Multimodal fusion diagnosis of depression and anxiety based on CNN-LSTM model

Wanqing Xie\textsuperscript{a,b,c,1}, Chen Wang\textsuperscript{d,1}, Zhixiong Lin\textsuperscript{e}, Xudong Luo\textsuperscript{e}, Wenqian Chen\textsuperscript{d}, Manzhu Xu\textsuperscript{f}, Lizhong Liang\textsuperscript{e,g}, Xiaofeng Liu\textsuperscript{a,h}, Yanzhong Wang\textsuperscript{i}, Hui Luo\textsuperscript{j}, Mingmei Cheng\textsuperscript{a,b,**}

\textsuperscript{a} Department of Intelligent Medical Engineering, School of Biomedical Engineering, Anhui Medical University, Hefei, China
\textsuperscript{b} Department of Psychology, School of Mental Health and Psychological Sciences, Anhui Medical University, Hefei, China
\textsuperscript{c} Suzhou Fanhan Information Technology Company, Ltd, Suzhou, China
\textsuperscript{d} College of the Mathematical Sciences, Harbin Engineering University, Harbin, China
\textsuperscript{e} Department of Psychiatry, Affiliated Hospital of Guangdong Medical University, Zhanjiang, China
\textsuperscript{f} Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
\textsuperscript{g} School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
\textsuperscript{h} Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA
\textsuperscript{i} School of Population Health \& Environmental Sciences, Faculty of Life Science and Medicine, King’s College London, London, UK
\textsuperscript{j} Marine Biomedical Research Institute of Guangdong Medical University, Zhanjiang 510240, China

ARTICLE INFO

Keywords:
Multimodal fusion
Depression
Anxiety
Convolutional neural network
Long short-term memory

ABSTRACT

Background: In recent years, more and more people suffer from depression and anxiety. These symptoms are hard to be spotted and can be very dangerous. Currently, the Self-Reported Anxiety Scale (SAS) and Self-Reported Depression Scale (SDS) are commonly used for initial screening for depression and anxiety disorders. However, the information contained in these two scales is limited, while the symptoms of subjects are various and complex, which results in the inconsistency between the questionnaire evaluation results and the clinician’s diagnosis results. To fully mine the scale data, we propose a method to extract the features from the facial expression and movements, which are generated from the video recorded simultaneously when subjects fill in the scale. Then we collect the facial expression, movements and scale information to establish a multimodal framework for improving the accuracy and robustness of the diagnosis of depression and anxiety.

Methods: We collect the scale results of the subjects and the videos when filling in the scales. Given the two scales, SAS and SDS, we construct a model with two branches, where each branch processes the multimodal data of SAS and SDS, respectively. In the branch, we first build a convolutional neural network (CNN) to extracts the facial expression features in each frame of images. Secondly, we establish a long short-term memory (LSTM) network to further embedding the facial expression feature and build the connections between various frames, so that the movement feature in the video can be generated. Thirdly, we transform the scale scores into one-hot format, and feed them into the corresponding branch of the network to further mining the information of the multimodal data. Finally, we fuse the embeddings of these two branches to generate inference results of depression and anxiety.

Results and conclusions: Based on the score results of SAS and SDS, our multimodal model further mines the video information, and can reach the accuracy of 0.946 in diagnosing depression and anxiety. This study demonstrates the feasibility of using our CNN-LSTM-based multimodal model for initial screening and diagnosis of depression and anxiety disorders with high diagnostic performance.

* Correspondence to: Southern Marine Science and Engineering Guangdong Laboratory, Zhanjiang, Guangdong 511458, China.
** Corresponding author at: Department of Intelligent Medical Engineering, School of Biomedical Engineering, Anhui Medical University, Hefei, China.
E-mail addresses: haihuo524023@gmail.com (H. Luo), chengmm68@gmail.com (M. Cheng).
1 These authors contribute equally to this manuscript.

https://doi.org/10.1016/j.compmedimag.2022.102128
Received 5 July 2022; Received in revised form 20 September 2022; Accepted 28 September 2022
Available online 4 October 2022
0895-6111/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

1.1. Background

With the fierce competition in modern society, people are facing challenges and pressures such as heavy work, accelerated life rhythm, and complex interpersonal relationships, which have resulted in more and more patients with mental and psychological diseases, among which depression and anxiety are two of the most prominent disorders. Globally, more than 264 million people suffer from depression and more than 374 million people suffer from anxiety disorders (Santomauro et al., 2021; Kroenke et al., 2001). In China, the incidence of anxiety is 4.98 %, and the incidence of depression is 4.06 % (Huang et al., 2019). These two diseases cause great suffering to patients. People with depression have some obvious characteristics, such as long-term depression, pessimism and even suicidal thoughts, which can last for several years and bring a heavy burden to patients (Machmutow et al., 2019). Anxiety patients are mainly panicked, nervous, and irritable. Patients can recover through drug treatment and psychological counseling (Sharrock et al., 2021; Kennedy et al., 2021). However, these symptoms are not easily detected, so we developed a model for automatic diagnosis of depression and anxiety.

Facial expression and movement are important for the diagnosis of depression and anxiety (Corneanu et al., 2016). However, simple self-assessment scales cannot provide such information for diagnosis. Inspired by this, we collect SAS and SDS scale data of patients and record the facial video while patients answer the scale. In order not to interfere with the subject’s answer, each subject fills in the scale in a separate room, and the software-defined camera (SDC) system (Yuan et al., 2020) records the facial video at the same time. Videos record many expressions and actions of subjects, such as crying, scratching the head, wiping the face, etc., which can provide rich information for clinical diagnosis (Pampouchidou et al., 2017; Zhu et al., 2020).

Based on the scale results and facial videos of the patient, we collect data from two modalities. Each person needs to answer 20 questions of SAS and 20 questions of SDS, so we record 40 videos for each people (Xie et al., 2021a; Jegede, 1977). Based on the multimodal data, we propose a multimodal fusion model with two branches. Each branch processes and learns facial features from SAS and SDS videos, and then fuses and extracts long-term video features through a convolutional neural network (CNN) (Xie et al., 2021a; Wang et al., 2021a; SravyaPranati et al., 2020) and long short-term memory network (LSTM) (Ogawa et al., 2018; Naem et al., 2021). Next, the model fuses the video features with scale scores, and finally gathers the multimodal features of the two scales in the decision layer to form the final inference.

1.2. Objective

Patients with depression and anxiety disorder increased rapidly during COVID-19 (Chopra et al., 2021; Hyland et al., 2020). However, due to policy restrictions, it is difficult for doctors to diagnose depression and anxiety disorders face to face. Therefore, our goal is to establish a model that can be deployed on a variety of devices, such as mobile phones, computers, and so on. This model can not only realize large-scale screening but also help doctors diagnose depression and anxiety.

Contributions of this paper:

2. Related work

2.1. Machine learning

With hand-craft features based on prior knowledge, machine learning technology can provide powerful support for classification decision-making. In the field of supervised learning, Naive Bayes is a simple and common classifier that can be used for numbers. It is a Bayesian network with only one parent node and several child nodes. Support vector machines are a good choice when dealing with high-dimensional data, overfitting can be avoided with appropriate kernels, and this method has high accuracy (Sen et al., 2020; Kotsiantis and Zaharakis, 2007).

2.2. Current status of diagnosis of depression and anxiety disorders

Liu et al. trained the extracted text features and extended features through Support Vector Machines to detect whether Sina Weibo users are prone to depression, with an accuracy rate of 83.27 %. However, this data does not have the important feature of human expressions (Liu et al., 2021). Orabi et al. used convolution to model Twitter Users, and the accuracy of diagnosing depression was 87.95 %, however, they only used text features for classification and no other features were used for multimodal fusion diagnosis (Orabi et al., 2018). Haque et al. used the DAIC-WOZ dataset for automatic inspection of severe depression, extracted video features through 3D, and then fused voice features for modeling, and the model accuracy reached 83.3 % (Haque et al., 2018). Y. Xie et al. used a convolutional neural network to diagnose depression and anxiety disorders by Electroencephalography (EEG), and the model accuracy rate was 67.67 % (Xie et al., 2020). Wang et al. made a diagnosis of depression and anxiety by integrating SDS and SAS video, with an accuracy rate of 80.22 %. However, this paper does not use the important feature of scale scores (Wang et al., 2021b). Nayak et al. adopted visual and thermal face image sequences and proposed architecture with CCNN (cascaded Convolutional Neural Network) and HMM (hidden Markov model) for estimating the posterior probability of diagnosing depression and anxiety by clustering sequences of emotional states. The model has been validated through DASS-21 (Depression Anxiety Depression Scale) subjective analysis and realized the accuracy of 91.98 % (Nayak et al., 2021).

3. Material and methods

3.1. Data collection and data preprocessing

In order to collect data with higher quality, we provide a separate and quiet room for each subject. In order to avoid the influence of the camera, we hid the Huawei software-defined camera behind the one-way mirror to hide the shooting (Xie et al., 2021a, 2021b). As shown in Appendix Fig. 1. Through this software, we collect data of 303 subjects (Among them, normal was 106, anxiety was 103 and depression was 94), each of whom is diagnosed by the professional doctor. The research protocol was approved by the Ethics Committee of the Affiliated Hospital of Guangdong Medical University (No. PJ2021–026).

Subjects need to answer SDS and SAS questions. Each scale contains 20 questions. We divide the scale scores into 4 levels. We take SAS scores as an example: scores less than 45 are normal, between 45 and 59 are moderate anxiety, between 60 and 74 are severe anxiety, and more than 75 are extreme anxiety (Dunstan and NJBp, 2020). The SDS score is also divided in a similar way (Zung, 1965).

SAS can only assess anxiety, while SDS can only assess depression, which may lead to errors in the diagnosis of depression and anxiety when only adopting these two scales for diagnosis. As shown in the first row of Table 1, SDS assesses the subject as the model depression, and SAS assesses the subject as the mild annex disorder, while the doctor’s diagnosis was an anxiety disorder. To solve this problem, we propose a

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Partial diagnostic results.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor diagnosis</td>
<td>SDS score diagnosis</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>Moderate depression</td>
</tr>
<tr>
<td>Depressive disorder</td>
<td>Moderate depression</td>
</tr>
<tr>
<td>Depressive disorder</td>
<td>Severe depression</td>
</tr>
</tbody>
</table>
multimodal diagnosis model of anxiety and depression that integrates SAS and SDS scales results and video information of the subjects.

In order to better integrate the video and scale information, we preprocess the scale data. Concretely, we partition the SAS assessment results into four categories, where 0 represents normal, 1 represents mild anxiety, 2 represents moderate anxiety, and 3 represents severe anxiety. SDS score assessments are handled in a similar way. The doctor’s final diagnosis is served as the final label with 0 being normal, 1 being depressed, and 2 being anxious (as shown in Table 2).

The preprocessing of video data is shown in Fig. 1 (In order to protect the privacy of patients, we code the face). We first crop the facial features of interest, and then scale the video frame size to $224 \times 224$. Rich and important information is contained in the expressions and actions of the human face, including crying, covering the face, shaking the head, etc., which are all recorded in the video.

3.2. Classification methods applied to scale score

In this paper, three different classification methods are first selected to establish a classification model of depression and anxiety disorders based only on the scales: support vector machine (SVM), Naive Bayes, and AdaBoost algorithm.

SVM is a supervised classification method and one of the classic algorithms in machine learning (Shia and Chen, 2021). The main advantages of this model: First, compared to other classification algorithms, SVM does not require too many samples. Since SVM introduces kernel functions and slack variables, it can handle high-dimensional functions. Second, SVM solves the problem by penalizing variables and kernel functions in the case where the sample data is linearly inseparable (VJSoE, 2006; Cusick et al., 2021). We first digitize the scale diagnosis and then use the doctor’s diagnosis as a label. In this classification, we choose the linear kernel function and the penalty coefficient is set to 10.

Naive Bayes is a widely used algorithm in Bayesian taxonomy. The algorithm itself is derived from Bayes’ theorem. When determining the target, it is considered that the attribute features of each part are independent of each other, and the dimensions of the feature vector of each object are also independent of each other and are not related to each other. In this paper, depression, anxiety, and normal are the items to be classified (Brahim et al., 2019). To find out the probability of each category appearing under the conditions of the occurrence of this item, and to identify which item this category belongs to, our prior probability is a multinomial distribution (Islam et al., 2007).

The AdaBoost algorithm is ensemble learning. It combines multiple machine learning models to form a more accurate model (Ying et al., 2013). We use the AdaBoost algorithm for classification, predicting the true condition of the subjects based on the SAS and SDS scores (Jia and Zhuang, 2021).

3.3. Video and score fusion model

To construct our multimodal model, we collect 20 videos of each scale for each subject. Long video time brings huge computing costs. In order to expand the data and avoid missing expression information, we set the overlap rate to 0.5, that is, a video is divided into 1–10 frames, and then 5–15 frames, so an n-frame video can be divided into $M = 2 \cdot \left( \frac{n}{15} \right) - 1$ frames. To obtain video features, we first sparsely extract frames from the videos, as shown in Fig. 2. We use $I$ to represent the SAS video frame, and $J$ to represent the SDS video frame. For SDS videos, the superscript $m$ represents the question answered, and the subscript $n$ represents the number of video frames, $P_n^m$ represents the video of the subject answering the SAS, $J_n^m$ represents the video of the subject answering the SDS.

Taking the SDS video as an example, the data feature size is $10 \times 3 \times 224 \times 224$. We simultaneously feed 20 videos of the subjects to the network (see Appendix Table 1 for specific parameters), which is represented by the following formula:

$$ F_c = f(P_n^m) $$ (1)

The subscript $c$ of $F_c$ represents SAS video or SDS video. In order to track the temporal features of the video, we input the extracted $F_c$ into the LSTM (Lu et al., 2020; Mårtensson et al., 2019), as shown in Fig. 3.

First, the feature $F_{c-1}^c$ extracted at the last time-stamp is processed by the forget gate. If the value is 1, the information will be passed down. If the value is 0, the information will be lost. The formula is as follows:

$$ F_{c} = \sigma_i (\omega_i \cdot [H_{c-1}^c, F_c] + b_i) $$ (2)

In order to serialize the video features, the features need to be updated, the features to be updated are generated through the tanh layer, and the sigmoid function is used to determine which video features need to be updated, as follows:

$$ r_c^t = \delta_i (\omega_i \cdot [H_{c-1}^c, F_c] + b_i) $$ (3)

$$ C_c^t = \tanh (\omega_i \cdot [H_{c-1}^c, F_c] + b_i) $$ (4)

Then, we can update the extracted video sequence features. Let $C_{c-1}^c$ represent the result of the forget gate. The input gate can be represented by the following formula:

$$ C_c^t = f_c^t \times C_{c-1}^c + r_c^t \times \tilde{C}_c^t $$ (5)

Finally, we need to determine the output value of the model. First, we use the Sigmoid function to determine which part of the video feature needs to be output, and then use the tanh function to process the update value $C_c^t$. Finally, the multiplication of these two parts is the output value, which can be depicted as:

$$ e_c^t = \sigma_i (\omega_i \cdot [H_{c-1}^c, F_c] + b_i) $$ (6)

$$ H_c^t = e_c^t \times \tanh (e_c^t) $$ (7)

Then, we fuse the extracted video sequence features with SAS and SDS scores. The fusion details are shown in Fig. 4. The 20 video features answered by the subjects are fused together, and the fusion dimension is 2560. Then the 2560-dimension feature is stretched to 2048 dimension. To fuse the score with video feature of human face, we convert the score to One-Hot format and resize it to a 2048-dimensional tensor. Next, we fuse the scale score with the video features, and the 20 video features and digital diagnostic information answered by the subject become 4096 dimensions. After that, the information from the two branches was fused into 8192 dimensions, which were then stretched to 3 dimensions for diagnosing depression and anxiety (Zhang et al., 2020). The final loss function is

$$ L(y', T) = \frac{1}{n} \sum_{i=1}^{n} \left[ y'_i \log(T_i) + (1 - y'_i) \log(1 - T_i) \right] $$ (8)

where $n$ represents the number of categories and $s$ represents the scale diagnostic digitized score.

4. Experiment

Based on subject SAS and SDS scores, we first apply traditional algorithms Support Vector Machines (SVM), Naive Bayes and AdaBoost to...
the diagnosis of depression and anxiety disorders. Furthermore, we design a fusion model of SAS and SDS videos and SAS and SDS scores on the basis of simulating doctors diagnosing subjects, and only diagnose anxiety disorders based on the fusion model of SAS video and SAS scores.

We use $FP$ for false positives, $FN$ for false negatives, $TP$ for true positives, and $TN$ for true negatives. We evaluate the model using accuracy, specificity, sensitivity, and precision. The formula for accuracy is $Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$. The formula for sensitivity is $Sensitivity = \frac{TP}{TP + FN}$. The formula for specificity is $Specificity = \frac{TN}{TN + FP}$. The formula for precision is $Precision = \frac{TP}{TP + FP}$.

4.1. Implementation details

Our training set, validation set, and test set have a ratio of 7:2:1. In order to make the model have better results, we randomly rotate the image by $-14$–$15$ degrees, randomly adjust the image brightness, randomly adjust the color saturation, and random shearing. The momentum optimizer parameter is set to 0.9 with the learning rate of 0.01 and training time for about 13 h.

4.2. Baseline and results

We use the doctor’s final diagnosis as the label. The accuracy of SDS in diagnosing depression is 0.8, while the accuracy of SAS in diagnosing anxiety is 0.7895. We use three supervised machine learning methods
for classification based only on the scale scores. As shown in Table 3, the best classification accuracy (3-class classification of normal, depression and anxiety) of these three methods is 83.78%. However, this method only uses the scale scores and does not use the video features of the subjects, which is not in line with the doctor’s diagnostic process.

To mimic a doctor’s diagnostic process, we fuse video and scale scores to diagnose depression and anxiety. As shown in Table 3, the accuracy rate of the fusion model is the highest at 0.946. Furthermore, as shown in Table 4, we also use only the SDS scores and videos to diagnose depression and anxiety.

Table 3
Results of 3-class classification.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy (3-class)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD (SDS and SAS scores)</td>
<td>0.838 ± 0.005</td>
</tr>
<tr>
<td>Naive Bayes (SDS and SAS scores)</td>
<td>0.757 ± 0.005</td>
</tr>
<tr>
<td>Adaboost algorithm (SDS and SAS scores)</td>
<td>0.838 ± 0.005</td>
</tr>
<tr>
<td>SDS and SAS (Score + Video) (OURS)</td>
<td>0.946 ± 0.005</td>
</tr>
<tr>
<td>SDS and SAS (Score + Video) (RNN)</td>
<td>0.87 ± 0.005</td>
</tr>
</tbody>
</table>

Table 4
Score and video fusion.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDS score + video (OURS)</td>
<td>0.929 ± 0.005</td>
<td>0.885 ± 0.005</td>
<td>0.958 ± 0.005</td>
<td>0.906 ± 0.005</td>
</tr>
<tr>
<td>SAS score + video (OURS)</td>
<td>0.931 ± 0.005</td>
<td>0.96 ± 0.005</td>
<td>0.889 ± 0.005</td>
<td>0.968 ± 0.005</td>
</tr>
<tr>
<td>SDS score + video (RNN)</td>
<td>0.861 ± 0.005</td>
<td>0.808 ± 0.005</td>
<td>0.913 ± 0.005</td>
<td>0.848 ± 0.005</td>
</tr>
<tr>
<td>SAS score + video (RNN)</td>
<td>0.868 ± 0.005</td>
<td>0.88 ± 0.005</td>
<td>0.815 ± 0.005</td>
<td>0.903 ± 0.005</td>
</tr>
</tbody>
</table>

Fig. 3. The features extracted by the convolutional neural network are then processed by LSTM for video feature extraction.

Fig. 4. Different video and score fusion structure details.
depression. The model has an accuracy of 0.929. Using SAS scores and videos to diagnose anxiety, the model was 0.931 accurate. We also replace LSTM with Recurrent Neural Network (RNN) (Mikolov et al., 2010) for comparison, and the LSTM module works better. The fusion video of different scale scores is shown in Fig. 5, and the fusion results of different modules are shown in Fig. 6.

It can be seen that the multimodal model based on two scales, SDS and SAS, has higher accuracy than the multimodal model of a single scale, and our model realizes better performance than the simple RNN model with the same multimodal data.

5. Discussion

5.1. Clinical prospect

Anxiety and depression have seriously affected the quality of life of patients, and even caused behaviors such as suicide. During the COVID-19 period, these two diseases have become more and more obvious. For example, Choi et al. found that 19% of the respondents suffered from depression through interviews in Hong Kong people, and 14% of the respondents suffered from anxiety disorders (Choi and Hui, 2020), however, in order to cope with COVID-19, the government has issued a stay-at-home order, reduced public transportation, reduced social interaction and other measures, making it difficult for patients to find a doctor (Ciotti et al., 2020). Based on our method, preliminary screening and diagnosis of depression and anxiety can be performed at home, at the clinic or in other places, such as drug rehabilitation centers, because our system can be embedded in a variety of terminals. In Fig. 1 of the Appendix, we illustrate the clinical operation on a computer with a camera. We have now upgraded the video capture device, which can synchronously capture the video of subjects answering scales through the embedded camera of computers, mobile phones and other devices. Benefiting from our multimodal model, we can analyze the facial expression and actions, and combine them with scales to achieve a more convenient and effective screening and diagnosis method. Besides, preliminary screening by our method can free doctors from repetitive labor and provide richer information for clinical diagnosis, thereby greatly reducing the consumption of medical resources and facilitating the large-scale screening of depression and anxiety.

5.2. Limitations and future directions

In order to imitate doctors to diagnose symptoms based on subjects’ expressions and scale scores, we studied how to integrate SAS and SDS videos and scores to diagnose depression and anxiety disorders. However, this study has certain limitations. Firstly, this study was based on a self-rating scale and did not incorporate its rating scale information. Secondly, information such as voice is missing. Thirdly, besides SAS and SDS, there are other scales widely used in the clinical diagnosis of depression and anxiety, HAMD (Hamilton depression scale) and HAMA (Hamilton anxiety scale). SDS and SAS are self-rating scales, while HAMD and HAMA require professional clinicians to score, so the applicable scenes of HAMD and HAMA are more limited, and the results of these two scales are more dependent on the raters. By comparison, SDS and SAS have wider application scenarios and convenience, which are more conducive to large-scale initial screening of depression and anxiety disorder. Benefiting from the good expansibility of our method, we can further collect the data from different scales and modalities to mine more diverse and deep information, and thereby building a more stable and accurate system with powerful assistance in the clinical diagnosis of depression and anxiety. Finally, our data is only from China, and it is best to populate the sample with subjects from different countries.

The scales applicable to the diagnosis of depression and anxiety may be different in various regions with different populations (De Man et al., 2021). In China, SDS and SAS are the common diagnosing scales, while in the U.S., GAD-7, PHQ-2 and PHQ-9 are the most common screening tools for generalized anxiety and depression disorder. Collecting data from other scales and applying our model to different types of scales from different regions is our promising future direction.

Since its rating scale can also assist doctors in diagnosis, in order to improve the accuracy of the model, the self-rating scale and its rating scale video and score and speech can be integrated into the model, or a better data set can be collected.

6. Conclusion

This paper first uses three supervised machine learning to classify SAS and SDS scores to diagnose depression and anxiety, with the highest classification accuracy of 83.78%. However, this method ignores the facial expressions in the subjects’ videos. In the diagnosis process of people’s expressions and scale scores, we extract features through a convolution neural network, maintain the time series features of video through long-short memory network, and then fuse the diagnostic scores of the scale. The accuracy of our model in diagnosing depression and anxiety is 94.67%, which can not only help doctors diagnose diseases, but also assist in large-scale screening.

CRediT authorship contribution statement

Wanqing Xie: Conceptualization, Methodology, Software, Funding acquisition. Chen Wang: Software, Writing – original draft, Visualization. Zhixiong Lin: Investigation, Data curation. Xudong Luo:
Investigation, Resources. Wenqian chen: Investigation, Formal anal-
ysis. Manzhu Xu: Investigation, Resources. Lizhong Liang: Investiga-
tion, Data curation. Xiaofeng Liu: Methodology, Software, Writing – review & editing. Yanzhong Wang: Methodology. Hui Luo: Conceptu-
alization, Data curation, Project administration. Mingmei Cheng: Methodology, Writing – original draft, Writing – review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data availability

We will establish our open-source dataset, and the data will be released.

Acknowledgements

This work was partially supported by HEU focused supporting direc-
tion of AI [XK22400210], Jiangsu Natural Science Foundation Youth Programme [BK20200238], National Key Research and Development Program under Grant [2021YFC3300500] and Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang) [JZW-2019-007].

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.compmedimag.2022.102128.

References


