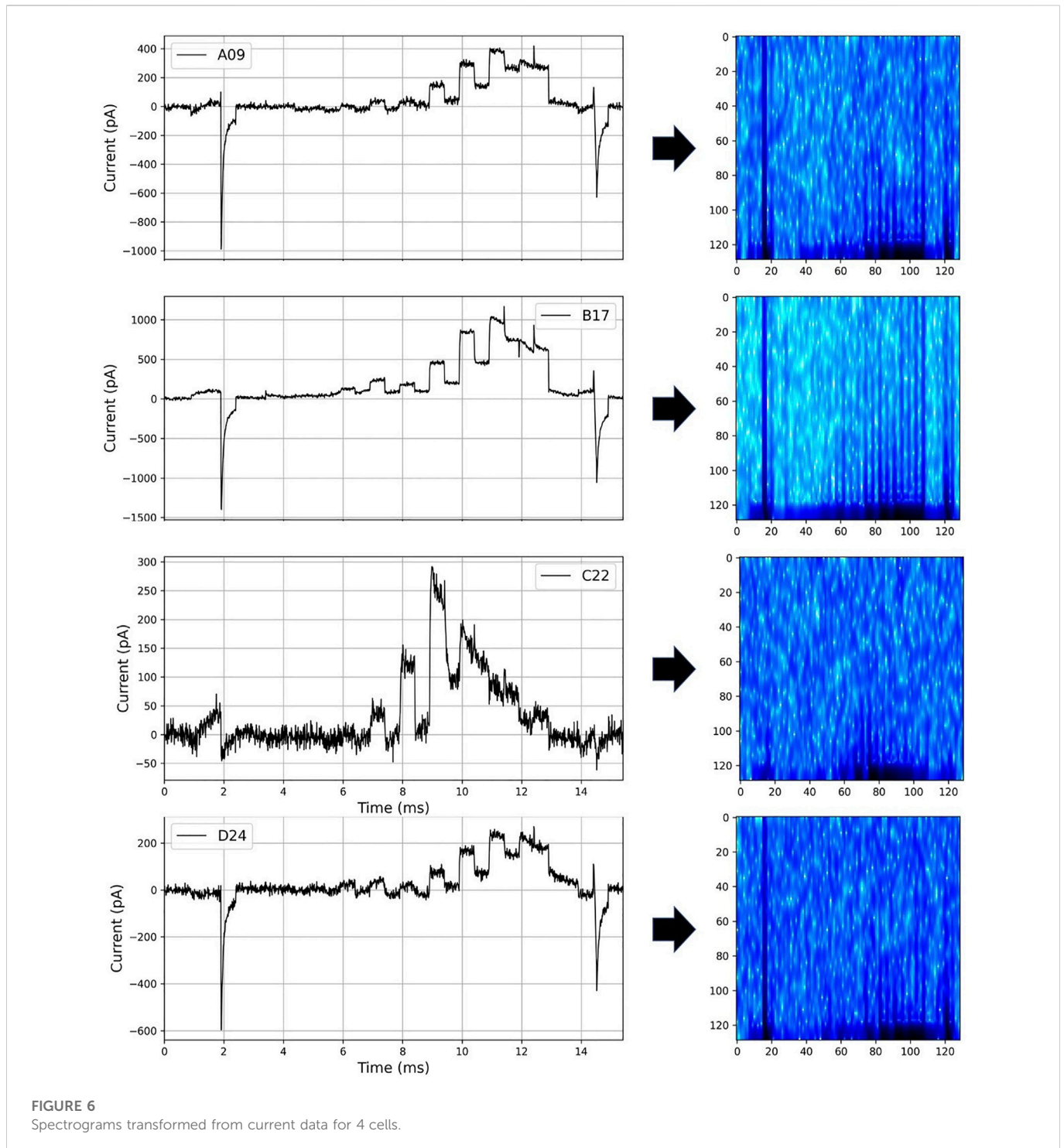
We thought our goal was closer to finding changes in the characteristics and patterns at specific times than predicting the current change over time. Therefore, we adopted a CNN-based model rather than a transformer-based model, and among them, an EfficientNet (Tan and Le, 2019) type model. We called our model hERGNet. Our hERGNet is very simple, as shown in Figure 5. It consists of a CNN-encoder network and a pretrained EfficientNetV2-M (Tan and Le, 2021). The CNN-encoder network extracts the features of the spectrogram and increases the number of channels to three to obtain an image-like shape, thereby making it possible to use the pretrained EfficientNetV2-M.

EfficientNetV2-M is responsible for finding parameters by extracting features from encoded data.

We converted the current into a spectrogram with a frequency perspective, as shown in Figure 6, so that hERGNet can better learn the characteristics of the current data. By adding frequency features to current data consisting of only time and intensity, it has the advantage of increasing information about data and transforming it into a two-dimensional form like an image, making it easier for a 2D CNN-based model to learn. In this study, the parameters "n_fft," "hop_length," and "win_length" of Short-Time Fourier Transform (Owens and
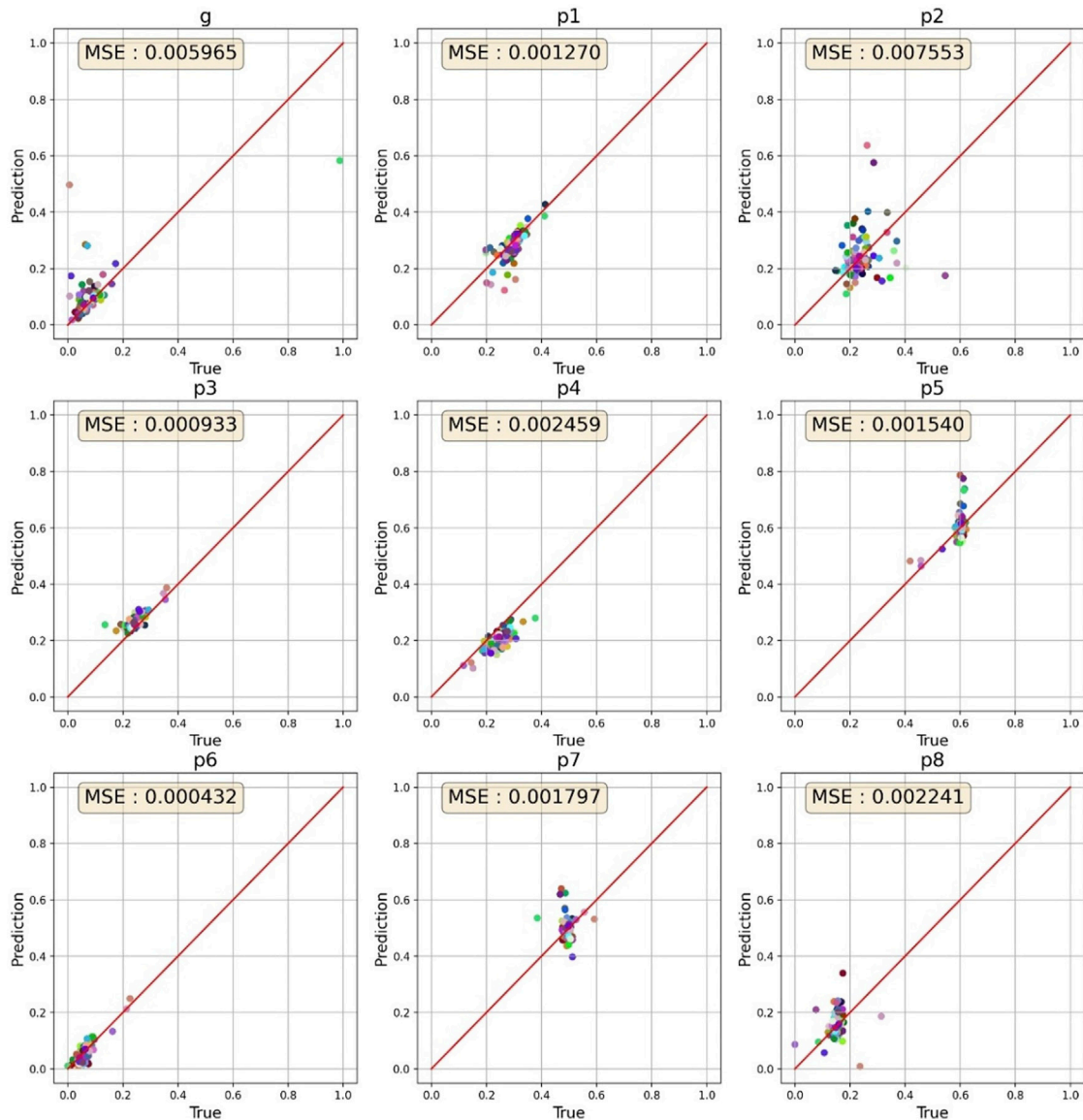
FIGURE 7
Prediction results for 100 cells. The parameters for most cells are clustered in a specific range. p_6 had the best predictions, while p_2 had the worst predictions.

Murphy, 1988) for spectrogram transformation were set to 256, 12, and 48, respectively.

2.4 Training

Mean Squared Error (MSE) was used as the loss function, so the cost function J(θ) is the same as Eq. (6)

$$J(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left(f_\theta(x_i) - y_i \right) \tag{6}$$

where x_i is the spectrogram of the current I_{Kr}, f_θ is the neural network, and y_i is the true parameters of the ion channel.

In general, it is known that the higher the resolution of the input, the higher the performance. However, the higher the resolution, the more memory is required, which is time-consuming to train due to the small batch size. Also, if you train all the generated 500,000 data from the beginning, the time will increase even more. Therefore, we first trained the hERGNet with a small resolution and a small number of data and then proceeded with transfer learning by increasing the resolution and the number of data. First, learning was performed on 300,000 spectrograms with a 97 × 97 resolution to 200 epochs. Then, we increased the resolution to 129 × 129 and performed transfer learning to 140 epochs using all 500,000 data.
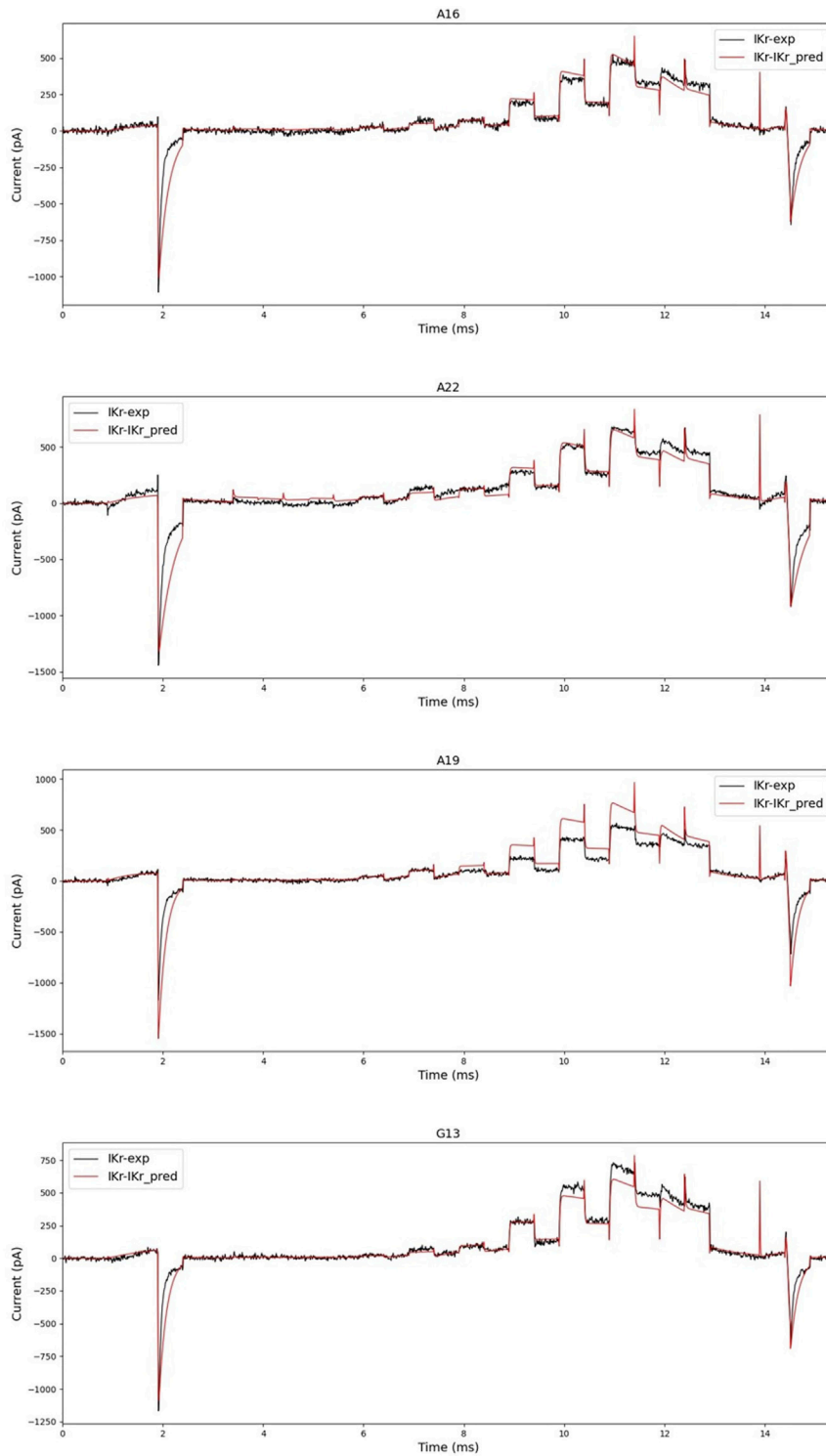
FIGURE 8
Black is experimental data, and red is simulated current data with parameters predicted by hERGNet. There are some very close predictions, such as A16 and A22, and results showing differences, such as A19 and G13. However, the flow and shape of the current are somewhat predictable.

2.5 Parameter inference

All fitting operations were the same as those in (Lei et al., 2019), and their open-source library PINTS (Clerx et al., 2019) was used. Furthermore, the CMA-ES algorithm (Hansen, 2006; Khan, 2018) was used as a global optimization algorithm to fit the model to the experimental data, and Markov Chain Monte Carlo (Jasra et al., 2007) with adaptive Metropolis (Haario et al., 2001) was used to explore the posterior probability distribution. Parameter inference was performed three times based on the initial value for each cell. The first was when the

TABLE 1 When the predicted value by hERGNet was used, there was a great improvement in the fitting speed.

	Predicted	Prior	Random
Average Iteration	341.8	546.4	601.3
Average Time (s)	396.3	630.9	686.0

initial value was given as a parameter predicted by our hERGNet, the second was a prior value, and the last was given as a value randomly extracted from the previously described parameter distribution.

3 Results

When hERGNet was trained on 300,000 data with a 97 × 97 resolution until 200 epochs, the best MSE for 100 experimental data (validation data) was 0.00781. After 140 additional learnings on 500,000 data with an increased resolution to 129 × 129, we identified MSEs lowered to 0.003035. Then, we tested hERGNet on 100 experimental data not used for learning. The MSE was recorded as 0.002688, which is better than the results for the validation dataset. Figure 7 shows the prediction results for 100 experimental data. The fitting test was performed on 50 experimental data out of the 100 test data.

First, when we generated current data with the predicted parameters, we compared how different it was from the experimental data as shown in Figure 8. As shown in the two figures above in Figure 8, the prediction was accurate enough that no fitting work was required for a significant number of cells. However, as shown in the two figures below in Figure 8, the prediction was not perfect, necessitating a fitting operation.

Next, as shown in Table 1, we compared the results when the initial values were given as parameters predicted by hERGNet, prior parameters, and random parameters. The results confirmed that our method significantly improved the fitting operation. The initial value of hERGNet did not result in a single failure in the fitting operation for 50 cells. However, the prior parameter caused one local minima problem, and in the random parameter, 16 failures occurred out of 50 fittings, and 11 local minima problems occurred. We compared the fitting rates for cells that succeeded in parameter inference. As shown in Table 1, the average iteration was 341.8 in the predicted parameter, 546.4 in the prior parameter, and 601.3 in the random parameter. The average time was 396.3 s for the predicted parameters, 630.9 s for the prior parameter, and 686.0 s for the random parameters. Of the 50 cells, all but two, D17 and G13, showed faster fitting rates when using the predicted parameters.

4 Discussion and conclusion

Parameter inference is an important part of the toxicity evaluation of drugs because it is possible to understand and predict physiological changes in cells caused by drugs by predicting the parameters of ion channels. However, the difficulty of the fitting and the time-consuming problem make us hesitant to use *in silico*. In this study, we propose a method for improving the fitting operation for the hERG channel model by setting the parameters predicted by hERGNet as initial values. The test results showed a clear improvement in the hERG model fitting. There was

no fitting failure, and the time-consuming problem was also improved. Depending on the range of parameters, training the neural network required a lot of data generation and was time-consuming. However, if experiments are conducted with the same voltage protocol for other cells in the future, our method could be very useful for inferring the parameters of ion channels.

Our method still has a lot to improve. The first is to improve the fitting method rather than simply presenting initial values. This is because stochastic methods, such as CMA-ES, may not immediately find optimal parameters due to the characteristics of the method, even if parameters close to the correct answer are presented. The second is to increase the similarity between experimental and simulation data, which is, after all, the most important factor for AI to predict parameters. We trained our hERGNet only with simulation data. If the similarity between experimental and simulation data can be increased through noise removal, etc., the predicted parameters will be closer to the correct answer.

In fact, our ultimate goal is to predict parameters that are very close to the correct answer, eliminating the need for model fitting. If this is possible, a new paradigm will be presented in drug development or drug toxicity assessment. To this end, we will first conduct a study on parameter prediction in multiple ion channels. Parametric prediction for multiple ion channels may aid in greatly reducing the amount and cost of experiments performed in the non-clinical stage.

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Author contributions

Design and conceptualization of the study: JS and CL; Data analysis: JS, YK; Software development: JS; Writing and editing: JS and CL; All authors contributed to the revision of the manuscript, read, and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aghasafari, P., Yang, P. C., Kernik, D. C., Sakamoto, K., Kanda, Y., Kurokawa, J., et al. (2021). A deep learning algorithm to translate and classify cardiac electrophysiology. *Elife* 10, e68335. doi:10.7554/ELIFE.68335

Alhusseini, M. I., Abuzaid, F., Rogers, A. J., Zaman, J. A. B., Baykaner, T., Clopton, P., et al. (2020). Machine learning to classify intracardiac electrical patterns during atrial fibrillation: Machine learning of atrial fibrillation. *Circ. Arrhythm. Electrophysiol.* 13, e008160. doi:10.1161/CIRCEP.119.008160

Beattie, K. A., Hill, A. P., Bardenet, R., Cui, Y., Vandenberg, J. I., Gavaghan, D. J., et al. (2018). Sinusoidal voltage protocols for rapid characterisation of ion channel kinetics. *J. Physiol.* 596, 1813–1828. doi:10.1113/JP275733

Clerx, M., Collins, P., de Lange, E., and Volders, P. G. A. (2016). Myokit: A simple interface to cardiac cellular electrophysiology. *Prog. Biophys. Mol. Biol.* 120, 100–114. doi:10.1016/j.pbiomolbio.2015.12.008

Clerx, M., Robinson, M., Lambert, B., Lei, C. L., Ghosh, S., Mirams, G. R., et al. (2019). Probabilistic inference on noisy time series (PINTS). *J. Open Res. Softw.* 7, 23. doi:10.5334/jors.252

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. Available At https://arxiv.org/abs/2010.11929.doi:10.48550/arXiv.2010.11929

Food and Drug Administration (2005). International conference on harmonisation; guidance on S7B nonclinical evaluation of the potential for delayed ventricular repolarization (QT interval prolongation) by human pharmaceuticals; availability. Notice. *Fed. Regist.* 70, 61133–61134.

Friedrichs, G. S., Patmore, L., and Bass, A. (2005). Non-clinical evaluation of ventricular repolarization (ICH S7B): Results of an interim survey of international pharmaceutical companies. *J. Pharmacol. Toxicol. Methods* 52, 6–11. doi:10.1016/j.vascn.2005.05.001

Grisetti, G., Guadagnino, T., Aloise, I., Colosi, M., della Corte, B., and Schlegel, D. (2020). Least squares optimization: From theory to practice. *Robotics* 9, 51. doi:10.3390/robotics9030051

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* 7, 223–242. doi:10.2307/3318737

Hansen, N. (2006). "The cma evolution strategy: A comparing review," in *Towards a new evolutionary computation* (Berlin, Heidelberg: Springer Berlin Heidelberg), 75–102. doi:10.1007/3-540-32494-1_4

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016 (IEEE), 770–778. doi:10.1109/CVPR.2016.90

Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E., et al. (2005). Sundials. *ACM Trans. Math. Softw.* 31, 363–396. doi:10.1145/1089014.1089020

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural comput.* 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735

Jasra, A., Stephens, D. A., and Holmes, C. C. (2007). On population-based simulation for static inference. *Stat. Comput.* 17, 263–279. doi:10.1007/s11222-007-9028-9

Jeong, D. U., and Lim, K. M. (2021). Convolutional neural network for classification of eight types of arrhythmia using 2D time–frequency feature map from standard 12-lead electrocardiogram. *Sci. Rep.* 11, 20396. doi:10.1038/s41598-021-99975-6

Jurkiewicz, N. K., and Sanguinetti, M. C. (1993). Rate-dependent prolongation of cardiac action potentials by a methanesulfonanilide class III antiarrhythmic agent. Specific block of rapidly activating delayed rectifier K$^+$ current by dofetilide. *Circ. Res.* 72, 75–83. doi:10.1161/01.RES.72.1.75

Kelleni, M. T., and Abdelbasset, M. (2018). "Drug induced cardiotoxicity: Mechanism, prevention and management," in *Cardiotoxicity* (London, UK: InTech). doi:10.5772/intechopen.79611

Khan, N. (2018). A parallel implementation of the covariance matrix adaptation evolution strategy. Available At https://arxiv.org/abs/1805.11201.doi:10.48550/arXiv.1805.11201

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi:10.1145/3065386

Lasser, K. E., Allen, P. D., Woolhandler, S. J., Himmelstein, D. U., Wolfe, S. M., and Bor, D. H. (2002). Timing of new Black box warnings and withdrawals for prescription medications. *JAMA* 287, 2215–2220. doi:10.1001/jama.287.17.2215

Lei, C. L., Clerx, M., Gavaghan, D. J., Polonchuk, L., Mirams, G. R., and Wang, K. (2019). Rapid characterization of hERG channel kinetics I: Using an automated high-throughput system. *Biophys. J.* 117, 2438–2454. doi:10.1016/j.bpj.2019.07.029

Lin, T., Wang, Y., Liu, X., and Qiu, X. (2022). A survey of transformers. *AI Open* 3, 111–132. doi:10.1016/j.aiopen.2022.10.001

Malik, M., and Camm, A. J. (2001). Evaluation of drug-induced QT interval prolongation: Implications for drug approval and labelling. *Drug Saf.* 24, 323–351. doi:10.2165/00002018-200124050-00001

O'Shea, K., and Nash, R. (2015). An introduction to convolutional neural networks. Available At https://arxiv.org/abs/1511.08458.doi:10.48550/arXiv.1511.08458

Owens, F. J., and Murphy, M. S. (1988). A short-time Fourier transform. *Signal Process.* 14, 3–10. doi:10.1016/0165-1684(88)90040-0

Pathmanathan, P., Shotwell, M. S., Gavaghan, D. J., Cordeiro, J. M., and Gray, R. A. (2015). Uncertainty quantification of fast sodium current steady-state inactivation for multi-scale models of cardiac electrophysiology. *Prog. Biophys. Mol. Biol.* 117, 4–18. doi:10.1016/j.pbiomolbio.2015.01.008

Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., and Dollár, P. (2020). "Designing network design spaces," in Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, June 2020, 10428–10436. doi:10.48550/arXiv.2003.13678

Rogers, A. J., Selvalingam, A., Alhusseini, M. I., Krummen, D. E., Corrado, C., Abuzaid, F., et al. (2021). Machine learned cellular phenotypes in cardiomyopathy predict sudden death. *Circ. Res.* 128, 172–184. doi:10.1161/CIRCRESAHA.120.317345

Ruder, S. (2016). An overview of gradient descent optimization algorithms. Available At https://arxiv.org/abs/1609.04747.doi:10.48550/arXiv.1609.04747

Sanguinetti, M. C., and Tristani-Firouzi, M. (2006). hERG potassium channels and cardiac arrhythmia. *Nature* 440, 463–469. doi:10.1038/nature04710

Sevakula, R. K., Au-Yeung, W. M., Singh, J. P., Heist, E. K., Isselbacher, E. M., and Armoundas, A. A. (2020). State-of-the-Art machine learning techniques aiming to improve patient outcomes pertaining to the cardiovascular system. *J. Am. Heart Assoc.* 9, e013924. doi:10.1161/JAHA.119.013924

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D.* 404, 132306. doi:10.1016/j.physd.2019.132306

Tan, M., and Le, Q. v. (2019). "EfficientNet: Rethinking model scaling for convolutional neural networks," in International conference on machine learning (PMLR), 6105–6114. doi:10.48550/arXiv.1905.11946

Tan, M., and Le, Q. v. (2021). EfficientNetV2: Smaller models and faster training. *Proc. Mach. Learn. Res. (PMLR)* 139, 10096–10106. doi:10.48550/arXiv.2104.00298

ten Tusscher, K. H. W. J., Noble, D., Noble, P. J., and Panfilov, A. v. (2004). A model for human ventricular tissue. *Am. J. Physiology-Heart Circulatory Physiology* 286, H1573–H1589. doi:10.1152/ajpheart.00794.2003

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., et al. (2017). "Attention is all you need," in Proceedings of the Neural Information Processing Systems (NIPS), Long Beach, CA, USA, December 2017. doi:10.48550/arXiv.1706.03762

Zeng, J., Laurita, K. R., Rosenbaum, D. S., and Rudy, Y. (1995). Two components of the delayed rectifier K$^+$ current in ventricular myocytes of the Guinea pig type. *Circ. Res.* 77, 140–152. doi:10.1161/01.RES.77.1.140