2024 6th International Conference on Electronics and Communication, Network and Computer Technology (ECNCT) | 979-8-3503-6614-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ECNCT63103.2024.10704467

Towards Continually Evolved Heart Sound Classifiers with Class-Incremental Learning

Zhongyu Luo[1]
[1]School of Electronic Information Engineering
Foshan University
Foshan, China
e-mail: luozhongyu2799@163.com

Sijie Wang[2]
[2]School of Electronic Information Engineering
Foshan University
Foshan, China
e-mail: heyyuan03@foxmail.com

Junjie He[3]
[3]School of Electronic Information Engineering
Foshan University
Foshan, China
e-mail: graversama@outlook.com

Aiguo Wang[4,*]
[4]School of Electronic Information Engineering
Foshan University
Foshan, China
[*] Corresponding author: agwang@fosu.edu.cn

Abstract—**Heart sound analysis provides a non-invasive and cost-effective method for identifying and studying various heart diseases. However, when new categories emerge in dynamic environments, traditional heart sound classification models generally require retraining and suffer from catastrophic forgetting. To address this, this study proposes a heart sound classification model via class-incremental learning (HSCIL) that integrates knowledge distillation, data replay, and dynamic network expansion techniques to reuse an existing classifier. Specifically, residual temporal convolutional networks are used as the backbone and dynamically expanded to include new classes. Additionally, MFCC features rather than raw heart sound samples are used for data replay. Knowledge distillation is also adopted to maintain consistency between outputs of the new model and old model. Finally, HSCIL is evaluated on a dataset with eight categories and compared with other four classification models. Experimental results show that HSCIL achieves 97.03% accuracy and 1.11% forgetting rate.**

Keywords-heart sound classification; incremental learning; knowledge distillation; data replay

I. Introduction

Cardiovascular disease (CVD) remains the leading cause of death worldwide, claiming over 17 million lives each year. Therefore, early detection and timely intervention of CVD are crucial for alleviating the burden of these diseases. Heart sound analysis is a non-invasive and cost-effective diagnostic tool in revealing heart conditions such as arrhythmias, valvular diseases, and congenital heart defects [1]. However, the accuracy of heart sound analysis heavily depends on the clinician's expertise, which is often limited, especially in resource-constrained settings. Therefore, automating this process using artificial intelligence has great significance.

Though recent advances in machine learning and deep learning have significantly improved automatic heart sound classification [2], current methods face challenges when applied to dynamic and open environments (e.g., new heart sound categories may emerge over time). Considering that most of existing heart signal analysis models are static and have fixed output categories, if a model is trained to recognize "normal" and "arrhythmic" heart sounds, it cannot recognize new categories such as "coronary artery disease" without retraining. Retraining with new data not only requires substantial data but also risks catastrophic forgetting (that is, the model forgets previously learned knowledge) [3]. To address these challenges, incremental learning offers a promising solution, allowing the model to gradually adapt to new classes while retaining knowledge of previously learned classes. This approach mimics human learning, enabling the integration of new information without extensive retraining.

Class-incremental learning methods can be broadly categorized into regularization-based, data replay-based, knowledge distillation-based, and network architecture-based methods. Regularization-based methods add constraints to loss functions to prevent models from forgetting old tasks. For example, elastic weight consolidation (EWC) reduces catastrophic forgetting by estimating the importance of each parameter and penalizing significant changes to these parameters during training on new tasks [4]. Data replay stores and replays subsets of old data during training on new tasks to refresh the model's memory of prior knowledge. For example, Rebuffi et al. proposes the incremental classifier and representation learning (iCaRL) method that uses a nearest sample classification strategy and stores old class samples to prevent forgetting [5]. Matthias et al. combines the nearest center classifier and data replay to address class incremental learning tasks in the presence of concept drift [6]. Knowledge distillation-based methods maintain the consistency of the outputs between the new and old classification models while handling new tasks. For example, learning without forgetting (LwF) uses knowledge distillation to retain prior knowledge while learning new tasks [7]. Learning without memorizing (LwM) uses attention mechanisms to select important features [8]. Network architecture-based methods modify network architecture. For example, dynamic expandable networks (DEN) selectively grow network capacity by adding neurons or layers as needed to effectively learn new tasks. DEN dynamically determines its network capacity during training to learn compact overlapping knowledge-sharing structures between tasks, preventing semantic drift by splitting and replicating units and adding timestamps [9].

Despite its great potential, there is currently a lack of research on incremental learning for heart sound classification. Motivated by previous studies, we in this study propose a class-incremental learning based continually evolved heart sound classification model, named HSCIL, that combines network expansion, knowledge distillation and data replay to reuse an existing classifier. The main contributions of this study are as follows:

(1) We propose a class-incremental learning framework for heart sound classification that combines knowledge distillation, data replay, and dynamic network expansion to learn new categories of heart sound signals while preventing catastrophic forgetting. Particularly, MFCC features are used as an example set of old classes during the incremental learning phase, and knowledge distillation is employed to ensure consistency between outputs of the new model and the old model.

(2) The effectiveness of the HSCIL model is validated through extensive experiments. Experimental results show that HSCIL outperforms its competitors in terms of accuracy and forgetting rate.

II. Method

Fig. 1 presents the proposed heart sound classification model via class-incremental learning (HSCIL), where knowledge distillation [7], data replay [5], and dynamically expanding network structure [9] are integrated into the model. First, during the incremental learning process, we maintain the model's memory of old classes by constructing and using an old class exemplar set (i.e., data replay, DR). For each old class, we do not store raw samples of the old class but store Mel-frequency cepstral coefficients (MFCC) features. Particularly, to facilitate feature learning, we extract first-order difference (ΔMFCC) and second-order difference (Δ^2MFCC) of MFCC to encode the heart sound signals. Then, the residual temporal convolutional network (TCN) is used as the backbone network to better learn complex spatial-temporal dependencies among the heart sound signals.

Second, to maintain consistency between outputs of the new model and old model, we combine the cross-entropy loss (L_{ce}) and distillation loss (L_{kl}). The purpose of distillation loss is to force the student model (new model) to learn new tasks while retaining knowledge of old tasks by introducing the soft targets of the teacher model (old model) during training. The loss L of a training sample x is the weighted sum of the classification loss L_{ce} and distillation loss L_{kl}.

$$L = \alpha \times L_{kl} + (1-\alpha) \times L_{ce} \tag{1}$$

where α is a weighting factor to balance the distillation loss and cross-entropy loss.

L_{ce} is the standard cross-entropy loss, as shown in (2).

$$L_{ce}(x) = -\sum_{i=1}^{|C|} y_i \log(p_i) \tag{2}$$

, where C is the set of observed classes so far, y is the one-hot ground-truth label, and p is the prediction probability.

L_{kl} measures the difference between prediction distributions of the student model and the teacher model, as shown in (3).

$$L_{kl} = \tau^2 \times \sum_{i=1}^{|C_{old}|} p_i(x) \log \frac{p_i(x)}{q_i(x)} \tag{3}$$

$$p_i(x) = \exp(\frac{\tilde{z}_i}{\tau}) \Big/ \sum_j \exp(\frac{\tilde{z}_j}{\tau}) \tag{4}$$

$$q_i(x) = \exp(\frac{z_i}{\tau}) \Big/ \sum_j \exp(\frac{z_j}{\tau}) \tag{5}$$

, where C_{old} is the set of old classes, and the temperature parameter τ is used to soften the probability distribution. The higher the temperature is, the smoother the distribution is. p_i is the prediction distribution of teacher model and q_i is the prediction distribution of student model. The use of τ^2 is to balance the loss function and restore the impact of the scaled temperature parameter on the gradient.

Third, considering that new heart sound classes are continuously learned, we expand the network structure by adding a new layer for every N new classes to improve its feature representation capability ($N = 3$ in our study).

III. Evaluation and Results

A. Datasets and data preprocessing

To evaluate the effectiveness of the proposed model, we compile and merge publicly available heart sound datasets along with our private heart sound dataset. These public datasets include the 2012 Classifying Heart Sounds Challenge sponsored by PASCAL, the normal and abnormal heart sound library in Frontiers in Bioscience, the Murmur Quiz database on the heart sound auscultation training website, and the Yaseen dataset. The following steps are performed to preprocess heart sound signals. First, the raw signals are processed using a fifth-order Butterworth filter with a frequency range of 25 to 400 Hz to smooth the signals. Next, the filtered signals are downsampled to 2000 Hz and segmented into 5-second audio segments using a sliding window method with a window step size of 2.5 seconds. Then, MFCC and its first-order and second-order differences are extracted from each segment to form feature vectors. The experimental dataset, as shown in Table 1, consists of 8 types of heart sound signals, including 1040 Normal samples, 255 mitral stenosis (MS), 307 mitral regurgitation (MR), 360 aortic stenosis (AS), 271 mitral valve prolapse (MVP), 246 Atrial Fibrillation (AF), 208 Extra Heart Sounds (EHS), and 356 Murmur cases.

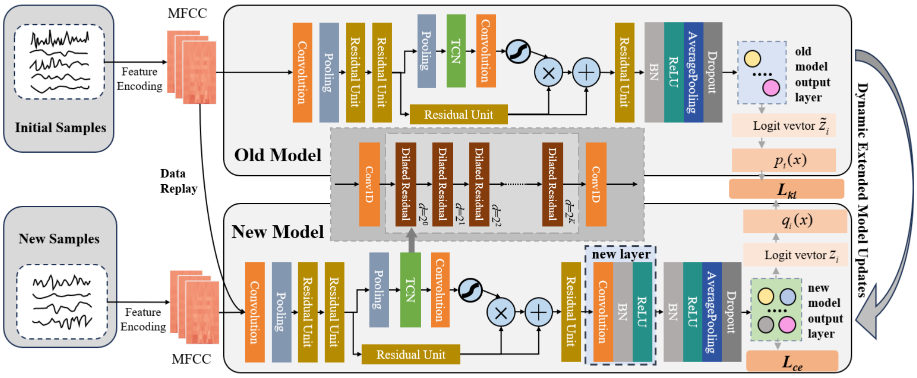


Figure 1. Network structure of the proposed model.

TABLE 1. STATISTICS OF THE HEART SOUND DATASETS

Classes	Normal	MS	MR	AS	MVP	AF	EHS	Murmur
Total	1040	255	307	360	271	246	208	356

B. *Experimental setup*

Four classic class-incremental learning algorithms are used as competitors: Finetune (FT) that directly fine-tunes the model, iCaRL, Learning without Forgetting (LwF), and Retraining that retrains the model. These algorithms use resnet32 as the classifier. We train the models on a server equipped with an NVIDIA GeForce RTX 4090 GPU and an Intel (R) Core (TM) i7-13700KF 3.42GHz CPU. The Adam optimizer is used to update the network parameters, which are initialized by the Xavier normal initializer. The batch size is empirically set to 32, and the learning rates for both the old model and the main model are 0.001. The network is trained from scratch for 50 epochs. For data replay, we store ten exemplar samples for each category. All categories in the dataset are shuffled randomly, with 2 categories in the first task, and the remaining categories are for incremental learning with a step size of 1 (that is, only one new class arrives for each incremental learning). Performance evaluation metrics are average incremental accuracy $\bar{A}$, average incremental forgetting rate $\bar{F}$, incremental accuracy curve, and forgetting curve.

$$\bar{A} = \frac{1}{T}\sum_{i=1}^{T} A_i \tag{6}$$

, where A_t is the incremental accuracy in the t^{th} incremental stage and T is the total number of incremental learning stages. Incremental accuracy refers to the classification accuracy of the current model on all seen classes.

$$\bar{F} = \frac{1}{T}\sum_{i=1}^{T} F_i \tag{7}$$

$$F_t = \frac{1}{t-1}\sum_{i=1}^{t-1} f_t^i \tag{8}$$

, where F_t is the forgetting rate in the t^{th} incremental stage. $f_t^i = \max_{t \in 1,\cdots,k-1}(a_{t,i} - a_{k,i})$ refers to the definition of forgetting for the i^{th} task after learning k tasks ($i < k$), where $a_{m,n}$ is the accuracy of the model on the n^{th} task after completing the m^{th} task.

C. *Experimental results*

In this section, we first present the experimental results comparing our method with baseline methods and perform an ablation study on the components of HSCIL.

1) *Comparison experiment*

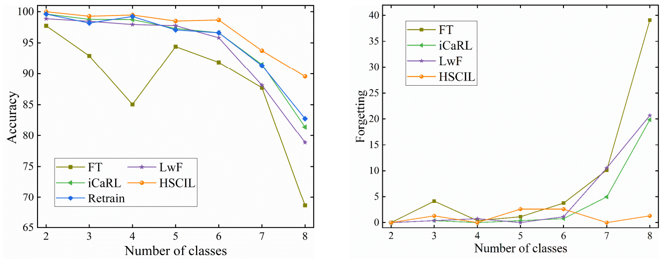
In this study, we compare the performance of our method with other four classic methods. Table 2 lists the average incremental accuracy and average incremental forgetting rate, with the best results shown in bold. Fig. 2 reports the setting where two classes are taken as the first task, and the remaining classes are divided into incremental tasks of one class each.

From Table 2 and Fig. 2, we can see that using Finetune to fine-tune in class incremental learning tasks leads to the model focusing only on new class information while ignoring old class information, with average accuracy and average forgetting rates of 88.33% and 8.38%, respectively. That is, Finetune suffers from catastrophic forgetting, resulting in the worst performance. LwF introduces knowledge distillation loss during the model updating process, establishing supervision of the new model by the old model to resist catastrophic forgetting. iCaRL uses exemplar sets on the basis of LwF to augment the training set samples at each stage and better utilizes the preserved old class samples. The average incremental accuracy of LwF and iCaRL methods is 93.69% and 94.86%, respectively, which greatly improves the model's performance compared to Finetune, demonstrating the effectiveness of data replay and knowledge distillation loss in incremental learning processes. Retraining the prediction model with both new class sample data and existing class sample data achieves the average accuracy of 94.98%, higher than the other three methods at the cost of higher training time. Fourth, we can see that HSCIL performs better than iCaRL. This is mainly because HSCIL combines knowledge distillation loss and dynamically adjusting the model structure.

In the forgetting rate curve in Fig. 2(b), the stability of HSCIL in incremental tasks can be seen. The average accuracy and average forgetting rate in incremental tasks are 97.03% and 1.11%, respectively, achieving the best performance in all experimental settings. This is mainly because HSCIL maintains the model's memory of old classes, uses distillation loss to guide the new model's output, and dynamically adjusts the network structure to enhance its feature representation capability.

TABLE 2. EXPERIMENTAL RESULTS OF DIFFERENT ALGORITHMS

Model	Average accuracy (%)	Average forgetting rate (%)	training time (min)
FT	88.33	8.38	75.27
iCaRL	94.86	3.76	91.08
LwF	93.69	4.78	52.84
HSCIL (ours)	97.03	1.11	102.62
Retraining	94.98	-	196.84



(a) Accuracy

(b) Forgetting

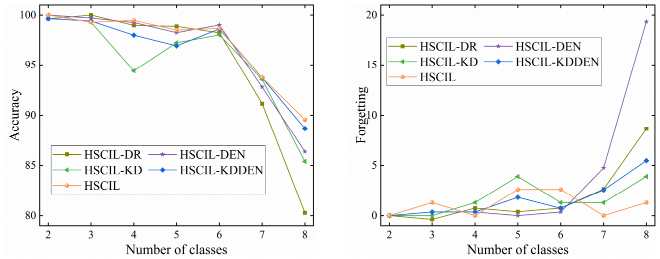
Figure 2. Performance comparison of different algorithms.

2) *Ablation study*

In this section, we perform ablation experiments to evaluate the effectiveness of each component of HSCIL: using only data replay (HSCIL-DR), using only knowledge distillation (HSCIL-KD), and using only dynamically expanding network (HSCIL-DEN) structure. For the ablation experiments, we use the temporal convolutional network as the backbone. Besides, considering that data may be unavailable due to data loss or privacy issues in real-world environments, we can only use knowledge distillation and dynamically expanding network (HSCIL-KDDEN) structure.

TABLE 3. PERFORMANCE COMPARISON OF DIFFERENT COMPONENTS OF HSCIL

Model	Average accuracy (%)	Average forgetting rate (%)
HSCIL-DR	95.30	1.83
HSCIL-KD	95.41	1.67
HSCIL-DEN	95.57	2.85
HSCIL-KDDEN	96.42	1.62
HSCIL	97.03	1.11



(a) Accuracy

(b) Forgetting

Figure 3. Performance comparison of different components of HSCIL.

Table 3 and Fig. 3 give the experimental results. These results indicate that HSCIL achieves the best performance in terms of average accuracy and average forgetting rate. For example, HSCIL-DR achieves average accuracy of 95.30%, lower than the performance of other HSCIL components. Second, HSCIL achieves average accuracy of 97.03%, higher than 95.30% of HSCIL-DR, 95.41% of HSCIL-KD, and 95.57% of HSCIL-DEN. Third, considering practical environments, HSCIL-KDDEN is our best choice, with average accuracy 0.61% lower than HSCIL and average forgetting rate 0.51% higher than HSCIL. We can also see that HSCIL-KDDEN performs stably in incremental tasks. The average accuracy of HSCIL-KDDEN is 1.01%, higher than HSCIL-KD; the average forgetting rate is 1.23%, lower than HSCIL-DEN.

IV. CONCLUSION

To better adapt existing heart sound analysis models to an "open world", we propose a heart sound classification via class-incremental learning model that combines knowledge distillation, data replay, and dynamic network expansion techniques towards better learn new heart sound categories of heart sound signals while preventing catastrophic forgetting. Specifically, the residual temporal convolutional network is used and the MFCC features are used as the example set of old classes for data replay. Knowledge distillation loss is also employed and the network structure is dynamically expanded to enhance its feature representation capabilities. Finally, comparative experiments are conducted and compared with other four models. Results show that the proposed model outperforms its competitors.

ACKNOWLEDGMENT

We thank the anonymous reviewers for their constructive comments.

REFERENCES

[1] Netto, A.N., Abraham, L., Philip S. (2024) HBNET: A blended ensemble model for the detection of cardiovascular anomalies using phonocardiogram. Technol. Health Care, 1-21.

[2] Mao, S., Sejdić, E. (2022) A review of recurrent neural network-based methods in computational physiology. IEEE Trans. Neural Netw. Learn. Syst., 34: 6983-7003.

[3] Zhang, J., Zhang, J., Ghosh, S., Li, D., Tasci, S., Heck, L., Zhang, H. Kuo, C.C.J. (2020) Class-incremental learning via deep model consolidation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. Snowmass Village. pp. 1131-1140.

[4] Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G. Rusu, A.A., Milan, K., Quan, J. Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran D., Hadsell, R. (2017) Overcoming catastrophic forgetting in neural networks. Proc. Nat. Acad. Sci., 114: 3521−3526.

[5] Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H. (2017) iCaRL: In cremental classifier and representation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii. pp. 5533−5542.

[6] Rolnick, D., Ahuja, A., Schwarz, J., Lillicrap, T., Wayne, G. (2019) Experience replay for continual learning. Proc. Adv. Neural Inf. Process. Syst., 32: 350-360.

[7] Li, Z.Z., Hoiem, D. (2018) Learning without forgetting. IEEE Trans. Pattern Anal. Mach. Intell., 40,: 2935−2947.

[8] Dhar, P., Singh, R.V., Peng, K.C., Wu, Z.Y., Chellappa, R. (2019) Learning without memorizing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach. pp. 5133−5141.

[9] Yoon, J., Yang, E., Lee, J., Hwang, S.J. (2017) Lifelong learning with dynamically expandable networks. arXiv preprint arXiv:1708.01547.