Assistant Diagnosis of Insanity Based on Infrared Thermal Image Analysis and Deep Learning Algorithm

Yeping Wang¹, Xi Wang¹, Peng Xiao¹, Cheng Chen¹, Ruoxiu Xiao¹*, Yanshen Wang², Yuanyuan Lu³ and Zhiliang Wang¹

¹School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China
²Beijing Wholelife Medical Science Ltd, Beijing, China
³Department of Ultrasound, Chinese PLA General Hospital, Beijing, China

*Corresponding author email: xiaoruoxiu@ustb.edu.cn

Abstract. Considering the lack of in-depth research on the cases of insanity and other mental diseases in the context of traditional Chinese medicine, a deep learning based algorithm which can realize auxiliary judgment function of insanity are proposed in this paper. First, the original image set is screened and then an improved U-net network is used to realize the division of the trunk and limbs of the human body, thus preventing interference from disease-independent areas to affect subsequent disease judgment. Finally, based on the classification of insanity, the reference function of visual analysis is added. 1508 IR images are divided into two groups to test the proposed method. And experimental results show that accuracy of the classification of insanity diseases can reach 0.92, which has a high reference value for the clinical diagnosis of insanity.

Keywords: Insanity; Infrared thermography; Image segmentation; Deep learning.

1. Introduction

In traditional Chinese medicine (TCM), mental disorders belong to the category of mental illness. Serious mental disorders have adverse effects on the healthy development of society, because of the severity, harmfulness, and heavy burden of such diseases [1]. To date, there are insufficient data and in-depth research regarding the cases of mental diseases in TCM. Furthermore, standardized and very objective standards for their diagnosis and treatment are also lacking, thereby affecting the effective development of TCM intervention with potential advantages. Recently, deep learning algorithms are widely used in speech recognition, image recognition and natural language processing, and they are gradually playing an increasingly important role in medical fields [2]. Such algorithms have great potential in auxiliary medicine, medical imaging, drug mining, and health management applications. The invention and improvement of medical infrared thermal imager ensure the accuracy and availability of data through scientific means to a certain extent. Therefore, it is possible to diagnosis of patient by analysis of IR thermal image.

For example, Chun et al. [3] conducted a bibliometric research on the medical applications of IR thermal imaging technology, which are mainly distributed in the fields of TCM, rehabilitation, and pain and are mostly focused on the application of Chinese medical herbs. Shen et al. [4] explored the characteristics of IR thermography of patients with insanity. Through the statistical analysis of the relative temperature difference in the temperature measurement area, they found that the Shangjiao, Governor-pulse, and Xuli in patients with insanity are lower than those in normal people, while those of Youxie are higher.
than those in normal people. Thus, their statistical results lay a theoretical foundation for the diagnosis of insanity by IR thermal imaging technology. In the processing of IR thermal images of the human body, Li et al. [5] proposed a method of automatically extracting and dividing the human body region. When determining the cross-part feature points, the inguinal bright band must be determined; however, such a bright band cannot be easily found in some IR thermal images. Hence, its extraction effect is not ideal. Chen et al. [6] carried out two kinds of preprocessing through cross-part fixed angle segmentation. However, traditional image processing algorithms are vulnerable to pose interference. Furthermore, in practical application, whole body image is usually collected in the image acquisition stage, and no stitching or other processes are needed. Thus, it is not flexible and practical.

Based on the above analysis, the paper uses the deep learning algorithm to study insanity in the category of psychosis in TCM, using a large sample size of IR data collected by medical IR thermal imager. The goal is to achieve the functions of screening, segmentation, classification, and visualization of IR images of insanity, thus providing auxiliary judgment and scientific basis for the clinical diagnosis of insanity and improving the accuracy of diagnosis.

2. Methods

We obtained IR data of insane and normal users through the original database and restored them into images. First, a simple Convolution Neural Network (CNN) network is used to classify the images according to the shooting posture, after which the pictures are screened to remove the non-standard, extremely blurred, and highly repetitive images. Next, according to the characteristics indicating that insanity is mainly related to the viscera and head, U-net network is used to segment it and extract the parts of human trunk and head, thus avoiding the interference of limbs in the subsequent judgment of a disease. After the completion of the segmentation phase, the training and tuning of the classification model with or without insanity are conducted, on the basis of which the visualization function of the activation region is further realized. The processing framework for judging insanity in human IR thermal images according to the procedure is described as figure 1.

![Figure 1. Overall flow chart.](image)

2.1. Data Pre-processing

With the help of the collection location number, the data information of the users in the insanity group and the normal users in the control group can be obtained. According to the regulations, there were 9 male postures and 11 female postures (including thymus examination). The original data contained the information uploaded by each collection point, so the image quality cannot be guaranteed. Moreover, the resolution of the IR image collected by the system is 320 pixels × 240 pixels, which is not that high. Therefore, this system intends to use a CNN classification model to classify the data set according to different positions. And then remove the shooting of non-standard, extremely blurred, high repeatability and other problems to facilitate subsequent processing.

2.2. Part Segmentation

Experts believe that the manifestation of insanity is mainly concentrated in the head and viscera areas and has less to do with other parts. Therefore, before making the classification judgment of insanity, the image should first be segmented and the limbs and other areas must be removed to avoid interference in the subsequent disease judgment.

In recent years, deep learning network has been extensively used in medical image segmentation; however, it is difficult to label image segmentation data sets. According to the traditional way of image
segmentation and labelling, each region to be distinguished by the marking tool must be circled separately. In view of the limited manpower and time, we have adopted a fast approach in the form of key point annotation. Here, we select several key points along the edge of the image, which are then connected to divide the human body into several large regions. Then, through the traditional threshold segmentation method, the region is refined and segmented, and finally, the corresponding ground truth image of an IR human body image is formed. In the choice of segmentation model, U-net [7] is utilized. U-net is a codec-type network, which takes the image as the input and outputs the probability score of pixel classification with Cross Entropy (CE) as the loss function. Murugesan et al. [8] proposed a method of multi-task joint training based on U-net network, which can achieve better results in segmentation, boundary and shape measurement. Inspired by this article, and also in order not to waste labelling resources, we add the prediction of key points to the network for training and use the joint loss function to train the network. On one hand, we want to study whether the training method of multi-loss in this kind of application scenario can improve the effect of network training. On the other hand, we also want to use it as an interface for subsequent function development in order to realize human acupoint prediction.

The loss function consists of two components, namely, the pixel classification CE loss of the mask and the Mean Square Error (MSE) loss of the regression of the predicted key points. The total loss is given as follows:

$$L_{loss} = L_{mask} + L_{keypoint}$$

The specific losses are formulated as below:

$$L_{mask} = -\sum_{x \in \Omega} \log(p(x)g(x))$$  \hspace{1cm} (2)

$$L_{keypoint} = \sum_{x \in \Omega} (\hat{D}(x) - D(x))^2$$  \hspace{1cm} (3)

In the above, $$L_{mask}$$ represents pixel classification error, $$x$$ is the pixel position in image space $$\Omega$$, $$p(x)g(x)$$ represents the prediction probability of the real label $$g(x)$$ after the softmax activation function, $$L_{keypoint}$$ represents the MSE of a pixel, $$\hat{D}(x)$$ is the predicted value, and $$D(x)$$ is the true value.

Through double-layer tagging, we send the data set into the U-net network for training, as shown in figure 2. The encoder used ResNet18, because it did not require the use of a complex network, thus saving on computing resources. Given that the number of data sets is small, and the image structure is
relatively unified, the data enhancement method is used. Moreover, only the random flipping is used, and no rotation, clipping, or other methods are added to increase the convergence speed.

2.3. Insanity Judgment and Visual Analysis

2.3.1. Insanity judgment. After conducting part segmentation, we obtain images of the main torso and head of the IR thermal map of the human body, which are divided into two groups according to whether they belong to insanity or not. Then, we send the images to the CNN for training. The small number of data sets is expanded by data enhancement methods, such as flipping, rotating, scaling, cropping, translation, and so on. In training, we use the ResNet50, Xception, MobileNetV2, and NASNetMobile models to compare the classification effects of different models in human IR thermal images.

2.3.2. Visualization Analysis. When a neural network is displayed visually, its interpretability can be vastly improved. Therefore, we study the visualization technology of deep neural network to verify whether the features learned by the network are pathological features. The visual display of the focus area provides a reference direction for doctors.

At present, the three methods of neural network visualization are the intermediate output of visual convolution neural network, the visual convolution neural network filter, and the class activation mapping (CAM) in visual image [9]. Among them, the thermal map helps people understand which pixels are mainly used by the neural network to judge the objects in the picture as the corresponding categories, thereby better highlighting the active area.

The original CAM method has achieved good interpretation results, but it requires the modification of the structure of the original model, resulting in the need to retrain the model; thus, we use the improved Grad-CAM method [10]. The basic idea of Grad-CAM is the same as that of CAM, which is to obtain the corresponding weight of each feature graph and finally find a weighted sum. However, the main difference between the Grad-CAM and CAM methods is the process of finding the weight $\omega$. Specifically, CAM retrains the weight by replacing the fully connected layer with the global average pooling layer, whereas Grad-CAM uses the global average of the gradient to calculate the weight.

3. Experiments and Results

3.1. Data Pre-processing Works

In original database, there were 39533 images of users in the normal group and 6261 images of the users in the insanity group. First, they were divided into seven categories according to posture: front body, back body, upper body front, upper body back, main torso, face, and other postures (Figure 3). After the preliminary classification, many repeated and blurred pictures were found in each category, which must be screened manually. Finally, we selected the clear images in the whole front body position for follow-up processing. There were 661 and 847 images in the disease group and the control group, respectively.

![Figure 3. Preliminary classification type.](image1)

![Figure 4. Marking key points and generating effects.](image2)
3.2. Training of Segmentation

We selected 22 key points of human body contours. These key points are connected, and the human body is divided into 5 regions. After processing, the resulting mask is illustrated in figure 4. The key points occupy a channel, and the segmentation occupies a channel. We then labeled 460 images and sent them to the multi-loss based U-net segmentation network for training.

The comparison of the model training curves between the multi-task training of the segmentation and key point locations and the single-task training of image segmentation can be seen in figure 5. According to the observation curve, the convergence speed of the network training using multi-loss is slower than that of the single segmentation task network. In the test set, the mIoU values of the multi-task and single-task training are both 0.90. As the former does not achieve better results, this indicates that the regression training of highly scattered key points does not improve the segmentation effect. Thus, it is only used as the basis for subsequent human acupoint prediction.

The segmentation result of the model is displayed in figure 6. Compared with the traditional segmentation method, deep learning based segmentation is more flexible and stable.

![Figure 5. Model training curve.](image)

![Figure 6. Image segmentation results.](image)

3.3. Judgment and Visualization

Table 1 shows the training accuracy of using four classification models to judge the image classification of patients with insanity, and good results have been achieved. The first three are models that contain different network structures or pursue different goals, while the last NASNet model is a non-artificially designed model. Among them, the accuracy of the Xception model on the test set is 0.92, which proves that the judgment result of deep learning has a strong reference. From the perspective of practical application, MobileNetV2 as a lightweight network, has the smallest model and is easy to deploy while ensuring accuracy.

| Model          | Val_acc | Test_acc | Model size(MB) |
|----------------|---------|----------|----------------|
| Resnet50       | 0.91    | 0.90     | 271            |
| Xception       | 0.97    | 0.92     | 240            |
| MobileNetV2    | 0.92    | 0.91     | 27             |
| NASNetMobile   | 0.91    | 0.89     | 52             |

On the basis of the classification and judgment of insanity diseases, the output of the trained model is modified, and the classification results are explained by Grad-CAM. The effect is presented in figure 7, which clearly and intuitively shows the reference areas for classification and judgment, and provides ideas for the accurate clinical diagnosis and focus analysis of doctors.
4. Conclusion

The whole processing flow of assessing the IR thermal image data for insanity is proposed in this paper, which is mainly designed by deep learning. Further, a simple and effective data annotation method is used to create the data set, complete the division of the IR image of the human body, and further realize the judgment of insanity and regional visual reference. On the whole, this work has realized the effective application of deep learning in the field of psychosis in TCM. In the future, we may further work on improving the accuracy of segmentation and classification, such as continuing to label data sets, adding attention mechanism to improve segmentation accuracy, and so on.

Acknowledgments

This work was supported by the National Key Research and Development Program of China [2017YFB1002804], The Major Program of National Social Science Foundation of China [17ZDA331], and the Fundamental Research Funds for the Central Universities [FRF-TP-19-015A2].

References

[1] Cuijpers P. Targets and outcomes of psychotherapies for mental disorders: an overview. World Psychiatry, 2019, 18(3):276-285.
[2] Fourcade A, Khonsari R H. Deep learning in medical image analysis: A third eye for doctors. Journal of Stomatology, Oral and Maxillofacial Surgery, 2019, 120(4):279-288.
[3] Xiao Chun, Liyan Pan, Yan Lin. Bibliometrics Research on Medical Application of Infrared Thermal Imaging Technology Based on Wanfang and CNKI Database. Chinese Medical Equipment, 2019, 34(07):165-169.
[4] Yujuan Shen, Hongmei Xu, Chunsheng Wang. Application of Infrared Thermal Imaging in Patients with Insanity. Infrared, 2019, 40(10):36-41.
[5] Mingrui Li, Chi Zhang, Ning Ding, Zhe Li, Deyu Li. Automatic segmentation on infrared images of human body based on feature points. Beijing Biomedical Engineering, 2019, 36(06):576-583.
[6] Chentao Chen, Zhiwei Pan, Huiliang Shen, Yunfang Zhu. Image stitching and partition algorithms for infrared thermal human-body images. Opto-Electronic Engineering, 2019, 46(09):67-75.
[7] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
[8] Murugesan Balamurali, Sarveswaran Kaushik, Shankaranarayana Sharath M, et al. Psi-Net: Shape and boundary aware joint multi-task deep network for medical image segmentation. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019: 7223-7226.
[9] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba. Learning Deep Features for Discriminative Localization. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 2921-2929.
[10] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 618-626.