BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) .

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

BMJ Open

Protocol for a cross-sectional study in automatic assessment of the severity of mitral regurgitation with AI assisted digital auscultation

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-074288
Article Type:	Protocol
Date Submitted by the Author:	11-Apr-2023
Complete List of Authors:	Zhang, Li; Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Department of Cardiology Cheng, Zhenfeng; Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Department of Cardiology Xu, Dongyang; Huzhou Institute of Zhejiang University, Center for Intelligent Acoustics and Signal Processing Wang, Zhi; Huzhou Institute of Zhejiang University, Center for Intelligent Acoustics and Signal Processing; Zhejiang University, State Key Laboratory of Industrial Control Technology Cai, Shengsheng; Huzhou Institute of Zhejiang University, Center for Intelligent Acoustics and Signal Processing; Suzhou Melodicare Medical Technology Co., Ltd. Hu, Nan; Soochow University, School of Electronics and Information Engineering Ma, Jianming; Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Administration Office Mei, Xueqin; Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Department of Medical Engineering
Keywords:	Valvular heart disease < CARDIOLOGY, Telemedicine < BIOTECHNOLOGY & BIOINFORMATICS, Information technology < BIOTECHNOLOGY & BIOINFORMATICS

SCHOLARONE™
Manuscripts

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Protocol for a cross-sectional study in automatic assessment of the severity of mitral regurgitation with AI assisted digital auscultation

Li Zhang[1#], Zhenfeng Cheng[1#], Dongyang Xu[2], Zhi Wang[2, 3], Shengsheng Cai[2, 4], Nan Hu[5], Jianming Ma[6], Xueqin Mei[7,*]

[1]Department of Cardiology, Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Huzhou 313000, China

[2]Center for Intelligent Acoustics and Signal Processing, Huzhou Institute of Zhejiang University, Huzhou 313000, China

[3]State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China

[4]Suzhou Melodicare Medical Technology Co., Ltd., Suzhou 215151, China

[5]School of Electronics and Information Engineering, Soochow University, Suzhou 215006, China

[6]Administration Office, Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Huzhou 313000, China

[7]Department of Medical Engineering, Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Huzhou 313000, China

[#]These authors contributed equally to this work and should be considered as co-first authors.

[*]Corresponding Author

Email addresses: Li Zhang (774624211@qq.com); Zhenfeng Cheng (czf918@qq.com); Dongyang Xu (xudongyang@ustc.edu); Zhi Wang (zjuwangzhi@zju.edu.cn); Shengsheng Cai (caishengsheng@mintti.cn); Nan Hu (hunan@suda.edu.cn); Jianming Ma (hzxzmjm@sina.cn); Xueqin Mei (meixueqin1970@163.com).

ABSTRACT

Introduction Mitral regurgitation (MR) is the most common cardiovascular disorder with morbidity rate of 2.5%. Though echocardiography is the most important modality in assessing MR, it has many limitations, especially for large-scale MR screening. Cardiac auscultation with electronic stethoscope and artificial intelligence (AI) can be fast and economical modality for assessing MR severity. Our objectives are: (1) to establish a deep neural network (DNN) based cardiac auscultation method for assessing the severity of MR; (2) to quantitatively measure the performance of the developed AI-based MR assessment method by virtual clinical trial.

Methods and analysis In the cross-sectional design, phonocardiogram (PCG) will be recorded at mitral-valve auscultation area of outpatient volunteers. The enrolled patients will be checked by echocardiography to confirm the diagnosis of MR, and the controls are those diagnosed as no-MR. Echocardiographic parameters will used as gold standard to assess the severity of MR, classified into four degrees: none, mild, moderate, and severe. The study consists of two stages. First, data will collected to form an MR-related cardiac sound database, for training of a DNN-based MR severity classifier. The automatic MR severity classifier will be integrated at the smartho-D2

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

electronic stethoscope. Second, the smart device will be used to automatically assess the MR severity of outpatients, by cardiologists as well as patients themselves. Sensitivity, specificity, precision, accuracy, and F1-score will be evaluated for the developed smart MR assessment device. The agreement between the performances of the smart device used by cardiologists and patients themselves will also be inspected.

Ethics and dissemination The study protocol was approved by the Medical Ethics Committee of Huzhou Central Hospital, China (registration number: 202302009-01). Dissemination will be through conference presentations and peer-reviewed journals.

Trial registration number ChiCTR2300069496

Keywords Valvular heart disease; Telemedicine; Information technology.

ARTICLE SUMMARY

Strengths and limitations of this study

- Though MR has high morbidity rate and may cause adverse outcomes, screening of MR in a large-population and regular evaluation are usually infeasible, due to scarcity of medical resources and skilled echocardiography operators.
- The feasibility of assessing the MR severity by automatically analyzing phonocardiogram (PCG) with AI assisted electronic stethoscope will studied, and the potential application of the smart device in remote healthcare will also be considered.
- Combination of EROA, RVol and RF will be used to label the severity of MR into four categories: none, mild, moderate, and severe.
- A DNN-based MR severity classification model will be trained and refined with the NAS, and the performance of the resultant smart electronic stethoscope will be evaluated, when users are either cardiologists or patients themselves.
- In this study, only PCG will be utilized, without simultaneously recorded electrocardiogram (ECG), with was usually used as reference for PCG cycle segmentation.

Word count 3974

INTRODUCTION

Mitral regurgitation (MR) is recognized as one kind of commonly occurring valvular heart disorder, due to its high morbidity rate. An early study of global burden of valvular heart disorder[1] reported that the morbidity rate of MR was as high as 2.5% among all population, and this rate can increase substantially with the growth of age. In the United States, more than 2 million of adults have had MR, and this number was predicted to be doubled in 2030[2]. A large-scale-community cross-sectional epidemiologic study from Europe[3] revealed that the new diagnostic rate of moderate or severe MR was 2.3%, much larger than that of aortic stenosis (0.7%). A

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Superieur (ABES) .

multicenter investigation study reported that MR was the most common valvular heart disorder in China[4].

MR consists of two categories: primary/organic MR and secondary/functional MR. Organic MR is due to disruption of the valve leaflets, chordae tendineae and annulus, caused by degenerative changes such as Barlow's disease or rheumatic heart diseases. Compared with organic MR, functional MR (FMR) is more prevalent[5]. FMR is secondary to abnormalities in cardiac function or structure, resulting in imbalance between mitral valve tethering and closure force, hence poor coverage. Patients with ischemic heart disease, dilated cardiomyopathy, heart failure with preserved ejection fraction (HFpEF), or atrial fibrillation (AF), are apt to have FMR. Pharmacologic therapy, cardiac resynchronization therapy (CRT), or surgical/percutaneous interventions can be applied in treatment for MR patients[6], depending on the MR assessment results. It is very important that MR can be detected and assessed in time, as moderate or severe MR is highly related to the morbidity, hospitalization, and long-term prognosis of left ventricular dysfunction and congestive heart failure[7]. However, in the United States around 49% of patients having moderate or severe MR did not go to hospital to be diagnosed[8], and this rate may be increased in developing countries. Therefore, a reliable, flexible and economical tool that can detect MR and assess the severity is essential.

Echocardiography is a key imaging technique for assessing the severity of MR. The severity level of MR is commonly categorized into three levels: mild, moderate, and severe. Qualitative, semi-quantitative and quantitative methods were used to assess the severity level of MR[9]. Quantitative methods have been well validated, while there is still no single echocardiographic parameter or optimal criteria to define the severity level of MR. In the recommendations for noninvasive evaluation of native valvular regurgitation from the American society of echocardiography[10], effective regurgitant orifice area (EROA), regurgitant volume (RVol) and regurgitant fraction (RF) can be used to assess the severity level of MR. EROA < 0.2 cm^2, RVol < 30 mL/beat or RF < 30% was considered as mild MR; 0.2 cm^2 $\leqslant$ EROA $\leqslant$ 0.39 cm^2, 30 mL/beat $\leqslant$ RVol $\leqslant$ 59 mL/beat or 30% $\leqslant$ RF $\leqslant$ 49% was considered as moderate MR; and EROA $\geqslant$ 0.4 cm^2, RVol $\geqslant$ 60 mL/beat or RF $\geqslant$ 50% was considered as severe MR. These parameters can be calculated by proximal isovelocity surface area (PISA) method. Though echocardiographic parameters are commonly used to level the severity of MR, there are some drawbacks. Firstly, most of the patients may not choose to go to hospital to assess the valvular heart disease, due to lack or over-concentration of medical resource. Secondly, the assessment performance of echocardiography is highly operator-dependent, as it is easy to be affected by transducer frequency, pulsed repetition frequency, and inappropriate gain settings[11]. Lastly, the color Doppler device is cumbersome and complex to use, hence not suitable for long-term monitoring of MR.

Cardiac auscultation has been an important tool which was often used in first-time cardiac examination. Though cardiac auscultation has the advantage of quick examination, it requires plenty of experiences and skills. For different users, their perceptive acoustic frequency ranges are also different. In comparison to traditional

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Superieur (ABES) .

auscultation, smart auscultation with the emerging electronic stethoscopes[12] manipulates digitalized phonocardiogram (PCG) with digital signal processing and pattern recognition techniques. In abnormal PCG recordings, apart from the first heart sound (S1) and the second heart sound (S2), other acoustic components such as the third heart sound (S3), the fourth heart sound (S4), murmurs, gallops, clicks, or opening snaps may appear. Features of the PCG signal corresponding to specific heart disease can be extracted and utilized for PCG classification. Dwivedi *et al.* gave a systematic review of automatic PCG analysis and classification[13], where conventional steps involved in this task was concluded as: preprocessing (denoising), segmentation, feature extraction, and classification. The review has now been updated[14], where deep learning algorithms used for heart sound classification in recent years have been supplemented. For cardiac auscultation denoising, various digital signal processing methods[15] have been tried, and the current trend is to combine these conventional methods with deep neural networks (DNNs)[16]. The necessity of PCG cycle segmentation is controversial, as in some studies reasonable PCG classification performances have also been achieved without segmentation[17]. Feature extraction is important for PCG signal classification, where handcrafted features[18] including time-frequency spectrum, Mel frequency cepstrum coefficients (MFCCs), and discrete wavelet transform (DWT) coefficients have been considered, and in recent years automatic feature extraction from the PCG spectra or waveform by deep learning method has been prevalent[19]. For handcrafted feature extraction, it is always followed by traditional machine learning methods including support vector machine (SVM), random forest, Gaussian mixture model, etc. In deep learning methods, such as convolutional neural network (CNN) or recurrent neural network (RNN), the classifiers are typically trained simultaneously with feature extraction step.

Though various PCG signal classification methods are coming forth, most of them were confined to distinguishing normal PCG recordings from abnormal ones, restricted by the task setup in two commonly used public database: the Physionet heart sound challenge 2016 database[20] and the PASCAL classifying heart sounds challenge 2011 database[21]. Recently, a database for detection of heart valve disorders was shared in the GitHub repository[22]. This database includes 1000 PCG segments, each one lasting for around 2 seconds, truncated from several PCG recording sources in books and websites, and they were divided evenly into 5 categories: normal, aortic stenosis (AS), MR, mitral stenosis (MS) and mitral valve prolapse (MVP). Several works were carried out in handling this 5-category classification task[23-29]. Time-frequency magnitude and phase features were used[23], and graph-based feature was developed[24], with traditional machine learning methods used for classification. Avanzato *et al.* discovered that with the raw data as the input to 1D-CNN, reasonable classification performance can still be achieved[25]. Karhade *et al.* used both time-domain polynomial chirplet transform (TDPCT) and frequency-domain polynomial chirplet transform (FDPCT)[26] for CNN-based classification. Convolutional and recurrent neural networks were developed by Alkhodari *et al.*[27], and attention mechanism was introduced by Chowdhury *et al.*[28] to address this issue.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

Recursive feature elimination (RFE) algorithm was applied by Arslan[29] to select the most distinctive deep features for heart valve disorder classification. Though the above-mentioned methods may show potential values in MR detection, there is no study of assessing the severity of MR using PCG signals, due to the scarcity from data to methods. In spite of this, the existing studies on PCG-based assessment of severity of other heart valve disorders, such as aortic stenosis[30] and tricuspid regurgitation[31], can be used for reference.

Huzhou Institute of Zhejiang University and Melodicare developed an electronic stethoscope device named Smartho-D2, which has been certificated by FDA 510k, NMPA, and CE. Smartho-D2 has two sound picking-up channels, one for auscultation and another one for background noise collection and hence noise cancellation. Our developed two-stage noise cancellation algorithm[16] has been applied in this device, to eliminate the background noise contamination. Smartho-D2 is portable, Bluetooth connected, and has self-contained software, supporting PCG waveform displaying, data storage and download from local and the cloud, online remote auscultation, etc. Secondary software development can be easily carried out in Smartho-D2, to add more machine-learning-based heart disease assessment functionalities. In this study, Smartho-D2 will be used for PCG recording from MR patients. With the recorded PCG signals, labeled by echocardiographic examination results, an automatic MR severity assessment algorithm will be developed and integrated in Smartho-D2**.** The performance of the developed smart stethoscope in automatically assessing the severity of MR will be evaluated via clinical experiments.

Objectives and hypotheses

The first objective of this study is to establish a DNN-based cardiac auscultation method that can automatically assess the severity of MR. Based on the findings of previous studies[23-31], the hypothesis in this study is that the severity of MR can be classified by putting PCG signals into a trained DNN with high sensitivity and specificity. The second objective is to quantitatively measure the performance of the developed AI-based MR assessment method when the AI-aided electronic stethoscope smartho-D2 is used by cardiologists as well as patients. It hypothesizes that the developed smart stethoscope will yield high classification performance indicators in automatically assessing the severity of MR, and the agreement between the performances of the device used by cardiologists and patients themselves will be high.

METHODS AND ANALYSIS

Study design and sample selection

The study and related experiments will be carried out at Department of Cardiology, Huzhou central hospital, which is located at the north of Zhejiang Province in China, serving for the healthcare of about 4 million people. In each year, about 50,000 outpatients come to seek help from a cardiologist, and over 50% of them suffer from heart valve diseases. By a cross-sectional design, adult patients diagnosed as MR or no-MR will be considered to be enrolled into experiments. The MR patient inclusion

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Superieur (ABES) .

criterions are: (1) the patient has been diagnosed to have MR by echocardiographic examination; (2) the patient did not take any medicine that can influence cardiac function in 24 hours; (3) the patient did not have history of heart valve surgery. If the quality of PCG signal recorded from a participant is too low to be utilized, it will be excluded. Gender and age of a MR patient will not be limited. Randomization will not be used for MR patients. The patients diagnosed as no-MR will be matched to that of MR patients by propensity score. The study is conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Written informed consent will be obtained from a patient before the data recording experiments.

There will be two stages in implementing the study. The first stage of research aims at building an automatic MR severity classifier with PCG signal as input. A PCG database for training the MR severity classifier will be built. The database will include PCG recordings collected from recruited patients mentioned above, and the data will be labeled according to the severity of MR, classified into four degrees: none, mild, moderate, and severe. Echocardiography will be used as gold standard to label the severity of MR. A DNN will be built to form the automatic MR severity classifier, and trained based on our MR-related PCG database. The trained MR severity classifier will be integrated at the Smartho-D2 electronic stethoscope. In the second stage, the performance of AI-aided electronic stethoscope Smartho-D2 will be evaluated in real clinical usage for automatic MR severity measurement. In this stage, the gold standard will still be given by echocardiographic examination. As bias may occur when a smart stethoscope is used by different types of operators, the agreement between automatic MR severity assessment results by a cardiologist and a patient will be evaluated. The overall study flow is displayed in Fig. 1.

For each patient, a cardiologist will collect one PCG recording, no shorter than 10 seconds, with the electronic stethoscope Smartho-D2, from the mitral valve auscultation area, at the cardiac apex, in the fifth intercostal space on the midclavicular line. Smartho-D2 is Bluetooth connected, and the Smartho APP is used to monitor and control the data recording. The recorded data can be labeled and stored in local mobile device, and can be transferred to a password-protected database server located at Huzhou central hospital. A sketch of the PCG data recording and storage in this study is given in Fig. 2. The recruited patients accept a routine echocardiographic examination for labeling the PCG recording, i.e. the severity level of MR. According to the recommendations for noninvasive evaluation of native valvular regurgitation from the American society of echocardiography[10], EROA, RVol and RF calculated by PISA method will be used for labeling the severity level of MR. Mild MR is considered when at least two of the following conditions are satisfied: 0 cm^2 < EROA < 0.2 cm^2, 0 mL/beat < RVol < 30 mL/beat, 0 < RF < 30%. Moderate MR is considered when at least two of the following conditions are satisfied: 0.2 cm^2 $\leqslant$ EROA $\leqslant$ 0.39 cm^2, 30 mL/beat $\leqslant$RVol $\leqslant$ 59 mL/beat, 30% $\leqslant$ RF $\leqslant$ 49%. Severe MR is considered when at least two of the following conditions are satisfied: ROA $\geqslant$ 0.4 cm^2, RVol $\geqslant$ 60 mL/beat or RF $\geqslant$ 50%. Otherwise no-MR will be considered. The examinations beyond routine ones in this study are cost-free. After

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

data mask by medical staffs to remove the patients' personal information, the data can be used by non-medical members for further processing and analysis.

The automatic MR severity classifier to be built will further be integrated at the Smartho-D2 electronic stethoscope. For practical usage, it will consist of adaptive noise cancellation, usable data extraction and the trained DNN. Smartho-D2 has two sound picking-up channels, and a two-stage noise cancellation algorithm[16] has been applied in this device. As reasonable PCG classification performances have been achieved without cardiac cycle segmentation in some previous work[17], in this study we will only divide the PCG recording into segments with sliding windows. A 2-second length window will be used, with sliding stride as 1 second. In real applications, PCG recording may be contaminated by inappropriate operations such as pressure with excessive force, clothing friction and auscultation leaving the body surface, especially when the stethoscope is used by patient themselves. In this study, before being put into the trained DNN, each PCG segment will be evaluated by a former developed usable data extraction module[16]. With a simple CNN structure, it judges a PCG segment as usable or unusable, where an unusable PCG segment implies that its quality is too low to be used for MR severity assessment. Only usable PCG segments will be put into the DNN trained from the established dataset. There are two considerations in building the DNN: (1) the DNN should extract and utilize features in data as much as possible to classify the level of MR severity; (2) the DNN should be lightweight to be integrated in a mobile device. In this study, a DNN with a clique net structure[32, 33] will be established, which take both feature reuse and lightweight into account. Furthermore, the neural architecture search (NAS)[34] will be used to refine the performance and model size of the DNN. For each usable PCG segment, the trained DNN will output a 4-class classification result. To composite the classification results from all usable PCG segment, a majority voting strategy will be applied to give the final decision in automatic assessment of the level of MR severity. The flowchart of PCG data processing is displayed in Fig. 3.

Interventions to be measured

There will be no intervention to be implemented to patients in this study. The subjects in this study are those MR or no-MR patients diagnosed by cardiologists, and our objective relies on exploiting the value of using PCG signal in automatic assessment of the level of MR. The study will not influence the diagnosis or treatment to the patient.

Sample size calculation

In the first stage of study, the main task is to collect PCG data and train the automatic MR severity classifier. The established database will be divided into 3 pieces equally, and a 3-fold cross-validation will be performed. In each round of model training, 2 pieces are used as training set and 1 piece is used as testing set, and finally the model with mediate values of classification metrics among 3 trained ones will be adopted. Let a one-sample proportion test, the z-test using s(Phat) with continuity correction, be used to assess whether the accuracy of the trained classifier

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Superieur (ABES) .

$P1$ is significantly larger than a hypothesized value $P0$. To detect $P1 = 0.95$ compared with $P0 = 0.90$, with one-sided type I error of 0.05 and power of 0.80 calculated by binomial enumeration, the required sample size in a testing set is 179. Hence, the total required sample size in the first stage of study is $179 \times 3 = 537$.

In the second stage of study, all collected PCG data will be used as testing set. In this sense, the required sample size seems to be 179. Cohen's kappa will be used to evaluate the agreement between automatic MR severity assessment results by a cardiologist and a patient. When the value of Kappa under the null hypothesis $\kappa 0 = 0.5$, the one under the alternative hypothesis $\kappa 1 = 0.6$ and the proportions of subjects assigned to 4 categories are 0.3:0.3:0.2:0.2, for a one-sided alternative hypothesis test of $\kappa 1 > \kappa 0$ with type I error of 0.05 and power of 0.80, the required sample size is calculated as 286. As 286 > 179, the required sample size in the second stage of study should be 286.

Finally, the total required sample size in this study will be 537 + 286 = 823. Power Analysis and Sample Size (PASS) software V2021 was used to calculate the size of sample to be collected.

Statistical analysis

The aim of the first stage of study is to build an automatic MR severity classifier. Blinding will be performed in the testing sets. Classification metrics including sensitivity, specificity, precision, accuracy and F1-score will be calculated and assessed. As a multi-category classification task is involved in this study, macro-average, micro-average and weighted-average metrics will be calculated. Macro-average means direct average of binary-classification metrics across classes. The micro-average metrics are calculated by using aggregate outcomes across all classes. The weighted-average metrics are given by weighted average of binary-classification metrics across classes, with proportions of classes as weights. Their one-versus-rest receiver operating characteristic (ROC) curves and corresponding area under curve (AUC) with confidence interval will also be given.

The above-mentioned classification metrics will also be assessed in the second stage, where the performance of established MR severity assessment device will be evaluated in real clinical usage. In this stage, to measure the bias when a smart stethoscope is operated by different types of users, the agreement between MR severity assessment results by a cardiologist user and a patient user will be evaluated, with the metric of Cohen's kappa. Kappa score > 0.8 will be deemed as good agreement.

Data collection and analysis plan

The data collection has been started on 15 March 2023, and it is planned to last for 1 year. The first stage of study is planned to be finished on 15 November 2023. Considering about 50,000 outpatient visits in one year, we believe that a one-year data collection plan will easily meet the sample size requirement, even with a conservative MR morbidity rate of 2.5% for these outpatients. The Smartho APP, deployed in a password-protected android pad, is used to monitor and control the PCG data

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Superieur (ABES) .

recording. Echocardiographic examination reports given by a skilled expert with above 10-year experiences will be used in data labeling. The labeled data will be transferred to a password-protected database server located at Huzhou central hospital from local device, and anonymized data files can be further processed and analyzed by non-medical members. Data analysis will be started on 1 July 2023, for classification model establishing and integration on the device. With the growth of recorded data size, through training, the weights in the classification model will be updated and the classification performance will be refined.

Patient and Public Involvement

Patients or the public are not involved in the design, or conduct, or reporting, or dissemination plans of this study.

Ethics and dissemination

Ethical approval was obtained from the Medical Ethics Committee of Huzhou Central Hospital, China (registration number: 202302009-01). This study was registered with Chinese Clinical Trial Registry, ChiCTR (trial registration number: ChiCTR2300069496). Written informed consent will be obtained from the patient before PCG recording. Access to collected data will be restricted to cardiologists from the research team treating the participants, and anonymized data files are shared to non-medical members of the research team. Dissemination will be through conference presentations and peer-reviewed journals.

DISCUSSION

Though it was realized a long time ago that quantitatively analyzing PCG signal may provide a convenient and fast way to assess heart valve diseases, there is no PCG-based method or device to automatically assess the severity of MR, a most common valvular heart disorder. In recent years, by using a public shared heart valve disorder PCG database[22], many studies[23-29] have been published, which tried to distinguish MR from normal and the other 3 valvular heart disorders. From these existing studies, MR related PCG signals show specific pattern that can be employed in MR detection. Though assessing the severity of MR using PCG signal has not been studied in the past, the similar methodology was applied in automatically grading the severity level of aortic stenosis[30] and tricuspid regurgitation[31]. By studying the PCG signals and echocardiographic parameters of 41 aortic stenosis patients[30], it was found that there was a good correlation between the duration of spectra above 300 Hz and the Doppler derived peak pressure gradient. Rujoie *et al.* studied the PCG signals of 22 subjects (divided into three categories: non, mild, and moderate to severe)[31], using MFCC and wavelet transform methods for feature extraction, genetic algorithms for feature selection, and KNN to implement the triple-classification task. The data recording has been started recently, and Fig. 4 shows typical PCG cases corresponding to different levels of MR severity. It seems that even the waveforms may reflect some information for classification, *e.g.*, the pattern of systolic murmur. Different from the research based on handcraft features and small-size database, our

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

study considers larger-size data collection and automatic feature extraction via DNN.

The MR severity level classifier will be integrated on the Smartho-D2, an electronic stethoscope device certificated by FDA 510k, NMPA, and CE. In recent years, some AI-aided auscultation products are emerging for cardiac healthcare. In 2019, the eMurmur Heart AI received FDA clearance and CE marking for detecting the presence of a murmur in PCG recording and determining if it is likely innocent or pathological. Evaluation from 603 outpatient visits[35] showed that the eMurmur had 93% (CI 90-95%) sensitivity and 81% (CI 85-91%) specificity for detection of pathologic cases. In 12 July 2022, FDA cleared the Eko Murmur Analysis Software (EMAS) for detecting and characterizing murmurs found in adults and pediatric patients. The EMAS was reported to identify structural murmurs with a sensitivity of 90.2% and specificity of 90.6% among adults, and have an overall sensitivity of 85.6% and specificity of 84.4%[36]. These existing smart cardiac auscultation applications showed feasibility of developing valvular heart disorder assessment methods on electronic stethoscope devices.

The limitation of this study is that only PCG will be utilized, without simultaneously recorded electrocardiogram (ECG). In many studies, ECG was used for reference of PCG cycle segmentation[37]. However, in recent years some studies showed that segmentation may not be necessary when deep learning is used for cardiac sound classification. By using Shapley values and occlusion maps[17], Dissanayake *et al*. showed that deep learning methods can automatically emphasize the components in PCG data that make significant contribution to classification. In our study, we consider to classify the PCG signals without PCG cycle segmentation. The collected large-size data will endow the established DNN-based MR severity level classifier with the ability of automatically selecting the useful PCG components. Multicenter study and involving more subjects in virtual clinical trial will be our future work, to refine and evaluate the performance of the developed method and smart device.

Author Contributions Li Zhang and Zhenfeng Cheng conceptualized and designed the study, and wrote the original draft. Dongyang Xu, Zhi Wang and Nan Hu will analysis the recorded data using digital signal processing algorithms and design the DNN-based classifier. Shengsheng Cai will be responsible for technical supports for the device. Jianming Ma will be responsible for project administration. Xueqin Mei provides mentorship to this study, edited and revised the draft, and will be responsible for project administration. All authors critically reviewed and approved the final manuscript before submission.

Funding This work was supported by Cardiovascular Discipline Group Funds of Huzhou under Grant XKQ-HT-202102A.

Competing interests Shengsheng Cai is an employee and owner of Suzhou Melodicare Medical Technology Co., Ltd., which developed the Smartho-D2 device.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Enseignement Superieur (ABES) .

Li Zhang, Zhenfeng Cheng Dongyang Xu, Zhi Wang, Nan Hu, Jianming Ma and Xueqin Mei declare no conflict of interest.

Data statement The dataset will shared at a Github repository after the relevant research being published.

REFERENCES

[1] Nkomo VT, Gardin JM, Skelton TN, Gottdiener JS, Scott CG, Enriquez-Sarano M. Burden of valvular heart diseases: a population-based study. *Lancet* 2006;368(9540):1005–1011.

[2] de Marchena E, Badiye A, Robalino G, *et al*. Respective prevalence of the different carpentier classes of mitral regurgitation: a stepping stone for future therapeutic research and development. *J Card Surg* 2011;26(4):385-392.

[3] d'Arcy JL, Coffey S, Loudon MA, *et al*. Large-scale community echocardiographic screening reveals a major burden of undiagnosed valvular heart disease in older people: The OxVALVE population cohort study. *Eur Heart J* 2016;37(47):3515-3522.

[4] Zhang B, Xu H, Zhang H, *et al*. Prognostic value of N-terminal pro-B-type natriuretic peptide in elderly patients with valvular heart disease. *J Am Coll Cardiol* 2020;75(14):1659-1672.

[5] Dziadzko V, Dziadzko M, Medina-Inojosa JR, *et al*. Causes and mechanisms of isolated mitral regurgitation in the community: clinical context and outcome. *Eur Heart J* 2019;40(27):2194-2202.

[6] Otto CM, Nishimura RA, Bonow RO, *et al*. 2020 ACC/AHA Guideline for the management of patients with valvular heart disease: Executive summary: A report of the American College of Cardiology American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* 2021;143(5):e35–71.

[7] 2nd Sannino A , Smith RL , Schiattarella GG, *et al*. Survival and cardiovascular outcomes of patients with secondary mitral regurgitation: A systematic review and meta-analysis. *JAMA Cardiol* 2017;2(10):1130–1139.

[8] Head SJ, van Leeuwen WJ, Van Mieghem NM, *et al*. Surgical or transcatheter mitral valve intervention: Complex disease requires complex decisions. *EuroIntervention* 2014;9(10):1133-1135.

[9] 3rd Otto CM, Nishimura RA, Bonow RO, *et al*. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: Executive summary: A report of the American college of cardiology American heart association joint committee on clinical practice guidelines. *Circulation* 2021;143(5):e35–71.

[10] Zoghbi WA, Adams D, Bonow RO, *et al*. Recommendations for noninvasive evaluation of native valvular regurgitation: a report from the American society of echocardiography developed in collaboration with the society for cardiovascular magnetic resonance. *J Am Soc Echocardiogr* 2017;30(4):303-371.

[11] Stevenson JG. Two-dimensional color Doppler estimation of atrioventricular valve regurgitation: important effects on instrument gain setting, pulse repetition frequency and carrier frequency. *J Am Soc Echocardiogr* 1989;2:1–10.

[12] Bank I, Vliegen HW, and Bruschke AVG. The 200th anniversary of the stethoscope: Can this low-tech device survive in the high-tech 21st century? *Eur Heart J* 2016;37(47):3536-3543.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Superieur (ABES) .

[13] Dwivedi AK, Imtiaz SA, and Odriguez-Villegas E. Algorithms for automatic analysis and classification of heart sounds - A systematic review. *IEEE Access* 2019;7:8316-8345.

[14] Chen W, Sun Q, Chen X, *et al*. Deep learning methods for heart sounds classification: A systematic review. *Entropy* 2021;23:667.

[15] Mondal A, Saxena I, Tang H, *et al*. A noise reduction technique based on nonlinear kernel function for heart sound analysis. *IEEE J Biomed Health Inform* 2018;22(3):775-784.

[16] Yang C, Dai N, Wang Z, Cai S, Wang J, and Hu N. Cardiopulmonary auscultation enhancement with a two-stage noise cancellation approach. *Biomed Signal Process Control* 2023;79:104175.

[17] Dissanayake T, Fernando T, Denman S, Sridharan S, Ghaemmaghami H, Fookes C. A robust interpretable deep learning classifier for heart anomaly detection without segmentation. *IEEE J Biomed Health Inform* 2020;99:2162-2171.

[18] Xu W, Yu K, Ye J, *et al*. Automatic pediatric congenital heart disease classification based on heart sound signal. *Artif Intell Med* 2022;126:102257.

[19] Li S, Li F, Tang S, and Luo F. Heart sounds classification based on feature fusion using lightweight neural networks. *IEEE Trans Instrum Meas* 2021;70:4007009.

[20] Clifford GD, Liu C, Moody B, *et al*. Classification of normal/abnormal heart sound recordings: The PhysioNet/computing in cardiology challenge 2016. *Proc. Comput. Cardiology Conf. (CinC)* 2016;43:609-612.

[21] Bentley P, Nordehn G, Coimbra M, *et al*. The PASCAL Classifying Heart Sounds Challenge 2011. Available: http://www.peterjbentley.com/heartchallenge/

[22] Yaseen GS and Kwon S. Classification of heart sound signal using multiple features. *Appl Sci* 2018;8:2344.

[23] Ghosh SK, Tripathy RK, Ponnalagu RN, and Pachori RB. Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features. *IEEE Sen Lett* 2019;3(12):7002604.

[24] Tuncer T, Dogan S, Tan RS, Acharya UR. Application of Petersen graph pattern technique for automated detection of heart valve diseases with PCG signals. *Inform Sciences* 2021;565:91-104.

[25] Avanzato R and Beritelli F. Heart sound multiclass analysis based on raw data and convolutional neural network. *IEEE Sen Lett* 2020;4(12): 7004104.

[26] Karhade J, Dash S, Ghosh SK. Time–frequency-domain deep learning framework for the automated detection of heart valve disorders using PCG signals. *IEEE Trans Instrum Meas* 2022;71:2506311.

[27] Alkhodari M and Fraiwan L. Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. *Comput Meth Prog Bio* 2021;200:105940.

[28] Chowdhury S, Morshed M, and Fattah SA. SpectroCardioNet: An attention-based deep learning network using triple-spectrograms of PCG signal for heart valve disease detection. *IEEE Sensors J* 2022;22(23):22799-22807.

[29] Arslan Ö. Automated detection of heart valve disorders with time-frequency and deep features on PCG signals. *Biomed Signal Process Control* 2022;78:103929.

[30] Kim D and Tavel ME. Assessment of severity of aortic stenosis through time-frequency analysis of murmur. Chest 2003;124:1638-1644.

[31] Rujoie A, Fallah A, Rashidi S, Khoshnood ER, and Ala TS. Classification and evaluation of the severity of tricuspid regurgitation using phonocardiogram. *Biomed Signal Process Control* 2020;57:101688.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

For peer review only

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

[32] Yang Y, Zhong Z, Shen T, and Lin Z. Convolutional neural networks with alternately updated clique. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)* 2018:2413-2422.

[33] Xiao B, Xu Y, Bi X, *et al.* Follow the sound of children's heart: A deep-learning-based computer-aided pediatric CHDs diagnosis system. *IEEE Internet Things J* 2020;7(3):1994-2004.

[34] Zoph B and Le QV. Neural architecture search with reinforcement learning. *Proc. Int. Conf. Learn. Representations* 2017.

[35] Thompson WR, Reinisch AJ, Unterberger MJ, and Schriefl AJ. Artificial intelligence-assisted auscultation of heart murmurs: Validation by virtual clinical trial. *Pediatr Cardiol* 2019;40:623-629.

[36] Eko Murmur Analysis Software User Manual. https://www.ekohealth.com

[37] Zhang H, Wang X, Liu C, *et al*. Discrimination of patients with varying degrees of coronary artery stenosis by ECG and PCG signals based on entropy. *Entropy* 2021;23:823.

For peer review only

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Superieur (ABES) .

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Figures:

Fig. 1 The overall study flow

Fig. 2 A sketch of the PCG data recording and storage

Fig. 3 The flowchart of PCG data processing in the automatic MR severity classifier

Fig. 4 Typical PCG cases corresponding to different levels of MR severity, collected in this study started on 15 March 2023

For peer review only

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

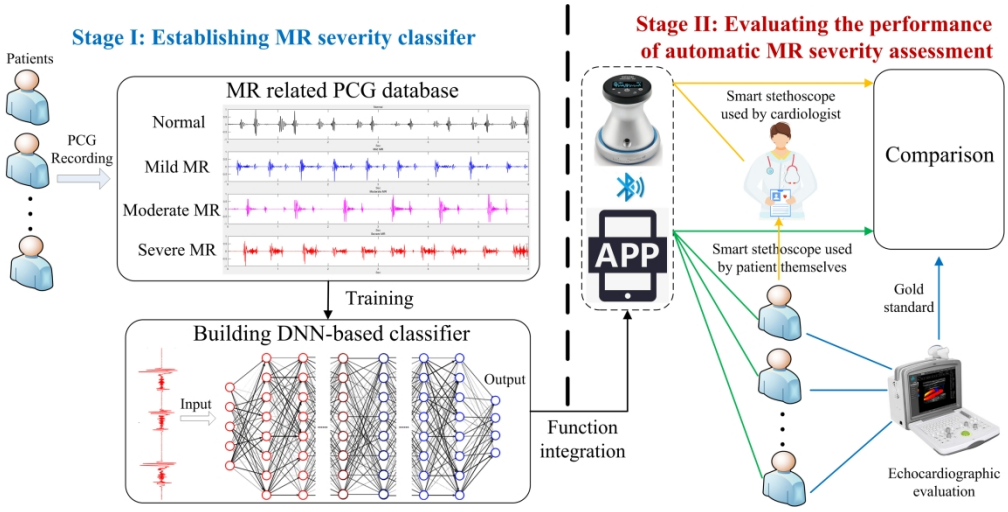

Fig. 1. The overall study flow

283x142mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Fig. 2. A sketch of the PCG data recording and storage

260x82mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

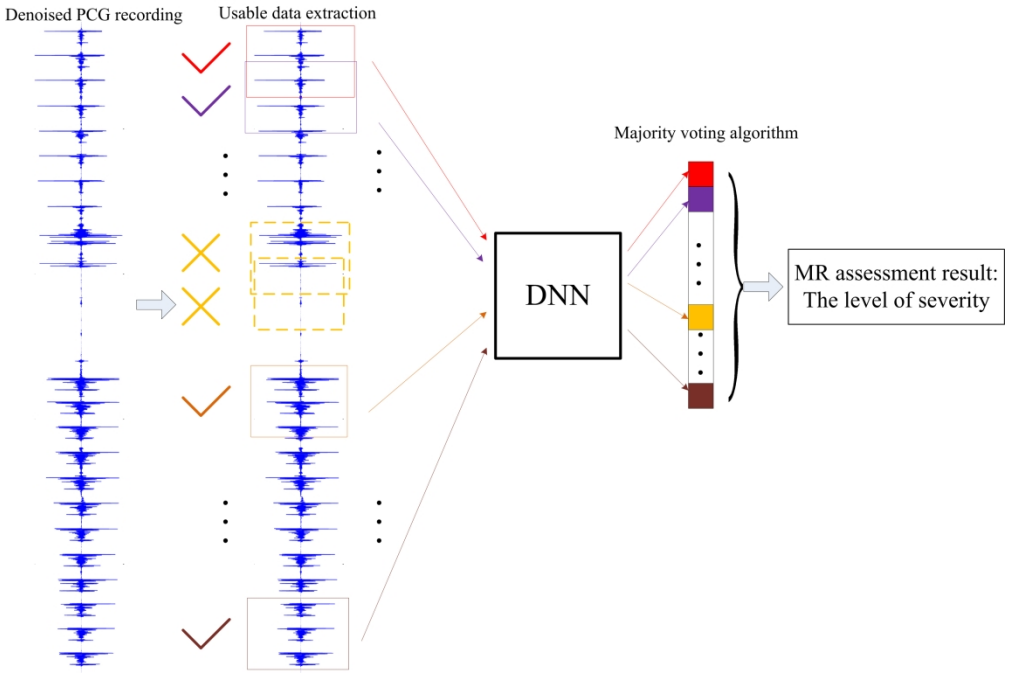

Fig. 3. The flowchart of PCG data processing in the automatic MR severity classifier

279x188mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

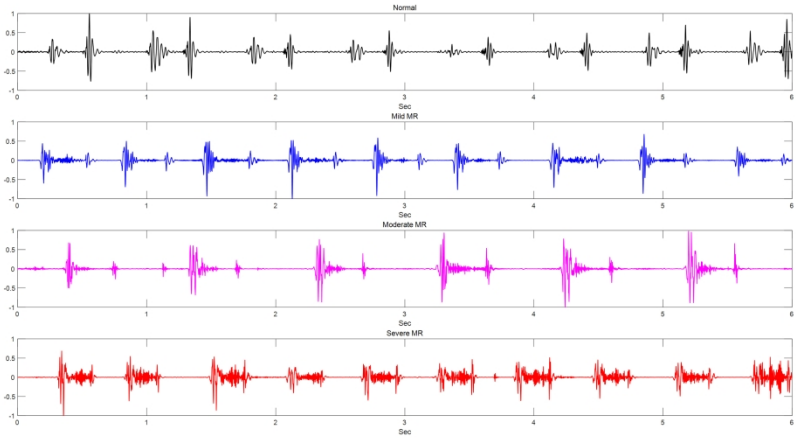

Fig. 4. Typical PCG cases corresponding to different levels of MR severity, collected in this study started on 15 March 2023

285x141mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5-7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-9
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	5-6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	7-8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8
		(*b*) Describe any methods used to examine subgroups and interactions	6
		(*c*) Explain how missing data were addressed	6
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7-8
		(b) Give reasons for non-participation at each stage	6
		(c) Consider use of a flow diagram	6
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	6
Outcome data	15*	Report numbers of outcome events or summary measures	8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8

For peer review only

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

		(*b*) Report category boundaries when continuous variables were categorized	8
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	8
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8
Discussion			
Key results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10
Generalisability	21	Discuss the generalisability (external validity) of the study results	9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	10

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

For peer review only

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

BMJ Open

Developing an AI-assisted digital auscultation tool for the automatic assessment of the severity of mitral regurgitation: protocol for a cross-sectional, non-interventional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-074288.R1
Article Type:	Protocol
Date Submitted by the Author:	19-Jan-2024
Complete List of Authors:	Zhang, Li; Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Department of Cardiology Cheng, Zhenfeng; Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Department of Cardiology Xu, Dongyang; Huzhou Institute of Zhejiang University, Center for Intelligent Acoustics and Signal Processing Wang, Zhi; Huzhou Institute of Zhejiang University, Center for Intelligent Acoustics and Signal Processing; Zhejiang University, State Key Laboratory of Industrial Control Technology Cai, Shengsheng; Huzhou Institute of Zhejiang University, Center for Intelligent Acoustics and Signal Processing; Suzhou Melodicare Medical Technology Co., Ltd. Hu, Nan; Soochow University, School of Electronics and Information Engineering Ma, Jianming; Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Administration Office Mei, Xueqin; Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Department of Medical Engineering
Primary Subject Heading:	Cardiovascular medicine
Secondary Subject Heading:	Health informatics
Keywords:	Valvular heart disease < CARDIOLOGY, Telemedicine < BIOTECHNOLOGY & BIOINFORMATICS, Information technology < BIOTECHNOLOGY & BIOINFORMATICS

SCHOLARONE™
Manuscripts

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Developing an AI-assisted digital auscultation tool for the automatic assessment of the severity of mitral regurgitation: protocol for a cross-sectional, non-interventional study

Li Zhang[1#], Zhenfeng Cheng[1#], Dongyang Xu[2], Zhi Wang[2, 3], Shengsheng Cai[2, 4], Nan Hu[5], Jianming Ma[6], Xueqin Mei[7,*]

[1]Department of Cardiology, Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Huzhou 313000, China

[2]Center for Intelligent Acoustics and Signal Processing, Huzhou Institute of Zhejiang University, Huzhou 313000, China

[3]State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China

[4]Suzhou Melodicare Medical Technology Co., Ltd., Suzhou 215151, China

[5]School of Electronics and Information Engineering, Soochow University, Suzhou 215006, China

[6]Administration Office, Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Huzhou 313000, China

[7]Department of Medical Engineering, Huzhou Central Hospital, the Affiliated Huzhou Hospital, Zhejiang University School of Medicine, Huzhou 313000, China

[#]These authors contributed equally to this work and should be considered as co-first authors.

[*]Corresponding Author

Email addresses: Li Zhang (774624211@qq.com); Zhenfeng Cheng (czf918@qq.com); Dongyang Xu (xudongyang@ustc.edu); Zhi Wang (zjuwangzhi@zju.edu.cn); Shengsheng Cai (caishengsheng@mintti.cn); Nan Hu (hunan@suda.edu.cn); Jianming Ma (hzxzmjm@sina.cn); Xueqin Mei (meixueqin1970@163.com).

ABSTRACT

Introduction: Mitral regurgitation (MR) is the most common valvular heart disorder with morbidity rate of 2.5%. While echocardiography is commonly used in assessing MR, it has many limitations, especially for large-scale MR screening. Cardiac auscultation with electronic stethoscope and artificial intelligence (AI) can be a fast and economical modality for assessing MR severity. Our objectives are: (1) to establish a deep neural network (DNN)-based cardiac auscultation method for assessing the severity of MR; (2) to quantitatively measure the performance of the developed AI-based MR assessment method by virtual clinical trial.

Methods and analysis: In a cross-sectional design, phonocardiogram (PCG) will be recorded at mitral-valve auscultation area of outpatients. The enrolled patients will be checked by echocardiography to confirm the diagnosis of MR or no-MR. Echocardiographic parameters will used as gold standard to assess the severity of MR, classified into four levels: none, mild, moderate, and severe. The study consists of two

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

stages. First, an MR-related cardiac sound database will be created, upon which a DNN-based MR severity classifier will be trained. The automatic MR severity classifier will be integrated with the Smartho-D2 electronic stethoscope. Second, the performance of the developed smart device will be assessed in an independent clinical validation dataset. Sensitivity, specificity, precision, accuracy, and F1-score will be evaluated for the developed smart MR assessment device. The agreement between the performances of the smart device used by cardiologists and by patients themselves will be inspected. The interpretability of the developed model will also be studied with statistical comparisons of occlusion-map-guided variables among the four severity groups.

Ethics and dissemination: The study protocol was approved by the Medical Ethics Committee of Huzhou Central Hospital, China (registration number: 202302009-01). Informed consent is required from all participants. Dissemination will be through conference presentations and peer-reviewed journals.

Study registration: ChiCTR2300069496.

Keywords: Valvular heart disease; Telemedicine; Information technology.

ARTICLE SUMMARY

Strengths and limitations of this study

- An AI-assisted automatic mitral regurgitation severity classifier will be developed and integrated at the Smartho-D2 electronic stethoscope.
- An independent clinical validation data set will be established to enhance the reliability and generalisability of the experiment.
- This study is non-interventional.
- Simultaneously recorded electrocardiogram is not used as reference for PCG cycle segmentation.

INTRODUCTION

Mitral regurgitation (MR) is recognized as one kind of commonly occurring valvular heart disorder, due to its high morbidity rate. An early study of global burden of valvular heart disorder [1] reported that the morbidity rate of MR was as high as 2.5% among all population, and this rate can increase substantially with the growth of age. In the United States, more than 2 million of adults have had MR, and this number was predicted to be doubled in 2030 [2]. A large-scale-community cross-sectional epidemiologic study from Europe [3] revealed that the new diagnostic rate of moderate or severe MR was 2.3%, much larger than that of aortic stenosis (0.7%). A multicentre investigation study reported that MR was the most common valvular heart disorder in China [4].

MR consists of two categories: primary/organic MR and secondary/functional MR.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

Caused by degenerative changes such as Barlow's disease or rheumatic heart diseases, organic MR is due to disruption of the valve leaflets, chordae tendineae and annulus. Compared with organic MR, functional MR (FMR) is more prevalent [5]. FMR is secondary to heart functional or structural abnormalities, resulting in imbalance between mitral valve tethering and closure force and hence poor coverage. Patients with ischemic heart disease, dilated cardiomyopathy, heart failure with preserved ejection fraction (HFpEF), or atrial fibrillation (AF), are apt to have FMR. Pharmacologic therapy, cardiac resynchronization therapy (CRT), or surgical/percutaneous interventions can be applied in treatment for MR patients [6], depending on the MR assessment results. It is very important that MR can be detected and assessed in time, as moderate or severe MR is highly related to hospitalization and long-term prognosis of left ventricular dysfunction and congestive heart failure [7]. However, in the United States around 49% of patients with moderate or severe MR remain undiagnosed [8], and this rate may be increased in developing countries. Therefore, a reliable, flexible and economical tool that can detect MR and assess the severity is essential.

Echocardiography is a key imaging technique for assessing the severity of MR. The severity level of MR is commonly categorized into three levels: mild, moderate, and severe. Qualitative, semi-quantitative and quantitative methods were used to assess the severity level of MR [9]. Quantitative methods have been well validated, while there is still no single echocardiographic parameter or optimal criteria to define the severity level of MR. In the recommendations for non-invasive evaluation of native valvular regurgitation from the American society of echocardiography [10], effective regurgitant orifice area (EROA), regurgitant volume (RVol) and regurgitant fraction (RF) can be used to assess the severity level of MR. EROA < 0.2 cm^2, RVol < 30 mL/beat or RF $< 30\%$ was considered as mild MR; 0.2 cm^2 $\leq$ EROA ≤ 0.39 cm^2, 30 mL/beat $\leq$RVol $\leq$ 59 mL/beat or $30\% \leq$ RF $\leq 49\%$ was considered as moderate MR; and EROA ≥ 0.4 cm^2, RVol ≥ 60 mL/beat or RF $\geq 50\%$ was considered as severe MR. These parameters can be calculated by proximal isovelocity surface area (PISA) method. Though echocardiographic parameters are commonly used to determine the severity level of MR, there are some drawbacks. Firstly, most of the patients may not choose to go to hospital to assess the valvular heart disease, due to lack or over-concentration of medical resource. Secondly, the assessment performance of echocardiography is highly operator-dependent, as it is easy to be affected by transducer frequency, pulsed repetition frequency, and inappropriate gain settings [11]. Lastly, the colour Doppler device is cumbersome and complex to use, hence not suitable for long-term monitoring of MR.

Cardiac auscultation has been an important tool which was often used in first-time cardiac examination. Though cardiac auscultation has the advantage of quick examination, it requires plenty of experiences and skills. For different users, their perceptive acoustic frequency ranges are also different. In comparison to traditional auscultation, in the aid of the emerging electronic stethoscopes [12], smart auscultation manipulates digitalized phonocardiogram (PCG) with digital signal processing and pattern recognition techniques. In abnormal PCG recordings, apart

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Superieur (ABES) .

from the first heart sound (S1) and the second heart sound (S2), other acoustic components such as the third heart sound (S3), the fourth heart sound (S4), murmurs, gallops, clicks, or opening snaps may appear. Features of the PCG signal corresponding to specific heart disease can be extracted and utilized for PCG classification. Dwivedi *et al.* gave a systematic review of automatic PCG analysis and classification [13], where conventional steps in this task was concluded as: preprocessing (denoising), segmentation, feature extraction, and classification. The review has now been updated [14], where deep learning algorithms for heart sound classification in recent years have been supplemented. For cardiac auscultation denoising, various digital signal processing methods [15] have been implemented, and the current trend is to combine these conventional methods with deep neural networks (DNNs) [16]. The necessity of PCG cycle segmentation is controversial, as in some studies reasonable PCG classification performances have also been achieved without segmentation [17]. Feature extraction is important for PCG signal classification, where handcrafted features [18] including time-frequency spectrum, Mel frequency cepstrum coefficients (MFCCs), and discrete wavelet transform (DWT) coefficients have been considered, and in recent years automatic feature extraction from the PCG spectra or waveform by DNNs has been prevalent [19]. For handcrafted feature extraction, it is always followed by traditional machine learning methods including support vector machine (SVM), random forest, Gaussian mixture model, etc. In deep learning methods, such as convolutional neural network (CNN) or recurrent neural network (RNN), the classifiers are typically trained simultaneously with feature extraction step.

Though various PCG signal classification methods are coming forth, most of them were confined to distinguishing normal PCG recordings from abnormal ones, restricted by the task setup in two commonly used public databases: the Physionet heart sound challenge 2016 database [20], and the PASCAL classifying heart sounds challenge 2011 database [21]. Recently, a database for detection of heart valve disorders was shared in the GitHub repository [22]. This database includes 1000 PCG segments, each one lasting for around 2 seconds, truncated from several PCG recording sources in books and websites. The segments were divided evenly into 5 categories: normal, aortic stenosis (AS), MR, mitral stenosis (MS) and mitral valve prolapse (MVP). Several works were carried out in handling this 5-category classification task [23-29]. Time-frequency magnitude and phase features were used [23], and graph-based feature was developed [24], with traditional machine learning methods used for classification. Avanzato *et al.* discovered that with the raw data as the input to 1D-CNN, reasonable classification performance can still be achieved [25]. Karhade *et al.* used both time-domain polynomial chirplet transform (TDPCT) and frequency-domain polynomial chirplet transform (FDPCT) [26] for CNN-based classification. Convolutional and recurrent neural networks were developed by Alkhodari *et al.* [27], and attention mechanism was introduced by Chowdhury *et al.* [28] to address this issue. Recursive feature elimination (RFE) algorithm was applied by Arslan [29] to select the most distinctive deep features for heart valve disorder classification. Though the above-mentioned methods may show potential values in

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

MR detection, there are no studies aiming at assessing the severity of MR, due to the lack of data as well as methods. In spite of this, the existing studies on PCG-based assessment of severity of other heart valve disorders, such as aortic stenosis [30] and tricuspid regurgitation [31], can be used for reference.

Huzhou Institute of Zhejiang University and Melodicare developed an electronic stethoscope device named Smartho-D2, which has been certificated by FDA 510k, NMPA, and CE. Smartho-D2 has two sound picking-up channels, one for auscultation and another for background noise collection. Our developed two-stage noise cancellation algorithm [16] has been applied in this device, to eliminate the background noise contamination. Smartho-D2 is portable, Bluetooth connected, and has self-contained software, supporting PCG waveform displaying, data storage and download from local and the cloud, online remote auscultation, etc. Secondary software development can be easily carried out in Smartho-D2, to add more machine-learning-based heart disease assessment functionalities. In this study, Smartho-D2 will be used for PCG recording from MR patients. With the recorded PCG signals labelled according to echocardiographic examination, an automatic MR severity assessment algorithm will be developed and integrated in Smartho-D2**.** The performance of the developed smart stethoscope in automatically assessing the severity of MR will be evaluated via clinical experiments.

Objectives and hypotheses

The first objective of this study is to establish a DNN-based cardiac auscultation method that can automatically assess the severity of MR. Based on the findings of previous studies [23-31], the hypothesis in this study is that putting PCG signals into a trained DNN can lead to automatic classification of the severity of MR, with high sensitivity and specificity. The second objective is to quantitatively measure the performance of the developed AI-based MR assessment method, when the AI-aided electronic stethoscope Smartho-D2 is used by cardiologists as well as patients. It is hypothesized that the developed smart stethoscope will yield high classification performance metrics in automatically assessing the severity of MR, and the agreement between the performances of the device used by cardiologists and patients themselves will be high.

METHODS AND ANALYSIS

Sample selection and patient recruitment

The study and related experiments will be carried out at Department of Cardiology, Huzhou central hospital, which is located at the north of Zhejiang Province in China, serving for the healthcare of about 4 million people. In each year, about 50,000 outpatients come to seek help from a cardiologist, and over 50% of them suffer from heart valve diseases. By a cross-sectional design, adult patients diagnosed as MR or no-MR will be considered to be enrolled into experiments. The MR patient inclusion criterions are: (1) the patient has been diagnosed to have MR by echocardiographic examination; (2) the patient did not take any medicine that can influence cardiac

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

function in 24 hours; (3) the patient did not have history of heart valve surgery. If the quality of PCG signal recorded from a participant is too low to be utilized, it will be excluded. Gender and age of a MR patient will not be limited. Randomization will not be used for MR patients. The patients diagnosed as no-MR will be matched to that of MR patients by propensity score. The study is conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Written informed consent will be obtained from a patient before the data recording experiments.

Study design

There will be two stages in implementing the study. The first research stage aims at building an automatic MR severity classifier with PCG signal as input. A PCG database for training the MR severity classifier will be built. The database will include PCG recordings collected from recruited patients mentioned above, and the data labels are four levels of the severity of MR: none, mild, moderate, and severe. Echocardiography will be used as gold standard to label the severity of MR. A DNN will be built to form an automatic MR severity classifier, and trained based on our MR-related PCG database. The trained MR severity classifier will be integrated at the Smartho-D2 electronic stethoscope. In the second stage, the performance of AI-aided electronic stethoscope Smartho-D2 will be evaluated in real clinical usage for automatic MR severity assessment. In this stage, the gold standard will still be given by echocardiographic examination. As bias may occur when a smart stethoscope is used by different types of operators, the agreement between automatic MR severity assessment results by a cardiologist user and a patient user will be evaluated. The overall study flow is shown in Figure 1.

For each patient, a cardiologist will collect one PCG recording at the mitral valve auscultation area for at least 10 seconds, with the electronic stethoscope Smartho-D2. The mitral valve auscultation area is at the cardiac apex, located in the fifth intercostal space on the midclavicular line. Smartho-D2 has a wide frequency response range from 20 to 20000 Hz. Evaluated with testing acoustic sources, it was found that 13 dB gain was achieved among 100 ~ 500 Hz, 13 ~ 21 dB gain was achieved among 500 ~ 600 Hz, and 21 dB gain was maintained among 600 ~ 1000 Hz. The sampling rate is 8000 Hz. As primary heart sound components and murmurs typically have frequency ranges below 250 Hz and 600 Hz, respectively, the frequency response characteristics and sampling rate can meet the high-quality requirement of PCG recording. Smartho-D2 is Bluetooth connected, and the Smartho APP is used to monitor and control the data recording. The recorded data can be labelled and stored in a local mobile device, and also can be transmitted to a password-protected database server located at Huzhou central hospital. The storage format of each PCG recording is the uncompressed ".pcm", and no encoding or compression algorithms were used when sending data to the cloud server. An outline of the PCG data recording and storage in this study is shown in Figure 2. The recruited patients accept a routine echocardiographic examination for labelling of the severity level of MR. According to the recommendations for non-invasive evaluation of native valvular regurgitation

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

from the American society of echocardiography [10], EROA, RVol and RF calculated by PISA method will be used for labelling. Mild MR is considered when at least two of the following conditions are satisfied: 0 cm^2 < EROA < 0.2 cm^2, 0 mL/beat < RVol < 30 mL/beat, 0 < RF < 30%. Moderate MR is considered when at least two of the following conditions are satisfied: 0.2 cm^2 ≤ EROA ≤ 0.39 cm^2, 30 mL/beat ≤ RVol ≤ 59 mL/beat, 30% ≤ RF ≤ 49%. Severe MR is considered when at least two of the following conditions are satisfied: ROA ≥ 0.4 cm^2, RVol ≥ 60 mL/beat or RF ≥ 50%. Otherwise no-MR will be considered. The examinations beyond routine ones in this study are cost-free. After protected health information is deidentified, the data can be used by non-medical members for further processing and analysis.

The automatic MR severity classifier to be built will further be integrated at the Smartho-D2 electronic stethoscope. For practical usage, it will consist of adaptive noise cancellation, usable data extraction and the trained DNN. Smartho-D2 has two sound picking-up channels, and a two-stage noise cancellation algorithm [16] has been applied in this device. As reasonable PCG classification performances have been achieved without cardiac cycle segmentation in some previous work [17], in this study we will only divide the PCG recording into segments with sliding windows. A 2-seconds window will be used, with sliding stride as 1 second. In real applications, PCG recording may be contaminated by inappropriate operations such as pressure with excessive force, clothing friction and auscultation leaving the body surface, especially when the stethoscope is used by patient themselves. In this study, before being put into the trained DNN, each PCG segment will be evaluated by a former developed usable data extraction module [16]. With a simple CNN structure, it judges a PCG segment as usable or unusable, where an unusable PCG segment implies that its quality is too low to be used for MR severity assessment. Only usable PCG segments will be put into the DNN trained with the established dataset.

It was recognized that the frequency range of pathological murmur is often wider than that of primary heart sounds, so time-frequency spectrum or MFCCs were usually chosen as the input features to the DNN [18]. However, frequency-domain features may be only parts of useful features in classifying the severity level of MR. Benefited by the power of efficiency in automatic feature extraction, the DNN built in this study is expected to extract and utilize features from data as much as possible for the classification task, with PCG segment as the input. In this study, a clique-block-based DNN [32, 33] will be established, which was believed to take both feature reuse and lightweight into account. A feasible DNN structure is given at the top of Figure 3, where features are automatically extracted by the clique blocks in different levels and concatenated for final classification. 1-D convolutions are used in particular to tackle the processing of 1-D signals or 1-D features. The transition block involves channel-wise attention mechanism, and uses convolution and pooling to achieve dynamic feature calibration. As the MR severity classifier involving a trained DNN should be integrated in a mobile device, the neural architecture search (NAS) [34] will be used to refine the performance and model size of the DNN. For each usable PCG segment, the trained DNN will output a 4-class classification result. To composite the classification results from all usable PCG segments, a majority voting

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

strategy will be applied to give the final decision on the level of MR severity. The flowchart of PCG data processing is shown in Figure 3.

Interventions

There will be no intervention to be implemented to patients in this study. The subjects in this study are patients diagnosed as MR or no-MR by cardiologists, and our objective relies on exploiting the performance of automatically assessing the severity level of MR with PCG signals. The study will not impact the diagnosis or further clinical care.

Sample size calculation

In the first stage of study, the main task is to collect PCG data and train the automatic MR severity classifier. The established database will be divided into 3 pieces equally, and a 3-fold cross-validation will be performed. In each round of model training, 2 pieces are used as training set and 1 piece is used as testing set, and finally the model with mediate values of classification metrics among 3 trained ones will be adopted. Let a one-sample proportion test, the z-test using s(Phat) with continuity correction, be used to assess whether the accuracy of the trained classifier $P1$ is significantly larger than a hypothesized value $P0$. To detect $P1 = 0.95$ compared with $P0 = 0.90$, with one-sided type I error of 0.05 and power of 0.80 calculated by binomial enumeration, the required sample size in a testing set is 179. Hence, the total required sample size in the first stage of study is $179 \times 3 = 537$.

In the second stage of study, all collected PCG data will be used as testing set. In this stage, the PCG as well as the automatic assessment result by the MR severity classifier integrated on the electronic stethoscope will be recorded. This independent clinical validation data set aims at enhancing the reliability and generalisability of the experiment. Compared to the first stage, the required sample size in the second stage seems to be 179. Cohen's kappa will be used to evaluate the agreement between automatic MR severity assessment results by a cardiologist and a patient. When the value of Kappa under the null hypothesis $\kappa 0 = 0.5$, the one under the alternative hypothesis $\kappa 1 = 0.6$ and the proportions of subjects assigned to four categories are 0.3:0.3:0.2:0.2, for a one-sided alternative hypothesis test of $\kappa 1 > \kappa 0$ with type I error of 0.05 and power of 0.80, the required sample size is calculated as 286. As $286 > 179$, the required sample size in the second stage of study should be 286.

Finally, the total required sample size in this study will be $537 + 286 = 823$. Power Analysis and Sample Size (PASS) software V2021 was used to calculate the size of sample to be collected. The schedule of data collection is outlined in Table 1.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Superieur (ABES) .

Table 1. Schedule of data collection

Research stage	Class of patients	Sample size
Stage I: Building the training data set for the classifier (15 March 2023 ~ 15 September 2024)	No-MR	161
	Mild MR	161
	Moderate MR	107
	Severe MR	108
	Total	537
Stage II: Independent clinical validation (16 September 2024 ~ 15 March 2025)	No-MR	86
	Mild MR	86
	Moderate MR	57
	Severe MR	57
	Total	286
Total		823

Statistical analysis

The aim of the first research stage is to build an automatic MR severity classifier. Blinding will be considered in the testing sets. Classification metrics including sensitivity, specificity, precision, accuracy and F1-score will be calculated and assessed. As a multi-category classification task is involved in this study, macro-average, micro-average and weighted-average metrics will be calculated. Macro-average means direct average of binary-classification metrics across classes. The micro-average metrics are calculated by using aggregate outcomes across all classes. The weighted-average metrics are given by weighted average of binary-classification metrics across classes, with proportions of classes as weights. Their one-versus-rest receiver operating characteristic (ROC) curves and corresponding area under curve (AUC) with confidence interval will also be given.

The above-mentioned classification metrics will also be assessed in the second stage, where the performance of established MR severity assessment system will be evaluated in real clinical usage. In this stage, to measure the bias when a smart stethoscope is used by different types of users, the agreement between MR severity assessment results by a cardiologist user and a patient user will be evaluated, with the metric of Cohen's kappa. Kappa score > 0.8 will be deemed as good agreement.

The interpretability of the developed classification model will also be considered. Inspired by a robust interpretable deep-learning-based PCG classifier study [17], occlusion maps will be employed in this study to show which area (S1, systole, S2, or diastole) in each PCG cycle is automatically emphasized by the model for MR severity classification. After the emphasized areas are determined, three classes of computerized feature variables, including temporal-domain features, frequency-domain features and chaotic features, would be calculated from these areas for statistical comparison among the four severity groups (none, mild, moderate, and severe). One-way ANOVA would be used first, and then SNK-q would be used for pairwise comparison. $P < 0.5$ would be considered as statistical significance. The above deep-learning-guided data mining for significant variables would be expected

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Superieur (ABES) .

to show the interpretability of the developed model.

Data collection and analysis plan

The data collection started on 15 March 2023, and it is planned to last for 2 years. The first stage of study is planned to be finished on 15 September 2024. Considering about 50,000 outpatient visits in one year, we believe that a one-year 1st stage data collection plan will easily meet the sample size requirement for training, even with a conservative MR morbidity rate of 2.5% for these outpatients. The Smartho APP, deployed in a password-protected android pad, is used to monitor and control the PCG data recording. Echocardiographic examination reports given by a skilled expert with above 10-year experiences will be used in data labelling. The labelled data will be transferred from local device to a password-protected database server located at Huzhou central hospital, and anonymized data files can be further processed and analysed by non-medical members. Data analysis will be started on 15 March 2024, for classification model establishing and integration on the device. With the growth of recorded data size, the weights in the classification model will be updated through training, and the classification performance is expected to be refined.

Patient and public involvement

None.

ETHICS AND DISSEMINATION

Ethical approval was obtained from the Medical Ethics Committee of Huzhou Central Hospital, China (registration number: 202302009-01). This study was registered with Chinese Clinical Trial Registry, ChiCTR (ChiCTR2300069496). Before PCG recording, written informed consent will be obtained from all patients. Access to collected data will be restricted to cardiologists from the research team treating the participants, and anonymized data files are shared to non-medical members of the research team. Dissemination will be through conference presentations and peer-reviewed journals.

DISCUSSION

Though it was realized a long time ago that quantitatively analysing PCG signal may provide a convenient and fast way to assess heart valve diseases, there is no PCG-based method or device to automatically assess the severity of MR, a most common valvular heart disorder. In recent years, by using a public shared heart valve disorder PCG database [22], many studies [23-29] have been published, which tried to distinguish MR from normal and the other 3 valvular heart disorders. From these existing studies, MR related PCG signals show specific pattern that can be employed in MR detection. Though assessing the severity of MR using PCG signal has not been studied in the past, the similar methodology was applied in automatically grading the severity level of aortic stenosis [30] and tricuspid regurgitation [31]. By studying the PCG signals and echocardiographic parameters of 41 aortic stenosis patients [30], it was found that there was a good correlation between the duration of spectra above

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

300 Hz and the Doppler derived peak pressure gradient. Rujoie *et al.* studied the PCG signals of 22 subjects (divided into three categories: non, mild, and moderate to severe tricuspid regurgitation) [31], using MFCC and wavelet transform methods for feature extraction, genetic algorithms for feature selection, and KNN to implement the triple-classification task. The data recording has been started recently, and Figure 4 shows typical recorded PCG cases corresponding to different levels of MR severity. It seems that even the waveforms may reflect some information for classification, *e.g.*, the pattern of systolic murmur. Different from the research based on handcraft features and small-size database, our study considers larger-size data collection and automatic feature extraction via DNN.

The application of deep learning to analysis of PCG signals is becoming promising in recent years. The existing studies in this field, taking discrimination among 5-class valvular heart disorders [22] for instance, involved various input features or network structures. To mimic the processing of images, time-frequency images were usually derived as input features to 2-D CNNs. Karhade *et al*. used a two-layered 2-D CNN [26], and Chowdhury *et al*. introduced attention mechanism [28] to the 2-D CNN. Arslan used continuous wavelet transform (CWT) to generate scalogram images, and employed multilayer extreme learning machine (ML-ELM) and several pretrained DNNs [29], including VGG16, ResNet50 and MobileNetV2, to extract deep features. Handcraft features may cause loss of information for classification, and therefore several studies tried to use 1-D convolutions to extract features from a raw data segment. Avanzato *et al*. built a 5-layered 1-D CNN [25] to extract features from raw PCG data, while Alkhodari *et al*. used a 3-layered 1-D CNN cascaded by a BiLSTM [27]. In this study, we consider to directly process the noise-cancellated PCG segment with a clique-block-based DNN, where clique blocks and channel-wise attention are employed. The clique block has displayed its advantages in feature extraction and weights reuse, compared with other prevalent blocks such as residual blocks or dense blocks. Apart from the feature extraction issue, another major drawback of the existing deep-learning-based studies is that the developed "black boxes" may lead to ethically problematic outcomes [35]. Inspired by the pioneer research on the interpretability of the PCG classification model [17], we use the trained deep learning model to guide the way to the discovery of areas in a PCG cycle that are important for classification. The interpretability would be further enhanced by statistically comparing computerized variables derived from these areas among groups.

The MR severity level classifier will be integrated on the Smartho-D2, an electronic stethoscope device certificated by FDA 510k, NMPA, and CE. In recent years, some AI-aided auscultation products are emerging for cardiac healthcare. In 2019, the eMurmur Heart AI received FDA clearance and CE marking for detecting the presence of a murmur in PCG recording and determining if it is likely innocent or pathological. Evaluation from 603 outpatient visits [36] showed that the eMurmur had 93% (CI 90-95%) sensitivity and 81% (CI 85-91%) specificity for detection of pathologic cases. In 12 July 2022, FDA cleared the Eko Murmur Analysis Software (EMAS) for detecting and characterizing murmurs found in adult and paediatric patients. The EMAS was reported to identify structural murmurs with a sensitivity of

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

90.2% and specificity of 90.6% among adults, and have an overall sensitivity of 85.6% and specificity of 84.4% [37]. These existing smart cardiac auscultation applications showed feasibility of developing valvular heart disorder assessment methods on electronic stethoscope devices.

A limitation of this study is that only PCG will be utilized, without simultaneously recorded electrocardiogram (ECG). In many studies, ECG was used for reference of PCG cycle segmentation [38]. However, in recent years some studies showed that segmentation may not be necessary when deep learning is used for cardiac sound classification. By using Shapley values and occlusion maps [17], Dissanayake *et al*. showed that deep learning methods can automatically emphasize the components in PCG data that make significant contribution to classification. In our study, we consider to classify the PCG signals without PCG cycle segmentation. The collected large-size data will endow the established DNN-based MR severity level classifier with the ability of automatically selecting the useful PCG components. A multicentre study involving more subjects in a virtual clinical trial will be among our future work, to refine and evaluate the performance of the developed method and smart device.

For peer review only

Contributors

Li Zhang and Zhenfeng Cheng conceptualized and designed the study, and wrote the original draft. Dongyang Xu, Zhi Wang and Nan Hu will analyse the recorded data using digital signal processing algorithms and design the DNN-based classifier. Shengsheng Cai will be responsible for technical supports for the device. Jianming Ma will be responsible for project administration. Xueqin Mei provides mentorship to this study, edited and revised the draft, and will be responsible for project administration. All authors critically reviewed and approved the final manuscript before submission.

Funding

This work was supported by the Huzhou Science and Technology Project (grant 2022GZ46) and the Cardiovascular Discipline Group Funds of Huzhou (grant XKQ-HT-202102A).

Competing interests

Shengsheng Cai is an employee and owner of Suzhou Melodicare Medical Technology Co., Ltd., which developed the Smartho-D2 device. Li Zhang, Zhenfeng Cheng Dongyang Xu, Zhi Wang, Nan Hu, Jianming Ma and Xueqin Mei declare no conflicts of interest.

Data availability statement

The dataset will be shared via a Github repository after the study results are published.

REFERENCES

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

[1] Nkomo VT, Gardin JM, Skelton TN, Gottdiener JS, Scott CG, Enriquez-Sarano M. Burden of valvular heart diseases: a population-based study. *Lancet* 2006;368(9540):1005–1011.

[2] de Marchena E, Badiye A, Robalino G, *et al*. Respective prevalence of the different carpentier classes of mitral regurgitation: a stepping stone for future therapeutic research and development. *J Card Surg* 2011;26(4):385-392.

[3] d'Arcy JL, Coffey S, Loudon MA, *et al*. Large-scale community echocardiographic screening reveals a major burden of undiagnosed valvular heart disease in older people: The OxVALVE population cohort study. *Eur Heart J* 2016;37(47):3515-3522.

[4] Zhang B, Xu H, Zhang H, *et al*. Prognostic value of N-terminal pro-B-type natriuretic peptide in elderly patients with valvular heart disease. *J Am Coll Cardiol* 2020;75(14):1659-1672.

[5] Dziadzko V, Dziadzko M, Medina-Inojosa JR, *et al*. Causes and mechanisms of isolated mitral regurgitation in the community: clinical context and outcome. *Eur Heart J* 2019;40(27):2194-2202.

[6] Otto CM, Nishimura RA, Bonow RO, *et al*. 2020 ACC/AHA Guideline for the management of patients with valvular heart disease: Executive summary: A report of the American College of Cardiology American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* 2021;143(5):e35–71.

[7] 2nd Sannino A , Smith RL , Schiattarella GG, *et al*. Survival and cardiovascular outcomes of patients with secondary mitral regurgitation: A systematic review and meta-analysis. *JAMA Cardiol* 2017;2(10):1130–1139.

[8] Head SJ, van Leeuwen WJ, Van Mieghem NM, *et al*. Surgical or transcatheter mitral valve intervention: Complex disease requires complex decisions. *EuroIntervention* 2014;9(10):1133-1135.

[9] 3rd Otto CM, Nishimura RA, Bonow RO, *et al*. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: Executive summary: A report of the American college of cardiology American heart association joint committee on clinical practice guidelines. *Circulation* 2021;143(5):e35–71.

[10] Zoghbi WA, Adams D, Bonow RO, *et al*. Recommendations for noninvasive evaluation of native valvular regurgitation: a report from the American society of echocardiography developed in collaboration with the society for cardiovascular magnetic resonance. *J Am Soc Echocardiogr* 2017;30(4):303-371.

[11] Stevenson JG. Two-dimensional color Doppler estimation of atrioventricular valve regurgitation: important effects on instrument gain setting, pulse repetition frequency and carrier frequency. *J Am Soc Echocardiogr* 1989;2:1–10.

[12] Bank I, Vliegen HW, and Bruschke AVG. The 200th anniversary of the stethoscope: Can this low-tech device survive in the high-tech 21st century? *Eur Heart J* 2016;37(47):3536-3543.

[13] Dwivedi AK, Imtiaz SA, and Odriguez-Villegas E. Algorithms for automatic analysis and classification of heart sounds - A systematic review. *IEEE Access* 2019;7:8316-8345.

[14] Chen W, Sun Q, Chen X, *et al*. Deep learning methods for heart sounds classification: A systematic review. *Entropy* 2021;23:667.

[15] Mondal A, Saxena I, Tang H, *et al*. A noise reduction technique based on nonlinear kernel function for heart sound analysis. *IEEE J Biomed Health Inform* 2018;22(3):775-784.

[16] Yang C, Dai N, Wang Z, Cai S, Wang J, and Hu N. Cardiopulmonary auscultation enhancement with a two-stage noise cancellation approach. *Biomed Signal Process Control* 2023;79:104175.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

[17] Dissanayake T, Fernando T, Denman S, Sridharan S, Ghaemmaghami H, Fookes C. A robust interpretable deep learning classifier for heart anomaly detection without segmentation. *IEEE J Biomed Health Inform* 2020;99:2162-2171.

[18] Xu W, Yu K, Ye J, *et al*. Automatic pediatric congenital heart disease classification based on heart sound signal. *Artif Intell Med* 2022;126:102257.

[19] Li S, Li F, Tang S, and Luo F. Heart sounds classification based on feature fusion using lightweight neural networks. *IEEE Trans Instrum Meas* 2021;70:4007009.

[dataset] [20] Clifford GD, Liu C, Moody B, *et al*. Classification of normal/abnormal heart sound recordings: The PhysioNet/computing in cardiology challenge 2016. *Proc. Comput. Cardiology Conf. (CinC)* 2016;43:609-612. Data from: PhysioNet Repository, version March 1, 2016. https://archive.physionet.org/challenge/2016

[dataset] [21] Bentley P, Nordehn G, Coimbra M, *et al*. Data from: The PASCAL Classifying Heart Sounds Challenge, version November 1, 2011. http://www.peterjbentley.com/heartchallenge

[dataset] [22] Yaseen GS and Kwon S. Classification of heart sound signal using multiple features. *Appl Sci* 2018;8:2344. Data from: Github Repository, version June 11, 2018. https://github.com/yaseen21khan/Classification-of-Heart-Sound-Signal-Using-Multiple-Features-

[23] Ghosh SK, Tripathy RK, Ponnalagu RN, and Pachori RB. Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features. *IEEE Sen Lett* 2019;3(12):7002604.

[24] Tuncer T, Dogan S, Tan RS, Acharya UR. Application of Petersen graph pattern technique for automated detection of heart valve diseases with PCG signals. *Inform Sciences* 2021;565:91-104.

[25] Avanzato R and Beritelli F. Heart sound multiclass analysis based on raw data and convolutional neural network. *IEEE Sen Lett* 2020;4(12): 7004104.

[26] Karhade J, Dash S, Ghosh SK. Time–frequency-domain deep learning framework for the automated detection of heart valve disorders using PCG signals. *IEEE Trans Instrum Meas* 2022;71:2506311.

[27] Alkhodari M and Fraiwan L. Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. *Comput Meth Prog Bio* 2021;200:105940.

[28] Chowdhury S, Morshed M, and Fattah SA. SpectroCardioNet: An attention-based deep learning network using triple-spectrograms of PCG signal for heart valve disease detection. *IEEE Sensors J* 2022;22(23):22799-22807.

[29] Arslan Ö. Automated detection of heart valve disorders with time-frequency and deep features on PCG signals. *Biomed Signal Process Control* 2022;78:103929.

[30] Kim D and Tavel ME. Assessment of severity of aortic stenosis through time-frequency analysis of murmur. Chest 2003;124:1638-1644.

[31] Rujoie A, Fallah A, Rashidi S, Khoshnood ER, and Ala TS. Classification and evaluation of the severity of tricuspid regurgitation using phonocardiogram. *Biomed Signal Process Control* 2020;57:101688.

[32] Yang Y, Zhong Z, Shen T, and Lin Z. Convolutional neural networks with alternately updated clique. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)* 2018:2413-2422.

[33] Xiao B, Xu Y, Bi X, *et al.* Follow the sound of children's heart: A deep-learning-based computer-aided pediatric CHDs diagnosis system. *IEEE Internet Things J* 2020;7(3):1994-2004.

[34] Zoph B and Le QV. Neural architecture search with reinforcement learning. *Proc. Int. Conf. Learn. Representations* 2017.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

For peer review only

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Enseignement Superieur (ABES) .

[35] Char DS, Shah NH, and Magnus D. Implementing machine learning in health care - addressing ethical challenges. *N Engl J Med* 2018;15(378):981–983.

[36] Thompson WR, Reinisch AJ, Unterberger MJ, and Schriefl AJ. Artificial intelligence-assisted auscultation of heart murmurs: Validation by virtual clinical trial. *Pediatr Cardiol* 2019;40:623-629.

[37] Eko Murmur Analysis Software User Manual. https://www.ekohealth.com

[38] Zhang H, Wang X, Liu C, *et al*. Discrimination of patients with varying degrees of coronary artery stenosis by ECG and PCG signals based on entropy. *Entropy* 2021;23:823.

For peer review only

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Figure titles:

Figure 1. Study flowchart

Figure 2. Outline of the PCG data recording and storage

Figure 3. Flowchart of PCG data processing and structure of a feasible DNN to be used in the automatic MR severity classifier

Figure 4. Typical PCG cases corresponding to different levels of MR severity, collected in this study (data collected began on 15 March 2023)

For peer review only

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

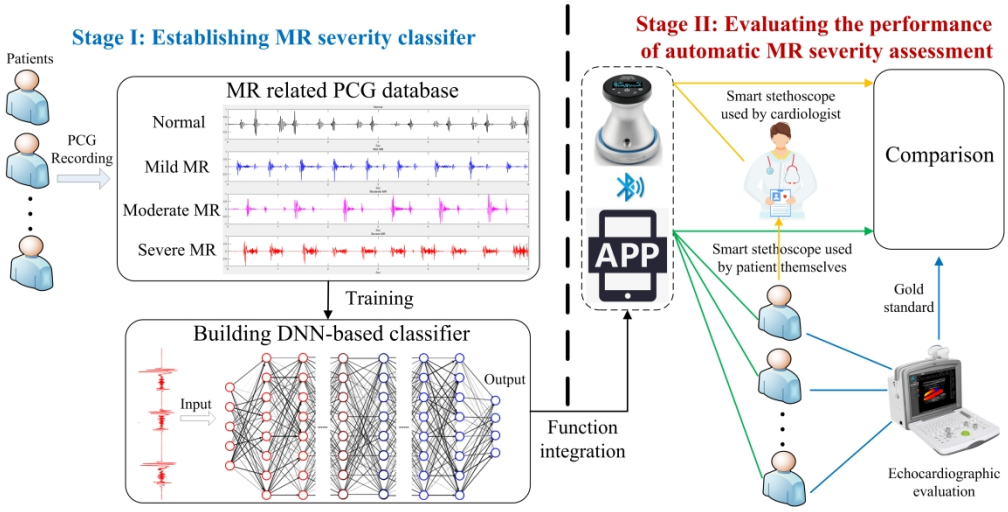

Fig. 1. The overall study flow

283x142mm (400 x 400 DPI)

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Fig. 2. A sketch of the PCG data recording and storage

260x82mm (500 x 500 DPI)

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

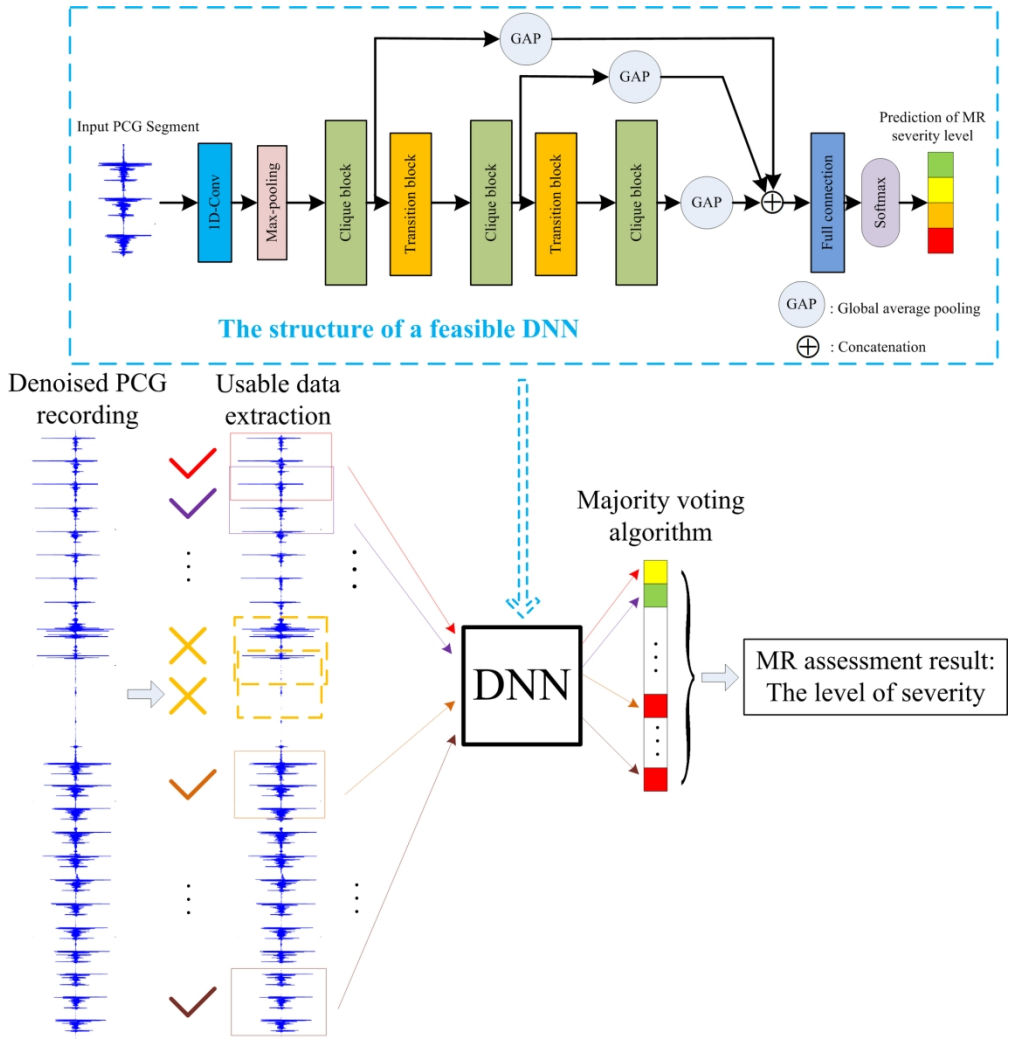

Fig. 3 The flowchart of PCG data processing and the structure of a feasible DNN to be used in the automatic MR severity classifier.

185x191mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

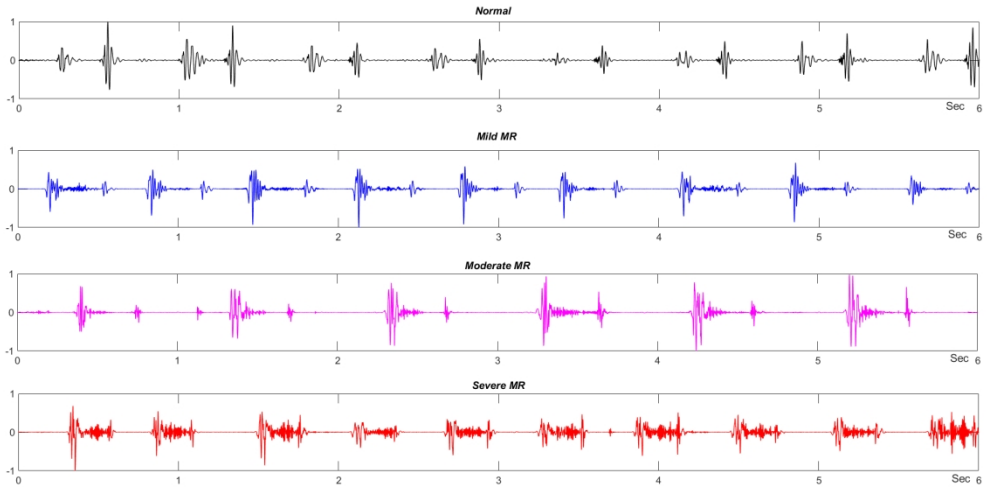

Fig. 4 Typical PCG cases corresponding to different levels of MR severity, collected in this study started on 15 March 2023

508x252mm (96 x 96 DPI)

BMJ Open: first published as 10.1136/bmjopen-2023-074288 on 29 March 2024. Downloaded from http://bmjopen.bmj.com/ on June 10, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) .
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5-7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-9
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	5-6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	7-8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8
		(*b*) Describe any methods used to examine subgroups and interactions	6
		(*c*) Explain how missing data were addressed	6
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	6
		(*e*) Describe any sensitivity analyses	8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7-8
		(b) Give reasons for non-participation at each stage	6
		(c) Consider use of a flow diagram	6
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	6
Outcome data	15*	Report numbers of outcome events or summary measures	8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Enseignement Superieur (ABES) .

		(*b*) Report category boundaries when continuous variables were categorized	8
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	8
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8
Discussion			
Key results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10
Generalisability	21	Discuss the generalisability (external validity) of the study results	9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	10

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

For peer review only

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.