國立台北大學資訊工程學系專題報告

Using Deep Learning in Ultrasound Image of the Long Head of Biceps Tendon to Grade the Severity of Inflammation

專題組員：杜奕萱、紀雯齡、林于晴、施雅馨
專題編號:PRJ-NTPUCSIE-107-002
執行時間:2018 年 08 月 至 2019 年 06 月

1. Abstract

Inflammation of long head of biceps tendon is a common reason of shoulder pain. Bicipital peritendinous effusion (BPE) is the most common biceps tendon abnormality and can be related to various shoulder injuries. The physicians usually use ultrasound Image to grade the severity of inflammation on long head of biceps tendon. However, the ultrasound Image which is clear and in correct position is not easy to obtain by the physicians without experiences. To reduce physicians' workload and miscarriage of justice, an automated bicipital peritendinous effusion recognition system on ultrasound Image was proposed to classify inflammation as normal plus mild, moderate, and severe. The proposed system firstly takes the ultrasound image as an input to determine whether the ultrasound picture is the biceps or not, and finally output the grade of bicipital peritendinous effusion prediction. In this study, two important factors were considered to solve computer-aided detection problems. First is to evaluate the influence of dataset scale and spatial image context on performance, which using faster regions with convolutional neural network (Faster R-CNN) to extract the region of interest (ROI) area. The other one is to explore and evaluate different convolutional neural network (CNN) architectures. The model performance was discussed according to various network configurations, parameters, and different amounts of training samples. The approach proposed in this study was applied to the three-class classification of bicipital peritendinous effusion images and achieved 75% accuracy. The results are competitive compared to the results of other state-of-the-art methods.

2. Introduction

Shoulder pain is a common musculoskeletal disorder, with prevalence of 6.7%-66.7% in the general population of their lifetime [1], and the long head of the biceps tendon (LHBT) is vulnerable to impingement clinically [2]. Bicipital peritendinous effusion (BPE) is the most common biceps tendon abnormality and can be related to various shoulder injuries [3]. High resolution ultrasound has been widely applied in diagnosing shoulder problems, and ultrasound is the first choice for the examination of lesions of tendons around the shoulder joint [4]. For painful BPE that is present, suspicious, or accidentally discovered by ultrasound, the interpretation of ultrasound results mainly depends on the knowledge, skill, and experience of the specialists, with profound subjectivity. In addition, increased number of patients has caused a reduction in the average diagnostic time spent on each case, which maybe affects the diagnostic quality and outcome [5].

Recently, the development of deep learning technology enables the possibility of image recognition to detect target areas in a medical image and classify the detected target features. The process for detection and classification is similar to the diagnostic flow charts of ultrasound findings by physicians and provides a brand-new idea for the solution of the previously mentioned issues.

Due to the recent revival of convolutional neural network (CNN), image recognition has made significant progress. CNN has been proven powerful in image object detection and classification tasks [6]. CNN has been applied extensively in several medical imaging applications [7]. In 2016, Anthimopoulos *et al.* proposed a study of lung texture to detect interstitial lung disease, and interstitial lung disease patterns can be detected and classified precisely by a deep CNN [8]. In 2016, Pereira *et al.* used CNN in magnetic resonance imaging to detect and classify brain tumor [9]. The use of small kernels allows designing a deeper architecture, as well having a positive effect against over-fitting, given the fewer number of weights in the CNN. By 2018, Wang *et al.* proposed a system for automatically recognizing skeletal maturity, which takes a one-handed X-ray as an input and ultimately outputs a prediction of bone age [10]. Hence, deep learning techniques especially CNN have been applied in many medical imaging analysis projects with promising results [7]. Nevertheless, the diagnostic application on ultrasound image of long head of biceps tendon has not been reported yet.

Referring to the latest deep learning technologies and according to requests of physicians of department of physical medicine and rehabilitation, a completely automated artificial intelligence- grading system on ultrasound image of the long head of biceps tendon was proposed in this study. This classifying system uses CNN to directly discriminate the level of BPE, assisting physicians in medical judgment. This system will reduce the probability of the physician not accurately distinguishing the severity of BPE. The research can also apply to medical teaching so as to enhance the skill and experience of medical students and resident doctors while using ultrasound imageon the biceps tendon. In order to improve the accuracy of the ultrasound images in the discrimination, the images were preprocessed by normalizing the images to improve the reference data needed for training the deep learning model. After enhancing the performance of the proposed system using CNN and preprocessing method, it is more convenient and immediate to understand the patient's problems of the biceps when the doctor makes a diagnosis. Then, this project will help clinical doctors only focus on the skill of operating ultrasound image on the lesions without any anisotropy or misdiagnosis.

3. Methods

The architecture of the proposed system is presented in Fig. 1. The system comprises four parts: data preparation, the proposed CNN, region of interest (ROI) detection, and data preprocessing. Two CNNs were adopted in this study, namely CNN-1 and CNN-2. In data preparation, 3801 ultrasound images were prepared for the proposed system. After the ultrasound images were prepared, those images were evaluated by CNN-1 to determine if the input images contain biceps. The ROI of the ultrasound image containing bicep was bounded by faster regions with convolutional neural network (Faster R-CNN). After the ROI was bounded, those images from the bounded area would be normalized to ensure the consistency of all images. After normalization, the CNN-2 was adopted to discriminate the severity of BPE.

A. Data Preparation

The ultrasound images were collected from the database in Chang Gung Memorial Hospital (CGMH). The present cross-sectional designed study analyzed the clinical sonographic data derived from patients referred for shoulder ultrasound examinations in CGMH. The research was approved by the institutional review board of CGMH (IRB No. 201800408B0) and the requirement for informed consent was not applied because this project was conducted by reviewing an existing data bank of routine ultrasound examinations in CGMH. The dataset includes 3801 images, containing 625 images from healthy subjects and 3176 images from subjects with various severities of BPE.

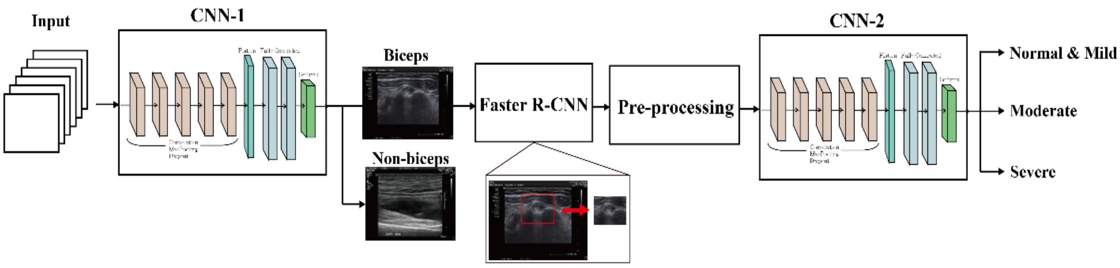
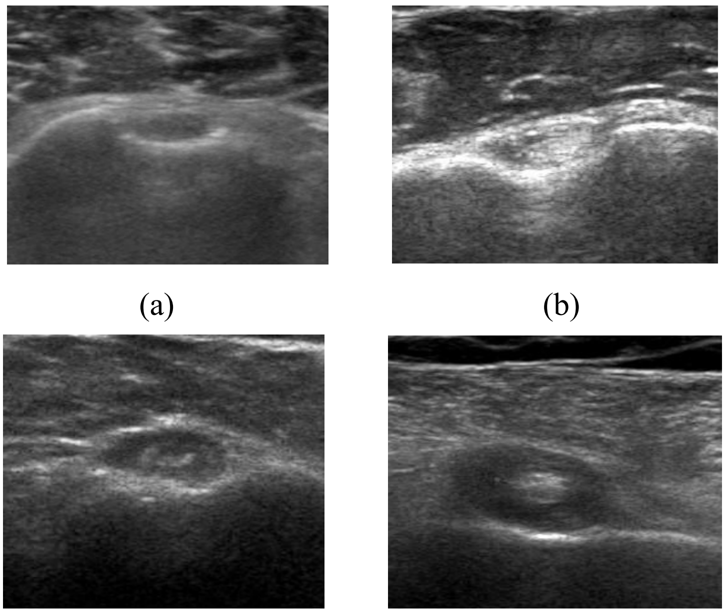


Fig. 1. Architecture of the proposed system.

Those ultrasound images were graded into four severities according to the thickness of the bicep's effusion by three physicians to enhance reliability of the inter-rating [3]. Fig. 2 illustrates the four severities of BPE. The four severities are normal, which the thickness of the effusion is less than 1 mm, and then mild (1–2 mm), moderate (2–3 mm) or severe (>3 mm)., which the thickness of the effusion is more than 3 mm.

The dataset was randomly separated into training and validation sets. The training set is used to train the deep learning model to discriminate the severity of BPE, and the validation set is used to choose the deep learning model with the best performance for classification. Table I shows the amount of images used in each subset. The ratio of training-validation is 7:2. Furthermore, to evaluate the performance of predicting the severity of BPE, 100 ultrasound images with biceps were adopted as the testing set to evaluate the final performance of classification.



(a) (b) (c) (d)

Fig. 2. The four severities of BPE. (a) normal, (b) mild, (c) moderate, and (d) severe.

TABLE I

DATASET OF DOFFERENT CLASSES

	Normal	**Mild**	**Moderate**	**Severe**
Training	467	1061	812	539
Validation	133	303	232	154
Testing	25	25	25	25
All	625	1389	1069	718

B. Proposed CNNs

In this study, two CNNs, including CNN-1 and CNN-2, were adopted for different purposes. CNN is a popular deep learning technique to automatically classify the medical images. Generally, the CNN architecture includes an input layer, several convolutional and pooling layers, one or more fully connected layers, and one output layer. The features of input images can be extracted and reduced through the convolutional layer and the pooling layer. The fully connected layer is used to select the significant features to map the final outputs by converting the two-dimensional feature map into one-dimensional output. The CNN architecture proposed in this study is shown in Fig. 3. The architectures of CNN-1 and CNN-2 are the same. For the parameter setting, the optimizer was set as Adam [11], and activation function was set as ReLu.

The aim of using CNN-1 is to identify whether the ultrasound image contains biceps. The purpose of identifying the biceps is to remove the images which may not contain the biceps and may take poor ultrasonographic pictures due to anisotrophy, in order to enhance the efficiency of classification. The aim of using CNN-2 is to discriminate the severity of BPE. However, in clinical situation, it is difficult for physicians to distinguish the differences between normal and the border of the biceps tendon with the mild effusion thickness visually [12]. Moreover, the patients whose severity of BPE is between mild and normal may leave hospital and will not be under treatment. Therefore, the subjects in normal and mild are defined as the same class in this study.

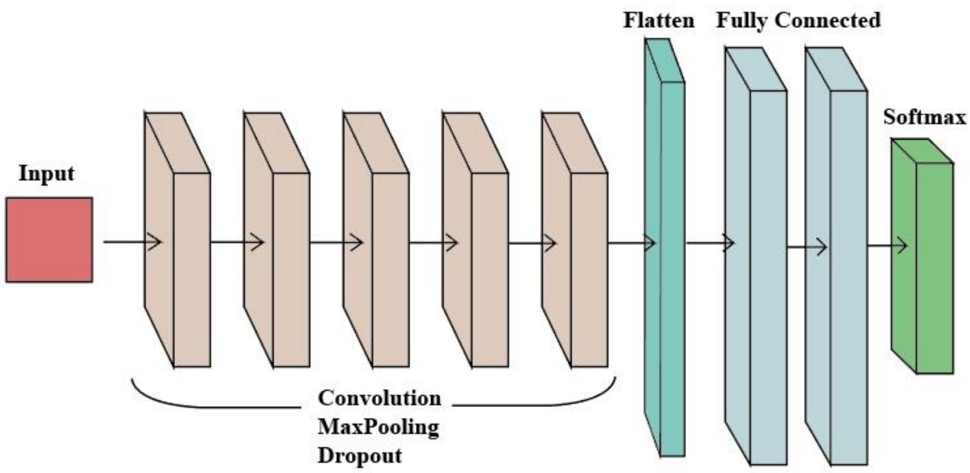
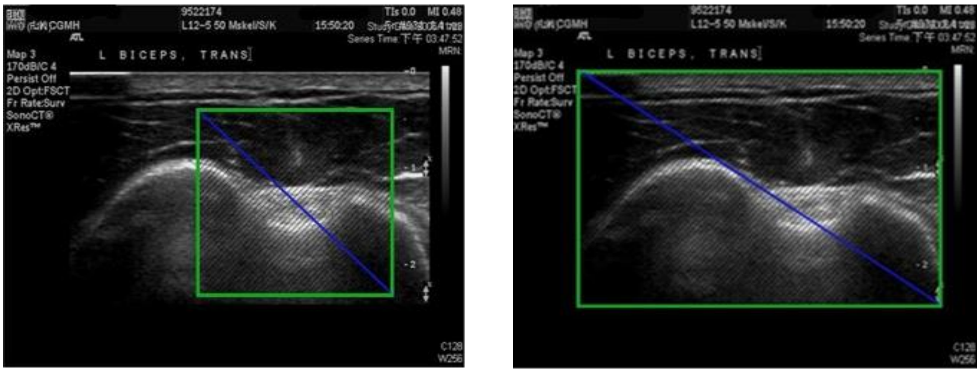


Fig. 3. CNN architecture for CNN-1 and CNN-2

C. ROI Detection

The original ultrasound images contain redundant information around the bicep area as shown in Fig. 4. The redundant information would reduce the accuracy of classification. To eliminate the redundant information and reserve the region only containing the biceps, the ROI of BPE in each image was selected for discriminating the severity of BPE in the classification phase. In this study, Faster R-CNN was adopted to automatically bound the ROI because it is the most common and effective method in the related research.

Before using Faster R-CNN, the original ultrasound images were annotated by using the bounding box to select the ROI area from the ultrasound images manually. The coordinate parameters were recorded and converted into the format of PASCAL VOC 2007. In this study, two different sizes of ROI were chosen for training the Faster-R-CNN model. One is the ROI in small size (ROI-S), which is the most specific area where physicians diagnose the severity of BPE. The green rectangle in Fig. 4(a) indicates the area of ROI-S. The other one is the ROI in big size (ROI-B), which is the area obtained by only removing the redundant information around the bicep area. The green rectangle in Fig. 4(b) indicates the area of ROI-B. The purpose of using the two different sizes of ROI is to evaluate which size of ROI can be used to obtain a higher accuracy while discriminating the severity of BPE.

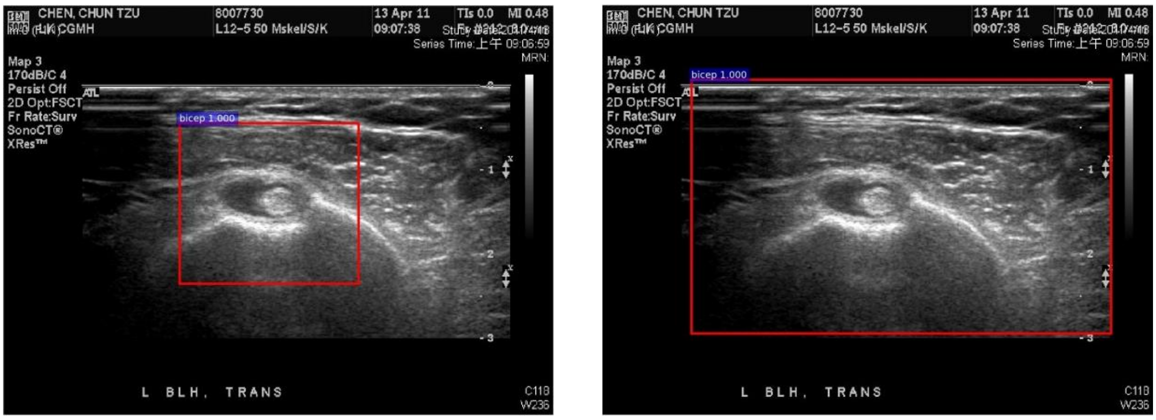


(a) (b)

Fig. 4. The annotation using (a) ROI-S, and (b) ROI-B.

The parameters for training Faster R-CNN were set as follows. The initial learning rate is 0.01. The weight decay is 0.0005. 40,000 iterations were performed during training. VGG16 network was used for feature extraction [13]. 1900 images were used for training, and the other 1901 images were used for validation. After the Faster R-CNN model was trained, the model could be used with the weights obtained from the training process. After receiving a test image, Faster R-CNN model provides the scores of all the candidate regions in the test image. The region which obtains the highest score, meaning the greatest probability, will be selected and marked with a red rectangle, as shown in Fig. 5. In this study, 1000 testing images were taken from 3,801 images and were found that the ROI with correct bounding were above the threshold of 0.9. Therefore, if the score of the region is lower than 0.9, the image will be removed from the dataset. Selecting the ROI from one image only takes Faster R-CNN 0.025 s, showing that using Faster R-CNN can reduce the time of bounding ROI comparing to manually bounding. After obtaining the bounding area, the pixels in the rectangle will be stored for further classification.

The ultrasonic images may have inconsistencies, such as different deflection depths and brightness, while scanning by the ultrasonic machine. The inconsistencies would affect the performance of classification. Therefore, image normalization is required to address the inconsistencies. Various research has proposed the method for normalizing medical images [14-16]. In this study, the method proposed by Mathur *et al.* [17] was adopted to normalize the image.



(a) (b)

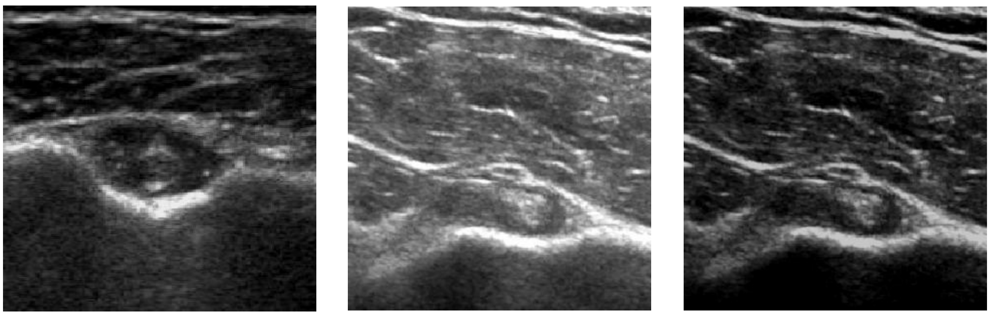
Fig. 5. Selected ROI in the test image. (a) ROI-S which score is 1.000; (b) ROI-B which score is 0.999.

The image which obtained the highest contrast ratio is selected as the target image. The target image was obtained as the reference for normalizing the other images, which was taken out 100 clear images from the 3,801 images by the physician and took the median image from these 100 images. Fig. 6 presents the result after applying the method on a biceps ultrasound image. The target image is shown in Fig. 6(a). Fig. 6(b) is the original image which is necessary to be normalized. Fig. 6(c) shows the image after normalization, which is clearer than the original image presented in Fig. 6(b).

4. Results

A. Experimental Configuration

In order to validate the performance of the proposed method in different settings, four experiments including model performance with different sizes of ROI, model performance with data augmentation, model performance with different CNN architectures, and model performance with different classes were conducted. Faster R-CNN framework implemented in Python was used to detect the ROI. MATLAB (R2018a, MathWorks, Inc., Natick, MA, USA) was used to preprocess the ultrasound images. CNNs were implemented by using TensorFlow (Version 1.8.0, Google Brain, Mountain View, CA, USA) and Keras (Version 2.2.4) in Python programming language [18]. All of the experiments were conducted by using a personal computer (Intel Core i7 3.40GHz, 24 GB DRAM) equipped with a graphics card with multi-core graphics processing units (GeForce RTX 2080 gaming OC 8G, GIGA-BYTE Technology Co., Ltd., New Taipei City, Taiwan).



(a) (b) (c)

Fig. 6. The result after image normalization: (a) The target image, (b) the original

B. Model Performance with Different Sizes of ROI

To evaluate the performance of using different sizes of ROI while predicting the severity of BPE, three different datasets including original image dataset, ROI-S dataset, and ROI-B dataset, were used for training CNN-2, and all data had been normalized. The performance of using the three different sizes of ROI are shown in Table II. The accuracy of original dataset, ROI-B, and ROI-S is 0.69, 0.72, and 0.74, respectively. The result shows that using ROI-S can obtain the highest accuracy while predicting the severity of BPE.

C. Model Performance with Different Classes

According to previous research [12], the BFE is usually categorized into four severities, including normal, mild, moderate, and severe. However, it is difficult to classify normal and mild clinically, since the different between normal and mild are indistinct. Moreover, many BPEs between normal and mild may be classified to normal or mild depending on the physician's subjectively judgement. So, the normal and mild cases may be considered as one class and further investigating the different combinations with normal-mild, mild-moderate, moderate-severe to prove further whether discriminating three severities of BPE (normal plus mild, moderate, and severe) can improve accuracy. Therefore, three binary classifiers, including normal-mild, mild-moderate, and moderate-severe were trained and tested. Normal-mild classifier is a classifier trained with the normal and mild datasets. The normal-mild classifier can discriminate severity of the ultrasound images between normal and mild. Mild-moderate classifier is a classifier trained with the mild and moderate datasets. The mild-moderate classifier can discriminate severity of the ultrasound images between mild and moderate. Moderate-severe classifier is a classifier trained with the moderate and severe datasets. The moderate-severe classifier can discriminate severity of the ultrasound images between moderate and severe.

TABLE II

ACCURACY OF DIFFERENT SIZES OF ROI

Size of ROI	Accuracy
Original image	0.69
ROI-B	0.72
ROI-S	0.74

Table III shows the accuracy of binary classification with the three classifiers. The accuracy of normal-mild classifier, mild-moderate classifier, and moderate-severe classifiers is 0.74, 0.80, and 0.77, respectively. The result shows that mild-moderate classifier obtained the highest accuracy, whereas normal-mild classifier obtained the lowest accuracy. The result indicates that the difference between normal and mild is small and would reduce the accuracy of prediction. Therefore, normal and mild datasets should be merged into one dataset according to the result in Table III.

Finally, the proposed system validated the performance of recognizing the BPE for three severities, which comprise normal plus mild, moderate, and severe. The system also validated the performances while discriminating the severities of BPE in other two situations, including two severities, namely normal and sick; and four severities, namely normal, mild, moderate, and severe. Accuracy, precision, recall, and F-score were used to evaluate the performance of each situation and architecture. In this experiment, all the data used to compare different classes' performance were normalized data of size of ROI-S and without data augmentation. The result of the experiment is presented in Table IV. The accuracy of CNN to discriminate the BFE into two severities, three severities, and four severities is 0.90, 0.75, and 0.61, respectively. The result indicates that the proposed method can accurately and highly discriminate two severities. Besides, discriminating three severities obtains higher accuracy than discriminating four severities.

TABLE III

ACCURACY OF DIFFERENT BINARY CLASSIFIERS

Binary Classifier	Accuracy
Normal-mild	0.74
Mild-moderate	0.80
Moderate-severe	0.77

TABLE IV

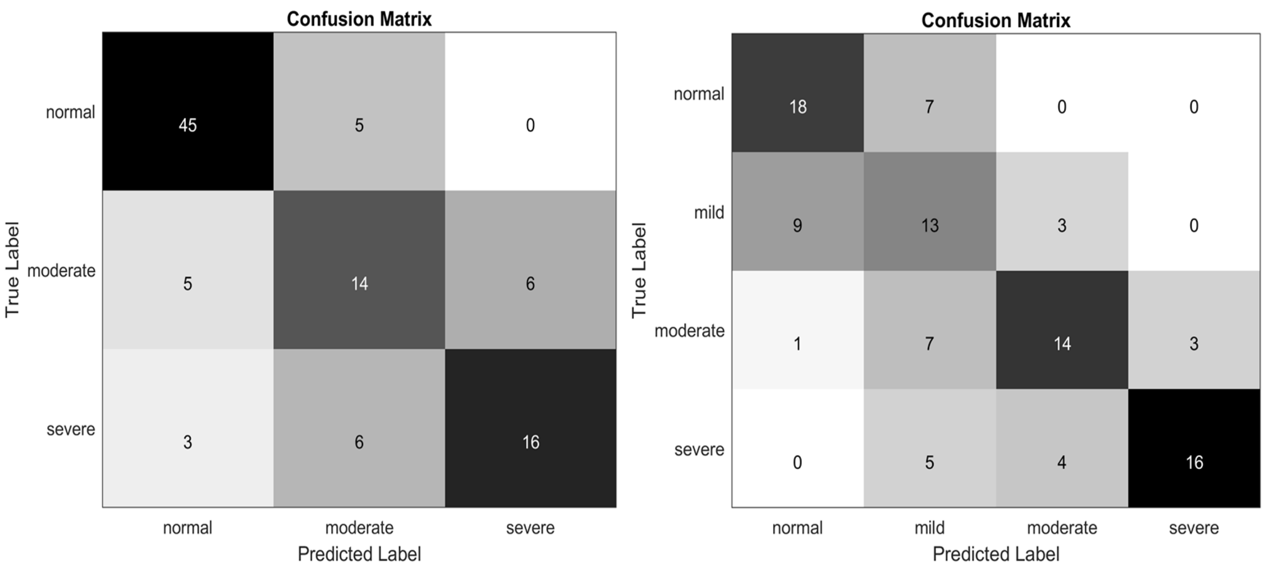
THE FULL RESULTS OF DIFFERENT CNN ARCHITECTURE AND CLASS AS CNN-2

Severities	Accuracy	Precision	Recall	F-score
2	0.900	0.898	0.900	0.897
3	0.750	0.746	0.750	0.747
4	0.609	0.640	0.610	0.618

In order to test the accuracy of the proposed system in the real clinical setting, 100 unused ultrasound images, consisting of 25 images for each severity, were adopted to validate the final accuracy of the proposed system. The confusion matrices of different classifying situation are presented in Fig. 7. For discriminating normal, moderate and severe, normal dataset contains 50 images, including 25 images from normal subjects and 25 images from mild subjects; moderate dataset contains 25 images from moderate subjects, and severe dataset contains 25 images from severe subjects. Fig. 7(a) shows that 5 normal subjects are misclassified as moderate subjects, 11 moderate subjects are misclassified as normal or severe subjects, and 9 severe subjects are misclassified as normal or moderate subjects. For discriminating normal, mild, moderate and severe, each dataset contains 25 images from the corresponding severity. Fig. 7(b) shows that 7 normal subjects are misclassified as mild subjects, 12 mild subjects are misclassified as normal or moderate subjects, 11 moderate subjects are misclassified as the other three severities, and 9 severe subjects are misclassified as mild or moderate subjects.

5. Discussion

An automated system for grading the severity of ultrasound image of biceps tendon was proposed in this study. Four experiments were conducted to evaluate the performance of the systems and select which setting of the deep learning can obtain the highest accuracy.



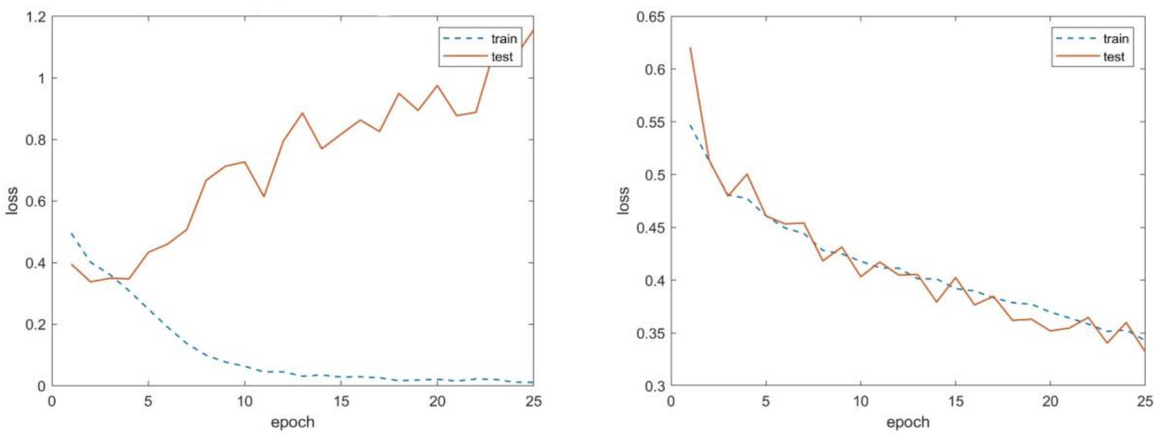
(a) (b)

Fig. 7. Confusion matrices of classification. (a) three severities classification, and (b) four severities classification

In the first experiment, original image, ROI-S, and ROI-B were evaluated to select which can obtain the highest accuracy. TABLE II shows that ROI-S obtained the highest accuracy while discriminating the severity of BPE. The reason is that ROI-S is the most specific area where physicians usually use to diagnose the severity of BPE. ROI-B obtained lower accuracy than ROI-S while discriminating the severity of BPE because ROI-B contained the other components around the biceps tendon, such as bone or subcutaneous tissue [12], which may affect the discrimination of severity of BPE. In all cases, using the original image shows the lowest accuracy because the redundant information around the bicep area will affect the performance of the system. Moreover, the redundant information is different among ultrasonic machines so the area of the redundant information should be removed to improve the accuracy of the automated system. In summary, the ROI-S is the best area for discriminating the severity of BPE.

Based on clinical cognition, the severity between normal and mild is indistinct and hard to classify by physicians. In the experiment of this study, the results shown in Tables III and IV also prove this cognition. Table III shows the accuracy of normal-mild is the lowest. The result in Table III also fits in with the result in Table IV. The three-severity classification which combined normal and mild as one class has better performance than four-severity classification.

The aim of this study is to obtain the CNN architecture with the highest accuracy while predicting the severity of BPE. However, parameters such as dropout function [19] of the CNN architectures will affect the performance of prediction. Fig. 8(a) shows that the over-fitting problem still exists in the proposed system. To solve the over-fitting problem, the dropout function was added after each convolutional layer. Fig. 8(b) shows the result after adding dropout functions solved the over-fitting problem and started to converge.



(a) (b)

Fig. 8. Loss of model (a) without dropout (b) with dropout

6. Conclusion

In this study, an automated system for grading the severity of BPE was proposed. Four experiments were conducted in this study to validate the performance of the grading system in different settings and situations. The result shows that grading the BPE in three severities, namely normal, moderate, and severe, can obtain the highest accuracy, which reaches 0.75. The contributions of the study are listed as follows. First, this is the first automated system for predicting the severity of BPE. Second, the system can automatically identify if an ultrasound image contains the bicep tendon, and the accuracy of identification is 0.94. Third, the grading system can detect the ROI automatically. The result also indicated that using ROI-S to detect the ROI can obtain the best performance. Fourth, the experiments were conducted to prove that the normal and mild subjects are difficult to be discriminated. Therefore, it is appropriate to grade the BPE into three severities. Last but not the least, several CNN architectures were validated to choose the optional CNN and parameters, and the accuracy can reach 0.75, which may be acceptable for helping clinical physicians to make the diagnosis and judge severity of BPE for the further treatment plan. This system not only can help to reduce the probability of misdiagnosis but also the physician's burden while judging the severity of BPE visually. In addition, this study can be served as an educating system for training the residents. The system combined to the ultrasound image can provide a referring and simple procedure to help physicians to diagnose the severity of BPE. In future works, more ultrasound images of bicep tendon are required to improve the overall accuracy and develop more robust automated systems for other tendonitis diseases.

7. Appreciation

Thanks to the advising professor of the instructor for a year, leading us when we have any problem, let us have a deeper understanding of the implementation process of the whole topic. Thank to Dr. Jean-Lon Chen from CGMH for giving us advice on medical treatment. We are also grateful to the senior sister in our lab, I-Jung Lee, for her help in giving us many suggestion during this study. Finally, we would like to thank all the people who have given us advice and help, so that we can successfully complete this study.

8. Reference

[1] J. J. Luime, B. W. Koes, I. J. Hendriksen, A. P. Verhagen, H. S. Miedema, and J. A. Verhaar, "Prevalence and incidence of shoulder pain in the general population; a systematic review," *Scand. J. Rheumatol.*, pp.73-81, 2004.

[2] M. W. Rodosky, C. D. Harner, and F. H. Fu, "The role of the long head of

the biceps muscle and superior glenoid labrum in anterior stability of the shoulder," Am. J. Sports Med., pp. 121-130, 1994.

[3] K.V. Chang, W.T. Wu, and L. Özçakar, "Association of bicipital peritendinous effusion with subacromial impingement: a dynamic ultrasonographic study of 337," *Sci. Rep.*, 2016.

[4] T. O. Smith, T. Back, A. P. Toms, and C. B. Hing, "Diagnostic accuracy of ultrasound for rotator cuff tears in adults: a systematic review and meta-analysis," Clin. Radiol. Imaging J., no. 66, pp. 1036-1048, 2011.

[5] C. S. Park, S. H. Kim, S. L. Jung, B. J. Kang, J. Y. Kim, and J. J. Choi, "Observer variability in the sonographic evaluation of thyroid nodules," *J Clin Ultrasound*, no. 38, pp. 287-93, 2010.

[6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998..

[7] H. Greenspan, B. V. Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153-1159, 2016.

[8] M. Anthimopoulos, S. Christodoulidis, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network", *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1207-1216, 2016.

[9] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240-1251, 2016.

[10] S. Q. Wang, Y. Y. Shen, C. H. Shi, P. Yin, Z. H. Wang, W. H. P. Cheung, J. P. Y. Cheung, K. D. K. Luk, and Y. Hu, "Skeletal Maturity Recognition Using a Fully Automated System with Convolutional Neural Networks," *IEEE Access*, vol. 6, pp. 29979-29993, 2018.

[11] D. P. Kingma and J. Ba. (2014), "Adam: a method for stochastic optimization," [Online]. Available: https://arxiv.org/abs/1412.6980.

[12] K. V. Chang, "Associations of sonographic abnormalities of the shoulder with various grades of biceps peritendinous effusion (Bpe)," *Ultrasound Med. Biol.*, vol. 40, no. 2, pp. 313-321, 2014.

[13] Muneeb ul Hassan (2018, Nov 20). Networks [Online]. Available: https://neurohive.io/en/popular-networks/vgg16/

[14] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserved

color normalization for histological images", *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, pp. 1012-1015, 2015.

[15] M. Macenko, M. Niethammer, J. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis", *Proc. IEEE Int. Symp. Biomed. Imag.: Nano Macro*, pp. 1107-1110, Boston, MA, USA, 2011.

[16] Y. Li, J. Wu, and Q. Wu, "Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning," *IEEE Access*, vol. 7, no. 2, pp. 21400-21408, 2019.

[17] P. Mathur, M. Ayyar, R. R. Shah , and S. Sharma, "Exploring classification of histological disease biomarkers from renal biopsy images," *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 81-90, 2019.

[18] Keras, 2018. (2.2.4 ed.) [Online]. Available: https://keras.io/.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *JMLR, Inc. and Microtome Publishing*, pp. 1929-1958, 2014.