Binary classification of non-specific low back pain condition based on the combination of B-mode ultrasound and shear wave elastography in multiple sites

Xiaocheng Yu  
Shenzhen University

Mou-Hsun Chiang  
The University of Hong Kong-Shenzhen Hospital

Guowen Zhu  
Shenzhen University

Faying Xu  
The University of Hong Kong-Shenzhen Hospital

Zhenhua Liu  
The University of Hong Kong-Shenzhen Hospital

Lin Su  
The University of Hong Kong-Shenzhen Hospital

Haiping Zheng  
The University of Hong Kong-Shenzhen Hospital

Chen Zhou  
The University of Hong Kong-Shenzhen Hospital

Qiuming Chen  
The University of Hong Kong-Shenzhen Hospital

Fen Gao  
The University of Hong Kong-Shenzhen Hospital

Mengting Lin  
The University of Hong Kong-Shenzhen Hospital

Xiaohua Xu  
The University of Hong Kong-Shenzhen Hospital

Yongjin Zhou (✉ yjzhou@szu.edu.cn)  
Shenzhen University

Research Article

Keywords: non-specific low back pain, ultrasound, shear wave elastography, machine learning, automatic classification
Abstract

Low back pain (LBP) is a problem that raises medical, social and economic concerns. Accurate and timely assessment and diagnosis of LBP, especially non-specific LBP (NSLBP), can help clinicians develop effective interventions and treatments for LBP patients. In this study, we integrated the B-mode ultrasound image feature from multiple sites with the shear wave elastography (SWE) feature of NSLBP patients, and then employed a support vector machines (SVM) model to classify NSLBP patients with the Visual Analogue Scale (VAS) as the ground truth. A total of 52 subjects were recruited from the University of Hong Kong-Shenzhen Hospital, and an optimal feature set with the size of 48 was obtained after feature extraction and feature selection, among which the SWE elasticity feature had most contribution in the classification task. The five-fold cross-validation accuracy, precision and sensitivity were 0.85, 0.89 and 0.86, respectively, higher than the previously reported values of MRI. The experimental results preliminarily demonstrate that the proposed methods can be applied to the automatic classification of NSLBP and find out the important site and position of the muscle in the NSLBP classification task.

1. Introduction

Low back pain (LBP) is defined as the pain between below the gluteal folds and above the vertebral protrusion that lasts for at least 1 day with or without leg pain. LBP represents one of the most common musculoskeletal disorders [1, 2], with a global prevalence of 7.8% in 2017 [3]. Meanwhile, the number of patients increases with the population growth and aging [4, 5]. LBP is also an important cause of physical disability [6], with an increasing financial burden [1, 4, 5, 7]. LBP can be acute, with a duration of less than 6 weeks, or chronic, which lasts more than 3 months [8]. Non-specific low back pain (NSLBP) refers to LBP without specific physiological or pathological causes, and its prevalence is 23%, which accounts for 90% of all LBP patients [9]. Currently, the diagnosis of LBP mainly focuses on NSLBP. Accurate and timely assessment and diagnosis of NSLBP can help the clinicians to develop effective interventions and treatments for the affected patients [10, 11].

The existing LBP prediction and assessment methods mainly focus on the subjective report of patients, biological signal and imaging studies. Kate et al. used a questionnaire to classify the severity of LBP in the enrolled patients [12]. Aoki et al. and Ogon et al. used a Visual Analogue Scale (VAS) to quantitatively assess NSLBP patients [13, 14]. However, these methods are limited by their subjective nature. Masoud et al. used kinematic data obtained from motion sensors combined with a STarT Back Screening Tool (SBST) questionnaire output to classify NSLBP patients into high risk and low-medium risk groups [15]. Jiang et al. used surface electromyography (sEMG) combined with the support vector machine (SVM) algorithm to classify and identify NSLBP patients who responded to functional recovery rehabilitation, which helps healthcare workers improve the efficiency of NSLBP rehabilitation [16]. However, the performance of these methods could be affected by the sources of their signals (sEMG or kinematic data), which are susceptible to interference from other human signals; this reduces the accuracy and stability of LBP classification.

Computed tomography (CT) was used in the work of Kamaz et al. to scan the paravertebral muscles of the L4-L5 level vertebrae in chronic LBP patients and healthy individuals [17]. Their results showed that chronic LBP resulted in different degrees of atrophy of the muscles (most prominent in the multifidus muscle (MF)). Furthermore, Juuso et al. used T2-weighted magnetic resonance imaging (MRI) of the intervertebral disc for
texture feature extraction. Then, NSLBP patients were divided into symptomatic and asymptomatic groups [18]. However, CT and MRI are relatively inflexible for the physicians to screen LBP and expensive for the patients [17–19].

Compared with CT and MRI, ultrasound is a non-invasive, low-cost and easy-to-use imaging tool. In recent years, it has been used to examine the skeletal muscle diseases [9]. Combined with the artificial intelligence (AI) technology, ultrasound has also been applied to the automatic classification of muscle states [20, 21].

A number of studies have performed ultrasound on the muscles of the relevant parts of LBP patients. Their results have shown that the MF is important in maintaining the stability of the spine [22], and most LBP patients have asymmetric MF [23] and larger fat area [24]. The thoracolumbar fascia (TLF) shear strain was reported to be lower in human subjects with chronic LBP [25]. LBP patients had a significantly smaller increase in transversus abdominis (TrA) thickness with isometric leg tasks compared with controls [26]. The mobility of the erector spinae (ES) in LBP patients was also reported to be decreased in endurance tasks [27]. In summary, LBP has complex causes, which involve multiple muscles and multiple sites. Hence, the diagnosis and prediction of LBP also need to combine more comprehensive information from multiple sites to achieve a more stable and robust prediction and diagnosis of LBP.

In addition to the B-mode ultrasound, shear wave elastography (SWE) has also been increasingly applied in LBP studies. Chan found that the elasticity of the MF is different between LBP patients and normal people in upright and forward-stooping positions [24]. Masaki et al. reported that the elasticity of the MF in LBP patients is significantly higher than that in normal people [28].

In summary, considering the complex etiology of NSLBP, it is particularly important to classify the severity or pain intensity of NSLBP and formulate corresponding treatment plans according to different classifications of NSLBP [29, 30]. However, an accurate manual quantification of NSLBP is difficult and time-consuming. Besides, it also relies on the physician's subjective judgment. An automatic classification of NSLBP could help the physicians conduct prompt interventions and formulate treatment plans for the patients.

Based on the previous report of the multiple images feature selection (MIFS) [21] framework, this study integrated the B-mode ultrasound image feature from multiple sites with the SWE elasticity feature of NSLBP patients, and then employed SVM to classify NSLBP patients using the VAS as ground truth. Furthermore, we explored the importance ranking of related muscles in the diagnosis of NSLBP.

2. Results

In this section, the results of feature selection using ANOVA and SVM to classify NSLBP are presented. At the same time, this section also exhibits the number of specific features in the optimal feature set after feature selection, the top 10 important features of the SVM model, the proportion of 4 muscles (MF, TLF, ES, TrA) in the optimal feature set, the proportion of different positions and sites in the optimal feature set, as well as the classification results between MIFS framework and single image feature selection (SIFS) framework.

2.1 Feature Analysis
In order to find out the best performance in different parameters, the \textit{percentile} parameter of \texttt{SelectPercentile} was set from 1 to 100 for feature selection. When the \textit{percentile} was set to 6, the model achieved the best performance, as shown in Fig. 1.

After feature selection, 48 features were finally obtained in the optimal feature set, which accounted for 6\% of the total feature set. Table 1 shows the number of different features in the optimal feature set.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|}
\hline
Feature type & Muscle morphological feature & Image frequency analysis feature & FOS feature & Haralick feature & Galloway feature & Local binary patterns feature & SWE elasticity feature & Total No. \\
\hline
Select No. & 0 & 2 & 3 & 27 & 12 & 0 & 4 & 48 \\
\hline
Total No. & 2 & 28 & 98 & 336 & 280 & 28 & 28 & 800 \\
\hline
\end{tabular}
\caption{Specific number of features in the optimal feature set}
\end{table}

Abbreviations: FOS, first-order statistical; SWE, shear wave elastography.

\subsection*{2.2 The classification performance}

Table 2 shows the classification results of different feature sets after performing five-fold cross-validation using SVM. It can be observed that the accuracy of using the SWE elasticity feature and B-mode ultrasound image feature to classify NSLBP was 0.65 and 0.81, respectively. When both were used to classify NSLBP patients, the accuracy reached 0.85.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Feature set & SWE elasticity & B-mode ultrasound image feature & Total \\
\hline
Accuracy & 0.65 & 0.81 & 0.85 \\
Sensitivity & 0.76 & 0.82 & 0.89 \\
Specificity & 0.54 & 0.80 & 0.80 \\
AUC & 0.66 & 0.80 & 0.88 \\
Precision & 0.69 & 0.86 & 0.86 \\
NPV & 0.70 & 0.82 & 0.89 \\
\hline
\end{tabular}
\caption{Classification results}
\end{table}

Abbreviations: AUC, area under curve; NPV, negative predictive value; SWE, shear wave elastography.

In the SVM model, the absolute value of the feature weight can indicate the importance of the feature in the classification task. In this experiment, the top 10 important features obtained in the SVM model are shown in Table 3. We conducted the Student’s t-test and Mann-Whitney test in the top 10 important features came from
the normal distribution or not, respectively. As shown in Table 3 (indicated by asterisks), all the 10 features showed statistically significant differences.

**Table 3** Top 10 important features in the optimal feature set.

<table>
<thead>
<tr>
<th>Feature Description</th>
<th>Position/Direction</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>L4-L5_L_MF_FirstOrderFeature.IOD</td>
<td>Prone position</td>
<td>*</td>
</tr>
<tr>
<td>L2-L3_R_MF_SWE.Std</td>
<td>Prone position</td>
<td>*</td>
</tr>
<tr>
<td>L2-L3_R_TLF_GallowayFeature.RLNU</td>
<td>Prone position, direction = 0º</td>
<td>*</td>
</tr>
<tr>
<td>L2-L3_R_TLF_HaralickFeature.Homogeneity</td>
<td>Prone position, direction = 90º</td>
<td>**</td>
</tr>
<tr>
<td>L2-L3_L_MF_GallowayFeature.RLNU</td>
<td>Tabletop position, direction = 90º</td>
<td>**</td>
</tr>
<tr>
<td>L4-L5_R_MF_SWE.Std</td>
<td>Tabletop position</td>
<td>**</td>
</tr>
<tr>
<td>L2-L3_L_ES_SWE.Std</td>
<td>Prone position</td>
<td>*</td>
</tr>
<tr>
<td>L2-L3_R_TLF_MFAF</td>
<td>Prone position, calculated by multi-window method</td>
<td>*</td>
</tr>
<tr>
<td>L2-L3_L_MF_GallowayFeature.GLNU</td>
<td>Tabletop position, direction = 0º</td>
<td>**</td>
</tr>
<tr>
<td>L2-L3_L_MF_GallowayFeature.RP</td>
<td>Tabletop position, direction = 0º</td>
<td>*</td>
</tr>
</tbody>
</table>

Abbreviations: L2-L3, L2-L3 lumbar spine muscle; L4-L5, L4-L5 lumbar spine muscle; L, left side of the lumbar spine muscle; R, right side of the lumbar spine muscle; SWE, shear wave elastography; Std, standard deviation; GLNU, Gray-Level Non-Uniformity; MF, multifidus muscle; ES, erector spinae; TLF, thoracolumbar fascia; TrA, transversus abdominis; *, p-value < 0.05; **, p-value < 0.01.

**Table 4** The number of features from different muscles in the optimal feature set.

<table>
<thead>
<tr>
<th>Muscle</th>
<th>MF</th>
<th>ES</th>
<th>TLF</th>
<th>TrA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select No.</td>
<td>20</td>
<td>2</td>
<td>26</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>Total No.</td>
<td>456</td>
<td>114</td>
<td>115</td>
<td>115</td>
<td>800</td>
</tr>
</tbody>
</table>

Abbreviations: MF, multifidus muscle; ES, erector spinae; TLF, thoracolumbar fascia; TrA, transversus abdominis.

**Table 5** The number of features at different positions and sites of the optimal feature set.

<table>
<thead>
<tr>
<th>Positions and sites</th>
<th>L2-L3</th>
<th>L4-L5</th>
<th>L</th>
<th>R</th>
<th>Prone position</th>
<th>Tabletop position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select No.</td>
<td>46</td>
<td>2</td>
<td>22</td>
<td>26</td>
<td>34</td>
<td>14</td>
</tr>
<tr>
<td>Total No.</td>
<td>572</td>
<td>228</td>
<td>399</td>
<td>399</td>
<td>572</td>
<td>228</td>
</tr>
</tbody>
</table>

Abbreviations: L2-L3, L2-L3 lumbar spine muscle; L4-L5, L4-L5 lumbar spine muscle; L, left side of the lumbar spine muscle; R, right side of the lumbar spine muscle.
Table 6 MIFS framework and SIFS of NSLBP classification performance.
<table>
<thead>
<tr>
<th>Muscle sites</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
<th>Precision</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2-L3_L_MF</td>
<td>0.67</td>
<td>0.93</td>
<td>0.38</td>
<td>0.61</td>
<td>0.64</td>
<td>0.87</td>
</tr>
<tr>
<td>L2-L3_R_MF</td>
<td>0.63</td>
<td>0.93</td>
<td>0.30</td>
<td>0.61</td>
<td>0.62</td>
<td>0.65</td>
</tr>
<tr>
<td>L4-L5_L_MF</td>
<td>0.68</td>
<td>0.65</td>
<td>0.71</td>
<td>0.66</td>
<td>0.71</td>
<td>0.66</td>
</tr>
<tr>
<td>L4-L5_R_MF</td>
<td>0.69</td>
<td>0.80</td>
<td>0.59</td>
<td>0.72</td>
<td>0.73</td>
<td>0.78</td>
</tr>
<tr>
<td>L2-L3_L_MF_TBT</td>
<td>0.77</td>
<td>0.75</td>
<td>0.79</td>
<td>0.79</td>
<td>0.81</td>
<td>0.76</td>
</tr>
<tr>
<td>L2-L3_R_MF_TBT</td>
<td>0.75</td>
<td>0.83</td>
<td>0.65</td>
<td>0.75</td>
<td>0.75</td>
<td>0.77</td>
</tr>
<tr>
<td>L4-L5_L_MF_TBT</td>
<td>0.67</td>
<td>0.67</td>
<td>0.66</td>
<td>0.69</td>
<td>0.76</td>
<td>0.67</td>
</tr>
<tr>
<td>L4-L5_R_MF_TBT</td>
<td>0.77</td>
<td>0.89</td>
<td>0.62</td>
<td>0.74</td>
<td>0.76</td>
<td>0.88</td>
</tr>
<tr>
<td>L2-L3_L_ES</td>
<td>0.66</td>
<td>0.72</td>
<td>0.58</td>
<td>0.61</td>
<td>0.71</td>
<td>0.62</td>
</tr>
<tr>
<td>L2-L3_R_ES</td>
<td>0.65</td>
<td>0.85</td>
<td>0.41</td>
<td>0.61</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>L2-L3_L_TLF</td>
<td>0.73</td>
<td>0.75</td>
<td>0.70</td>
<td>0.78</td>
<td>0.77</td>
<td>0.71</td>
</tr>
<tr>
<td>L2-L3_R_TLF</td>
<td>0.75</td>
<td>0.89</td>
<td>0.58</td>
<td>0.77</td>
<td>0.71</td>
<td>0.83</td>
</tr>
<tr>
<td>L2-L3_L_TrA</td>
<td>0.62</td>
<td>0.67</td>
<td>0.54</td>
<td>0.61</td>
<td>0.69</td>
<td>0.59</td>
</tr>
<tr>
<td>L2-L3_R_TrA</td>
<td>0.71</td>
<td>0.68</td>
<td>0.75</td>
<td>0.68</td>
<td>0.78</td>
<td>0.67</td>
</tr>
<tr>
<td>Total</td>
<td>0.85</td>
<td>0.89</td>
<td>0.80</td>
<td>0.88</td>
<td>0.86</td>
<td>0.89</td>
</tr>
</tbody>
</table>
3. Discussion

1) Performance comparison between the proposed method and the MRI work

This experiment combined the B-mode ultrasound image feature with the SWE elasticity feature to classify NSLBP patients. Compared with the MRI work in NSLBP classification [18], the accuracy and sensitivity after five-fold cross-validation of this experiment were 0.85 and 0.89, respectively, which were higher than MRI work (0.83 and 0.82, respectively), suggesting that the proposed method can better identify NSLBP patients with a moderate-severe pain. The precision was 0.86 in this experiment, which is much higher than the MRI work (0.56), indicating that the classification of the proposed method is more robust in classifying positives than the MRI work. NPV refers to the proportion of true negatives among all negative results, the NPV and specificity of this experiment were 0.89 and 0.80, respectively, lower than those of the MRI work (0.94 and 0.83, respectively), which means a relatively low proportion of classifying negatives and indicates that NSLBP patients with a mild pain are relatively more difficult to be identified in this experiment. The AUC of this experiment was 0.88, lower than MRI work (0.91), which suggests that the MRI work may be more robust in NSLBP classification.

Nevertheless, compared with MRI, ultrasound is easier to operate, and timesaving for the clinicians, as well as causes less financial burden for NSLBP patients. It should be noted that the NSLBP classification in this experiment is based on the pain intensity evaluated by VAS, which is a subjective variable. However, all the patients have experienced significant pain for more than 3 months, and we obtained a relatively low proportion of false positives in the classification results, which is desirable in medical research.

2) Involving SWE elasticity makes a difference

The classification results indicate that using the SWE elasticity feature or using B-mode ultrasound image feature for classification is not as good as combining both in the classification of NSLBP patients. Besides, the classification results also confirmed the importance of SWE in the clinical research [31].

3) A brief discussion from the perspective of features

In this study, an optimal feature set with the size of 48 was obtained after feature selection (as shown in Table 1). Among the total feature proportion in the optimal feature set, the proportion of the SWE elasticity feature was the highest, achieved 14.3% (4/28), which indicates that SWE features plays an important role in the NSLBP classification in this experiment. This also preliminarily confirms the previous experiments by Chan et al. and Masaki et al. [24, 28] that the SWE elasticity of LBP patients is different from that of normal people and the SWE elasticity of LBP patients could change.
By analyzing the top 10 important feature weights of the SVM classification model, the Galloway feature accounted for 40.0% of the entire classification task (as shown in Table 3), indicating that the Galloway feature calculated by the GLRLM could well show the difference in the B-mode ultrasound image features between NSLBP patients.

4) A brief discussion from the perspective of muscles

As shown in Table 4, among 48 selected features from 800 total features, the features of MF and TLF had the highest proportions, reaching 4.4% (20/456) and 22.6% (26/115), respectively, while ES and TrA had the lowest proportions, reaching 1.6% (2/114) and 0.0% (0/115), respectively. It can be seen that MF and TLF play an important role in the classification of NSLBP patients [17, 25], which indicates that more attention should be paid to the MF and TLF in NSLBP classification in future experiments.

5) A brief discussion from the perspective of positions and sites

As shown in Table 5, the feature of the L4-L5 lumbar spine muscle accounted for 0.9% (2/228) in 48 selected features from 800 total features, while the feature of L2-L3 lumbar spine muscle accounted for 8.0% (46/572), indicating that the L2-L3 lumbar spine muscle was more important than the L4-L5 lumbar spine muscle in the classification of NSLBP in this experiment. The feature proportion of the left side of the lumbar spine muscle was 5.5% (22/399), while the right side of the lumbar spine muscle was 6.5% (26/399), which indicates that both sides of the lumbar spine muscle contributed equally to the NSLBP classification in this experiment in some extent. As for the data acquisition position, the tabletop position had more contribution (accounted for 6.1% (14/228)) than the prone position (accounted for 5.9% (34/572)), which verifies the importance of the tabletop position in the imaging of related muscles [32].

In this experiment, the MIFS framework extracts features from all 14 images in multiple sites and searches for the best feature set. As shown in Table 6, compared with the SIFS framework extracts the feature from a single image in single site, the accuracy, specificity, AUC, precision and NPV of MIFS framework were higher than SIFS framework, which indicated that MIFS framework had a better performance in NSLBP classification.

In this experiment, the data acquisition repeatability test was performed on the SWE elasticity values. The SWE elasticity values of 10 people were measured by two clinical operators at different times, and significant differences were found in the SWE elasticity of the TLF between the two operators, which indicated possible differences between the operators in the SWE elasticity measurement of the TLF. Furthermore, the choice of the ROI of the B-mode ultrasound image in this experiment depended on the experience of the clinical operator, and there was no fixed size. In future experiments, with the use of automatic methods, such as deep learning-based methods, the subjective influence of the operators can be reduced. At the same time, with the increase of the dataset size in feature experiments, the classification model can be more robust.

4. Conclusion

In this study, an SVM model was used to classify NSLBP patients based on the MIFS framework combined with the B-mode ultrasound image feature and SWE elasticity feature. Previously, MIFS [21] was proven to
have a certain effect on the classification of the motion level. The etiology of LBP in this experiment is complicated, and related studies have also shown that the muscles of LBP patients are different from those of normal people in multiple sites of the muscle. Based on the above-mentioned points, this experiment combined multiple sites of the B-mode ultrasound image feature and SWE elasticity feature to conduct a preliminary classification of NSLBP patients and achieved a better performance than SIFS framework.

5. Materials And Methods

5.1 Participants

B-mode ultrasound images and SWE elasticity of 52 NSLBP patients were collected from August 2020 to April 2021 from the University of Hong Kong-Shenzhen Hospital. Before data collection, VAS was used to evaluate the pain intensity of the patients. Each patient was informed of the purpose and process of the experiments and an informed consent form was collected from all the subjects. Ethical approval was authorized by the University of Hong Kong-Shenzhen Hospital ([2020]178).

Table 7 Descriptive statistics from the data.

<table>
<thead>
<tr>
<th></th>
<th>Mild (VAS &lt;= 3)</th>
<th>Moderate-severe (VAS &gt; 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>Female</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>Weight (Kg)</td>
<td>65.52±11.52</td>
<td>64.38±10.82</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.69±0.10</td>
<td>1.69±0.07</td>
</tr>
<tr>
<td>BMI</td>
<td>22.91±2.72</td>
<td>22.54±3.09</td>
</tr>
<tr>
<td>Age</td>
<td>35.96±7.62</td>
<td>41.11±10.24</td>
</tr>
</tbody>
</table>

Notes: There were no significant differences among the groups in terms of the weight, height, age or body mass index (BMI).

Abbreviations: BMI, body mass index; std, standard deviation; Kg, kilogram; m, meter.

The patients who had LBP with a significantly intense pain during rest and/or daily activities (according to VAS ≥ 1) that lasted for more than 3 months were included in the study. The exclusion criteria included history of spinal or lower limb fractures, spinal surgery or spinal deformities. According to the pain intensity, the patients were divided into two groups: (1) NSLBP patients with a mild pain (VAS <= 3) (2) NSLBP patients with a moderate-severe pain (VAS > 3). Data statistics are shown in Table 7.

5.2 Experimental Design
5.2.1 Data acquisition

Mindray DC-80 (Mindray, China) with an 8.5 MHz linear array ultrasonic transducer was used for data acquisition, which can collect B-mode ultrasound images and the SWE elasticity values (mean and standard deviation) of the muscles. In addition, the quality of the obtained SWE elasticity values was checked using DC-80 self-checking module for the elasticity imaging quality, to make the SWE elasticity measurement more stable.

5.2.2 Data acquisition process

The data acquisition process included the following steps:

1) Taking the prone position, a thin pillow was put under the abdomen of the patient to make the low back flat. Then, the arms were placed flat on both sides (left side and right side) of the body, as shown in Fig. 2(a).

i. B-mode ultrasound images of the muscle (MF, ES, TrA and TLF) on both sides of the patient’s L2-L3 lumbar spine as well as the SWE elasticity values (mean and standard deviation) of the muscle region of interest (ROI) were acquired.

ii. B-mode ultrasound images of the MF on both sides of the patient’s L4-L5 lumbar spine as well as the SWE elasticity values (mean and standard deviation) of the muscle ROI were acquired.

2) Taking the tabletop position[32], the patient's low back was kept flat and relaxed, as shown in Fig. 2(b).

i. B-mode ultrasound images of the MF on both sides of the patient’s L2-L3 and L4-L5 lumbar spine and the SWE elasticity values (mean and standard deviation) of the muscle ROI were acquired.

Finally, we obtained the B-mode ultrasound images as well as the SWE elasticity values of a total of 14 images from the patient's 4 muscles (MF, ES, TLF and TrA), as shown in Table 8. The representative B-mode ultrasound images are shown in Fig. 3.

Table 8 Data collection sites and the patients’ positions.

<table>
<thead>
<tr>
<th>L2-L3</th>
<th>L4-L5</th>
<th>L&amp;R</th>
<th>Prone position</th>
<th>Tabletop position</th>
<th>SWE elasticity</th>
<th>Muscle thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ES</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>TLF</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>TrA</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Notes: The ✓ represents the imaging of the muscle in the corresponding position.

Abbreviations: L2-L3, L2-L3 lumbar spine muscle; L4-L5, L4-L5 lumbar spine muscle; L&R, left and right sides of the lumbar spine muscle; SWE, shear wave elastography; MF, multifidus muscle; ES, erector spinae; TLF,
5.3 LBP Framework

This experiment was based on the MIFS framework, as shown in Fig. 4, which was mainly composed of feature extraction and feature selection. First, 55 B-mode ultrasound image features were extracted from the ROI of muscles in multiple sites (details are shown below). Then, they were combined with the SWE elasticity feature from the ROI to form the total feature set. After feature standardization and selection, important features were selected from the total feature set to construct the optimal feature set. Finally, the optimal feature set was used to train the SVM model and classify NSLBP patients.

5.4 Data Processing

5.4.1 Feature Extraction

Based on the MIFS framework, this experiment extracted the features from the B-mode images as well as SWE elasticity values, including the following features:

1. Muscle morphological feature: composed of average thickness feature of the muscle (TLF and TrA). Thickness is usually defined as the distance between the midpoints of the upper and lower muscle and calculated by the mean of the left and right sides of the muscle (TLF and TrA) thickness. In order to ensure the accuracy and validity of the value, we double checked the thickness measurement.

2. Mean image frequency analysis features (MFAF) of the ROI: calculated separately as the maximum entropy method and multi-window method[33, 34].

3. Image texture features: composed of two features. One is the first-order statistical (FOS) feature derived from the gray-level histogram, including the integrated optical density, mean, standard deviation, variance, skewness, kurtosis and energy. The other is the high-order texture feature, including the Haralick feature calculated from the Gray-Level Co-occurrence Matrix (GLCM)[35], Galloway feature calculated from the Gray-Level Run-Length Matrix (GLRLM)[36] as well as local binary patterns feature (energy and entropy) [21].

4. SWE elasticity feature: composed of the mean and standard deviation of SWE elasticity from the muscle ROI.

Finally, we obtained 800 features (57 features per ROI of an image and 14 images per subject) and 2 average thickness features of the TLF and TrA of each subject. The details of the features are listed in Table 9.

Table 9 Details of the features.
<table>
<thead>
<tr>
<th>Feature type</th>
<th>Feature names</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle morphological feature</td>
<td>muscle (TLF and TrA) average thickness feature</td>
<td>N = 2</td>
</tr>
<tr>
<td>Image frequency analysis feature</td>
<td>MFAF</td>
<td>Calculated by the maximum entropy method and multi-window method. N = 2</td>
</tr>
<tr>
<td>FOS feature</td>
<td>Integrated optical density, mean, standard deviation, variance, skewness, kurtosis and energy</td>
<td>Derived from the gray-level histogram. N = 7</td>
</tr>
<tr>
<td>Haralick feature</td>
<td>Contrast, correlation, energy, entropy, homogeneity and symmetry</td>
<td>Calculated from the Gray-Level Co-occurrence Matrix (GLCM) with 4 directions: 0°, 45°, 90° and 135°. N = 24</td>
</tr>
<tr>
<td>Galloway feature</td>
<td>Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray-Level Non-Uniformity (GLNU), Run Length Non-Uniformity (RLNU) and Run Percentage (RP)</td>
<td>Calculated from the Gray-Level Run-Length Matrix (GLRLM)) with 4 directions: 0°, 45°, 90°, and 135°. N = 20</td>
</tr>
<tr>
<td>Local binary patterns feature</td>
<td>Energy and entropy</td>
<td>N = 2</td>
</tr>
<tr>
<td>SWE elasticity feature</td>
<td>Mean and std</td>
<td>N = 2</td>
</tr>
</tbody>
</table>

Abbreviations: N, the number of features; TLF, thoracolumbar fascia; TrA, transversus abdominis; FOS, first-order statistical; MFAF, mean image frequency analysis feature; SWE, shear wave elastography; std, standard deviation.

### 5.4.2 Feature Selection

1) The variance between the features of the total feature set obtained in this experiment was relatively large. To prevent the variance of some features from being much larger than the variance of other features, which could result in slow convergence or non-convergence of the model, this experiment standardized the total feature set to make the mean and variance of the feature equal to 0 and 1, respectively, as shown in Eq. 1.

\[
y = \frac{x - \mu}{\sigma} \quad (1)
\]
where $y$ is the feature after standardization, $x$ is the feature before standardization, $\mu$ is the mean of the feature set, and $\sigma$ is the standard deviation of the feature set.

2) In this experiment, the number of features obtained by each patient was 800. To prevent the model from overfitting or having difficulty converging, it was necessary to perform feature selection on the total feature set and to reduce the feature dimension. We used `SelectPercentile` of `sklearn.feature_selection` in the machine learning package `Scikit-learn (v. 0.22.1)` for feature selection. The `score_func` parameter contains several feature selection methods. In this experiment `score_func` chose the default `f_classif` as the feature selection method, and the kernel method of `f_classif` was the analysis of variance (ANOVA)[37]. Due to the ability to combine all the feature information, ANOVA does not only improve the efficiency, but also increases the reliability of the feature selection. After feature selection with ANOVA, the optimal feature set was finally extracted from the total feature set.

5.5 Classification

1) The `Scikit-learn (v. 0.22.1)` machine learning library was used in `Python (v. 3.7.6)` to build a machine learning-based pipeline to analyze the feature data.

2) SVM was used to classify NSLBP patients. SVM is a two-classification model, and its basic model is a linear classifier with the largest interval defined in the feature space. The basic idea behind SVM is to solve the separation hyperplane that can correctly divide the dataset and have the largest geometric interval. SVM has several complex kernel methods, which can make the data that are inseparable in the linear space separable in other dimensions. At the same time, the addition of regularization enhances the robustness of the SVM model and reduces the possibility of overfitting. The SVM model of this experiment used a linear kernel and L2 regularization.

3) When the amount of data is sufficient, datasets are usually divided into training set, validation set and test set, and the final performance results come from the test set. However, since the amount of data in this experiment is limited, the above-mentioned single test set could not reliably report the performance indicators of the classifier[21]. To overcome this disadvantage, five-fold cross-validation combined with grid search was used to optimize and evaluate the classification model in this experiment.

4) In the classification task of NSLBP of this experiment, NSLBP patients with a moderate-severe pain were treated as positive cases and those with a mild pain as negative ones. The metrics of accuracy, specificity, sensitivity, AUC (area under the receiver operator characteristic curve), precision and negative predictive value (NPV) were used to quantify the classification results. At the same time, the classification results of using the SWE elasticity feature, B-mode ultrasound image feature and SWE elasticity feature combined with B-mode ultrasound image feature were compared.
Declarations

- Ethical Approval and Consent to participate
Not applicable.

- Availability of data and materials
The data that support the findings of this study are available on request from the corresponding author, Yongjin Zhou. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

- Competing interests
None.

- Funding
None.

- Authors’ contributions
Y. Zhou and X. Xu designed the experiments, X. Yu and M. Hou did the experiments with other authors and wrote the manuscript. All authors have contribution on conclusion and introduction sections and approved the final manuscript.

References


Figures
Figure 1

The performance of feature selection with different values of the *percentile* parameter.
Figure 2

Data acquisition positions.
Figure 3

Representative images of 4 acquired muscles. a) B-mode ultrasound images of NSLBP patients with a mild pain. B) B-mode ultrasound images of NSLBP patients with a moderate-severe pain. MF, multifidus muscle; ES, erector spinae; TLF, thoracolumbar fascia; TrA, transversus abdominis.

Figure 4
The framework of this experiment. Mild, NSLBP patients with a mild pain (VAS \leq 3); Moderate-severe, NSLBP patients with a moderate-severe pain (VAS > 3); SVM, support vector machine.