Main disease classification of intermittent claudication via L1-regularized SVM

Main disease classification of intermittent claudication via L1-regularized SVM

Tetsuyou Watanabe[1], Takeshi Yoneyama[1], Yasumitsu Toribatake[2] and Hiroyuki Hayashi[2]

Abstract— **There are multiple diseases that cause intermittent claudication, including lumber spinal canal stenosis (LSS) and peripheral arterial disease (PAD). LSS is categorized on the basis of the diseased part: L4 and L5. The medical treatment for these groups is totally different and the differentiation is important. With this in mind, we examined walking-motion data for patients and derived several features for the differentiation in previous studies. However, these features were not specialized for classification, and there is no guarantee that the features are effective for real differentiation. The present study investigates the possibility of differentiation by gait analysis, via use of an L1-regularized support vector machine (SVM). An L1-regularized SVM can execute both classification and feature selections simultaneously. On the basis of this method, our paper presents the methodology for classifying the underlying disease of the intermittent claudication with an accuracy of 79.7%. In addition, new effective features for the differentiation are extracted.**

I. INTRODUCTION

There is a walking symptom called intermittent claudication [1], in which patients experience severe pain in the lower limbs while walking. However, they can walk again after taking a break. The main diseases that cause this pain are lumber spinal canal stenosis (LSS) and peripheral arterial disease (PAD). Toribatake et al. [1] pointed out the similarity between PAD and LSS groups. The medical treatment for both these groups is totally different. Then, it is very important to differentiate these diseases. LSS can be divided into L4, L5, and S1 radiculopathies based on the stenosis area. Among the 3 radiculopathies, S1 is rare and sometimes does not require any treatment. This paper focuses on L4 and L5. Herein, the differentiation of healthy controls (Normal), PAD, L4, and L5 patients is considered. Generally, to identify the main disease, several tests such as angiography, myelography, magnetic resonance imaging (MRI), and ankle brachial index (ABI) [2] are conducted. These tests are precise but are invasive and expensive. In addition, highly skilled professionals are required to conduct the tests. These drawbacks make it difficult to conduct these tests at small hospitals. Moreover, at-home differentiation might be useful. The best differentiation method would use the minimum number of simple instruments that could be easily used by even non-professionals.The affected area of legs are different from the diseases. Then it is possible that gait pattern is also kinematically different from the disease.

With this in mind, we developed another examination method for analyzing two-dimensional walking motion measured by commercially available cameras [5], and we derived several differentiation factors [6]. However, classification was not taken into account in the factor deviation, and the factors are not always suitable for classification. There are several studies on gait analysis of intermittent claudication by other groups [3], [4]. However, differentiation has not been considered in any study thus far.

This paper presents an approach for a classification methodology with L1-regularized SVM that extracts relevant features with the classification and the differentiation. L1-regularized SVM is a classifier for linear classification with a large number of instances/features, and can select the relevant instances/features simultaneously with the classification. On the basis of information obtained in previous studies [5], [6], we define candidates of the features (such as fundamental statistics), and make the features sparse with a Gaussian basis function. The sparseness of the input instances/features gives a highly accurate classification of the Normal, PAD, L4 and L5 groups of patients. The main difference of the extracted features with the differentiation features from the previous studies is that those are valid for not the entire group but a part of the group, and can deal with individual differences.

II. METHODOLOGY

A. Participants

The participants were 13 healthy individuals (Normal; 5 men and 8 women), 13 PAD patients (12 men and 1 woman), 11 LSS (L4) patients (7 men and 4 women), and 22 LSS (L5) patients (9 men and 13 women). Diagnoses of the participants were made by conducting MRA, ABI, MRI, contrast-enhanced CT, selective nerve root blocks, and radiological examinations, and considering any clinical features and surgical findings.

B. Motion capture

The following is the goal of the proposed system. People measure their walking by using a commercially available camera, although the obtained gait pattern is two-dimensional. Upon sending the data to the server, the server determines which disease most likely causes the intermittent claudication and which the individual should go to, vascular surgery or orthopedics. The system would be available at small hospitals and at home. The requirements for the walking measurement system are easy setup, easy measurement,

*This work was not supported by any organization

[1]T. Watanabe and T. Yoneyama are with the School of Mechanical Engineering, Kanazawa University, Kakuma-machi, Kanazawa, 920-1192, Japan (corresponding author to provide phone: +81-76-234-4682; fax: +81-76-234-4682; e-mail: twata@t.kanazawa-u.ac.jp)

[2]Y. Toribatake and H. Hayashi are with the Department of Orthopedic surgery, Kanazawa University

easy calibration, available for use in small spaces, and short measurement time. With these in mind, we set up the simple walking measurement system (Fig. 1). Participants attached LED makers and walked on the treadmill in semidarkness such that the LED lights can be easily seen. The positions of the markers are the acromion, anterior superior iliac, head of the fibula, lateral malleolus and the 5th metatarsal head of the participant (Fig. 2a). Note that the right side is facing forward, as shown in Fig. 2. The markers were attached to the impaired leg. To determine a safe treadmill speed for normal walking, the participants practiced walking on the treadmill before the actual experiment. During the experiment, if the participant felt pain, we stopped the measurement. For safety, doctors watched the participants to help in the event of an accident. As described later, we used only angles for the analysis, and did not do any calibrations (for example, for estimating leg length). For recording the gait pattern, we used a commercially available camera with a frame rate of 30 [frame/s].

C. Analysis

The angles used for analysis are shown in Fig. 2b. We detected the marker positions by our own algorithm [5] based on an LK filter [7]. We derived the angles from the marker positions. Note that the angles are not identical to the actual angles because they are mapped on the sagittal plane; we also gave them arbitrary names. The accuracy of this system depends on the resolution of the camera and the distance between the treadmill and camera. The resolution was 0.007∼0.04 [rad].

III. Differentiation via L1-regulated SVM

A. L1-regulated SVM

Here, L1-regulated SVM [8] is introduced. Consider the given pairs of instances and labels $(\boldsymbol{x}_1, y_1)$, $\cdots$, $(\boldsymbol{x}_m, y_m)$ where $\boldsymbol{x}_i \in \mathcal{R}^n$ and $y_i \in \{-1, 1\}$ is the label for $\boldsymbol{x}_i$.

We consider the following problem.

$$\min_{\boldsymbol{w}} \parallel \boldsymbol{w} \parallel_1 + C \sum_{i=1}^{m} \xi_i^2 \qquad (1)$$
$$\xi_i = \max(0, 1 - y_i \boldsymbol{w}^T \boldsymbol{x}_i)$$

Here $\parallel \cdot \parallel_1$ expresses L1 norm. Note that $\parallel \boldsymbol{w} \parallel_1$ is the regularization term and ξ_i is the hinge loss. The original SVM [9] can be regarded as the problem that minimizes the error (loss) with the regularization term expressed by L2 norm. By using the regularization term expressed by L1 norm instead of the original L2 norm, the problem can become a type of least absolute shrinkage and selection operator (LASSO) [10]. Thanks to the characteristic of the constraints, the part of the components of $\boldsymbol{w}$ is perfectly identical to zero. Then, LASSO can automatically select relevant instances (features), and can handle a large number of instances (n) compared to the number of the data pairs (m). Therefore, we can define a large number of candidates of features as the components of instance $\boldsymbol{x}_i$, without considering whether or not the candidates are relevant or irrelevant.

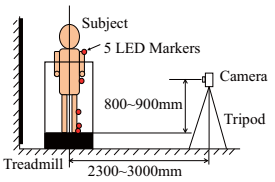
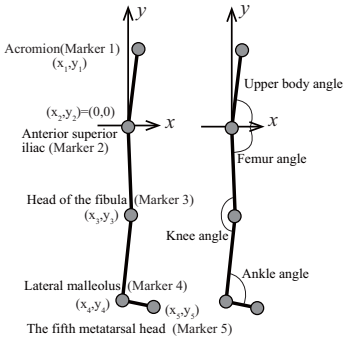


Fig. 1. Walking measurement system



(a) Marker positions (b) Angles

Fig. 2. The coordinate frame, the marker positions, and the angles

Let $\boldsymbol{d}_i \in \mathcal{R}^l$ be the original data corresponding to $\boldsymbol{x}_i$. The relationship between $\boldsymbol{d}_i$ and $\boldsymbol{x}_i$ can be represented by

$$\boldsymbol{x}_i = \boldsymbol{\phi}(\boldsymbol{d}_i) \qquad (2)$$

where $\boldsymbol{\phi} = [\phi_1\ \phi_2\ \cdots]^T$ is the basis function. Note that in general, $n \gg l$. Different from the original SVM, for $\boldsymbol{\phi}$, it is not necessary to construct kernel function like $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{\phi}(\boldsymbol{x}_i)^T \boldsymbol{\phi}(\boldsymbol{x}_i)$. Any type of nonlinear functions can be used as $\boldsymbol{\phi}$.

Given test data $\boldsymbol{d}_t$, the decision of the classifier is done by

$$\mathrm{sgn}(f(\boldsymbol{d}_t)) = \mathrm{sgn}(\boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{d}_t)) \qquad (3)$$

Note that we can add a bias term such as $f(\boldsymbol{d}) = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{d}) + b$ by letting one component of $\boldsymbol{\phi}$ be 1.

The goal of the differentiation is that given a (test) data set, we appropriately judge which group among the Normal, PAD, L4 and L5 groups the data belongs to. Then, a 4-class classifier is needed, although L1-regularized SVM is a binary classifier. A one-versus-the-rest (OVR) approach is used for constructing the 4-class classifier. OVR is used to train k independent binary classifiers where at every ith classifier, we split the data set into two classes: one is the data belonging to the ith class and the other is the remaining data. Letting the label for the former data set be positive and the label for the latter data set be negative, we conduct the training. Then, the ith classifier decision is done by

$$\mathrm{sgn}(f_i(\boldsymbol{d}_t)) = \mathrm{sgn}(\boldsymbol{w}_i^T \boldsymbol{\phi}(\boldsymbol{d}_t)) \qquad (4)$$

The overall decision is done by

$$i = \underset{i}{\mathrm{argmax}}\ f_i(\boldsymbol{d}) \qquad (5)$$

TABLE I

COMPONENTS OF DATA ($\boldsymbol{d}$) FOR INPUT OF THE L1-REGULARIZED SVM. NOTE THAT WE CONSIDERED THE LISTED VALUES WITH REGARD TO EVERY ANGLE SHOWN IN FIG.2B AND THE ANGLES CORRESPONDING TO THE QUADRICEPS AND GASTROCNEMIUS MUSCLES.

average	of angle at
maximum	1 cycle
minimum	swing phase
median	stance phase
range	
Average of 10/20 points before swing phase starts	
Average of 10/20 points after swing phase starts	
Average of 5/10 points before and after stance phase starts	
Average of 5/10 points before and after the maximum point	
Average of 5/10 points before and after the minimum point	

		Predicted class			
		Normal	PAD	L5	L4
Actual class	Normal	**12**	0	0	1
	PAD	1	**8**	3	1
	L5	0	0	**20**	2
	L4	1	2	1	**7**

Fig. 3. Confusion matrix of the L1-regulated SVM classifier. Total accuracy was 79.7%

TABLE II

THE AVERAGE NUMBER OF SELECTED FEATURES FOR EVERY CLASSIFIER

Classifier for Normal vs others	175.4
Classifier for PAD vs others	255.3
Classifier for L5 vs others	324.0
Classifier for L4 vs others	245.4

B. Application methodology

We took the first 10 cycles per participant from the recorded data. We approximated the gait pattern for every angle by spline curves and took 101 equally spaced data points (with respect to time) from one cycle of the gait pattern. Then, all of the data obtained for every angle was made into a 101 dimensional vector. Utilizing the data vector, we calculated the average, maximum, minimum, and median values with respect to 1 cycle and the swing and stance phases. Note that we focused on the motion of the marker attached on the 5th metatarsal head. We regard the time when the x component of the marker position is maximum as the stance phase start time. Similarly, we regarded the time when the marker position is minimum as the swing phase start time. In addition, we considered the stance phase start time, swing phase start time, the time when the angle is maximum and minimum as the special points, and calculated the average values around the special points. The details of the calculated values are shown in Table I. We calculated these values with regard to every angle shown in Fig. 2b, and Angle 5 = Upper body angle + Femur angle − Knee angle and Angle 6 = −Knee angle + Ankle angle. In our previous paper [6], we showed that with regard to the factors associated with the quadriceps and gastrocnemius muscles (biarticular muscles), there are statistically significant differences between some groups. Then, we additionally calculated the feature values with regard to Angle 5 corresponding to the quadriceps muscle and Angle 6 corresponding to the gastrocnemius muscle. These listed values are regarded as data $\boldsymbol{d} \in \mathcal{R}^l (= \mathcal{R}^{150})$.

L1-regularized SVM is normally for the case when the number of instances is much larger than the number of data, and the instances are supposed to be sparse. Then, in order to obtain the sparseness of the instances and good classification, we considered the following basis function

$$\phi_{jk}(\boldsymbol{d}) = \exp(-\frac{\parallel d_j - d_{jk_{avg}} \parallel^2}{2\sigma_{jk}^2}) \quad (j = 1, \cdots, 59,\ k = 1, \cdots, 150) \tag{6}$$

where $d_{jk_{avg}}$ and σ_{jk} are respectively the average and variance values for the jth original instance d_j (jth component of $\boldsymbol{d}$) of the kth participant. In total, $90000 (= 150 + 59 \times 150)$ instances were used where the number of participants is 59 and $\boldsymbol{d}$ itself is included ($\phi(\boldsymbol{d}) = \boldsymbol{d}$).

In order to investigate the possibility of differentiation, the one subject-leave-out method was used. Data for one participant were used as test data and the remaining data were used as training data. We had 59 participants, and repeated the analysis 59 times while changing the test data. Because there were 10 data sets for every participant, we had 10 results at every test. We decided which group the participant belongs to by a majority vote.

C. Results

Fig. 3 shows the results of the classification in the form of a confusion matrix. The total classification accuracy was 79.7%, which confirms the efficiency of the present approach.

The L1-generized SVM classifier can select relevant features. In order to see the effect, we investigated the number of selected relevant features for every classifier that is for classifying one specified group and the other groups. The results are shown in Table II. It can be seen that the relevant features are effectively selected, and the number of the relevant features are definitely much smaller than the number of the input features (instances).

In order to see which kinds of features are effective for classification and selection, we investigated the relevant selected features. Regarding the features/instances expressed by (6) as the jth component (feature) of $\boldsymbol{d}_i$, we counted the number of relevant features. Table III shows the relevant features for each classifier, which are the features selected most often.

D. Discussion

Although the total accuracy for classification was good, it should be improved. The classification accuracy for the Normal and L5 groups was very high while that for the PAD and L4 groups was not high. L1-regularized SVM is not suitable for dealing with the correlation between instances/features. In the present approach, we did not define the instances/features $\boldsymbol{\phi}$ considering the correlation between

TABLE III

THE RELEVANT FEATURES SELECTED MOST OFTEN

Classifier	Relevant features
Normal vs others	Minimum value at stance phase of Femur angle
PAD vs others	Range at swing phase of the angle corresponding to the gastrocnemius muscle
L5 vs others	Average of 20 points before the swing phase of the Ankle angle starts
L4 vs others	Median of the Femur angle at 1 cycle

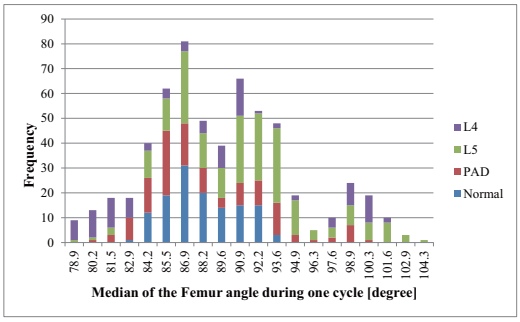


Fig. 4. Distribution of the median values of the Femur angle during one cycle. This feature is for classifying L4 and the other groups.

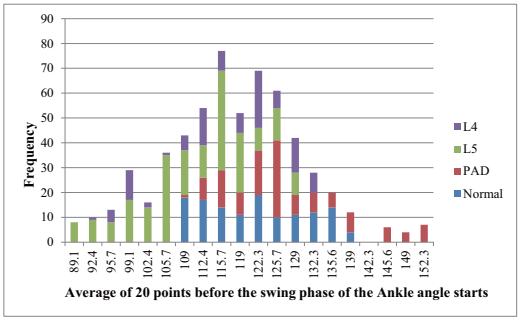


Fig. 5. Distribution of the average values of 20 points before the swing phase of the Ankle angle starts. This feature is for classifying L5 and the other groups.

the components of data $\boldsymbol{d}$, because the number of all possible combinations of the correlation is extremely large and consequently untreatable. This might be one reason why the classification accuracy for the PAD and L4 groups is not high. How to deal with the correlation is part of our future work. As shown in Table II, the number of relevant features for classification of the L5 group is the largest, which might affect the good classification result shown in Fig. 3. Note that, in general, classifying the Normal group is easier than classifying the other groups, and then the required number of relevant features was small.

Among the features listed in Table III, we focus on the features for classifying the L4 and L5 groups because that for the PAD group is close to the features presented in [6]. In order to see the details of the features, we investigated the distribution of the features, as shown in Fig. 4 and Fig. 5. We made 20 equally spaced intervals for the feature values. The scale of the horizontal axis shows the center value of each interval. We counted the elements/features included in every interval and showed the counted frequency in the figures. The vertical axis shows the frequency. From Fig. 4, it can be seen that most of the elements distributed around the lowest values belong to the L4 group. These elements are considered to have effectively worked for classification. The instances given by (6) are for every participant. This is the trick that gives features valid for not the entire group but for part of the group. Then, the present approach can deal with individual differences, which is the main difference from the previous studies. This discussion is true of the feature for classification of the L5 group, as shown in Fig. 5. From a clinical perspective, a low femur angle is indicative of a low hip position; this is considered to be the strategy for compensating the decrease of the quadriceps' muscle power in the L4 group. As for the feature shown in Fig. 5, it is related with the drop foot. The decrease in power of the tibialis anterior muscle can be seen in the L5 group, and their toes are likely to drop especially during the swing phase, which could cause stumbling. Then, for safe landing, they took the strategy of keeping the ankle angle as small as possible just before takeoff.

IV. CONCLUSION

This paper presented an approach for classifying the main disease of intermittent claudication via an L1-regularized SVM. The accuracy of the approach was 79.7%, and relevant differentiation features could be extracted with the classification. The main characteristics of the extracted features are that they are valid for not the entire group but for part of the group, and are able to deal with individual differences. Although the classification accuracy is high, it is not enough for practical differentiation. Improving the classification is our future work.

REFERENCES

[1] Y. Toribatake, "Classification and differential diagnosis of intermittent claudication," *J. of spine and spinal cord*, vol. 21, no. 4, pp. 333–340, 2008, (In Japanese).

[2] Y. Toribatake and N. Komine, "Usefulness of stress-loading test for ankle brachial index using an originally developed exercise device to detect peripheral arterial disease," *International angiology*, vol. 28, no. 2, pp. 100–105, 2009.

[3] S. A. Scherer, J. S. Bainbridge, W. R. Hiatt, and J. G. Regensteiner, "Gait characteristics of patients with claudication," *Archives of Physical Medicine and Rehabilitation*, vol. 79, no. 5, pp. 529–531, 1998.

[4] N. Yokogawa et. al., "Evaluation of differences in the gait characteristics of patients with lumbar spinal canal stenosis (l4 radiculopathy) and osteoarthritis of the hip by using a new motion analysis method," *Proc. of EuroSpine*, 2012.

[5] T. Watanabe, Y. Sanou, T. Yoneyama, Y. Toribatake, H. Hayashi, and N. Yokogawa, "Walking motion analysis of intermittent claudication and its application to medical diagnosis," *Proc. BioRob*, pp. 448–453, 2010.

[6] T. Watanabe, T. Yoneyama, Y. Toribatake, H. Hayashi, and N. Yokogawa, "Study on differentiation factors for main disease identification of intermittent claudication," *Proc. of EMBC*, pp. 4696 – 4699, 2012.

[7] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proc. of Imaging Understanding Workshop*, pp. 121–130, 1981.

[8] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

[9] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.