2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871518

A Meta-Transfer Learning Approach to ECG Arrhythmia Detection

Wuxia Chen, Taposh Banerjee, and Eugene John

Abstract— **Automatic classification of cardiac abnormalities is becoming increasingly popular with the prevalence of ECG recordings. Many signal processing and machine learning algorithms have shown the potential to identify cardiac abnormalities accurately. However, most of these methods heavily rely on a large amount of relatively homogeneous datasets. In real life, chances are that there is not enough data for a specific category, and regular deep learning may fail in this scenario. A straightforward intuition is to use the knowledge learned from previous data to solve the problem. This idea leads to learning-to-learn: extrapolating the knowledge accumulated from the old dataset and using it in a different but somewhat related dataset. In this way, we do not need to have considerable data to learn the new task because the underlying features of the old and new datasets resemble one another. In this paper, a novel machine learning method is introduced to solve the ECG arrhythmia detection problem with a limited amount of data. The proposed method combines the popular techniques of meta-learning and transfer learning. It is shown that our method achieves much higher accuracy in ECG arrhythmia classification with a few data and learns the new task faster than regular deep learning.**

I. INTRODUCTION

The electrocardiogram (ECG) is a non-invasive representation of the heart's electrical activity from electrodes placed on the skin. It is a fundamental tool in diagnosing a wide range of arrhythmia abnormalities, with more than 300 million ECGs obtained annually worldwide [1]. Automatic ECG interpretation has been used for over 50 years and has helped to predict cardiovascular morbidity and mortality [2] and reduce the laborious workflow of physicians. The early and correct diagnosis of cardiac abnormalities may increase the chances of successful treatments [3]. However, manual interpretation of the electrocardiogram is time-consuming and requires skilled personnel with a high degree of training [4].

The combination of the increasing availability of ECG data and the development of deep learning algorithms presents an opportunity to analyze ECG signals accurately and facilitates automated ECG interpretation [5]. The existing literature contains many arrhythmia detection algorithms with substantial preprocessing of the ECG signals [6].

Deep learning enables computational models composed of multiple processing layers to learn data representations with numerous levels of abstraction. In this way, deep learning can reveal complex patterns in large datasets. It outperforms other traditional machine learning techniques in that human engineers do not design its feature extractors; instead, they are learned from raw data using a general-purpose learning procedure. Also, a deep learning algorithm can learn the subtle details of high-dimensional data and has applications in a broad spectrum of fields [7].

Deep neural networks are effective at learning from a large amount of training data and can predict very accurately in test datasets. One shortcoming of the typical deep neural network is that it requires a massive amount of similar data points for any given category. But in the real world experience, we only see a small number of instances of each object throughout our lifetime. Another disadvantage of the deep neural network is that it begins with some random parameters and takes a long time to converge every time it learns the training data. Fast learning is the goal of computer-aided classification because it resembles human cognition: to learn new skills with a small training session or recognize new items with a few prior instances.

It is intuitive to take advantage of the knowledge acquired in previous learning sessions to achieve the goal of fast learning. The tricky part is how to combine the new task with the prior learning experience. An important practical approach is to give the neural network some good, meaningful starting parameters to have maximal performance on a new task after a bit of training, using a small amount of data from the new task [9]. The idea of finding good starting points leads to the notion of learning-to-learn. Two popular methods in this domain are meta-learning and transfer learning. The goal of meta-learning is to train a *launch model* using a family of tasks, such that it can later solve a new learning task from the same family using only a small number of training samples [8]. The goal of transfer learning is to fine-tune a trained model for a certain task so that it can perform better on a new and similar task. While both methods hold promise, neither of them is perfectly suitable for the problem of arrhythmia detection in ECG data. In practice, we only have access to a small collection of ECG datasets, certainly not enough to train a meta-learning launch model in the traditional sense. Also, we found in our studies that a traditional transfer learning approach leads to quite a poor accuracy.

In this paper, we propose a combined meta-transfer learning approach [21] that can be implemented with significantly lower computational overhead than traditional meta-learning. Also, our method performs better than classical transfer learning. The rest of this paper is organized as follows. In Section II, we briefly describe the principles of meta-learning and transfer learning. In Section III, we discuss our proposed methodology and pre-processing of the ECG

The authors are with the Department of Electrical and Computer Engineering at the University of Texas at San Antonio (UTSA). This work was supported in part by the National Science Foundation under Grant 2041327. The work of Wuxia Chen and Taposh Banerjee was also supported in part by the Cloud Technology Endowed Professorship IV at UTSA. Email: wuxia.chen@my.utsa.edu; taposh.banerjee@utsa.edu; eugene.john@utsa.edu

signals. In Section IV, we provide the experimental results and discussion.

II. Meta-learning and transfer learning

In this section, we provide a brief overview of meta-learning and transfer learning.

A. Meta-learning

Suppose we have a dataset with only a small amount of samples, and we want our model to learn it quickly. The idea behind meta-learning is to use prior knowledge to accomplish this goal so that the learning does not start from scratch. We first revisit some ideas from supervised machine learning.

Consider a model which maps observation x to label y. The model parameters are denoted as ϕ, and the training data are denoted as D. One way to visualize the training in supervised learning is to view it as a maximum likelihood problem, where we maximize the likelihood of the parameters given the training data: $\arg\max_\phi \log p(\phi|D)$. We redefine the problem as maximizing the probability of the data given the parameters $\arg\max_\phi \log p(D|\phi)$ and maximizing the marginal probability of our parameters $\arg\max_\phi \log p(\phi)$. Then we get some optimization over the dataset given $\sum_i \log p(y_i|x_i, \phi)$ and we have the regularizer $\log p(\phi)$(e.g., weight decay):

$$\begin{aligned}\arg\max_\phi \log p(\phi|D) &= \arg\max_\phi \log \frac{p(\phi, D)}{p(D)} \\ &= \arg\max_\phi \log p(\phi, D) \\ &= \arg\max_\phi \log p(D|\phi) + \log p(\phi) \\ &= \arg\max_\phi \sum_i \log p(y_i|x_i, \phi) + \log p(\phi),\end{aligned}$$

where $D = \{(x_1, y_1), \cdots, (x_k, y_k)\}$, where x_is are the input observations observations (e.g. an image or a signal track), and y_is are the input labels [9].

As mentioned previously, the art of meta-learning involves integrating the learned experience with a few shots of new data points. We would like to incorporate the prior knowledge with the new data points; hence the problem becomes $\arg\max_\Phi \log p(\phi|D, D_{meta-train})$, where $D_{meta-train}$ denotes the previous training data or meta-training data. But keeping $D_{meta-train}$ forever is laborious and impractical since the data occupies a substantial memory. The core idea behind meta-learning is to represent the meta-training dataset using a set of meta-parameters obtained via $\theta = \arg\max_\theta p(\theta|D_{meta-train})$. The meta-parameters θ are extracted from $D_{meta-train}$ to denote the prior knowledge we need to know to solve new tasks quickly [9].

The description of the new task is still a maximum likelihood problem $\arg\max_\phi \log p(\phi|D, D_{meta-train})$. We would like to maximize the likelihood of the parameters over the new data given the past meta-training data. The likelihood function can be viewed as an integration over the meta-parameter θ. We approximate this integral with a point estimate θ^* for the meta-parameters. $p(\theta^*|D_{meta-train})$ corresponds to meta-training, where we learn the meta-parameters based on the old meta-training data, and $p(\phi|D, \theta^*)$ corresponds to adaptation, where we learn the new parameters for a new task given the new data and the meta-parameters:

$$\begin{aligned}\log p(\phi|D, D_{meta-train}) &= \log \frac{p(\phi, D, D_{meta-train})}{p(D, D_{meta-train})} \\ &= \log \int \frac{p(\phi, D)p(\theta, D_{meta-train})}{p(D)p(D_{meta-train})} \, d\theta,\end{aligned}$$

here we assume ϕ is independent of $D_{meta-train}|\theta$,

$$\begin{aligned}&= \log \int \frac{p(\phi, D, \theta)p(\theta, D_{meta-train})}{p(D, \theta)p(D_{meta-train})} \, d\theta \\ &= \log \int p(\phi|D, \theta)p(\theta|D_{meta-train}) \, d\theta \\ &\approx \log p(\phi|D, \theta^*)p(\theta^*|D_{meta-train}),\end{aligned}$$

where $\theta^* = \arg\max_\theta p(\theta|D_{meta-train})$

In this way, the new task

$$\arg\max_\phi \log p(\phi|D, D_{meta-train})$$

can be approximated by

$$\arg\max_\phi \log p(\phi|D, \theta^*),$$

where ϕ denotes the task-specific parameters and θ^* acts as preliminary information shared across all tasks.

In summary, there are two steps to learn a new task. First, acquire previous experience with meta-learning parameter $\theta^* = \arg\max_\theta p(\theta|D_{meta-train})$, and then adapt with a few shots $\phi^* = \arg\max_\phi p(\phi|D, \theta^*)$. Fig. 1 shows how the meta-parameters θ^* participate in the new task, and the task-specific parameters ϕ get updated from meta-parameters θ with a few new input/output pairs. Finally, our new model parameters ϕ^* are used for prediction.

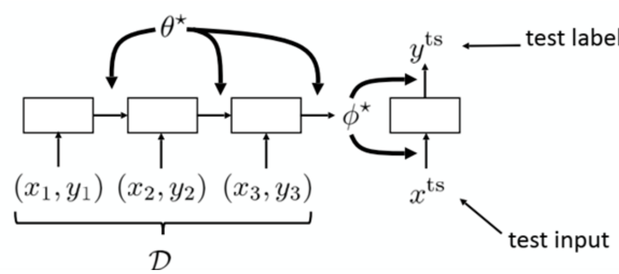

Fig. 1. Meta-parameter θ^* participate in the training of parameters ϕ^* in the new tasks [9].

B. Transfer Learning

The underlying logic of transfer learning is the model pre-trained on big data generalizes better than randomly initialized ones [21] [22]. Transfer learning also takes advantage of the pre-trained model. It freezes most trained layers and keeps the feature extractor unchanged for the new task. Then it substitutes the classifier (the top fully connected layers) according to the number of categories in the new task with a new trainable one (see Fig. 2). In this situation, the pre-trained parameters θ and the task-specific parameters ϕ are mostly the same except for their classifiers.

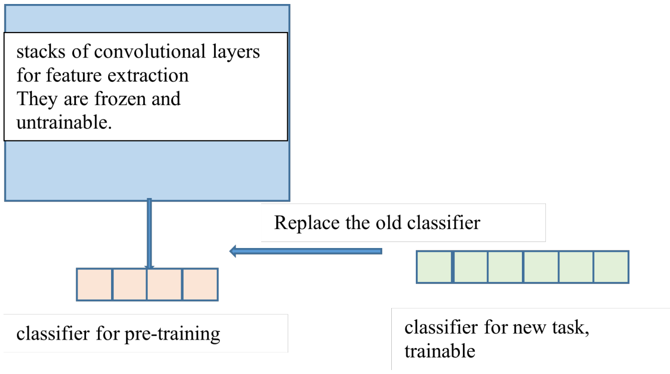

Fig. 2. The transfer learning methodology. Here the earlier part of the network is frozen while transferring to the new task.

III. A META-TRANSFER LEARNING METHOD TO CLASSIFY ARRHYTHMIA

In classical meta-learning [8], to learn the meta-parameter θ^*, it is important that not only the meta-training dataset $D_{meta-train}$ be large, but should also contain samples from various tasks from a large family of tasks. But this is not realistic in ECG arrhythmia detection because we often have access to only a few or even a single dataset for learning and prediction. A transfer learning approach is more suitable for ECG data because we only need one dataset to learn and then fine tune a small subset of the learned parameters. However, our studies have shown that a transfer learning approach leads to poor prediction and generalization error performance. In this section, we propose a methodology that combines the concepts of meta-learning and transfer learning. Specifically, we train the model as done in transfer learning using a single dataset. However, unlike transfer learning, we do not freeze any part of the learned network or coefficients and update all of them using data from the new task. The latter approach is employed in meta-learning. In the rest of the paper, we call this methodology a meta-transfer learning approach. Below, we discuss our approach in detail.

We use a neural network to automatically classify various cardiac arrhythmias in our experiments. The datasets we use in our experiments are: 2017 PhysioNet/Cinc Challenge database [10][11], China Physiological Signal Challenge (CPSC) 2018 dataset [12], Georgia 12-lead ECG Challenge (G12EC) Database [11], and MIT-BIH Arrhythmia Database [13][11]. All the ECG signals in these databases are carefully examined and labeled by fully experienced cardiologists.

We build our base learner in one dataset and then apply the trained launch model to learn the new data in another dataset. The meta-parameter θ is saved for later fine-tuning to adapt to the new task. The new task is to recognize and classify other arrhythmias from different databases with prior knowledge θ. We manually replace the old classifier with a new one according to the number of arrhythmia types in the new task. The classifier is a stack of fully connected dense layers followed by a softmax activation layer [14]. It gives the probability distribution of the class labels, and the class with the highest probability will be considered the correct one. Unlike typical transfer learning, during which the feature extractors are frozen and untrainable, we allow all the weights to be updated during the fine-tuning session, as is shown in Fig. 3. Each time, a small number of signals are fed into the neural network to fine-tune the task-specific parameters ϕ so that our model may quickly learn the new types of arrhythmia.

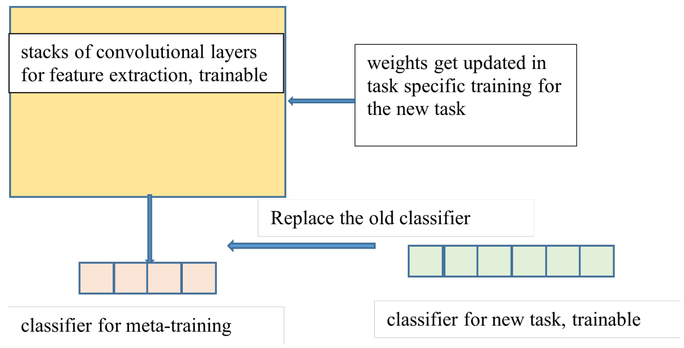

Fig. 3. Our proposed meta-transfer learning methodology. All the parameters are trainable in our fine-tuning session.

We now discuss the preprocessing of the data and the design of our network. Usually, when designing a neural network, it is required that the inputs are in the same format. It is natural to think about data standardization, giving the data a uniform format to better analyze and utilize it in further research [15]. After standardization, we put different variables onto the same scale, so comparing them becomes meaningful [16]. A typical operation in standardization is to subtract the mean and divide it by the standard deviation. For each recording, we do

$$signal_i \leftarrow \frac{signal_i - mean_i}{std_i}.$$

After this step, all the signals are standardized vertically. But horizontally, they still have different lengths. Next, we will pad the short signals with zeros to match the long signal in size. In this way, all the signals fed into the neural network are of the same length. The labels are also padded according to the number of waveforms each ECG signal has. For example, if one ECG signal has 23 waves, then there are 23 labels to be padded. Until this step, in the training data, one recording (signal-labels pair) contains a standardized, zero-padded ECG signal and a label vector indicating the number of waves in that signal.

We build a 1D neural network that classifies different arrhythmia types. During the meta-training, the network is shown an ECG signal and a label vector. We expect the desired category to have the highest score of all the types after training. Cross entropy is used as an objective function to measure the error between the output scores and the desired pattern of scores. The network modifies its internal parameters during the training to reduce the error. Our 1D neural network is end-to-end, so in the output end, the model will give a prediction vector telling the probability of each arrhythmia type [17].

IV. Experimental Results and Discussion

A good measure is necessary to evaluate the performance of the model. In statistical analysis, the F-score is used to measure the accuracy of a test. It is calculated with the precision and recall of the test [18]. A true positive (TP) is an outcome correctly labeled as positive. Similarly, a true negative (TN) is an outcome correctly labeled as negative. A false positive (FP) is an outcome that is mistakenly labeled as positive, and a false negative (FN) is an outcome mistakenly labeled as negative [19]. Accuracy denotes the ratio of correctly recognized instances over all testing instances. Specifically,

$$precision = \frac{TP}{TP+FP}, \qquad recall = \frac{TP}{TP+FN},$$

$$F_1 \quad score = \frac{2}{precision^{-1} + recall^{-1}},$$

$$accuracy = \frac{TN+TP}{TP+TN+FP+FN}.$$

We use the above metrics to report our numerical results.

A. Learning AF and NSR signals in CPSC 2018 database and G12EC database based on the knowledge from 2017 PhysioNet Challenge database

In the 2017 PhysioNet Challenge database, there are four categories of signals: normal sinus rhythm (NSR), atrial fibrillation(AF), other arrhythmia signals, and noise. We implement the meta-transfer learning idea with only AF and NSR signals in this experiment. We randomly select 657 AF signals and 4433 NSR signals from the 2017 PhysioNet Challenge database during the meta-training session and train our model to recognize those two signals. Then we use this trained model as a base to learn and classify the AF and NSR signals from CPSC 2018 database and G12EC database. We use a few new signals to train the model initially and then gradually increase the training samples. We selected 1000 signals from the target database randomly to test the performance of the model. The accuracy values are reported in Fig. 4 and Fig. 5. The detailed performances concerning each arrhythmia are shown in the appendix (see from Fig. 10 to Fig. 13). The experiment results show that our model predicts the signal labels in the test more accurately than regular deep learning.

B. Learning six arrhythmias in CPSC 2018 database and G12EC database with the knowledge from 2017 PhysioNet Challenge database

We expand the experiment to recognize some other arrhythmias the model has never seen in the previous meta-training. In this experiment, our model is trained to learn six types of signals: normal sinus rhythm (NSR), atrial fibrillation (AF), first-degree atrioventricular block (I-AVB), left bundle branch block (LBBB), right bundle branch block (RBBB) and premature atrial contraction (PAC), with the previous knowledge accumulated from the 2017 Physionet Challenge. We randomly select 1000 signals of these six arrhythmias from the target database CPSC 2018 to test our model's performance. During the new task training session, the 1D neural network is only trained with a few examples of the six arrhythmias. A comparison of our method and regular deep learning is presented in Fig. 6.

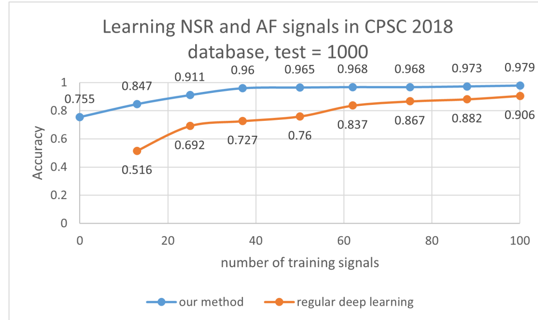

Fig. 4. The accuracy of our method vs. regular deep learning in CPSC 2018 database (NSR and AF signals only)

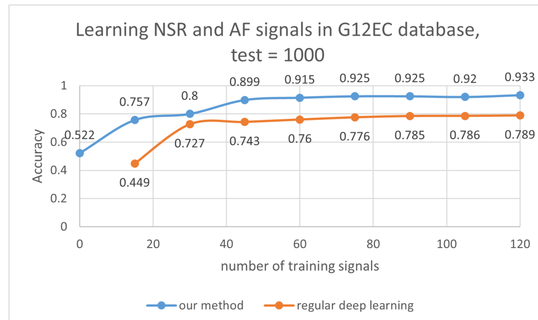

Fig. 5. The accuracy of our method vs. regular deep learning in G12EC database (NSR and AF signals only)

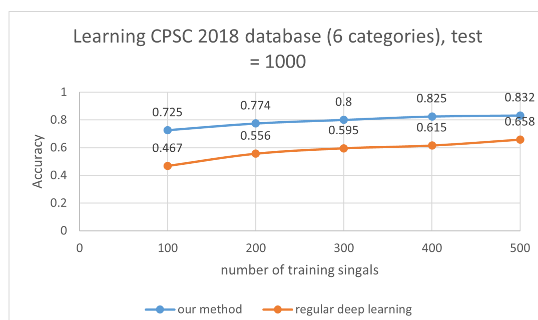

Fig. 6. The accuracy of our method vs.regular normal deep learning task: learning CPSC 2018, meta-trained in 2017 PhysioNet Challenge

C. Learning six arrhythmias in G12EC database with the knowledge from 2017 PhysioNet Challenge database

We do the same experiment in the previous subsection to G12EC database. Similarly, we select 500 signals of six arrhythmias in G12EC database randomly to test the performance of our model. The experiment result is shown in Fig. 7.

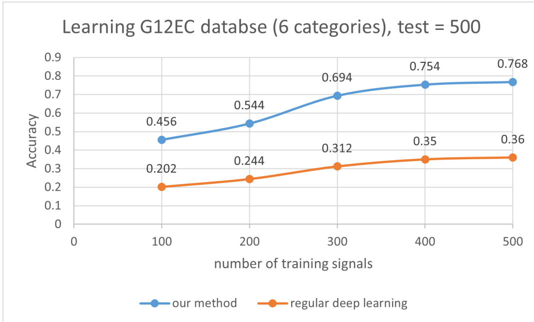

Fig. 7. The accuracy of our method vs. normal deep learning task: learning G12EC, meta-trained in 2017 PhysioNet Challenge

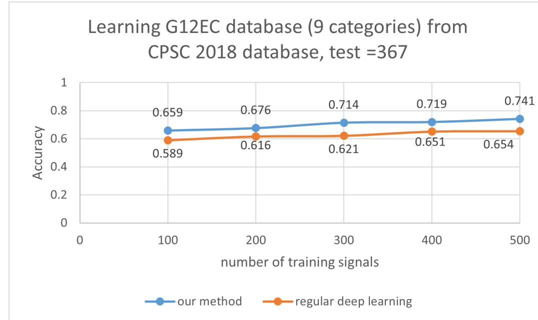

Fig. 9. The accuracy of our method vs. normal deep learning task: learning G12EC, meta-trained in CPSC 2018

D. Learning five arrhythmias in MIT-BIH Arrhythmia database with the knowledge from 2017 PhysioNet Challenge database

We want to learn five types signals in MIT-BIH database: N, L, R, V, and A. The meta-training is completed in 2017 PhysioNet Challenge database. We select 7751 samples of five arrhythmia types from MIT-BIH database randomly to test our model performance. The experiment result is shown in Fig. 8.

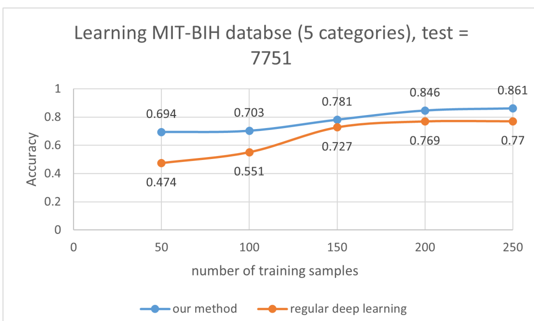

Fig. 8. The accuracy of our method vs. normal deep learning task: learning MIT-BIH database, meta-trained in 2017 PhysioNet Challenge

E. Learning arrhythmias in G12EC database with the knowledge from CPSC 2018 database

In this experiment, we would like to learn nine types of signals from G12EC database: normal sinus rhythm (NSR), atrial fibrillation (AF), first-degree atrioventricular block (I-AVB), right bundle branch block (RBBB), premature atrial contraction (PAC), sinus tachycardia (STach), left ventricular hypertrophy (LVH), sinus bradycardia (SB), and sinus arrhythmia (SA). We select 367 signals of nine arrhythmias from G12EC database randomly as the test dataset. The experiment result is shown in Fig. 9.

F. Transfer learning vs. meta-transfer learning

We also used the classical transfer learning for this problem. We found that this leads to quite poor prediction performance. Since we do not pre-train our model using a variety of data, simply transferring the prior knowledge to the new model does not help classify the new arrhythmia.

These experiments show that the meta-transfer learning approach achieves higher general accuracy than regular deep learning. In the appendix, there are more details about the performance of the proposed method. We show how well our model can predict each type of signal.

V. Conclusion and Future Work

This paper proposes a meta-transfer learning approach to utilize the previous learning experience in learning the new data so that the training does not start from scratch. The experimental results show that the proposed method achieves better accuracy with limited data and classifies faster than regular deep learning. We also prove that this method helps recognize various types of previously unseen arrhythmias based on prior knowledge with a bit of training. In the future, the meta-transfer learning approach may also help to detect arrhythmias in animals based on understanding human arrhythmia [20]. Another future research direction we will pursue is to collect ECG signals recorded in different ethnic groups, then take a proportion of signals from each data set. The conglomerated dataset can be used as the base for classical meta-training, and the neural network will be robust and not biased on any specific dataset.

Appendix

In this appendix, we provide detailed performances of our method concerning each category of signal in the first experiment of Section IV. We use the accuracy to measure the general performance of our model in the test (how many predictions are correct over all the test data). We use the F_1 score to measure how good our model is in the prediction of each category. The following pictures show that the meta-transfer learning method predicts better than regular deep learning in each class.

References

[1] Holst H, Ohlsson M, Peterson C, Edenbrandt L (1999). A confident decision support system for interpreting electrocardiograms. Clinical Physiology (Oxford, England), 19(5), 410–418.

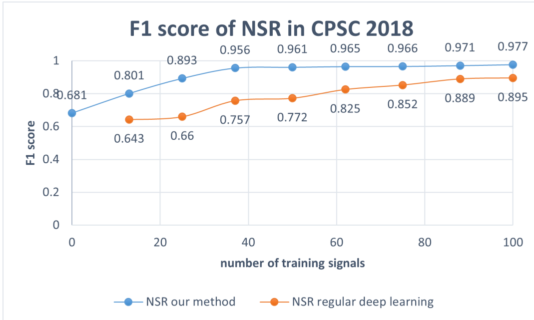

Fig. 10. The F_1 score of NSR signal in CPSC 2018 database

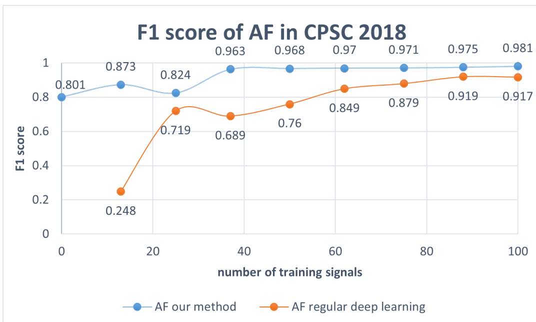

Fig. 11. The F_1 score of AF signal in CPSC 2018 database

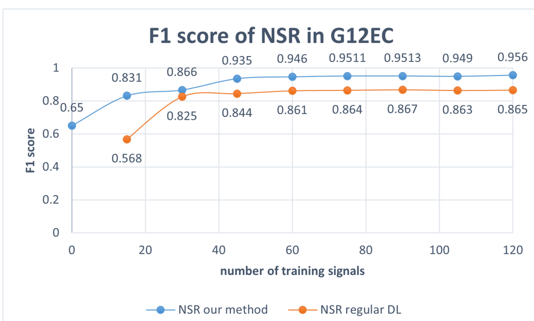

Fig. 12. The F_1 score of NSR signal in G12EC database

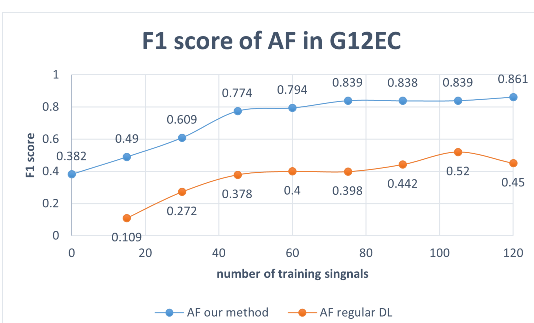

Fig. 13. The F_1 score of AF signal in G12EC database

[2] Kligfield P, Gettes L, Bailey J, Childers R, Deal B, Hancock, E, van Herpen G, Kors J, Macfarlane P, Mirvis D, Pahlm O, Rautaharju P (2007). Recommendations for the Standardization and Interpretation of the Electrocardiogram. Part I: The Electrocardiogram and Its Technology A Scientific Statement From the American Heart Association Electrocardiography and ArrhythmiasnCommittee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society Endorsed by Journal of the American Collegeof Cardiology, 49(10), 1109–1127.

[3] Adams H, del Zoppo G, Alberts M, Bhatt D, Brass L, Furlan A, Grubb R, Higashida R, Jauch E, Kidwell C, Lyden P, Morgenstern L, Qureshi A, Rosenwasser R, Scott P, Wijdicks E, (2007). Guidelines for the Early Management of Adults With Ischemic Stroke. Stroke, 38(5), 1655–1711.

[4] Bickerton M, Pooler A (2019). Misplaced ECG electrodes and the need for continuing training. British Journal of Cardiac Nursing, 14(3), 123–132.

[5] Hannun A, Rajpurkar P, Haghpanahi M, Tison G, Bourn C, Turakhia M,(2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine. 2019 Jan 1; 25(1):65–9.

[6] Shashikumar S, Clifford G, Shah A, Nemati S (2018). Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2018 Jul 19; 715–23.

[7] Lecun Y, Bengio Y, Hinton G (2015). Deep learning. Vol. 521, Nature. Nature Publishing Group; 2015. p. 436–44.

[8] Finn C, Abbell P, Levine S (2017). Model-agnostic meta-learning for fast adaptation of deep networks. Proceedings of the 34th International Conference on Machine Learning

[9] Finn C, Stanford CS 330 Lecture Notes, 2019.

[10] Clifford G, Liu C, Moody B, Lehman L, Silva I, Li Q, Johnson A, Mark R (2017). AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017. Computing in Cardiology, 44, 1–4.

[11] Goldberger A, Amaral L, Glass L, et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation. 101 (23), pp.e215–e220.

[12] Liu F, Liu* C, Zhao L, Zhang X, et al. (2018) An open access database for evaluating the algorithms of ECG rhythm and morphology abnormal detection. Journal of Medical Imaging and Health Informatics, 2018, 8(7): 1368–1373.

[13] Moody G, Mark R (2001). The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209)

[14] Krizhevsky A, Sutskever I, et al. (2012). Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25,1097–1105.

[15] Kreyszig E (1979). Advanced Engineering Mathematics. 1979. 880.

[16] Glantz S, et al (2016). Primer of Applied Regression and Analysis of Variance. 3rd ed. McGraw Hill.

[17] Hannun A, Rajpurkar P, Haghpanahi M, Tison G, Bourn C, Turakhia M, et al. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine. 2019 Jan 1;25(1):65–9.

[18] Sasaki Y, Fellow R, (2007). The truth of the F-measure.

[19] Fawcett, T. (2005). An introduction to ROC analysis.

[20] van Steenkiste G, van Loon G, Crevecoeur G (2020). Transfer Learning in ECG Classification from Human to Horse Using a Novel Parallel Neural Network Architecture. Scientific Reports 2020 10:1. 2020 Jan 13;10(1):1–12.

[21] Sun Q, Liu Y, Chua T, and Schiele B. (2019). Meta-transfer learning for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 403-412).

[22] Erhan D, Courville A, Bengio Y, Vincent P. (2010, March). Why does unsupervised pre-training help deep learning?. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (pp. 201-208). JMLR Workshop and Conference Proceedings.