

DeepMiceTL: a deep transfer learning based prediction of mice cardiac conduction diseases using early electrocardiograms

Ying Liao, Yisha Xiang, Mingjie Zheng and Jun Wang

Corresponding author: Yisha Xiang, Department of Industrial Engineering, University of Houston, Houston, Texas 77004, USA. Tel: 713-743-4180;

E-mail: yxiang4@Central.UH.EDU

Abstract

Cardiac conduction disease is a major cause of morbidity and mortality worldwide. There is considerable clinical significance and an emerging need of early detection of these diseases for preventive treatment success before more severe arrhythmias occur. However, developing such early screening tools is challenging due to the lack of early electrocardiograms (ECGs) before symptoms occur in patients. Mouse models are widely used in cardiac arrhythmia research. The goal of this paper is to develop deep learning models to predict cardiac conduction diseases in mice using their early ECGs. We hypothesize that mutant mice present subtle abnormalities in their early ECGs before severe arrhythmias present. These subtle patterns can be detected by deep learning though they are hard to be identified by human eyes. We propose a deep transfer learning model, DeepMiceTL, which leverages knowledge from human ECGs to learn mouse ECG patterns. We further apply the Bayesian optimization and k -fold cross validation methods to tune the hyperparameters of the DeepMiceTL. Our results show that DeepMiceTL achieves a promising performance (F1-score: 83.8%, accuracy: 84.8%) in predicting the occurrence of cardiac conduction diseases using early mouse ECGs. This study is among the first efforts that use state-of-the-art deep transfer learning to identify ECG patterns during the early course of cardiac conduction disease in mice. Our approach not only could help in cardiac conduction disease research in mice, but also suggest a feasibility for early clinical diagnosis of human cardiac conduction diseases and other types of cardiac arrhythmias using deep transfer learning in the future.

Keywords: cardiac conduction disease, electrocardiogram (ECG), mouse model, deep transfer learning, Bayesian optimization

INTRODUCTION

The cardiac conduction system (CCS) initiates and propagates electrical depolarization, which coordinates the synchronized contraction of cardiac muscle. Abnormalities in cardiac conduction cause various types of arrhythmias (e.g. sick sinus syndrome and atrioventricular block), which are the major causes of morbidity and mortality worldwide [1, 2]. There is considerable clinical significance and an emerging need of early detection of cardiac conduction diseases for preventive treatment success before more severe arrhythmias and symptoms occur. Recorded electrocardiograms (ECGs) can reflect the cardiac rhythm and electrical activity, which help in the diagnosis of cardiac conduction diseases. However, the early prediction of cardiac conduction diseases in humans can be particularly challenging because of the complexity of such diseases and the lack of early ECGs collected before the occurrence of symptoms in patients. Taking inherited cardiac conduction diseases as an example, even with disease family history, it is impractical for everyone to conduct genetic testing and frequent ECG screening. Cardiac conduction diseases often have no early symptoms and

patients with potential diseases commonly do not visit physicians for ECG until symptoms appear. Importantly, in addition to genetic defects, different factors such as injury, environment, aging and many other factors can also cause cardiac conduction diseases, making the diagnosis and treatment more challenging. For example, other than various diseases and medications that can cause sinus node dysfunction (also known as sick sinus syndrome, SSS), the incidence of SSS also significantly increases with aging. SSS are frequently the diseases of the elderly and the incidence of SSS is projected to double over the next 50 years [3]. SSS causes various arrhythmias such as sinus bradycardia, sinus arrest/pause and sino-atrial exit block [4]. Notably, patients with SSS can experience very mild symptoms all the way to severe cardiac events and even sudden death, while it is common that SSS patients are asymptomatic with normal ECGs in the early disease course [5, 6]. To advance the understanding of the pathogenesis, progression and mechanisms underlying cardiac arrhythmias such as cardiac conduction diseases, animal models (e.g. mouse model) are widely used in cardiac arrhythmia research [7–11]. There is an urgent need to develop models for predicting

Ying Liao is a Ph.D. candidate in the Department of Industrial, Manufacturing & Systems Engineering, Texas Tech University, Lubbock, Texas, USA. Her research interests include statistical machine learning with applications in healthcare and manufacturing.

Yisha Xiang is an associate professor in the Department of Industrial Engineering, University of Houston, Houston, Texas, USA. Her research interests include decision-making under uncertainty and statistical machine learning with applications in healthcare and manufacturing.

Mingjie Zheng is a postdoctoral researcher in the Department of Pediatrics, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, Texas, USA. His research interests include cardiac arrhythmias.

Jun Wang is an associate professor in the Department of Pediatrics, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, Texas, USA. Her research interests include disease biology.

Received: November 29, 2022. **Revised:** February 10, 2023. **Accepted:** March 1, 2023

© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

cardiac conduction disease tendencies in mice. However, few cardiac conduction disease predictive models for mice are readily available.

In this paper, we develop a deep transfer learning (TL) model to detect the subtle signature of future cardiac conduction diseases using early mouse ECGs. Our work is motivated by two streams of the literature. First, the utility of deep learning models in ECG analysis has been widely investigated [12–14]. For example, Hannun et al. [12] developed an end-to-end deep neural network to classify 12 types of ECG rhythms. The end-to-end approach takes raw ECGs as input and learns optimal features directly from data without any human guidance. Their results show that the proposed network achieves an average F1-score of 83.7%, exceeding the average diagnostic performance of cardiologists (78%). Attia et al. [13] proposed a deep convolutional neural network (CNN) to predict the atrial fibrillation (AF) using normal sinus rhythm ECGs. Their hypothesis is that the CNN can identify subtle patterns in normal ECGs that are due to structural changes associated with a history of (or impending) AF. Their results are promising—identifying AF patients with an overall accuracy of 79.4%.

Second, machine learning methods have been used in animal models. For example, machine learning methods are used to identify the genetic mutation in mice in [15, 16]. Both works consider the problem of detecting a *Scn5a*^{+/-} mutation in mice from mouse ECG signals. The *Scn5a*^{+/-} mutation has been implicated in the hereditary Brugada syndrome, which is associated with 4–12% of clinically reported sudden cardiac deaths [15]. Bonet-Luz et al. [15] extracted various groups of features, including ECG intervals and amplitudes and features derived from the symmetric projection attractor reconstruction (SPAR) method, and used the conventional machine learning algorithms (i.e. *k*-nearest neighbors and support vector machine) to distinguish the *Scn5a*^{+/-} mutant mice from the wild-type ones. Aston et al. [16] combined deep TL technique with the SPAR method to identify the *Scn5a*^{+/-} mutation in mice. They applied the SPAR method to generate attractor images from raw ECGs and then used these images as input in the networks. Their work includes a small number of mice, which are not sufficient to train deep models as these models often require large volume of training data to achieve the desirable performance. To address the challenge of insufficient mouse data, they used deep CNNs (e.g. AlexNet) pretrained on the ImageNet database [17] to provide initial weights for training the generated attractor images. In addition, recent other works on ECG analysis have focused on the end-to-end deep TL and pretrain a CNN directly on raw ECG data [18, 19]. Particularly, Steenkiste et al. [18] used deep TL to borrow knowledge from human ECGs for heartbeat classification in horses.

The satisfactory performance reported in the aforementioned papers makes it promising to investigate whether deep learning methods can detect subtle abnormalities in early ECGs of mice with cardiac conduction diseases. In this study, we propose a deep TL model, DeepMiceTL, to predict cardiac conduction diseases using early mouse ECGs. The CCS-injured *Hcn4*^{CreERT2}-DTA mouse model used in our study can induce the progression of various types of cardiac arrhythmias, including the SSS and atrioventricular blocks (AV blocks), and can ultimately cause cardiac arrest and sudden death. We first pretrain a CNN for AF classification on a publicly available human ECG dataset, i.e. PhysioNet/CinC challenge 2017 dataset [20, 21]. The pretrained CNN is then finetuned on the mouse ECG data to predict the occurrence of future cardiac conduction diseases. Figure 1 illustrates the data scarcity issue in this study. The knowledge learned

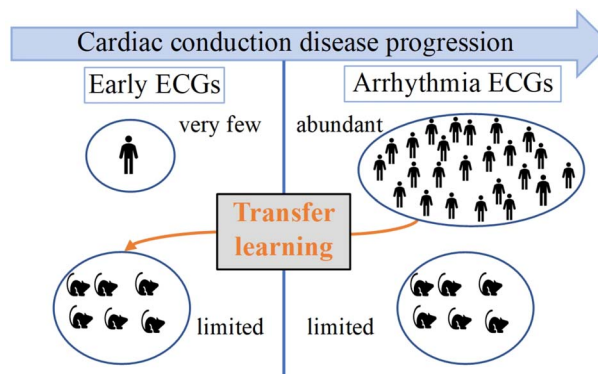


Figure 1. Illustration of the data scarcity issue in this study. The knowledge learned from the classification of human arrhythmias is transferred to the prediction of cardiac conduction disease tendencies in mice.

from the classification of human arrhythmias is transferred to the prediction of cardiac conduction disease tendencies in mice. To further improve the prediction performance, we use the Bayesian optimization (BO) and *k*-fold cross validation methods to determine the hyperparameters used in CNNs. Our results show that the deep TL approach that leverages knowledge from human ECG data achieves significantly better prediction performances on the target task compared with the method of training the CNN from scratch on the mouse ECGs. The proposed DeepMiceTL achieves a promising performance (F1-score: 83.8%, accuracy: 84.8%) in predicting cardiac conduction diseases in mice. Our findings establish a successful early-ECGs-based cardiac conduction disease prediction using the deep TL models, which also sheds light on the use of deep learning methods in early cardiac conduction disease prediction in humans.

MATERIALS AND METHODS

Mouse ECG dataset

The mouse ECG data (from Wang lab unpublished work) were collected from the mouse experiments that were designed to establish a CCS injury model at Wang's lab at McGovern Medical School, the University of Texas Health Science Center at Houston (UTHealth). All animal studies and procedures were performed in accordance with the approval by the UTHealth Animal Care and Use Committee and under the National Institutes of Health's Guide for the Care and Use of Laboratory Animals. Mice were housed under the standard 12-h light/dark cycle with water and chow. DTA (Diphtheria toxin fragment A) [22] is a pathogen toxin that efficiently kills the cell once introduced into a cell [23]. To ablate CCS cells in mice, a Cre-inducible DTA line was used [24]. Upon a Cre-mediated excision of a floxed stop sequence, the ROSA-eGFP-DTA transgenic line efficiently expresses DTA. The ROSA-eGFP-DTA transgenic mice [25] were bred with mice expressing *Hcn4*^{CreERT2}, a tamoxifen-inducible CCS-specific Cre recombinase [26]. Both male and female mice at 8 weeks of age were intraperitoneally injected with tamoxifen (10 mg/ml) 150 μ l/day for 2 days to induce Cre activity. After being anesthetized with 2–3% isoflurane in 100% oxygen, mice were secured in a supine position on a regulated heat pad while surface ECG was continuously obtained using a Rodent Surgical Monitor (Indus Instruments, Webster, TX). Lead-II ECG signals were daily collected from the mice using LabChart Software (ADInstruments) with sampling frequency 2000 Hz, which is the most commonly used lead and generally provides the best view of heart electrical activities. In

this experiment, $Hcn4^{CreERT2}$ -DTA mice, in which the CCS was injured after tamoxifen injection, are referred to as the mutant group. A total of 20 mice (10 control and 10 mutant mice) were included in this mouse experiment.

Human ECG dataset

Due to insufficient data in the mouse ECG dataset, we use the TL technique to transfer knowledge from the openly accessible human ECG data for our target prediction task. The PhysioNet/CinC challenge 2017 dataset [20, 21], which is collected to encourage the development of methods to classify AF from short ECG recordings, is considered in this study. Only the training set is included because the hidden test set still remains inaccessible to the public. The training set consists of 8528 labeled single-lead (lead-I) short ECG recordings (9 to 61 s), each of which is sampled at 300 Hz and belongs to one of the following classes: AF, Normal, Other or Noise. Each recording was taken from an individual participant.

Workflow

Our method contains four main steps (shown in Figure 2): (1) data collection and preprocessing, (2) pretraining a CNN on human ECG data (DeepHuman), (3) finetuning a CNN with TL (DeepMiceTL) on mouse ECG data for the target task and (4) training a CNN on the same mouse dataset without TL (DeepMice). The target task is to predict the occurrence of future cardiac conduction diseases (i.e. binary classification) using early ECGs of mice. In the fourth step, DeepMice is trained to investigate the impacts of TL on prediction performances of the target task. In addition, several important hyperparameters that are used in the CNNs, including the ones in the network architecture (e.g. number and size of the filters) and the ones in the training algorithm (e.g. learning rate), are tuned by BO and k -fold cross-validation methods. In the following subsections, we will describe in detail about data processing, network architecture and training algorithm, hyperparameters tuning and TL that are used to develop the proposed CNNs. We will

also introduce the experiment design and the model evaluation metrics.

Data preprocessing

In the mouse experiment, all mice in the mutant group were found to present cardiac conduction diseases on Day 6, and died suddenly between 7 and 15 days with severe cardiac arrhythmias. Figure 3 compares the ECG changes from Day 1 to Day 7 between two random pairs of mutant and control mice. It is challenging, if possible at all, to identify any subtle abnormalities in early ECGs collected before cardiac conduction diseases occur through human eyes, but deep learning has augmented diagnostic capability. In this study, we hypothesize that deep learning models can detect such subtle patterns which carry the signature of future cardiac conduction diseases. To investigate this hypothesis, we design a binary classification experiment to predict the occurrence of future cardiac conduction diseases using ECGs collected from Day 1 to Day 5. These recordings, ranging from 3 to 5 min, are first divided into segments of length 4.5 s (i.e. 9000 samples). To reduce overfitting and ensure good representation of the ECG recordings, we randomly select five segments on each day for each mouse. The selected raw ECG segments are directly used as the input in CNNs. Table 1 summarizes the numbers of extracted early ECG segments from the 20 mice. To build the binary classifier, all segments extracted from 10 mutant mice are labeled as positive cases (Class 1), whereas the segments extracted from 10 control ones are labeled as negative (Class 0).

To transfer knowledge from human ECGs, we pretrain a CNN on the PhysioNet/CinC challenge 2017 dataset. For computational simplicity, we only use the AF and Normal recordings and consider a binary classification problem for this pretraining task. The recordings that have length less than 30 s (i.e. 9000 samples) are excluded to match the input size of mouse ECGs for TL. For those with length more than 30 s, we only keep the first 30s signal. Finally, there are in total of 4570 Normal and 636 AF recordings included in this study. The choice for the length of ECG segments

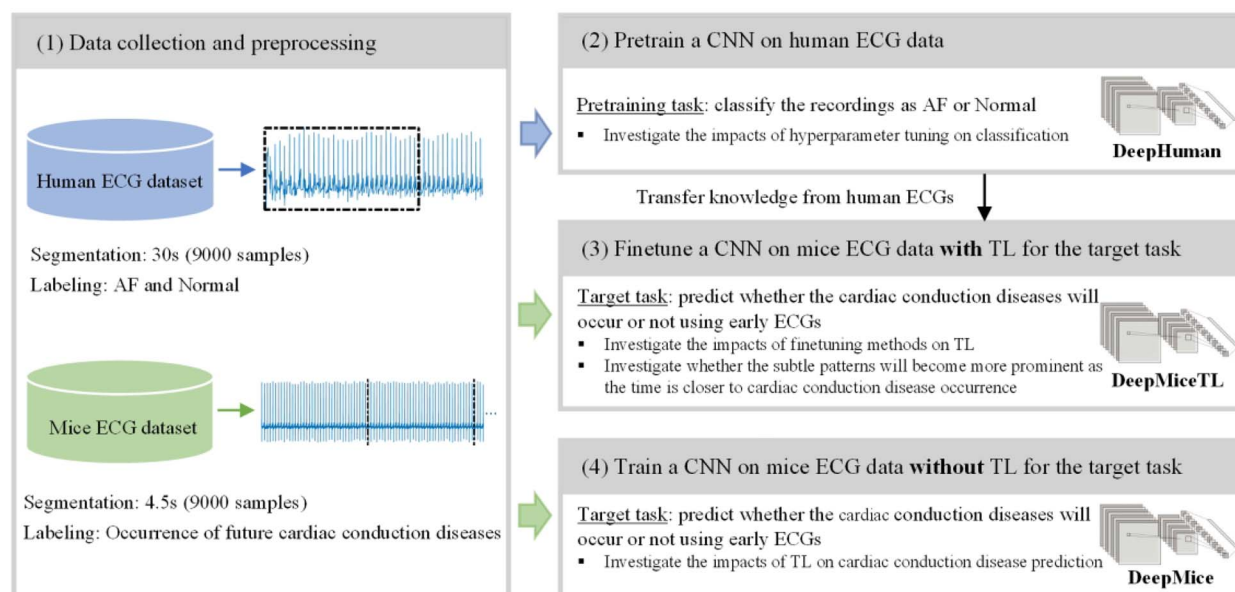


Figure 2. The workflow. The first step is to build mouse and human ECG datasets, including segmentation and labeling procedures. In the second step, we pretrain a CNN (DeepHuman) on the human ECG dataset to classify AF and Normal recordings. In the third step, the learned weights of the DeepHuman are then transferred to a new CNN (DeepMiceTL) for the target task. To further investigate the impacts of TL on the prediction performance of the target task, we also train a CNN without TL (DeepMice) on the same mouse dataset in the fourth step.

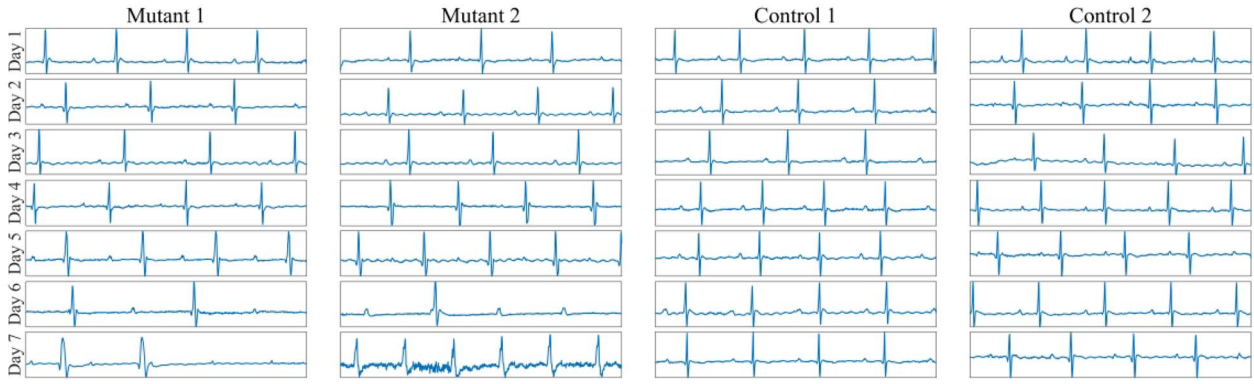


Figure 3. Random examples of 0.5s ECG signals collected during Day 1 to Day 7 for two pairs of mutant and control mice. For the mutant mice, severe cardiac conduction diseases occurred on Day 6 and finally led to their death. On the other hand, the control mice did not present arrhythmias and stayed alive during the experiment.

Table 1. Numbers of extracted mice early ECG segments

Group	Day					Total
	1	2	3	4	5	
Control	50	50	50	50	50	250
Mutant	50	50	50	50	40	240

(i.e. human: 30 s; mouse: 4.5 s) is reasonable because the mouse heart rate is approximately 10 times as humans [27]. In addition, the amplitude of ECG signals remains unchanged as it is similar between mice and humans.

CNN architecture and model training

We modify the end-to-end deep CNN proposed in [12] by tuning the architecture hyperparameters, including the number of residual blocks, the number of filters, the filter size and the dropout rate. Since TL typically uses the same architecture of the pretrained CNN for target problems, we first modify the architecture in the pretraining task (i.e. DeepHuman) and then use the modified architecture for the target task (i.e. DeepMiceTL and DeepMice). Details of hyperparameter tuning are provided in next subsection.

Figure 4 shows the architecture, which contains four stages, i.e. input, feature extraction, classification and output. It takes as input only the raw single-lead ECG signal with 9000 samples and no other features. The feature extraction stage consists of three convolutional (Conv) layers and R residual blocks with two Conv layers in each block, leading to a total of $2R + 3$ Conv layers. These layers have filters with a same size of $k \times 1$ but varying numbers of filters. Specifically, the first three Conv layers have N filters for each one. In the j th residual block, the Conv layer has $N' = N \times 2^{\lfloor j/3 \rfloor}$ filters, where $j = 1, \dots, R$. That is, N' starts at N and is then incremented by a factor of two every three blocks. Shortcut connections, using max pooling (Max Pool) layers, are applied in these residual blocks to make the optimization of such a deep network tractable [28]. Each residual block subsamples its inputs by a factor of two. Given the initial input size (9000×1) and this subsampling schedule, the maximum value of R is 12 to ensure valid inputs for each layer. The common rectified linear units (ReLU) activation function is used in the CNNs, allowing the network to generate a complex, nonlinear representation of the ECGs for automatic feature extraction. Moreover, dropout and batch normalization (BN) layers are applied to reduce over-fitting and improve generalization of CNNs. Dropout is a technique that

randomly drops neurons at each training epoch to prevent the network from using a small group of features for learning. All dropout layers have the same dropping rate of σ . The BN layers constantly normalize data for each mini-batch to reduce the internal covariant shift caused by progressive transforms [29]. Finally, in the classification stage, one fully connected (FC) layer is included, providing a mapping between the extracted features and the score s_i of assigning to class i , where $i = 1, \dots, C$. In other words, this FC layer has a total of C neurons with each neuron receiving inputs from all neurons in its previous layer. The softmax function then generates a probability distribution based on these scores.

The categorical cross-entropy is used as the loss function to measure the performance of CNNs for our classification tasks. To additionally reduce overfitting, an L2 regularization term is added to the loss function. The adaptive moment estimation (ADAM) optimizer, which is shown to work well with sparse gradients and does not require a stationary objective function [30], is used to train CNNs in this paper. Five hyperparameters in ADAM, i.e. the initial learning rate η , the decay rates β_1 and β_2 , the regularization coefficient λ and the mini-batch size B , are also included in the hyperparameter tuning process.

Bayesian optimization for hyperparameter tuning

The performance of a CNN significantly depends on a good setting of hyperparameters, including the ones in the network architecture and the ones in the training algorithm. In this paper, we consider the BO method, which has recently emerged as a powerful solution for tuning hyperparameters in deep learning models [31–34]. A total of nine hyperparameters are included in the tuning process with four in the architecture and five in the ADAM optimizer. Table 2 shows the ranges of these hyperparameters for pretraining and TL models, which are determined based on the default values (set in Matlab) and practitioner's experiences. Therefore, we aim to find the global optimum \mathbf{x}^* such that

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}), \quad (1)$$

where \mathcal{X} denotes the search space of hyperparameters $\mathbf{x} = (R, N, k, \sigma, \eta, \beta_1, \beta_2, \lambda, B)$ and $f(\mathbf{x})$ denotes the classification performance on the validation set. Recall that hyperparameters in the architecture (R, N, k, σ) are only tuned in the pretraining task and the selected values of these hyperparameters are then used for the target problem.

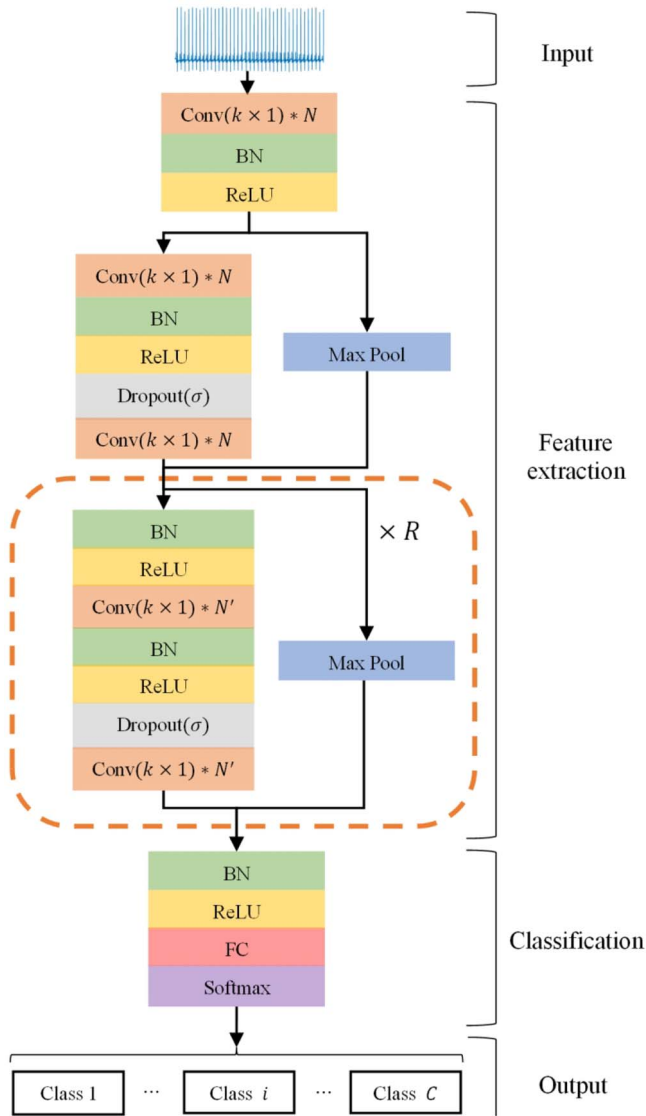


Figure 4. CNN architecture. The network takes raw ECG data (9000 samples) as input, and outputs a prediction of one out of C possible classes. The feature extraction stage consists of $2R + 3$ convolutional (Conv) layers. These layers have filters with a same size of $k \times 1$ but varying numbers of filters. The first three Conv layers have N filters for each one. In the j th residual block, the Conv layer has $N' = N \times 2^{j/3}$ filters, where $j = 1, \dots, R$. All dropout layers have the same dropout rate of σ . Finally, one fully connected (FC) layer is included in the classification stage. Hyperparameters (R, N, k, σ) are selected by BO in the pretraining task.

Algorithm 1. Bayesian Optimization

- 1: **for** $n = 1, 2, \dots$ **do**
- 2: Select \mathbf{x}_{n+1} by optimizing the acquisition function

$$\mathbf{x}_{n+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \alpha(\mathbf{x} | \mathcal{D}_n).$$
- 3: Sample the objective function to obtain y_{n+1} .
- 4: Augment the data $\mathcal{D}_{n+1} = \{\mathcal{D}_n, (\mathbf{x}_{n+1}, y_{n+1})\}$.
- 5: Update the posterior.
- 6: **end for**

The fundamental idea of BO is to construct a probabilistic model that defines a distribution over the objective function $f(\mathbf{x})$. The Gaussian process (GP) is typically used as the prior for $f(\mathbf{x})$ due to its flexibility and tractability. For any arbitrary point \mathbf{x} , the

Table 2. Hyperparameters ranges for pretraining and TL models

Hyperparameter	Range	
	Pretraining	TL
Number of residual blocks R	$\{3, \dots, 12\}$	-
Number of filters N	$\{8, 16, 32, 64\}$	-
Filter size k	$\{8, 16, 32, 64\}$	-
Dropout rate σ	$\{0.2, \dots, 0.8\}$	-
Initial learning rate η	$[10^{-5}, 10^{-1}]$	$[10^{-6}, 10^{-3}]$
Gradient decay rate β_1	$[0.8, 0.98]$	-
Squared gradient decay rate β_2	$[0.9, 0.999999]$	-
Regularization coefficient λ	$[10^{-6}, 10^{-2}]$	-
Mini-batch size B	$[100, 300]$	-

objective can be evaluated by training the CNN and generating a noisy observation y , where the noise is caused by the stochastic optimization algorithm in training the CNN. We denote the set of n observations of input-target pairs by $\mathcal{D}_n = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$. Based on the property of GP, the posterior of the random variable $f(\mathbf{x})$ conditioned on \mathcal{D}_n is also normally distributed. Given the the posterior distribution, BO uses an *acquisition function*, denoted by $\alpha(\mathbf{x} | \mathcal{D}_n)$, to determine the next candidate to evaluate via a proxy optimization. The *acquisition function* is designed to leverage the exploration (sampling from the areas with large uncertainty) and exploitation (sampling from the areas with high values), including the probability of improvement, the expected improvement and the GP upper confidence bound [31, 32]. The BO algorithm is summarized in Algorithm 1, where the computational details are referred to [31].

Transfer learning

We investigate whether the human ECG data can be leveraged to address the scarcity of early mouse ECGs for predicting cardiac conduction disease tendencies. We pretrain a CNN on human ECGs (i.e. PhysioNet/CinC challenge 2017 dataset) and transfer the learned features to the target task of predicting cardiac conduction diseases in mice. The pretrained CNN provides good initial weights that improve the learning of mouse ECGs on the feature extraction stage due to some already acquired knowledge from human ECGs. There are generally two ways to perform TL. The first one is to finetune the entire network, whereas the other method freezes the weights in feature extraction layers and only updates parameters in the classification stage [35, 36]. We examine the impacts of the finetuning methods on TL by comparing the prediction performance between these two methods. ADAM is used to finetune CNNs in TL similarly as the development of the initial models but with a reduced learning rate (as shown in Table 2). A reduction in learning rate is because the purpose of the TL is to finetune the CNN rather than retraining one. To ensure the performance of the target prediction task using new mouse data, we increase the learning rate of the FC layer (classification stage) by 10 times compared with those transferred layers (feature extraction stage) if the first method is used. Moreover, we similarly use the BO to tune the hyperparameters in the training algorithm to achieve better performance.

Experiment design and evaluation metrics

We randomly split the human ECG dataset into training (80%, 4165 recordings) and testing (20%, 1041 recordings) sets. The training set is first used to tune the hyperparameters based on the 5-fold cross validation method, which provides more robust model evaluation. Figure 5 illustrates the procedure of tuning hyperparameters for the pretrained CNN (DeepHuman) on human ECG data.

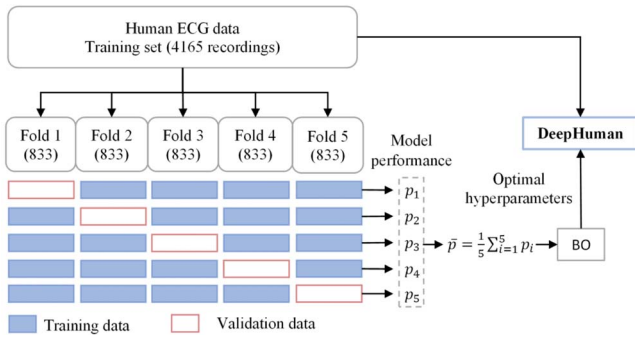


Figure 5. Illustration of the procedure of using BO and 5-fold cross-validation approaches to tune hyperparameters for the DeepHuman that is pretrained on human ECG data.

In the 5-fold cross validation, the training set is first randomly divided into five disjoint folds that have the same number (833) of ECG recordings. Each fold in turn plays the role of validating the performance of the CNN that is trained on the other four folds. The model performance is denoted by p_i that is evaluated on fold i . The average evaluation metric \bar{p} is then used as the objective $f(\mathbf{x})$ in BO to select the optimal hyperparameters. Given the selected hyperparameters, DeepHuman is fitted using the whole training set and its generalization ability is evaluated on the testing set.

For mouse data, we build the training and testing sets based on individual mouse to ensure that testing data are not exposed to CNN training. Specifically, we randomly select two control and two mutant mice and use the ECGs collected from these mice to build the testing set, and ECGs collected from the remaining eight control and eight mutant mice form the training set. Next, we divide the training set into four disjoint folds for cross validation to ensure equal number of mice in each fold, i.e. two control and two mutant mice. We first use the training set to optimize the hyperparameters of the proposed CNN (DeepMiceTL) based on the BO and 4-fold cross validation methods. Given the optimal hyperparameters obtained from BO, the DeepMiceTL is then finetuned on the whole training set and is evaluated on the testing set. To examine the impacts of leveraging human ECGs on cardiac conduction disease prediction in mice, we also train a CNN (DeepMice) on the same mouse ECG dataset without TL, given hyperparameters selected from BO.

For both the pretraining and target tasks, standard metrics of binary classifiers are used for performance evaluation, including accuracy (ACC), precision (PRE), sensitivity (SEN), specificity (SPE) and F1-score,

$$\begin{aligned} \text{ACC} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{PRE} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{SEN} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{SPE} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \\ \text{F1-score} &= \frac{2 \times \text{PRE} \times \text{SEN}}{\text{PRE} + \text{SEN}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}, \end{aligned}$$

where TP and TN are the numbers of positive and negative cases that are correctly classified, and FP and FN are the numbers of positive and negative cases that are misclassified. In particular, the F1-score is selected as the performance metric (i.e. p_i shown in Figures 5) in BO to optimize the hyperparameters because it is

the harmonic mean of the PRE and SEN and rewards the model that simultaneously maximizes both the PRE and SEN. F1-score is especially preferred over accuracy for classification problems with imbalanced data, in case the class that has a larger number of samples will dominate the model's classification result.

RESULTS

In this section, we assess the performance of the proposed deep TL model for the prediction of cardiac conduction diseases in mice. We first pretrain a CNN to classify AF in humans using the publicly available ECGs from the PhysioNet/CinC challenge 2017 dataset. We use the state-of-the-art BO method to tune the hyperparameters of the CNN because the performance of deep learning models are sensitive to hyperparameters. We compare the classification performance of the CNNs using hyperparameters with and without tuning to assess BO's benefits. The knowledge learned from the classification of human arrhythmias is embedded in the parameters of the pretrained CNN. We then use the novel TL method to leverage this knowledge and finetune the CNN for cardiac conduction disease prediction in mice. We compare the performance of the proposed deep TL model with that of the CNN trained without TL to examine the advantages of using TL technique to address the data scarcity issue in mouse ECGs. We further investigate whether the prediction performance will improve as the time of collecting ECGs from the mice gets closer to the occurrence of the severe cardiac conduction diseases. We use the Matlab R2021a to perform the CNN training, TL, and hyperparameter tuning at a power workstation that has Intel® Xeon® Gold 5122 CPU, 128 GB installed RAM and NVIDIA Quadro P400 GPU.

Using optimized hyperparameters from BO improves the AF classification performance

In the pretraining task, the CNNs with and without using BO to optimize the hyperparameters are trained on the same training set with randomly initialized weights. The ADAM algorithm terminates when it performs for 30 epochs or the automatic validation stopping criterion (i.e. the loss on the validation set has not decreased for five times of evaluation) is met, whichever comes first. After 20 epochs, the learning rate is scheduled to decrease by a factor of 0.5 for the trade-off between computational time and model performance. Table 3 presents the selected hyperparameters from the BO and the corresponding default values that are set in Matlab.

We compare the AF classification performance of DeepHuman on the testing set between using default and BO hyperparameters in Table 4. Mean values and standard errors (shown in brackets) of the performance metrics are reported based on five independent randomly initialized runs. We can see all performance metrics are above 90% using hyperparameters optimized by the BO method, showing improvements to different degrees, compared with the performances using hyperparameters without tuning. In particular, there is an approximate 3.1% increase for the mean value of F1-score. Moreover, applying BO to determine the hyperparameters of the CNN provides more stable model performance as the standard errors of the metrics are reduced compared with that using the default values. This indicates the necessity of tuning hyperparameters in CNNs for better and more stable performance. Therefore, we apply BO to optimize the hyperparameters in other CNNs in the following target task. The goal of the 2017 PhysioNet/CinC Challenge is to classify ECG recordings into one of the four rhythms: Normal, AF, Noise or

Table 3. Hyperparameters selected from BO for the pretrained CNN (DeepHuman)

Method	Architecture				ADAM				
	R	N	k	σ	η	β_1	β_2	λ	B
Default	-	-	-	0.5	1.00×10^{-3}	0.90	0.999	1.00×10^{-4}	128
BO	12	32	16	0.2	9.02×10^{-3}	0.82	0.928	2.42×10^{-5}	103

Table 4. AF classification performance (%) of the pretrained CNN (DeepHuman) on the testing set with default and BO hyperparameters

Method	ACC	PRE	SEN	SPE	F1-score
Default	97.7(0.4)	88.3(3.1)	90.4(3.7)	98.6(0.5)	89.2(1.8)
BO	98.3(0.2)	91.3(2.6)	92.7(1.9)	98.9(0.4)	92.0(0.6)

Note: mean values and standard errors (shown in brackets) of the performance metrics are reported based on five independent randomly initialized runs, and the best-performing method/model in each performance metric is bold (same in Tables 5 and 6).

Other. In one of the best performers of this challenge, the F1-scores for Normal, AF, Other and Noise are 91.2%, 81.3%, 75.1% and 56.7%, respectively. We can see that, in our simplified binary classification problem (AF or Normal), the proposed DeepHuman with BO hyperparameters provides satisfactory performance of learning human ECG patterns, achieving an F1-score of 92.0%.

Transferring knowledge learned from the AF classification in humans significantly improves the cardiac conduction disease prediction in mice

The target task in this study is to predict the occurrence of cardiac conduction diseases using early ECGs of mice. To examine the impacts of transferring knowledge from human ECGs, we compare the performance of the proposed deep TL model (DeepMiceTL) with that of the CNN (DeepMice) trained without TL. Both CNNs are trained using the hyperparameters optimized by BO. With randomly initialized weights, DeepMice is trained for 30 epochs or the automatic validation stopping criterion is met. The learning rate is decreased after 20 epochs with a factor of 0.5. For DeepMiceTL, we transfer weights from the pretrained DeepHuman and finetune this model on mouse ECGs, where the same automatic stopping method and learning rate dropping schedule are used as applied in training the DeepMice.

Tables 5 shows the prediction performance of the CNNs without and with TL on the testing set for the target task. In Table 5, the effects of finetuning methods used in TL are also examined by comparing the prediction performance between two methods, i.e. (a) finetuning the entire network and (b) freezing the weights in transferred layers and only updating parameters in the classification stage. We can see that finetuning the entire network in TL provides significantly better performance than the other one that finetunes only the classification stage. Moreover, the mean values of most performance metrics in DeepMiceTL-(a) are significantly improved, compared with those of the model without TL (DeepMice). Overall, the proposed DeepMiceTL-(a) achieves prediction accuracy of 84.8% and an F1-score of 83.8%. The precision (PRE) shows that among all the early ECG segments that are predicted as future cardiac conduction disease occurrence, 89.6% of them are truly collected from the mutant mice. We also observed that there is a small decrease in the sensitivity (SEN), which is the percentage of the correctly predicted cases among all cases with future cardiac conduction diseases. It can also be seen that the specificity (SPE) is significantly improved, showing that 90.8% of cases that will not present future cardiac

Table 5. Prediction performance (%) of CNNs without (DeepMice) and with (DeepMiceTL) TL and three feature-based logistic regression (LR) models on the testing set for the target task

Model	ACC	PRE	SEN	SPE	F1-score
DeepMice	77.2(8.3)	76.6(13.5)	82.8(8.3)	71.6(18.2)	78.7(6.1)
DeepMiceTL-(a)	84.8(0.8)	89.6(1.1)	78.8(1.1)	90.8(1.1)	83.8(0.9)
DeepMiceTL-(b)	48.4(8.6)	49.6(10.6)	58.4(40.6)	38.4(45.1)	46.8(22.8)
LR-(avg)	77.6(1.1)	85.9(1.1)	66.0(2.4)	89.2(1.1)	74.6(1.6)
LR-(avg+std)	75.8(1.1)	85.0(2.5)	62.8(2.3)	88.8(2.3)	72.2(1.4)
LR-(avg+std+min+max)	74.0(2.8)	83.8(5.0)	59.6(2.6)	88.4(3.8)	69.6(3.1)

(a): The TL is performed by finetuning the entire network. (b): The TL is performed by freezing the weights in transferred layers and only updating parameters in the classification stage. Avg, std, min and max represent the average, standard deviation, minimum and maximum values of the extracted RR, PR and QRS intervals.

conduction diseases are correctly predicted. These results show that human ECG data, even collected for a different pretraining task, can provide considerable information to improve the cardiac conduction disease prediction in mice. The decreased standard errors also show that the prediction performance is more stable with TL than that without TL.

To additionally examine the potential advantages of the proposed deep TL model over the traditional feature-based machine learning methods, we train logistic regression (LR) models on the RR, PR and QRS intervals that are extracted from the selected ECG segments using LabChart Software (ADInstruments). Specifically, we compute four common statistics of these intervals for each ECG segment, i.e. average (avg), standard deviation (std), minimum (min) and maximum (max), and then develop three different LR models based on these computed statistics, i.e. LR-(avg), LR-(avg+std) and LR-(avg+std+min+max). For example, LR-(avg) takes as predictors only the average values of the RR, PR and QRS intervals. We can see that the proposed DeepMiceTL-(a) outperforms these three LR models in all performance metrics, especially in the SEN. Compared with the best-performing LR model, i.e. LR-(avg), the proposed DeepMiceTL-(a) achieves 9.3% and 12.3% increase in ACC and F1-score, respectively. These results indicate that the deep TL technique performs better in detecting subtle abnormalities in early ECGs compared with feature-based machine learning methods.

In summary, our study provides promising evidence that the proposed model can detect the subtle signature of future cardiac conduction diseases in early ECGs of mice. This encouraging result in mice indicates the potential of using deep TL methods to develop an inexpensive clinical screening tool of identifying human patients who are at risk of future cardiac conduction diseases by examining their early ECGs.

Whether the prediction will be more accurate when the time ECGs are collected is closer to cardiac conduction disease occurrence

In the target task, it is also important to investigate the correlation between the subtle patterns associated with future cardiac conduction diseases and progression time. For this purpose, we

Table 6. Prediction performance (%) of DeepMiceTL-(a) trained on ECGs collected on different days

Day	ACC	PRE	SEN	SPE	F1-score
1	47.0(12.5)	48.7(11.3)	54.0(5.5)	40.0(21.2)	50.9(8.2)
2	63.0(9.1)	61.6(8.2)	70.0(21.2)	56.0(13.4)	64.4(11.3)
3	100.0(0.0)	100.0(0.0)	100.0(0.0)	100.0(0.0)	100.0(0.0)
4	94.0(4.2)	100.0(0.0)	88.0(8.4)	100.0(0.0)	93.5(4.7)
5	88.0(9.1)	100.0(0.0)	76.0(18.2)	100.0(0.0)	85.4(11.5)

compare the prediction performance of the proposed DeepMiceTL that is trained on ECGs collected on different days (shown in Table 6). For these five models, TL is conducted using the better performing finetuning method, i.e. DeepMiceTL-(a). We observe that, in general, ECGs collected closer to the occurrence of cardiac conduction diseases provides a better predictive performance. For example, using ECGs collected on Days 3–5 that are closer to the occurrence time of cardiac conduction diseases (Day 6) achieves better performance compared with using ECGs collected on Days 1–2. In particular, we can see that the PRE and SPE achieve 100% starting from Day 3 to Day 5, implying that all ECG segments that are collected from control mice are correctly predicted as no future cardiac conduction diseases. Our conjecture is that the closer to the time of cardiac conduction disease occurring, the subtle patterns in ECGs that are associated with future diseases become more prominent. To understand the detected subtle patterns and obtain more clinical insights, we will explore methods to improve the model transparency and interpretability in our future work.

DISCUSSION

Cardiac conduction disease is a major cause of morbidity and mortality worldwide. Though ECG-based early cardiac conduction disease prediction has considerable clinical significance and is an emerging need, developing early screening tools is challenging due to the lack of early ECGs availability before symptoms occur in patients. Mouse models have been widely used and have contributed significantly to cardiac diseases research. The primary goal of our work here is to investigate the hypothesis that subtle abnormalities are presented in early ECGs before severe cardiac conduction diseases occur in mice and that these subtle patterns can be detected by deep learning methods. In this paper, we develop a deep TL model to detect the subtle signature of future cardiac conduction diseases using early mouse ECGs, which provides significant and innovative insights in early screening and prediction of cardiac conduction disease. Our work is the first to investigate whether or not subtle patterns are present in early ECGs before severe cardiac conduction diseases occur in mice. This study resolves the data scarcity issues in mouse ECG analysis by leveraging the state-of-the-art TL technique. The deep TL model developed in this paper transfers the knowledge acquired from the classification of human arrhythmias using the publicly available database consisting of human arrhythmias and normal ECGs to the prediction of cardiac conduction disease in mice. BO has been used to effectively tune the hyperparameters used in the deep learning models and enhance the prediction results. The model that we develop for cardiac conduction disease prediction in mice has the potential to be further transferred to advance the development of cost-effective clinical screening tools for identifying patients with cardiac conduction disease tendencies when early human ECGs are available in the future.

Different factors such as genetics, injury, environment, aging and many other factors can cause cardiac conduction diseases. The goal of our study is to use deep learning methods to develop ECG-based early prediction models that can be used to predict the occurrence of various types of cardiac conduction diseases at an early stage. Therefore, we use the CCS-injured *Hcn4*^{CreERT2}-DTA mouse model in our study due to its multiple advantages compared with many other mouse models in cardiac disease research such as inherited diseases. First, abnormalities in the CCS are common reasons that cause cardiac conduction diseases and many facts can cause CCS defects [26]. Second, the cardiac conduction diseases that occurred in the *Hcn4*^{CreERT2}-DTA mice include various disease types such as SSS, AV blocks and cardiac arrests, which allows a broad scope of our predictive approach instead of limiting to only one specific type of cardiac conduction disease. Third, notably, the induction of cardiac conduction disease in *Hcn4*^{CreERT2}-DTA mice can be managed by tamoxifen, and their disease progression are quick and consistent. After tamoxifen induction, all *Hcn4*^{CreERT2}-DTA mice did not display any obvious phenotype nor visible cardiac arrhythmias on ECG signals during the early days, yet these mice later developed cardiac conduction diseases that became more and more severe quickly, leading ultimately to sudden death. It is common that in the early disease course, human populations are asymptomatic and have no visible arrhythmias on ECGs, but they likely develop cardiac conduction diseases with progression that may take a long time. For example, in human populations, though inherited SSS can occur in younger populations, this type of cardiac conduction disease is more often acquired in older populations. The *Hcn4*^{CreERT2}-DTA mouse model takes the advantage of the fast progression of comprehensive cardiac conduction diseases and makes the prediction and validation efficient within a relatively short time course.

DeepMiceTL, the deep TL model we develop in this study, achieves a promising prediction performance of cardiac conduction diseases in mice using their early ECGs. From a methodological point of view, we use the TL method to resolve the data scarcity issue in mouse models. Although it is much easier to collect mouse ECGs, the number of mice used in a mouse model is usually limited, and is not adequate for training a deep model. Inspired by the recent papers which apply the state-of-the-art TL technique to ECG analysis [16, 18, 19], TL allows us to leverage the openly available human ECG data by pretraining a CNN for AF classification to address the scarcity of mouse ECG data. The rationale of using TL is that mouse and human have shared considerable genetic conservation and arrhythmogenic features, and mouse models have been widely used for human cardiac disease research [37]. Our results show that transferring knowledge learned from AF classification in humans significantly improves the performance of predicting cardiac conduction disease in mice. Compared with the model without TL (i.e. DeepMice), our model provides improved prediction results (e.g. 9.8% increase in ACC and 6.5% increase in F1-score), showing the strength of leveraging human ECG data for cardiac conduction disease prediction in mice. Furthermore, we have applied BO and *k*-fold cross-validation methods to tune the hyperparameters of the CNNs. We observe that using optimized hyperparameters from BO achieves better and more stable model performance, compared with that of using hyperparameters without tuning. Our results show that all performance metrics of AF classification are above 90% using hyperparameters optimized by the BO method.

From a biomedical point of view, the overall cardiac conduction disease prediction accuracy is approximately 85%, implying

that with 85% chance, the model correctly predicts whether or not a mouse will present cardiac conduction diseases in the future. This level of accuracy is significant, demonstrating great utility for mouse ECG modeling and prediction. Our model also achieves approximately 90% PRE, which is another important performance metric, indicating that 90% ECG segments that are identified to have subtle abnormalities which will develop into severe cardiac conduction diseases in the future are collected from mutant mice that present these diseases later in the experiment. In addition, the achieved SPE shows that over 90% ECG segments that are collected from control mice have been correctly predicted as no future cardiac conduction diseases. We further examine whether the prediction results will improve as the day on which ECGs are collected gets closer to the occurrence of the severe cardiac conduction diseases. That is, we examine whether the subtle patterns that are associated with future cardiac conduction diseases will evolve with time. We compare the prediction performance of the proposed model that is trained on ECGs collected on different days before cardiac conduction diseases occur. The results indicate that the closer to the time of cardiac conduction disease occurring, the subtle patterns in early ECGs become more prominent. In particular, all ECG segments that are identified with subtle abnormalities are collected from mutant mice as the ECG collection time nears the occurrence of severe cardiac conduction diseases (i.e. 100% PRE from Day 3 to Day 5).

In conclusion, our work provides promising evidence that the state-of-the-art deep TL models are capable of identifying subtle signatures of future cardiac conduction diseases from early ECGs when mice are still asymptomatic, which is the first effort in ECG-based early prediction of cardiac conduction disease in mice. We have shown that these identified subtle patterns can serve as a critical indicator of increased risk of future cardiac conduction diseases. It will be valuable to test the performance of our proposed framework in different mouse models in future investigations. Importantly, we believe that the deep TL models used in this study can potentially be transferred to the prediction of human cardiac conduction diseases when early human ECGs are collected in the future. Our method could advance the development of cost-effective clinical screening tools for identifying patients with cardiac conduction disease tendencies, but further exploration of methods for improving the model transparency and interpretability to obtain more clinical insights are still needed, which could be an important and interesting area of future study.

Key Points

- This study is the first study that investigates whether or not subtle patterns presented in early ECGs before severe cardiac conduction diseases occur in mice.
- A state-of-the-art deep transfer learning model, DeepMiceTL, is proposed to identify subtle patterns in early ECGs of mice and predict future cardiac conduction diseases.
- DeepMiceTL can leverage human ECGs by pretraining a CNN on these data and transfer the learned knowledge to address the scarcity of mouse ECG data.
- Bayesian optimization and k -fold cross validation methods can effectively optimize the hyperparameters of CNNs for a better prediction performance.

FUNDING

The results are based on M.Z. and J.W.'s unpublished mice ECG data. Data collection is supported by funds from the National Institutes of Health (R01HL142704 to J.W.) and American Heart Association (970606 to J.W., 902940 to M.Z.).

References

1. Park DS, Fishman GI. The cardiac conduction system. *Circulation* 2011;**123**(8):904–15.
2. Burnicka-Turek O, Broman MT, Steimle JD, et al. Transcriptional patterning of the ventricular cardiac conduction system. *Circ Res* 2020;**127**(3):e94–106.
3. Jensen PN, Gronroos NN, Chen LY, et al. Incidence of and risk factors for sick sinus syndrome in the general population. *J Am Coll Cardiol* 2014;**64**(6):531–6.
4. Herrmann S, Fabritz L, Layh B, et al. Insights into sick sinus syndrome from an inducible mouse model. *Cardiovasc Res* 2011;**90**(1):38–48.
5. Semelka M, Gera J, Usman S. Sick sinus syndrome: a review. *Am Fam Physician* 2013;**87**(10):691–6.
6. Adan V, Crown LA. Diagnosis and treatment of sick sinus syndrome. *Am Fam Physician* 2003;**67**(8):1725–32.
7. Wang J, Klysik E, Sood S, et al. Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proc Natl Acad Sci* 2010;**107**(21):9753–8.
8. Zheng M, Li RG, Song J, et al. Hippo-yap signaling maintains sinoatrial node homeostasis. *Circulation* 2022;**146**:1694–711.
9. Swaminathan PD, Purohit A, Soni S, et al. Oxidized camkii causes cardiac sinus node dysfunction in mice. *J Clin Invest* 2011;**121**(8):3277–88.
10. Torrente AG, Zhang R, Zaini A, et al. Burst pacemaker activity of the sinoatrial node in sodium-calcium exchanger knockout mice. *Proc Natl Acad Sci* 2015;**112**(31):9769–74.
11. Dobrev D, Wehrens XHT. Mouse models of cardiac arrhythmias. *Circ Res* 2018;**123**(3):332–4.
12. Hannun AY, Rajpurkar P, Haghighpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019;**25**(1):65–9.
13. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ecg algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 2019;**394**(10201):861–7.
14. Somani S, Russak AJ, Richter F, et al. Deep learning and the electrocardiogram: review of the current state-of-the-art. *EP Europace* 2021;**23**:1179–91.
15. Esther B-L, Lyle JV, Huang CL-H, Zhang Y, Nandi M, Jeevaratnam K, and Aston PJ. Symmetric projection attractor reconstruction analysis of murine electrocardiograms: Retrospective prediction of scn5a+/- genetic mutation attributable to brugada syndrome. *Heart Rhythm* 2020;**17**(5):368–75.
16. Aston PJ, Lyle JV, Bonet-Luz E, et al. Deep learning applied to attractor images derived from ecg signals for detection of genetic mutation. In: *2019 Computing in Cardiology (CinC)*. IEEE, 2019, 1–4.
17. Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, 248–55.
18. Van Steenkiste G, Loon van G, Crevecoeur G. Transfer learning in ecg classification from human to horse using a novel parallel neural network architecture. *Sci Rep* 2020;**10**(1):1–12.

19. Weimann K, Tim OF Conrad. Transfer learning for ecg classification. *Sci Rep* 2021;**11**(1):1–12.
20. Clifford GD, Liu C, Benjamin Moody H, et al. Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017. In: *2017 Computing in Cardiology (CinC)*. IEEE, 2017, 1–4.
21. Goldberger AL, Amaral LAN, Glass L, et al. Physiobank, physiobook, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 2000;**101**(23):e215–20.
22. Sharma NC, Efstratiou A, Mokrousov I, et al. Diphtheria (primer). *Nat Rev Dis Primers* 2019;**5**.
23. Yamaizumi M, Mekada E, Uchida T, Okada Y. One molecule of diphtheria toxin fragment a introduced into a cell can kill the cell. *Cell* 1978;**15**(1):245–50.
24. Ivanova A, Signore M, Caro N, et al. In vivo genetic ablation by cre-mediated expression of diphtheria toxin fragment a. *Genesis* 2005;**43**(3):129–35.
25. Heallen T, Zhang M, Wang J, et al. Hippo pathway inhibits wnt signaling to restrain cardiomyocyte proliferation and heart size. *Science* 2011;**332**(6028):458–61.
26. Liang X, Wang G, Lin L, et al. Hcn4 dynamically marks the first heart field and conduction system precursors. *Circ Res* 2013;**113**(4):399–407.
27. Boukens BJ, Rivaud MR, Rentschler S, Coronel R. Misinterpretation of the mouse ecg: 'musing the waves of mus musculus'. *J Physiol* 2014;**592**(21):4613–26.
28. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: *European conference on computer vision*. Springer, 2016, 630–45.
29. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International conference on machine learning*. PMLR, 2015, 448–56.
30. Kingma DP, Ba J. Adam: a method for stochastic optimization arXiv preprint arXiv:1412.6980. 2014.
31. Jia W, Chen X-Y, Zhang H, et al. Hyperparameter optimization for machine learning models based on bayesian optimization. *Jo Electron Sci Technol* 2019;**17**(1):26–40.
32. Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. *Adv Neural Inf Process Syst* 2012;**25**.
33. Cho H, Kim Y, Lee E, et al. Basic enhancement strategies when using bayesian optimization for hyperparameter tuning of deep neural networks. *IEEE Access* 2020;**8**:52588–608.
34. Shahriari B, Swersky K, Wang Z, et al. Taking the human out of the loop: a review of bayesian optimization. *Proc IEEE* 2015;**104**(1):148–75.
35. Yosinski JASON, Clune JEFF, Bengio YOSHUA, and Lipson HOD. How transferable are features in deep neural networks? *arXiv preprint arXiv:1411.1792*, 2014.
36. Fawaz HI, Forestier G, Weber J, et al. Transfer learning for time series classification. In: *2018 IEEE international conference on big data (Big Data)*. IEEE, 2018, 1367–76.
37. Martin CA, Zhang Y, Grace AA, Huang CL-H. In vivo studies of scn5a+/- mice modeling brugada syndrome demonstrate both conduction and repolarization abnormalities. *J Electrocardiol* 2010;**43**(5):433–9.