EMPIRICAL RESEARCH

Battling with the low-resource condition for snore sound recognition: introducing a meta-learning strategy

Jingtan Li[1,2†], Mengkai Sun[1,2†], Zhonghao Zhao[1,2], Xingcan Li[1,2], Gaigai Li[1,2], Chen Wu[1,2], Kun Qian[1,2*], Bin Hu[1,2*], Yoshiharu Yamamoto[3] and Björn W. Schuller[4,5]

Abstract

Snoring affects 57 % of men, 40 % of women, and 27 % of children in the USA. Besides, snoring is highly correlated with obstructive sleep apnoea (OSA), which is characterised by loud and frequent snoring. OSA is also closely associated with various life-threatening diseases such as sudden cardiac arrest and is regarded as a grave medical ailment. Preliminary studies have shown that in the USA, OSA affects over 34 % of men and 14 % of women. In recent years, polysomnography has increasingly been used to diagnose OSA. However, due to its drawbacks such as being time-consuming and costly, intelligent audio analysis of snoring has emerged as an alternative method. Considering the higher demand for identifying the excitation location of snoring in clinical practice, we utilised the Munich-Passau Snore Sound Corpus (MPSSC) snoring database which classifies the snoring excitation location into four categories. Nonetheless, the problem of small samples remains in the MPSSC database due to factors such as privacy concerns and difficulties in accurate labelling. In fact, accurately labelled medical data that can be used for machine learning is often scarce, especially for rare diseases. In view of this, Model-Agnostic Meta-Learning (MAML), a small sample method based on meta-learning, is used to classify snore signals with less resources in this work. The experimental results indicate that even when using only the ESC-50 dataset (non-snoring sound signals) as the data for meta-training, we are able to achieve an unweighted average recall of 60.2 % on the test dataset after fine-tuning on just 36 instances of snoring from the development part of the MPSSC dataset. While our results only exceed the baseline by 4.4 %, they still demonstrate that even with fine-tuning on a few instances of snoring, our model can outperform the baseline. This implies that the MAML algorithm can effectively tackle the low-resource problem even with limited data resources.

Keywords Computer audition, Snore sound classification, Meta-learning, Low-resource, Obstructive sleep apnoea, Digital health

†Jingtan Li and Mengkai Sun contributed equally to this work.

*Correspondence:
Kun Qian
qian@bit.edu.cn
Bin Hu
bh@bit.edu.cn
Full list of author information is available at the end of the article

Li *et al. EURASIP Journal on Audio, Speech, and Music Processing*   (2023) 2023:43

Page 2 of 9

1 Introduction

In the UK, more than 40 % of people frequently snore [1]. Moreover, research has indicated that snoring significantly affects the quality of the bed partner's sleep quality [2]. Related research shows that snoring is closely associated with obstructive sleep apnoea (OSA) [3]. Fifteen million American adults suffer from OSA by estimation [4]. The presence of comorbidities such as excessive daytime sleepiness and heightened risk of cardiovascular disease are commonly associated with this disorder [5]. Furthermore, obesity, increased risk of mental illness, endocrine system imbalances, and sexual dysfunction have all been confirmed to be associated with OSA [6–9]. At present, the clinical diagnosis of OSA heavily depends on the analysis of monitoring data obtained from polysomnography (PSG) and the personal expertise of physicians [10]. On the one hand, despite its unique value in the clinical diagnosis of OSA, PSG has limitations including high cost, inconvenience in terms of portability, high application difficulty, and unsuitability for large-scale population screening, as well as its primary focus on snoring loudness and frequency rather than analysis of snoring characteristics [11, 12]. Snoring, on the other hand, can provide detailed information on a person's respiratory status, highlighting its potential for use as a valuable diagnostic tool in OSA [13]. Early studies demonstrate that acoustic-based methods can be used to diagnose respiratory disorders such as OSA [14–16]. With the development of artificial intelligence, machine learning (ML), and deep learning (DL), algorithms have been shown to be effective in audio signal processing [17–19]. However, accurately determining the location of snore excitation is essential for the clinical surgical management of OSA [20, 21]. In view of this, an open snoring sound dataset, the Munich-Passau Snore Sound Corpus (MPSSC) [22], that classifies snoring into four different types, naming velum (V), oropharyngeal (O), tongue (T), and epiglottis (E), is considered in this work. Figure 1 illustrates the locations of four types of snoring in the upper airway.

Nonetheless, due to the private nature of medical data such as snoring and the difficulty in accurately labelling, medical data for machine learning remains scarce [23]. Therefore, this paper considers using the Model-Agnostic Meta-Learning (MAML) algorithm (a meta-learning strategy), which has already achieved success in small-sample image processing, to tackle the issue of scarce medical data [24]. The aim of MAML is to learn the best initialisation parameter θ from tasks constructed from the training set, which can be quickly adapted to different new tasks [25]. This research focuses on one question: how to use less data for training, but still achieve good performance on the test set. Therefore, in this study, the
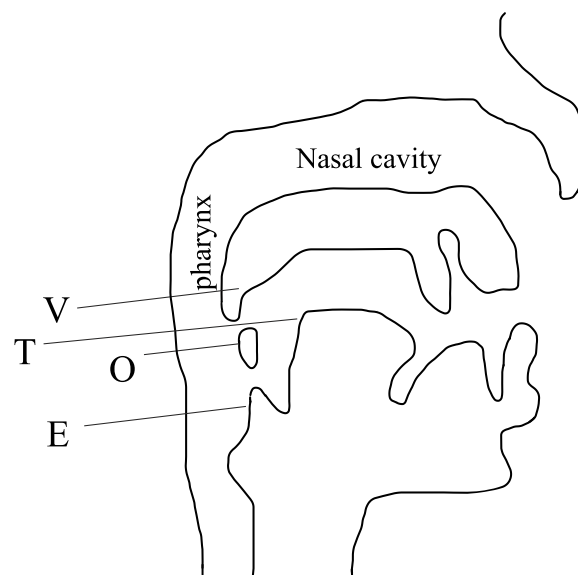

Fig. 1 A diagram of the upper airway showing the location where VOTE snoring is triggered

ESC-50 sound dataset and the MiniImageNet dataset are used as the training data. Then, we test and compare the model on the test partition of the MPSSC dataset.

The main contributions of this work are as follows: First, we are able to achieve a UAR of 60.2 % that exceeds the benchmark of 55.8 % [22]. Second, the performance improvement was achieved by using a limited amount of fine-tuning data (only 36 snoring sounds, less than 7 % of the original training dataset). The paper's structure is organised as follows: Section 2 summarises previous research related to this work. Section 3 provides a detailed description of the MPSSC dataset and the methods used in this study. Section 4 presents the findings and outcomes obtained from the conducted experiments. Section 5 shows the limitations and prospects for future research of this paper. Finally, we conclude our work in the "Conclusion" section.

2 Related work

Cosztolya et al. [26] use the 'ComParE' features of the openSMILE toolkit and a Support Vector Machine (SVM) classifier on MPSSC achieving a UAR of 62 %. Amiriparian et al. [27] propose employing deep spectrum features and an SVM for snoring classification and achieve a UAR of 67 %. Qian et al. [28] improve the low-level wavelet features extracted from snoring data with a bag-of-audio-words approach and get a UAR of 69.4 %. Demir et al. [29] obtained a UAR of 72.6 % by utilising a histogram of local binary patterns and a histogram of oriented gradients to describe snore sounds. Li et al. [30] first treat the MPSSC dataset

Li *et al. EURASIP Journal on Audio, Speech, and Music Processing* (2023) 2023:43

Page 3 of 9

as a few-shot learning task, and they apply one of the meta-learning strategies named prototypical network to the MPSSC dataset and yield a UAR of 77.13 %.

In recent years, meta-learning has achieved significant breakthroughs in the domain of acoustic events. Shi et al. [31] found that meta-learning models can achieve superior performance in acoustic event detection compared to supervised baselines. By incorporating self-supervised learning with MAML (a meta-learning strategy), Lemkhenter et al. [32] significantly improved the performance of the sleep scoring model compared to standard supervised learning. Heggan et al. [33] demonstrated that gradient-based meta-learning methods consistently outperformed baseline methods across seven audio datasets.

In comparison to Li et al. [30], we did not use the test set from the original MPSSC partition to extract snoring sounds as the support set for fine-tuning during testing. Instead, we used the development set from the original MPSSC partition to extract snoring sounds as the support set. This ensures that the testing data was not used before testing, leading to more accurate test results. Therefore, there is a considerable difference between our results and that of Li et al.

Although previous studies have achieved promising UAR results on the MPSSC dataset, they have all utilised the entire MPSSC dataset for training. Hence, we are contemplating utilising a smaller quantity of MPSSC training data to train the model.

3 Datasets and Methods

3.1 Datasets

3.1.1 Munich-Passau Snore Sound Corpus

The MPSSC dataset is a publicly available collection of snore sounds from 219 subjects who underwent drug-induced sleep endoscopy (DISE) at three different medical centres. Snoring can be classified into four distinct types, namely velum (V), oropharyngeal (O), tongue (T), and epiglottis (E), based on the respective locations of their excitation within the upper airway [22]. In MPSSC, the number of T-type and E-type snoring samples is less than that of V-type and O-type snoring samples in each division (Train: V: 168, O: 76, T: 8, E: 30). The detailed information of the MPSSC dataset is presented in Table 1. In this paper, we only use 36 snoring samples from the development portion of MPSSC as the fine-tuning support data during the meta-testing, while the remaining 529 snoring samples (including training and development) are not used. Meanwhile, we use the test portion of the MPSSC dataset with the original split to test our model in this work.

Table 1 Detailed information on the MPSSC dataset. The snoring sounds of 219 subjects were equitably allocated among three partitions, specifically the training, development, and testing sets

	Train	Dev	Test	$\sum$
V	168	161	155	484
O	76	75	65	216
T	8	15	16	39
E	30	32	27	89
$\sum$	282	283	263	828
No. of patients	73	73	73	219

3.1.2 ESC-50

The ESC-50 dataset encompasses 2000 environmental sound recordings that have been labelled with corresponding tags. Each recording has a duration of 5 s and can be assigned to one of 50 distinct semantic classes, with 40 exemplary instances per class [34]. We employ Mel-spectrograms to extract the acoustic features of sound from the ESC-50 dataset. For model training, we utilise 35 out of the 50 distinct sound categories present in the ESC-50 dataset as our training data, while the remaining 15 categories serve as our validation data.

3.1.3 MinilmageNet

The DeepMind team has used the MiniImageNet dataset for few-shot learning research for the first time [35]. Therefore, we aim to investigate whether employing non-audio datasets such as MiniImageNet for training purposes can still yield commendable results on new tasks, such as those presented in the MPSSC dataset. By doing so, we hope to demonstrate the universality and robustness of MAML in the field of audio classification.

3.2 Methods

3.2.1 Mel-spectrogram

Mel spectrograms provide visualised information on the auditory system of human hearing, making them a viable input for convolutional neural networks [36].

In view of this, we extract Mel spectrograms from four distinct snoring types, as well as the ESC-50 sound dataset. Due to the fact that the image size in the MiniImageNet dataset is 84×84×3, we set the size of the spectrograms extracted from the MPSSC and ESC-50 datasets also to 84×84×3, in order to maintain consistency of image size. The spectrograms of the different categories of snoring sounds are displayed in Fig. 2. To preserve the genuine and efficacious portions of the spectrogram, we implement a cropping mechanism, which
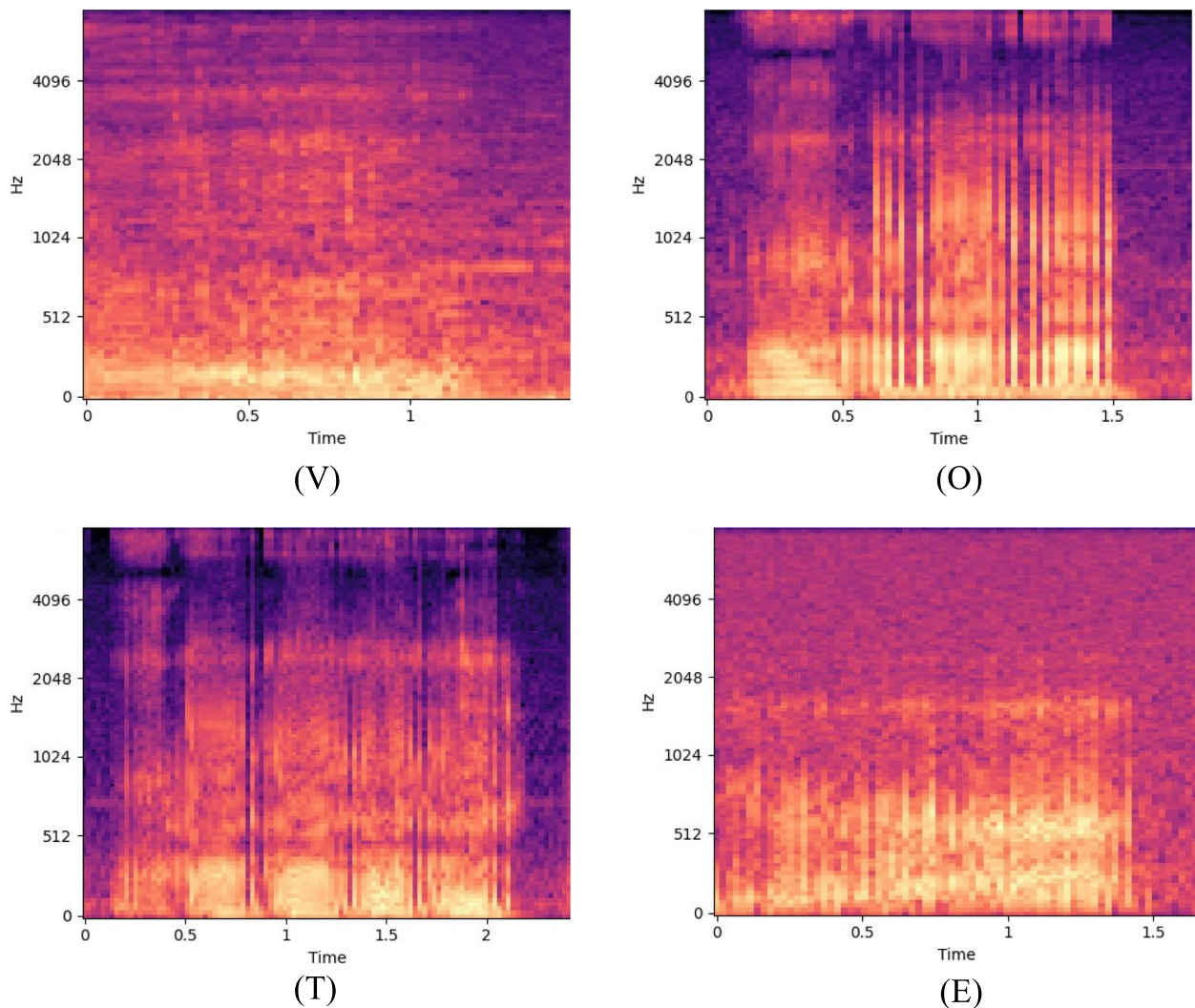
Fig. 2 Spectrograms of the four distinct types of snoring sounds V, O, T, and E. By analysing the spectrograms, we can extract distinctive features of the four different types of snoring sounds

involves removing the upper segment of the spectrogram image beyond the 10,000 Hz threshold.

3.2.2 Model-Agnostic Meta-Learning

MAML divides the train set and test set into N-way, K-shot, and Q-query problems. This indicates that N categories are randomly selected from the data set each time, and $K + Q$ samples are selected for each category as one task (in this paper, all random functions utilise a Python random function with a seed value 400), which means that each task contains $N \times (K + Q)$ sampled data [25]. Specifically, we randomly initialise a parameter θ and assign this parameter θ to each task$_i$ in a batch as θ_i. In each task$_i$, we update the parameter θ_i using K support images for each

task (using the inner learning rate) to obtain θ_i' – the computation formula is presented in Eq. (1).

$$\theta_i' = \theta - \alpha \nabla_\theta L_{\tau_i}(f_\theta), \tag{1}$$

where L_{T_i} denotes the loss obtained on the support set of task$_i$ by the model and α represents the inner learning rate. Then, we test on Q query images and obtain the loss$_i$ for this task. After that, we sum up the batch of loss$_i$ to obtain the loss and use this loss to update the outer parameter θ to θ' (using the outer learning rate). The detailed calculation formula is given in Eq. (2) below. β is the outer learning rate.

$$\theta_i' \leftarrow \theta - \beta \nabla_\theta \sum_{\tau_i \sim p(\tau)} L_{\tau_i}(f_{\theta_i'}). \tag{2}$$

For the next batch of tasks, we initialise the parameters for each task using θ' and repeat the above process until completion. The framework of MAML used in this paper is shown in Fig. 3.

By training and adjusting model parameters on one task distribution within a given dataset, the MAML algorithm enables the resultant model to quickly adapt to new tasks through one or a few updates on the support set. This also means that the MAML algorithm can adapt to different new learning tasks with greater universality and robustness.

In this paper, as the MPSSC dataset is a four-classification problem, we set the N of N-way to 4 in MAML. Additionally, taking into account the number of samples in the T category (the smaller category) in the test set, we set the parameter values to $K = 5$ and $Q = 9$ (the value of K-shot and Q-query mentioned above). Therefore, for each training task, we randomly select 14 images from each of the 4 selected categories, totalling 56 images within each task distribution.

During the meta-testing phase, we got 9 images for each category of snoring sounds and fine-tuned the meta-trained model using a total of 36 snoring samples (4 categories * 9 samples each) as the support set for meta-testing. Meanwhile, the query set comprised the testing portion of the entire MPSSC dataset.

3.2.3 Experiment

In this work, we have devised experiments in two distinct directions, and the detailed demonstration of experiments is presented in Table 2. The first set use 64 classes from MiniImageNet that are unrelated to snore sound as the meta-training data, with 16 classes as meta-validation, and the MPSSC test data was used for meta-testing. In the second set, we extract Mel-spectrograms from different sound data in the ESC-50 dataset, using 35 classes as meta-training, and 15 classes as meta-validation. The meta-testing was performed on the original test partition of the MPSSC data, similar to MiniImageNet. To ensure the test data is only used for testing, during the meta-testing, we apply the 36 snoring samples from the development set of the original MPSSC dataset to the support set of the new snoring classification task and use the entire test set of the MPSSC dataset's original split for prediction on the query set. In the two directions, we use a four-layer convolutional neural network with a

Table 2 The training data used includes MiniImageNet and ESC-50, with testing conducted on the original test set of MPSSC

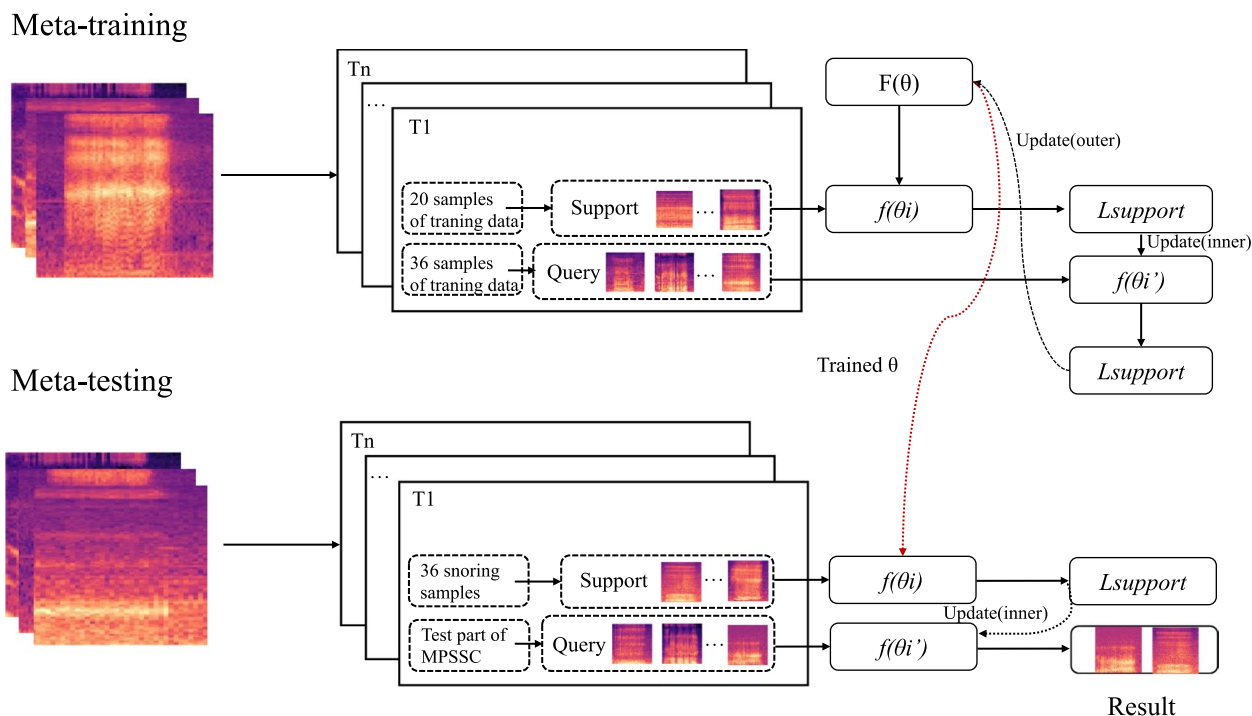| Training set	Development set	Testing set
MiniImageNet	MiniImageNet	MPSSC
ESC-50	ESC-50	MPSSC

Fig. 3 The framework of MAML is employed in this work. After training, the optimal model parameters θ are obtained and used as the initial parameters for the meta-testing stage

Li *et al. EURASIP Journal on Audio, Speech, and Music Processing*   (2023) 2023:43

Page 6 of 9

ReLU activation function and an Adam optimiser, with a meta-learning inner learning rate of 0.01 and an outer learning rate of 0.001. Furthermore, we utilised an FFT window size of 1024, a frameshift size of 512, a quantity of 128 Mel filters, and a power of 2 when computing the Mel spectrogram of the audio. The detailed architecture of the CNN utilised in this paper is illustrated in Fig. 4.

4 Experimental results

As the MPSSC dataset is imbalanced, we use UAR to assess the performance of the model. As mentioned above, we define $N = 4$ as a four-class classification problem. To compute the UAR, we calculate the recall for each class and obtain the UAR by computing the unweighted average of the recall of the four classes. Specifically, the formula for calculating UAR is defined as follows:

$$UAR = \frac{\sum_{i=1}^{N_c} Recall_i}{N_c}, \tag{3}$$

The formula for recall in Eq.(3) refers to Eq. (4).

$$Recall = \frac{TP}{TP + FN}, \tag{4}$$

where TP refers to the number of samples correctly predicted as positive by the model, while FN refers to the number of samples that the model should have predicted as positive but were incorrectly predicted as negative.

According to the calculated UAR, the experimental results are shown in Table 3. We have observed that, although there are considerable differences between the MiniImageNet data set and the MPSSC snore

Table 3 Performing meta-training using the MiniImageNet and ESC-50 datasets, respectively, and conducting meta-testing on the test set of the MPSSC dataset using the original partition

Meta-training dataset	UAR
MiniImageNet	41.2%
ESC-50	60.2%

spectrogram, the results show that MAML still learns some features from the MiniImageNet data set and achieves 41.2 % UAR, which exceeds the chance level of 25.0 % by 16.2 %. This result is also confirmed in the study of Heggan et al. [33]. In addition, the other set of experiments using the ESC-50 sound dataset's mel spectrogram as training achieved a UAR of 60.2 % on the test set. The confusion matrix is displayed in Fig. 5.

Moreover, the UAR of 60.2 % on the test data indicates that we have surpassed the MPSSC baseline using only 36 instances of non-test snoring data. In other words, we have successfully addressed the low-resource challenge for snoring detection. This supports that the MAML algorithm can learn how to learn through other tasks and can fine-tune with a small amount of labelled snoring data for good performance on unlabelled data. Furthermore, using non-snoring sounds for training also indicates that our model has better generalisability.

5 Discussion

In this study, we achieved a UAR of 60.2 % on the MPSSC test set using ESC-50 as the meta-training data. This indicates our success in addressing the low-resource
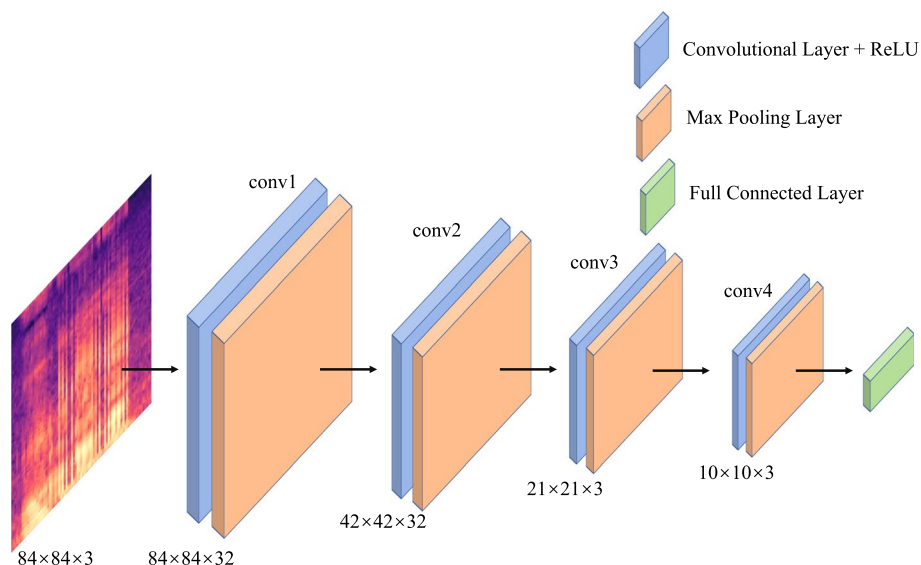

Fig. 4 The detailed architecture of the four-layer convolutional neural network with pooling used in this paper

Li *et al. EURASIP Journal on Audio, Speech, and Music Processing*   (2023) 2023:43
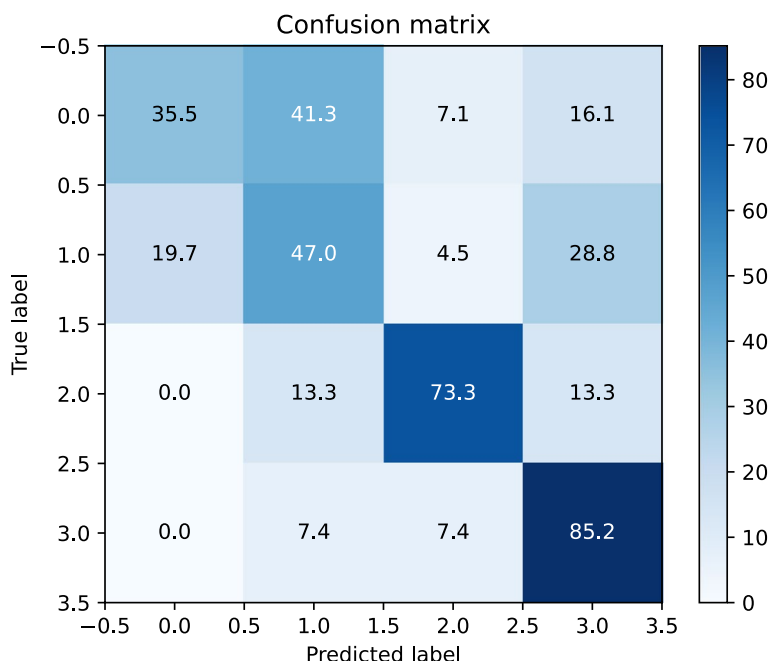
Page 7 of 9

Fig. 5 The confusion matrix of the prediction results for the test set of the MPSSC dataset, using the fine-tuned model on 36 snoring samples from the Dev portion of the MPSSC dataset. The digits in the matrix represent the percentage recognised as a particular category, where the numbers on the diagonal represent the probability of correctly predicting that class

challenge for snoring detection. Since no snoring data was used during the entire training process, our results also suggest that the MAML algorithm can be applied to solve other low-resource problems in medical data, particularly for rare diseases. However, the limitation of our strategy is that the result only surpasses the benchmark by 4.4 %. In the future, we plan to use larger sound datasets such as AudioSet and UrbanSound8K as meta-training data, along with better meta-learning strategies and audio denoising techniques, to improve the performance of the model and achieve higher UAR after small-scale fine-tuning. Furthermore, we will incorporating additional model comparisons to render the experimental outcomes more comprehensive.

6 Conclusion

In order to battle with the challenge of low resources, this paper proposes the use of the MAML algorithm and the design of two experiments to recognise snoring sounds in the MPSSC snoring recognition problem. The MAML algorithm updates the parameters during the meta-training by performing tasks and quickly adapts to new tasks through several updates in the meta-testing. In this study, we use spectrograms of snoring sounds and natural sounds, as well as images from the MiniImageNet dataset, as inputs for the MAML algorithm. The outcome indicates that by utilising solely the

ESC-50 dataset as meta-training data and subsequently fine-tuning 36 instances of snoring sounds (less than 7 % of the original training dataset) from the original partitioned development section through MPSSC, a UAR of 60.2 % was achieved on the test section of MPSSC. This result surpasses the benchmark of 55.8 % UAR for this dataset. Furthermore, although the non-sound dataset MiniImageNet did not perform well on the test set, it also indicates that the model learned useful information. This suggests that our model can quickly adapt to similar new classification tasks with very few new examples and achieve considerable results on testing. This achievement places the MAML algorithm as a promising solution for low-resource problems.

Authors' contributions

JL and MS conceived the study; these two authors provided general guidance to the drafting of the survey and contributed to the "Abstract" and "Introduction" sections. KQ, CW, and GL contributed to the "Related work"; KQ, JL, and XL, MS contributed to the "Datasets and Methods" section; ZZ and JL contributed to the "Experimental results" section; MS and KQ contributed to the "Discussion" section; JL, MS, and ZZ contributed to the "Conclusion" section; JL and MS contributed to the all the figures in the manuscript; BS contributed to the expertise of AI. KQ, BH, BS, YY, and MS generally reviewed and revised the manuscript.

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Key Laboratory of Brain Health Intelligent Evaluation and Intervention, Ministry of Education (Beijing Institute of Technology), Beijing, China. [2]School of Medical Technology, Beijing Institute of Technology, Beijing, China. [3]Educational Physiology Laboratory, The University of Tokyo, Tokyo, Japan. [4]GLAM - Group on Language, Audio, & Music, Imperial College London, London, 2AZ, UK. [5]Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Augsburg, Germany.

References

1. M.M. Ohayon, C. Guilleminault, R.G. Priest, M. Caulet, Snoring and breathing pauses during sleep: telephone interview survey of a united kingdom population sample. Bmj **314**(7084), 860 (1997)
2. I. Sharief, G.E. Silva, J.L. Goodwin, S.F. Quan, Effect of sleep disordered breathing on the sleep of bed partners in the sleep heart health study. Sleep **31**(10), 1449–1456 (2008)
3. J. Arnold, M. Sunilkumar, V. Krishna, S. Yoganand, M.S. Kumar, D. Shanmugapriyan, Obstructive sleep apnea. J Pharm Bioallied Sci **9**(Suppl 1), S26 (2017)
4. V.K. Somers, D.P. White, R. Amin, W.T. Abraham, F. Costa, A. Culebras, S. Daniels, J.S. Floras, C.E. Hunt, L.J. Olson, T.G. Pickering, R. Russell, M. Woo, T. Young, Sleep apnea and cardiovascular disease. Circulation **118**(10), 1080–1111 (2008)
5. D.J. Eckert, A. Malhotra, Pathophysiology of adult obstructive sleep apnea. Proc Am Thorac Soc **5**(2), 144–153 (2008)
6. N. Kuvat, H. Tanriverdi, F. Armutcu, The relationship between obstructive sleep apnea syndrome and obesity: A new perspective on the pathogenesis in terms of organ crosstalk. Clin Respir J **14**(7), 595–604 (2020)
7. C.N. Kaufmann, R. Susukida, C.A. Depp, Sleep apnea, psychopathology, and mental health care. Sleep Health **3**(4), 244–249 (2017)
8. A. Lavrentaki, A. Ali, B.G. Cooper, A.A. Tahrani, Mechanisms of endocrinology: Mechanisms of disease: the endocrinology of obstructive sleep apnea. Eur J Endocrinol **180**(3), R91–R125 (2019)
9. S. Skoczyński, K. Nowosielski, Ł Minarowski, G. Brożek, A. Oraczewska, K. Glinka, K. Ficek, B. Kotulska, E. Tobiczyk, R. Skomro et al., Sexual disorders and dyspnoea among women with obstructive sleep apnea. Adv Med Sci **65**(1), 189–196 (2020)
10. N. Ahmadi, G.K. Shapiro, S.A. Chung, C.M. Shapiro, Clinical diagnosis of sleep apnea based on single night of polysomnography vs. two nights of polysomnography. Sleep Breathing **13**, 221–226 (2009)
11. J. Mantua, N. Gravel, R.M. Spencer, Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors **16**(5), 646 (2016)
12. S. Kwon, H. Kim, W.H. Yeo, Recent advances in wearable sensors and portable electronics for sleep monitoring. Iscience **24**(5), 102461 (2021)
13. F. Dalmasso, R. Prota, Snoring: analysis, measurement, clinical implications and applications. Eur Respir J **9**(1), 146–159 (1996)
14. K. Qian, X. Li, H. Li, S. Li, W. Li, Z. Ning, S. Yu, L. Hou, G. Tang, J. Lu et al., Computer audition for healthcare: Opportunities and challenges. Front Digit Health **2**, 5 (2020)
15. H. Xu, W. Song, H. Yi, L. Hou, C. Zhang, B. Chen, Y. Chen, S. Yin, Nocturnal snoring sound analysis in the diagnosis of obstructive sleep apnea in the chinese han population. Sleep Breathing **19**, 599–605 (2015)
16. J. Fiz, J. Abad, R. Jane, M. Riera, M. Mananas, P. Caminal, D. Rodenstein, J. Morera, Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea. Eur Respir J **9**(11), 2365–2370 (1996)
17. K. Qian, C. Janott, M. Schmitt, Z. Zhang, C. Heiser, W. Hemmert, Y. Yamamoto, B.W. Schuller, Can machine learning assist locating the excitation of snore sound? a review. IEEE J Biomed Health Inf **25**(4), 1233–1246 (2020)
18. G. Sharma, K. Umapathy, S. Krishnan, Trends in audio signal feature extraction methods. Appl Acoust **158**, 107020 (2020)
19. H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.Y. Chang, T. Sainath, Deep learning for audio signal processing. IEEE J Sel Top Signal Process **13**(2), 206–219 (2019)
20. K.K. Li, Surgical therapy for adult obstructive sleep apnea. Sleep Med Rev **9**(3), 201–209 (2005)
21. H.D. Ephros, M. Madani, S.C. Yalamanchili et al., Surgical treatment of snoring & obstructive sleep apnoea. Indian J Med Res **131**(2), 267 (2010)
22. C. Janott, M. Schmitt, Y. Zhang, K. Qian, V. Pandit, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert et al., Snoring classified: the munich-passau snore sound corpus. Comput Biol Med **94**, 106–118 (2018)
23. T. Kohlberger, Y. Liu, Generating diverse synthetic medical image data for training machine learning models. Google AI Blog **1**, 1–1 (2020), https://blog.research.google/2020/02/generating-diverse-synthetic-medical.html?m=1#:~:text=Generating%20Diverse%20Synthetic%20Medical%20Image%20Data%20for%20Training%20Machine%20Learning%20Models,-Wednesday%2C%20February%2019&text=The%20progress%20in%20machine%20learning,of%20large%2C%20meticulously%20labeled%20datasets
24. P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, G. Neubig, Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. ACM Comput Surv **55**(9), 1–35 (2023)
25. C. Finn, P. Abbeel, S. Levine, in *Proceedings of the 34th International Conference on Machine Learning(ICML)*. Model-agnostic meta-learning for fast adaptation of deep networks. (PMLR, Sydney, 2017), p. 1126–1135
26. G. Gosztolya, R. Busa-Fekete, T. Grósz, L. Tóth, in *Proceedings of the 18th Annual Conference of the International Speech Communication Association(INTERSPEECH)*. DNN-Based Feature Extraction and Classifier Combination for Child-Directed Speech, Cold and Snoring Identification. (ISCA, Stockholm, 2017), p. 3522–3526
27. S. Amiriparian, M. Gerczuk, S. Ottl, N. Cummins, M. Freitag, S. Pugachevskiy, A. Baird, B. Schuller, in *Proceedings of Interspeech 2017*. Snore Sound Classification Using Image-Based Deep Spectrum Features (2017), pp. 3512–3516. https://doi.org/10.21437/Interspeech.2017-434
28. K. Qian, M. Schmitt, C. Janott, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, B. Schuller, A bag of wavelet features for snore sound classification. Ann Biomed Eng **47**(4), 1000–1011 (2019)
29. F. Demir, A. Sengur, N. Cummins, S. Amiriparian, B. Schuller, in *Proceedings of 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Low level texture features for snore sound discrimination. (IEEE, Honolulu, 2018), p. 413–416
30. L. Ding, J. Peng, Automatic classification of snoring sounds from excitation locations based on prototypical network. Appl Acoust **195**, 108799 (2022)

Li *et al. EURASIP Journal on Audio, Speech, and Music Processing*   (2023) 2023:43

Page 9 of 9

31. B. Shi, M. Sun, K.C. Puvvada, C.C. Kao, S. Matsoukas, C. Wang, in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Few-shot acoustic event detection via meta learning. (IEEE, 2020), p. 76–80

32. A. Lemkhenter, P. Favaro, in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Towards sleep scoring generalization through self-supervised meta-learning. (IEEE, Glasgow, 2022), p. 2961–2966

33. C. Heggan, S. Budgett, T. Hospedales, M. Yaghoobi, Metaaudio: A few-shot audio classification benchmark. arXiv preprint arXiv:2204.02121 (2022)

34. K.J. Piczak, in *Proceedings of the 23rd ACM international conference on Multimedia*. Esc: Dataset for environmental sound classification (Brisbane, 2015), pp. 1015–1018

35. S. Ravi, H. Larochelle, in *International conference on learning representations*. Optimization as a model for few-shot learning. (ICLR, Toulon, 2017)

36. B. Zhang, J. Leitner, S. Thornton, *Audio recognition using mel spectrograms and convolution neural networks* (Noiselab University of California, San Diego, 2019)

Publisher's Note