IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 October 2023, accepted 20 November 2023, date of publication 23 November 2023, date of current version 1 February 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3335948

RESEARCH ARTICLE

Simultaneous Learning of Erector Spinae Muscles for Automatic Segmentation of Site-Specific Skeletal Muscles in Body CT Images

MASAHIRO KAWAMOTO[1], NAOKI KAMIYA[1], (Member, IEEE), XIANGRONG ZHOU[2], HIROKI KATO[3], TAKESHI HARA[2,4], AND HIROSHI FUJITA[2]

[1]Graduate School of Information Science and Technology, Aichi Prefectural University, Nagakute, Aichi 480-1198, Japan
[2]Faculty of Engineering, Gifu University, Gifu 501-1193, Japan
[3]Department of Radiology, Graduate School of Medicine, Gifu University, Gifu 501-1193, Japan
[4]Center for Healthcare Information Technology (C-HiT), Tokai National Higher Education and Research System, Nagoya, Aichi 464-8601, Japan

Corresponding authors: Masahiro Kawamoto (im233002@cis.aichi-pu.ac.jp) and Naoki Kamiya (n-kamiya@ist.aichi-pu.ac.jp)

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (C) (#21K12731).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Review Committees of Gifu University (28-120, December 9, 2022) and Aichi Prefectural University (Jo2022-049, January 4, 2023).

ABSTRACT Skeletal muscle segmentation of the L3 slice can be used to estimate total body skeletal muscle mass. However, site-specific three-dimensional (3D) segmentation of each region, such as the erector spinae, quadratus lumborum, psoas major, oblique muscle, and rectus abdominis, in the L3 slice has not yet been achieved for the accurate measurement of skeletal muscle mass. Herein, we propose a method for site-specific 3D segmentation of skeletal muscles in the L3 slice from body CT images. We focused on the characteristics of the erector spinae muscle (ESM), which can be simultaneously observed with other skeletal muscles on craniocaudal slices and can be accurately segmented using machine learning. We introduce a segmentation method with ESM (w/ESM) using 2D U-Net for the simultaneous learning of the erector spinae and skeletal muscles, which are recognized as targets. In a three-fold cross-validation using 30 CT image cases, the mean Dice value of the baseline method, without ESM (wo/ESM), was 0.637, whereas that for the segmentation of skeletal muscles by w/ESM was 0.864, an improvement of 0.227. Our method improves the accuracy of site-specific segmentation of skeletal muscles within the L3 slice and helps evaluate the skeletal muscles through 3D imaging. This effect of w/ESM was confirmed for the skeletal muscles within the L3 slice as well as the trapezius and supraspinatus muscles. These results demonstrate the effectiveness of simultaneously learning the erector spinae and skeletal muscles in improving the accuracy of site-specific segmentation of skeletal muscles.

INDEX TERMS Erector spinae muscles, L3 slice, simultaneous learning, skeletal muscles, U-Net.

I. INTRODUCTION

The evaluation of skeletal muscle mass and adipose tissue within skeletal muscles helps understand the skeletal muscle health and function [1], [2], [3], [4], [5], [6]. For example, sarcopenia is a condition wherein the skeletal muscle mass significantly decreases with age, leading to reduced muscle strength and functionality [1]. This may further result in physical and functional disabilities [2]. In addition, decreased skeletal muscle mass in lung cancer patients is caused by cancer-associated cachexia (CAC) [3]. In patients with chronic obstructive pulmonary disease (COPD), a correlation has been observed between disease severity and decrease in the cross-sectional area of the erector spinae [4]. Moreover, the erector spinae plays a crucial role in spinal disorders [5]. These findings illustrate the significance of measuring skeletal muscle characteristics such as muscle mass, muscle density, and fat infiltration (FI) [6]. These skeletal muscle

The associate editor coordinating the review of this manuscript and approving it for publication was R. K. Tripathy.

features can be measured using CT images and can be useful in disease prediction and patient prognosis through diagnostic imaging [6]. The skeletal muscle mass index (SMI), which is the amount of skeletal muscle mass in the third lumbar vertebra (L3) section normalized by the square of the height, is used to quantify skeletal muscle mass [7]. This index is used because the skeletal muscle mass in the L3 cross-section correlates with the total body skeletal muscle mass [8].

Previously, the semi-automatic segmentation of skeletal muscles in L3 cross-sections has been performed using software-based thresholding [9], and automatic segmentation has been performed using deep learning [10], [11], [12]. In the automatic segmentation of skeletal muscles using deep learning, multiple skeletal muscles within the L3 cross-section are segmented as a single region [10], [11]. In other words, the erector spinae, quadratus lumborum, psoas major, external oblique abdominis, internal oblique abdominis, transverse abdominis, and rectus abdominis were segmented into the same skeletal muscle region without any distinction. Additionally, multiple skeletal muscles within the L3 cross-section have been site-specifically segmented [12]. Moreover, these techniques are limited to segmentation within the L3 cross-section and do not allow a three-dimensional (3D) evaluation of individual skeletal muscles.

In contrast, the method for site-specific 3D segmentation of skeletal muscles involves automatic segmentation by constructing shape models that mathematically represent the general form of skeletal muscles [13]. In such shape-model-based muscle segmentation, origins and insertions, which are the anatomical attachment points of muscles, are used to locate individual skeletal muscles. These points must be recognized as landmarks (LMs) in a computational environment. The skeletal muscle is segmented by placing a shape model along the centerline connecting these LMs. However, muscle segmentation based on shape models presents two challenges. First, it is limited to skeletal muscles, where LM recognition is possible. Second, the accuracy of skeletal muscle segmentation depends on the accuracy of the shape model placed on the centerline, which indicates the direction of the muscle fibers generated from the obtained LMs [13]. Therefore, machine learning-based skeletal muscle segmentation that does not rely on these handcrafted features has been investigated. For example, while segmenting the erector spinae using random forests, iterative learning with multiple resolutions resulted in an accuracy of 93.0% [14]. Additionally, the use of deep learning technology has led to the segmentation of the supraspinatus and erector spinae muscle (ESM) using a two-dimensional convolutional neural network [15], [16]. In the lower limbs, the simultaneous segmentation of 19 skeletal muscle regions has been achieved [17]. Thus, the possibility of site-specific skeletal muscle segmentation has increased significantly compared to shape-model-based methods because of the automatic acquisition of features through machine and deep learning. Moreover, while 3D recognition of skeletal muscles has been conducted, albeit for limited targets, there is a lack of research elucidating the effects of simultaneously learning multiple skeletal muscles.

In this study, we propose a method for 3D site-specific segmentation of skeletal muscles in an L3 cross-section using deep learning. We focused on the erector spinae, which is the largest skeletal muscle located on the back [18]. Many skeletal muscles can be simultaneously observed with the erector spinae in the craniocaudal cross-section. Moreover, machine learning-based methods have achieved high accuracy in segmenting the erector spinae (93.0% [14] and 93.4% [16]). Therefore, we expect to use the erector spinae to improve the 3D segmentation accuracy of site-specific skeletal muscles in the L3 cross-section. We compared the accuracy of segmenting each skeletal muscle in the L3 cross-section, including the quadratus lumborum, psoas major, oblique muscle, and rectus abdominis, by using the proposed method of simultaneous learning with the erector spinae (w/ESM) and individual learning without the erector spinae (wo/ESM).

II. METHODS

A. OVERVIEW

In this study, we propose a method for the automatic segmentation of skeletal muscles in the L3 cross-section in body CT images. We used 2D U-Net [19]—which is widely used for automatic segmentation in medical imaging [20]—as our deep learning framework. An outline of the proposed method is shown in Fig. 1. It is a general training test model wherein the ESM and a target skeletal muscle are selected, and both skeletal muscle regions are learned simultaneously. Consequently, a segmentation network of the erector spinae and target skeletal muscle is constructed. In the segmentation results, only the target skeletal muscle is extracted and integrated, which enables site-specific segmentation of the skeletal muscles depicted in the L3 cross-section. A detailed explanation of the proposed method is given below.

B. SITE-SPECIFIC SEGMENTATION WITH SIMULTANEOUS LEARNING OF ERECTOR SPINAE MUSCLES (w/ESM)

We performed site-specific segmentation of the quadratus lumborum, psoas major, oblique muscles (internal oblique, external oblique, and transverse abdominal muscles), and rectus abdominis in the L3 cross-section. In the site-specific segmentation of skeletal muscles, the key point is to simultaneously learn the erector spinae and skeletal muscles in the L3 cross-section. During training, simultaneous learning was performed by selecting a region of the skeletal muscle to be segmented from the body CT images and one of quadratus lumborum, psoas major, oblique muscles, and rectus abdominis, and then inputting the two regions together with the ESM into 2D U-Net. Each skeletal muscle was then segmented on the input CT image, which was unknown during testing. The 2D U-Net consists of an encoder and a decoder. In the encoder, 3×3 convolution is applied twice and max pooling once, and this process is repeated four times followed by a 3×3 convolution twice to the feature map. In the decoder,

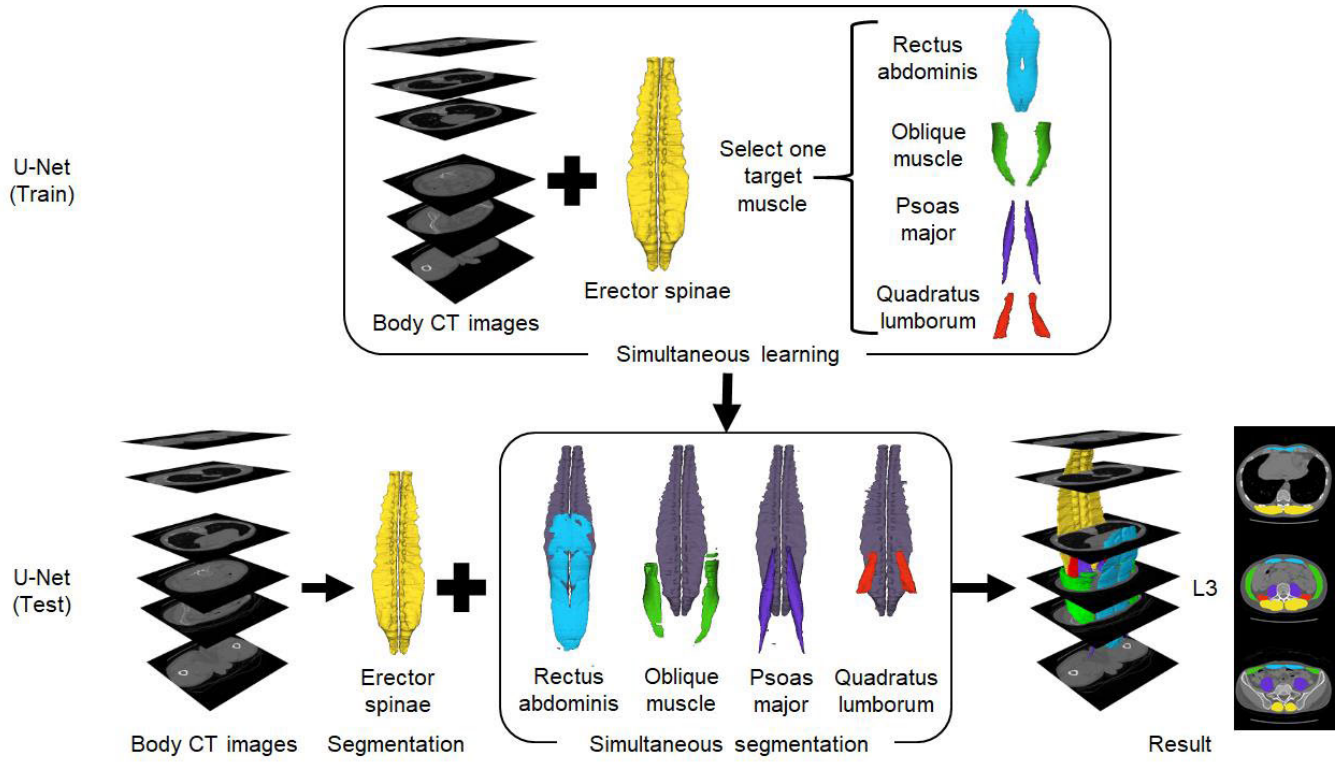

FIGURE 1. Overview of proposed method for automatic segmentation of skeletal muscles via simultaneous learning of erector spinae in body CT images.

the feature map is combined with the corresponding feature map of the encoder using up convolution, and a 3 × 3 convolution is repeated twice. In the sequence, batch normalization [21] and rectified linear unit (ReLU) [22] are applied. Finally, a 1 × 1 convolution and the softmax activation function are applied. In this network, the target skeletal muscles and erector spinae are segmented. Therefore, only the region of the target skeletal muscle is selected from the segmentation results and used as the site-specific segmentation result of the skeletal muscle. The target skeletal muscles and erector spinae are assigned different label values and could be distinguished, thus configuring a multiclass classification problem that classifies the two skeletal muscle areas and background. The loss function employed is the same as that used in [15]; we combined Cross Entropy (CE) and Dice, with weights of 0.5 and 1 assigned to CE and Dice, respectively, as shown below:

$$loss_{CE} = -\frac{\sum_{j=0}^{C}\sum_{i=0}^{N} y_{ij}\log\left(\hat{y}_{ij}\right)}{N}, \tag{1}$$

$$loss_{Dice} = -\frac{1}{N}\frac{2\times\sum_{j=0}^{C}\sum_{i=0}^{N} y_{ij}\hat{y}_{ij} + smooth}{\sum_{j=0}^{C}\sum_{i=0}^{N} y_{ij} + \sum_{j=0}^{C}\sum_{i=0}^{N}\hat{y}_{ij} + smooth}, \tag{2}$$

$$loss = 0.5\times loss_{CE} + loss_{Dice}, \tag{3}$$

where N refers to the number of pixels, C refers to the number of classes to be classified, y_{ij} and $\hat{y}_{ij}$ indicate the ground truth and prediction probability, respectively, at coordinate i in class j, and *smooth* is a constant to avoid division by zero, which was set to 1 in this study.

III. EXPERIMENTS

A. IMAGE DETAILS AND EXPERIMENTAL ENVIRONMENT

This study was approved by the ethical review committees of Gifu University (28-120, December 9, 2022) and Aichi Prefectural University (Jo2022-049, January 4, 2023). In this study, we used 30 unenhanced body CT images obtained using a LightSpeed Ultra 16 (GE Healthcare, Chicago, IL, USA) at Gifu University Hospital. The image size was 512 × 512 × 802 - 1104 [voxel], and the spatial resolution was 0.625 × 0.625 × 0.625 [mm]. The ground truth was created by manually segmenting the erector spinae, quadratus lumborum, psoas major, oblique muscle, and rectus abdominis.

Experiments were conducted on a computer with four Tesla V100 (32 GB) graphics processing units (GPUs). We used TensorFlow 2.6.0 [23] and Keras 2.6.0 as machine learning libraries. Furthermore, during learning, the parameters of 2D U-Net were set as follows: 50 for the number of epochs, 3×10^{-4} for the learning rate, 4 for the batch size, and Adam [24] was used as the optimization function. In this study, we used data augmentation for the training images during learning. The shear transformations of random angles from $-\pi/8$ to $\pi/8$, rotations from $-10°$ to $+10°$, enlargements and reductions to ensure that the length of one side of an image changes from 0.65 times to 1.35 times, and a horizontal

TABLE 1. Evaluation results.

Method		Quadratus lumborum			Psoas major			Oblique muscle			Rectus abdominis			Average		
		DC	RC	PC	DC	RC	PC	DC	RC	PC	DC	RC	PC	DC	RC	PC
Baseline (wo/ESM)	Mean	0.381	0.330	0.522	0.901	0.911	0.893	0.746	0.705	0.855	0.823	0.829	0.853	0.637	0.626	0.684
	SD	0.328	0.312	0.400	0.043	0.044	0.059	0.187	0.246	0.065	0.162	0.179	0.066	0.371	0.379	0.366
Proposed (w/ESM)	Mean	0.832	0.829	0.848	0.913	0.914	0.914	0.849	0.859	0.865	0.829	0.824	0.862	0.864	0.860	0.883
	SD	0.049	0.087	0.093	0.029	0.039	0.043	0.145	0.167	0.053	0.138	0.164	0.063	0.098	0.124	0.066

SD: Standard deviation, DC: Dice, RC: Recall, PC: Precision

flipping were applied. The volume of training data was set to eight times the number before augmentation.

B. EVALUATION

We used Dice, precision, and recall to evaluate the similarity between the segmentation results of the skeletal muscles in the L3 slice and ground truth. The metrics are defined as follows:

$$Dice = \frac{2 \times |Res \cap GT|}{|Res| + |GT|}, \tag{4}$$

$$Recall = \frac{|Res \cap GT|}{|GT|}, \tag{5}$$

$$Precision = \frac{|Res \cap GT|}{|Res|}, \tag{6}$$

where Res is the segmentation result, GT is the ground truth, and $|\ |$ returns the number of voxels contained in the region. The segmentation accuracy of the skeletal muscles was evaluated using three-fold cross-validation using 20 of the 30 images as training data and 10 images as test data.

The proposed method was compared with the learning of skeletal muscles recognized within the L3 slice alone (wo/ESM). The segmentation of each skeletal muscle represents a two-class classification between the skeletal muscle and the background. A sigmoid function was applied as the activation function of the 2D U-Net. The loss function used was a combination of binary cross entropy (BCE) and Dice.

$$loss_{BCE} = -\frac{\sum_{i=1}^{N} \left[y_i \log\left(\hat{y}_i\right) + (1 - y_i) \log\left(1 - \hat{y}_i\right) \right]}{N}, \tag{7}$$

$$loss_{Dice} = -\frac{2 \times \sum_{i=0}^{N} y_i \hat{y}_i + smooth}{\sum_{i=0}^{N} y_i + \sum_{i=0}^{N} \hat{y}_i + smooth}, \tag{8}$$

$$loss = 0.5 \times loss_{BCE} + loss_{Dice}, \tag{9}$$

where $\boldsymbol{y_i}$ and $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ indicate the ground truth and prediction probability at coordinate $\boldsymbol{i}$, respectively, and the other variables have the same meaning as that described in (1)–(3).

IV. RESULTS

Table 1 presents a baseline comparison of the site-specific segmentation accuracy of each skeletal muscle by w/ESM and wo/ESM. The mean Dice value for all skeletal muscles obtained using the proposed method was 0.864 (SD:0.098). By contrast, the mean Dice value for all skeletal muscles using the baseline was 0.637 (SD:0.371). The distributions and densities of the Dice values for w/ESM and wo/ESM are shown in the violin plot in Fig. 2. Fig. 2 shows that the mean Dice values for all skeletal muscles (quadratus lumborum, psoas major, oblique muscle, and rectus abdominis) were higher in the w/ESM group, and the distribution of data in the w/ESM group was more concentrated than in the wo/ESM group, indicating less variability in the distribution. Particularly, the w/ESM group showed a higher distribution peal in the quadratus lumborum and oblique muscles, narrower width, and greater improvement. The outliers were also improved by w/ESM for all skeletal muscles, indicating the effectiveness of the proposed method.

Fig. 3 shows the site-specific segmentation results on the L3 section for each skeletal muscle (a) and the areas of over- and under-extraction (b). In this example, under the proposed method (w/ESM), the under-extraction of the quadratus lumborum and oblique muscles observed at baseline (wo/ESM) was improved.

V. DISCUSSION

The results demonstrated that w/ESM improved the segmentation accuracy of the skeletal muscles in the L3 section. First, we discuss the recognition accuracy of the ESM, which was trained simultaneously with a recognition target. The mean Dice value of the ESM was 0.930 (SD:0.015) when trained alone (wo/ESM). By contrast, the mean Dice values of the ESM when trained simultaneously with the quadratus lumborum, psoas major, oblique muscle, and rectus abdominis (w/ESM) were 0.922 (SD:0.023), 0.922 (SD:0.023), 0.924 (SD:0.028), and 0.928 (SD:0.016), respectively. These results indicate that when the ESM, which has a high segmentation accuracy and small variance, is solely used for training, it can be stably recognized and can improve the segmentation accuracy of other skeletal muscles.

Next, we conducted simultaneous learning of the erector spinae muscle (w/ESM) for skeletal muscles other than those in the L3 section and examined its effects. We targeted the trapezius, which is a large muscle on the back, and the supraspinatus—a small muscle in the scapular region. The skeletal muscles were used for site-specific 3D segmentation [13], [15]. From the simultaneous learning of the erector spinae muscle (w/ESM), the mean Dice values of the trapezius and supraspinatus muscles determined using the proposed

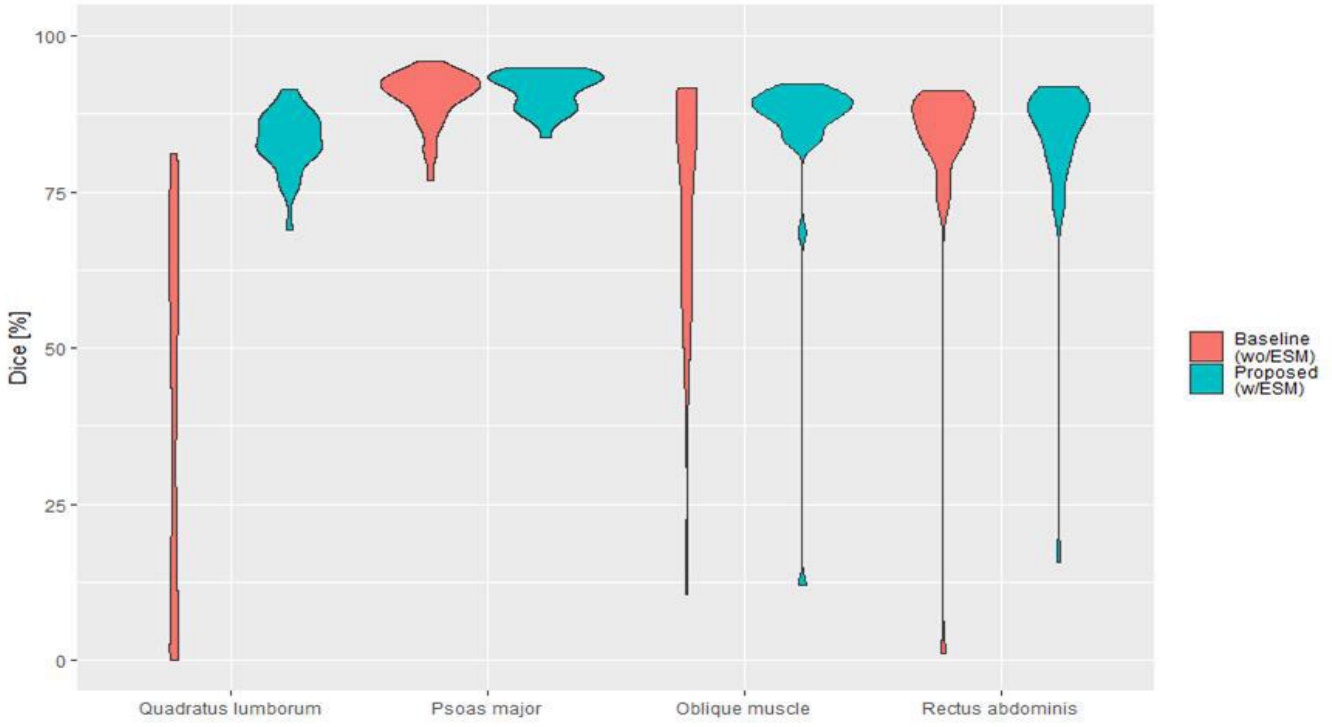

FIGURE 2. *Violin plots of Dice values for the quadratus lumborum, psoas major, oblique muscle, and rectus abdominis from the baseline (wo/ESM) and the proposed method (w/ESM). Dice values of the baseline and proposed methods are shown in red and blue, respectively.*

method were 0.876 (SD:0.080) and 0.883 (SD:0.068), respectively. However, for the baseline without the erector spinae muscle (wo/ESM), the mean Dice values for the trapezius and supraspinatus muscles were 0.527 (SD:0.384) and 0.259 (SD:0.375), respectively. Therefore, w/ESM improves the site-specific segmentation accuracy of skeletal muscles compared to that of wo/ESM and is not limited to the skeletal muscles in the L3 section. Fig. 4 shows the 3D rendered images of the site-specific segmentation results, the ground truth image obtained by the proposed method, and the baseline in the same case. The proposed method improved the under-extraction of the trapezius, quadratus lumborum, and supraspinatus, with Dice values of 0.904, 0.816, and 0.892, respectively. By contrast, the baseline showed an under-extraction area in the trapezius with a Dice value of 0.638, and the quadratus lumborum and supraspinatus were recognized as background area.

Next, we show that in the case of w/ESM, the segmentation accuracy of each skeletal muscle improved by simultaneously learning the ESMs. Table 2 presents the skeletal muscle segmentation results when the model was trained on each skeletal muscle simultaneously with skeletal muscles other than the ESMs. The segmentation accuracy for w/ESM was the highest among all skeletal muscles. Compared to the baseline, the segmentation accuracy improved for all combinations for the quadratus lumborum and trapezius. However, for some combinations of the psoas major, rectus abdominis, oblique muscle, and supraspinatus, the segmentation accuracy decreased. In particular, the psoas major and rectus abdominis muscles showed decreased segmentation accuracy owing to simultaneous learning for all combinations except the ESM. Thus, using the ESM as a target for simultaneous learning is optimal when aiming to improve the segmentation accuracy of each skeletal muscle.

Finally, we compared the accuracy of skeletal muscle segmentation in the L3 cross-section with those achieved in related studies. The proposed method aims to segment individual muscles in the L3 cross-section, and it also enables the segmentation of all muscles in the L3 cross-section. In other words, the segmentation results of the skeletal muscles in the L3 cross-section obtained using w/ESM can be treated as one region. The mean Dice value was 0.921 (SD: 0.058), which has a similar accuracy to that of the Dice value of 0.93 (SD: 0.03) achieved for skeletal muscles by the method by Castiglione et al. [10]. The proposed method's site-specific segmentation in the L3 cross-section achieved mean Dice values of 0.869 (SD:0.085), 0.916 (SD:0.087), 0.884 (SD:0.182), and 0.837 (SD:0.208) for the quadratus lumborum, psoas major, oblique muscle, and rectus abdominis, respectively, and the mean Dice value for all skeletal muscles was 0.891 (SD = 0.141). Furthermore, the mean Dice value for the 3D evaluation when the skeletal muscles within the L3 cross-section were integrated into one region based on the proposed method was 0.902 (SD:0.036), and

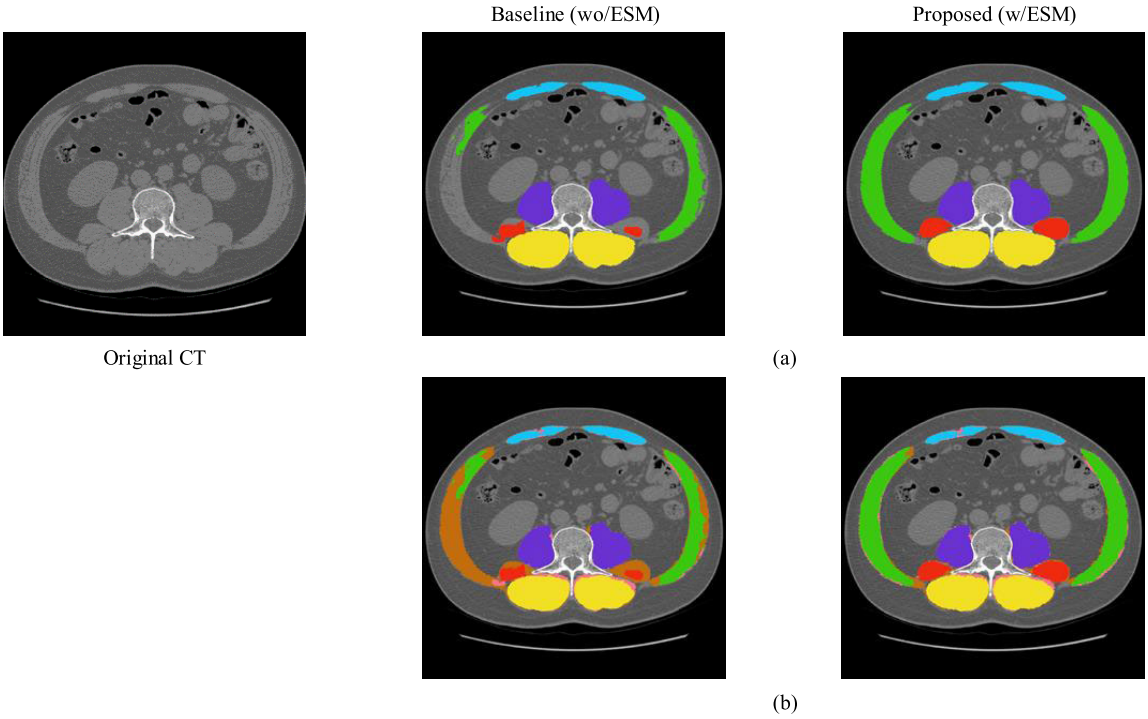

FIGURE 3. Segmentation results for each skeletal muscle in the L3 cross-section. (a) Coincident regions with the correct region for erector spinae, quadratus lumborum, psoas major, oblique muscle, and rectus abdominis are shown in yellow, red, purple, green, and light blue, respectively. (b) Over- and under-extraction are shown in pink and brown, respectively.

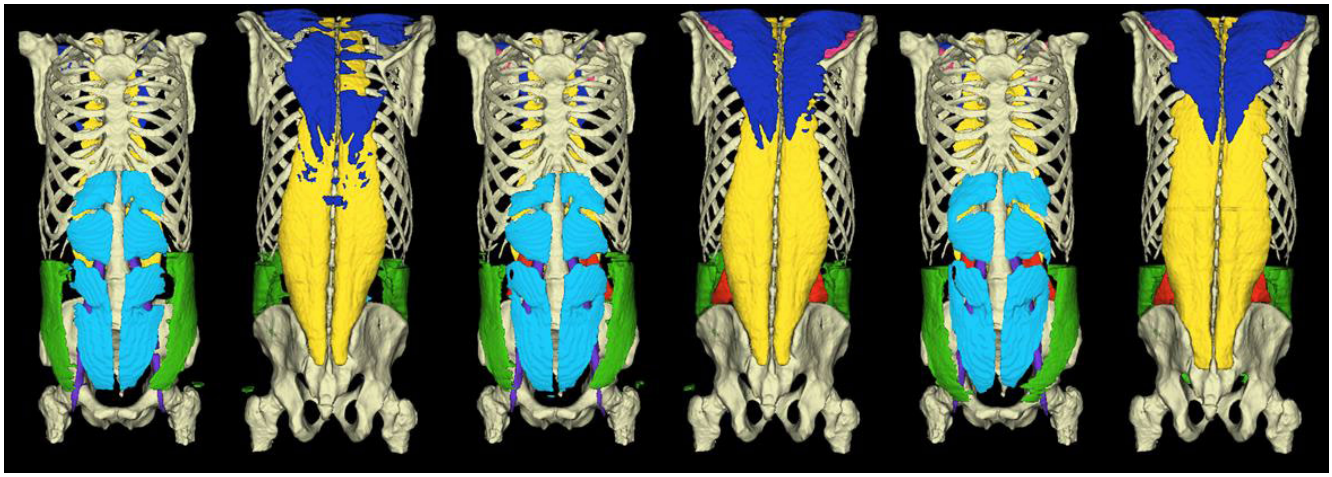

FIGURE 4. 3D volume rendered images of the segmentation results for each skeletal muscle. (a) Baseline (wo/ESM) segmentation results and Dice values. (b) Segmentation result using the proposed method (w/ESM) and the Dice value. Erector spinae, quadratus lumborum, psoas major, oblique muscle, rectus abdominis, trapezius, and supraspinatus muscles are shown in yellow, red, purple, green, light blue, blue, and pink, respectively.

the mean Dice value for all skeletal muscles in site-specific segmentation was 0.864 (SD:0.098). These results indicate that the proposed method not only achieves an accuracy equivalent to those achieved in related research on skeletal muscles in the L3 cross-section but also achieves site-specific segmentation accuracy and enables site-specific skeletal muscle segmentation in 3D [12].

This study has some limitations; the improvement due to w/ESM was evident in the segmentation of skeletal muscles, which were the target of site-specific segmentation; however,

TABLE 2. Evaluation results for each skeletal muscle combination.

		Q: Quadratus lumborum			P: Psoas major			O: Oblique muscle			R: Rectus abdominis			T: Trapezius			S: Supraspinatus		
		DC	RC	PC	DC	RC	PC	DC	RC	PC	DC	RC	PC	DC	RC	PC	DC	RC	PC
Baseline (wo/ESM)	Mean	0.381	0.330	0.522	0.901	0.911	0.893	0.746	0.705	0.855	0.823	0.829	0.853	0.527	0.476	0.662	0.259	0.276	0.250
	SD	0.328	0.312	0.400	0.043	0.044	0.059	0.187	0.246	0.065	0.162	0.179	0.066	0.384	0.351	0.411	0.375	0.403	0.365
Proposed (w/ESM)	Mean	0.832	0.829	0.848	0.913	0.914	0.914	0.849	0.859	0.865	0.829	0.824	0.862	0.876	0.849	0.915	0.883	0.886	0.891
	SD	0.049	0.087	0.093	0.029	0.039	0.043	0.145	0.167	0.053	0.138	0.164	0.063	0.080	0.106	0.061	0.068	0.115	0.038
Q & P	Mean	0.643	0.625	0.809	0.608	0.602	0.737	-	-	-	-	-	-	-	-	-	-	-	-
	SD	0.289	0.333	0.172	0.426	0.428	0.310	-	-	-	-	-	-	-	-	-	-	-	-
Q & O	Mean	0.701	0.645	0.835	-	-	-	0.828	0.871	0.819	-	-	-	-	-	-	-	-	-
	SD	0.182	0.209	0.124	-	-	-	0.155	0.173	0.060	-	-	-	-	-	-	-	-	-
Q & R	Mean	0.634	0.627	0.819	-	-	-	-	-	-	0.548	0.511	0.762	-	-	-	-	-	-
	SD	0.275	0.330	0.112	-	-	-	-	-	-	0.366	0.373	0.234	-	-	-	-	-	-
Q & T	Mean	0.386	0.372	0.479	-	-	-	-	-	-	-	-	-	0.837	0.811	0.881	-	-	-
	SD	0.322	0.340	0.395	-	-	-	-	-	-	-	-	-	0.073	0.123	0.075	-	-	-
Q & S	Mean	0.767	0.785	0.766	-	-	-	-	-	-	-	-	-	-	-	-	0.503	0.569	0.466
	SD	0.067	0.102	0.104	-	-	-	-	-	-	-	-	-	-	-	-	0.371	0.422	0.354
P & O	Mean	-	-	-	0.883	0.846	0.932	0.821	0.815	0.864	-	-	-	-	-	-	-	-	-
	SD	-	-	-	0.050	0.088	0.044	0.141	0.172	0.075	-	-	-	-	-	-	-	-	-
P & R	Mean	-	-	-	0.715	0.719	0.773	-	-	-	0.613	0.596	0.795	-	-	-	-	-	-
	SD	-	-	-	0.313	0.355	0.191	-	-	-	0.358	0.376	0.206	-	-	-	-	-	-
P & T	Mean	-	-	-	0.827	0.792	0.906	-	-	-	-	-	-	0.846	0.796	0.924	-	-	-
	SD	-	-	-	0.140	0.206	0.063	-	-	-	-	-	-	0.102	0.141	0.037	-	-	-
P & S	Mean	-	-	-	0.872	0.870	0.885	-	-	-	-	-	-	-	-	-	0.509	0.505	0.524
	SD	-	-	-	0.083	0.123	0.049	-	-	-	-	-	-	-	-	-	0.374	0.374	0.390
O & R	Mean	-	-	-	-	-	-	0.840	0.842	0.866	0.815	0.807	0.870	-	-	-	-	-	-
	SD	-	-	-	-	-	-	0.137	0.166	0.049	0.169	0.013	0.070	-	-	-	-	-	-
O & T	Mean	-	-	-	-	-	-	0.814	0.801	0.843	-	-	-	0.610	0.584	0.855	-	-	-
	SD	-	-	-	-	-	-	0.121	0.167	0.084	-	-	-	0.382	0.392	0.238	-	-	-
O & S	Mean	-	-	-	-	-	-	0.709	0.684	0.796	-	-	-	-	-	-	0.250	0.218	0.465
	SD	-	-	-	-	-	-	0.221	0.269	0.118	-	-	-	-	-	-	0.360	0.319	0.475
R & T	Mean	-	-	-	-	-	-	-	-	-	0.719	0.655	0.862	0.826	0.790	0.901	-	-	-
	SD	-	-	-	-	-	-	-	-	-	0.227	0.251	0.170	0.154	0.179	0.065	-	-	-
R & S	Mean	-	-	-	-	-	-	-	-	-	0.774	0.757	0.821	-	-	-	0.179	0.132	0.290
	SD	-	-	-	-	-	-	-	-	-	0.149	0.179	0.080	-	-	-	0.261	0.193	0.421
T & S	Mean	-	-	-	-	-	-	-	-	-	-	-	-	0.767	0.703	0.919	0.000	0.000	-
	SD	-	-	-	-	-	-	-	-	-	-	-	-	0.185	0.228	0.055	0.000	0.000	-

SD: Standard deviation, DC: Dice, RC: Recall, PC: Precision

variability existed in the effects between skeletal muscle sites, making it challenging to predict the effects of w/ESM quantitatively on unknown subjects. Thus, although the effect of w/ESM on site-specific skeletal muscle segmentation in this study was most pronounced for the supraspinatus, quadratus lumborum, and trapezius muscles, the effect of w/ESM on improving the segmentation accuracy was limited to the psoas major and rectus abdominis. Therefore, in future works, in site-specific skeletal muscle segmentation for skeletal muscles within the L3 cross-section proposed in this study and for skeletal muscles other than the trapezius and supraspinatus, the simultaneous learning of the ESM and the target skeletal muscle by w/ESM should be considered first, along with a comparison of accuracy with wo/ESM. Furthermore, the relationship between the ESM and target skeletal muscle must be quantified. The proposed method utilized 2D U-Net for carrying out 3D recognition. However, in the realm of medical image segmentation, organ recognition has been conducted using 3D networks [25]. Therefore, it is essential to explore whether the insights garnered through w/ESM in

skeletal muscle segmentation are effective when applied to 3D networks. This has facilitated the need for further investigation to ascertain the applicability of the knowledge gained from w/ESM in a 3D network setting for skeletal muscle segmentation.

VI. CONCLUSION

In this study, we proposed a method for the 3D site-specific segmentation of skeletal muscles within the L3 slice from body CT images using simultaneous learning of the erector spinae muscle (w/ESM). This method conducts site-specific segmentation of skeletal muscles by simultaneously learning two muscle regions, the erector spinae and target skeletal muscles, using 2D U-Net. We demonstrated that the w/ESM achieved a higher segmentation accuracy with a mean Dice value of 0.864, compared to the 0.637 achieved when the erector spinae was not simultaneously
learned.

Additionally, the effect of w/ESM was not limited to the skeletal muscles within the L3 slice but also improved the site-specific segmentation accuracy of the back muscles, specifically the trapezius and supraspinatus. Among the combinations of simultaneous learning, the choice of the ESM resulted in the most significant improvement in the segmentation accuracy of each skeletal muscle. Driven by the ability to recognize each skeletal muscle depicted in the L3 cross-section in 3D, which is not just limited to the L3 cross-section, we believe that it is possible to explore effective cross-sections other than the L3 for estimating total body skeletal muscle volume and to consider the use of 3D region-specific skeletal muscle volumes.

However, although w/ESM was effective in all regions, its effect varied among the different muscle sites. A quantitative prediction of the effects of w/ESM in unknown subjects is challenging. Future studies should focus on quantifying the effects of w/ESM on the erector spinae and targeting skeletal muscles for site-specific segmentation of skeletal muscles.

ACKNOWLEDGMENT

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (C) (#21K12731).

REFERENCES

[1] A. J. Cruz-Jentoft, J. P. Baeyens, J. M. Bauer, Y. Boirie, T. Cederholm, F. Landi, F. C. Martin, J.-P. Michel, Y. Rolland, S. M. Schneider, E. Topinková, M. Vandewoude, and M. Zamboni, "Sarcopenia: European consensus on definition and diagnosis: Report of the European working group on sarcopenia in older people," *Age Ageing*, vol. 39, no. 4, pp. 412–423, Apr. 2010, doi: 10.1093/ageing/afq034.

[2] I. Janssen, S. B. Heymsfield, and R. Ross, "Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability," *J. Amer. Geriatrics Soc.*, vol. 50, no. 5, pp. 889–896, May 2002, doi: 10.1046/j.1532-5415.2002.50216.x.

[3] O. Al-Sawaf, "Body composition and lung cancer-associated cachexia in TRACERx," *Nature Med.*, vol. 29, no. 4, pp. 846–858, Apr. 2023, doi: 10.1038/s41591-023-02232-8.

[4] K. Tanimura, S. Sato, Y. Fuseya, K. Hasegawa, K. Uemasu, A. Sato, T. Oguma, T. Hirai, M. Mishima, and S. Muro, "Quantitative assessment of erector spinae muscles in patients with chronic obstructive pulmonary disease. Novel chest computed tomography–derived index for prognosis," *Ann. Amer. Thoracic Soc.*, vol. 13, no. 3, pp. 334–341, Mar. 2016, doi: 10.1513/annalsats.201507-446oc.

[5] M. Suo, J. Zhang, T. Sun, J. Wang, X. Liu, H. Huang, and Z. Li, "The association between morphological characteristics of paraspinal muscle and spinal disorders," *Ann. Med.*, vol. 55, no. 2, Sep. 2023, Art. no. 2258922, doi: 10.1080/07853890.2023.2258922.

[6] K. Engelke, O. Museyko, L. Wang, and J.-D. Laredo, "Quantitative analysis of skeletal muscle by computed tomography imaging—State of the art," *J. Orthopaedic Transl.*, vol. 15, pp. 91–103, Oct. 2018, doi: 10.1016/j.jot.2018.10.004.

[7] C. M. M. Prado, J. C. K. Wells, S. R. Smith, B. C. M. Stephan, and M. Siervo, "Sarcopenic obesity: A critical appraisal of the current evidence," *Clin. Nutrition*, vol. 31, no. 5, pp. 583–601, Oct. 2012, doi: 10.1016/j.clnu.2012.06.010.

[8] W. Shen, M. Punyanitya, Z. Wang, D. Gallagher, J. Albu, S. B. Heymsfield, and S. Heshka, "Total body skeletal muscle and adipose tissue volumes: Estimation from a single abdominal cross-sectional image," *J. Appl. Physiol.*, vol. 97, no. 6, pp. 2333–2338, Dec. 2004, doi: 10.1152/japplphysiol.00744.2004.

[9] J. L. A. van Vugt, S. Levolger, A. Gharbharan, M. Koek, W. J. Niessen, J. W. A. Burger, S. P. Willemsen, R. W. F. de Bruin, and J. N. M. Ijzermans, "A comparative study of software programmes for cross-sectional skeletal muscle and adipose tissue measurements on abdominal computed tomography scans of rectal cancer patients," *J. Cachexia, Sarcopenia Muscle*, vol. 8, no. 2, pp. 285–297, Nov. 2016.

[10] J. Castiglione, E. Somasundaram, L. A. Gilligan, A. T. Trout, and S. Brady, "Automated segmentation of abdominal skeletal muscle on pediatric CT scans using deep learning," *Radiol., Artif. Intell.*, vol. 3, no. 2, Mar. 2021, Art. no. e200130.

[11] H. Shen, P. He, Y. Ren, Z. Huang, S. Li, G. Wang, M. Cong, D. Luo, D. Shao, E. Y.-P. Lee, R. Cui, L. Huo, J. Qin, J. Liu, Z. Hu, Z. Liu, and N. Zhang, "A deep learning model based on the attention mechanism for automatic segmentation of abdominal muscle and fat for body composition assessment," *Quant. Imag. Med. Surgery*, vol. 13, no. 3, pp. 1384–1398, Mar. 2023, doi: 10.21037/qims-22-330.

[12] R. Kreher, M. Hinnerichs, B. Preim, and S. Saalfeld, "Deep-learning-based segmentation of skeletal muscle mass in routine abdominal CT scans," *In Vivo*, vol. 36, no. 4, pp. 1807–1811, Aug. 2022, doi: 10.21873/invivo.12896.

[13] N. Kamiya, "Deep learning technique for musculoskeletal analysis," in *Deep Learning in Medical Image Analysis: Challenges and Application*, G. Lee and H. Fujita, Eds. Cham, Switzerland: Springer Nature, 2020, pp. 165–176, doi: 10.1007/978-3-030-33128-3_11.

[14] N. Kamiya, J. Li, M. Kume, H. Fujita, D. Shen, and G. Zheng, "Fully automatic segmentation of paraspinal muscles from 3D torso CT images via multi-scale iterative random forest classifications," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 13, no. 11, pp. 1697–1706, Sep. 2018, doi: 10.1007/s11548-018-1852-1.

[15] Y. Wakamatsu, N. Kamiya, X. Zhou, H. Kato, T. Hara, and H. Fujita, "Automatic segmentation of supraspinatus muscle via bone-based localization in torso computed tomography images using U-net," *IEEE Access*, vol. 9, pp. 155555–155563, 2021, doi: 10.1109/ACCESS.2021.3127565.

[16] Y. Wakamatsu, N. Kamiya, X. Zhou, T. Hara, and H. Fujita, "Relationship between number of annotations and accuracy in segmentation of the erector spinae muscle using Bayesian U-Net in torso CT images," *Int. Forum Med. Imag. Asia*, vol. 11792, Apr. 2021, Art. no. 1179207, doi: 10.1117/12.2590780.

[17] Y. Hiasa, Y. Otake, M. Takao, T. Ogawa, N. Sugano, and Y. Sato, "Automated muscle segmentation from clinical CT using Bayesian U-Net for personalized musculoskeletal modeling," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1030–1040, Apr. 2020, doi: 10.1109/TMI.2019.2940555.

[18] R. L. Drake et al., *Conceptual Overview—GRAY's Anatomy for Students*. Amsterdam, The Netherlands: Elsevier, 2011, pp. 56–121.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, 2015, pp. 234–241.

[20] J. S. Suri, M. Bhagawati, S. Agarwal, S. Paul, A. Pandey, S. K. Gupta, L. Saba, K. I. Paraskevas, N. N. Khanna, J. R. Laird, A. M. Johri, M. K. Kalra, M. M. Fouda, M. Fatemi, and S. Naidu, "UNet deep learning architecture for segmentation of vascular and non-vascular images: A microscopic look at UNet components buffered with pruning, explainable artificial intelligence, and bias," *IEEE Access*, vol. 11, pp. 595–645, 2023, doi: 10.1109/ACCESS.2022.3232561.
[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[22] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, Haifa, Israel, no. 432, Jun. 2010, pp. 807–814.
[23] M. Abadi, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," in *Proc. 12th USENIX Symp. Operating Syst. Design Implement.*, Atlanta, GA, USA, Nov. 2016, pp. 265–283.
[24] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. of 3rd Int. Conf. Learn. Represent.*, San Francisco, CA, USA, May 2015, pp. 1–15. [Online]. Available: https://arxiv.org/pdf/1412.6980v8.pdf
[25] C. You, "Implicit anatomical rendering for medical image segmentation with stochastic experts," in *Proc. Med. Image Comput. Comput. Assist. Intervent.*, vol. 14222, Oct. 2023, pp. 561–571.

MASAHIRO KAWAMOTO received the B.S. degree in information science and technology from Aichi Prefectural University, Japan, in 2023, where he is currently pursuing the M.S. degree. His current research interest includes medical image processing.

NAOKI KAMIYA (Member, IEEE) received the B.S. degree in information science from Gifu University, Japan, in 2005, and the M.S. and Ph.D. degrees in regeneration and advanced medical sciences from the Graduate School of Medicine, Gifu University, in 2007 and 2011, respectively. In 2010, he was an Assistant Professor with the Department of Information and Computer Engineering, Toyota National College of Technology. He was an Assistant Professor with the School of Information Science and Technology, Aichi Prefectural University, in 2015, where he was a Lecturer, in 2017. He was a Visiting Scholar with the Institute for Surgical Technology and Biomechanics, University of Bern, in 2017. He is currently an Associate Professor with the School of Information Science and Technology, Aichi Prefectural University.

XIANGRONG ZHOU received the M.S. and Ph.D. degrees in information engineering from Nagoya University, Japan, in 1997 and 2000, respectively. From 2000 to 2002, he continued his research in medical image processing as a Postdoctoral Researcher with Gifu University, Japan, where he is currently an Associate Professor with the Department of Electrical, Electronic and Computer Engineering. His current research interests include medical image analysis, medical image visualization, and pattern recognition.

HIROKI KATO received the degree from the School of Medicine, Gifu University, Japan, in 1999, and the Ph.D. degree from the Department of Radiology, Gifu University, in 2007. In 1999 and 2000, he was a Resident with the Department of Radiology, Gifu University Hospital, and Jichi Medical University Hospital. In 2002 and 2004, he was a Clinical Fellow and an Assistant Professor with the Department of Radiology, Gifu University Hospital. He is currently an Associate Professor with the Department of Radiology, Graduate School of Medicine, Gifu University. His current research interests include diagnostic imaging using computed tomography and magnetic resonance imaging. His specialized fields include head, neck, genitourinary, and musculoskeletal imaging.

TAKESHI HARA received the B.S. and M.S. degrees from the Faculty of Engineering, Gifu University, Japan, in 1994 and 1995, respectively, and the Ph.D. degree from Gifu University, in 2000. In 1995, he was a Research Assistant with the Faculty of Engineering, Gifu University, where he was an Associate Professor, in 2001 and 2017. He was an Associate Professor with the Department of Intelligent Image Information, Gifu University. From 2008 to 2009, he was a Visiting Associate Professor with the Department of Radiology, University of Chicago. Since 2019, he has been a Professor with the Faculty of Engineering and the Chair of the Artificial Intelligence Research Promotion Center, Medical Research Department, Gifu University.

HIROSHI FUJITA received the B.S. and M.S. degrees in electrical engineering from Gifu University, Japan, in 1976 and 1978, respectively, and the Ph.D. degree from Nagoya University, in 1983. In 1978, he was a Research Associate with the Gifu National College of Technology, where he was an Associate Professor, in 1986. From 1983 to 1986, he was a Visiting Researcher with the K. Rossmann Radiological Image Laboratory, University of Chicago. He was an Associate Professor with the Faculty of Engineering, Gifu University, in 1991, where he was a Professor, in 1995. Since 2002, he has been a Professor and the Chair of intelligent image information with the Graduate School of Medicine, Gifu University. He is currently a Research Professor with Gifu University and a Visiting Professor with Zhengzhou University, China.