RESEARCH   **Open Access**

Development of an artificial intelligence-based application for the diagnosis of sarcopenia: a retrospective cohort study using the health examination dataset

Chang-Won Jeong[1,2], Dong-Wook Lim[1], Si-Hyeong Noh[1], Sung Hyun Lee[3] and Chul Park[4*]

Abstract

Background Medical imaging techniques for diagnosing sarcopenia have been extensively investigated. Studies have proposed using the T-score and patient information as key diagnostic factors. However, these techniques have either been time-consuming or have required separate calculation processes after collecting each parameter. To address this gap, we propose an artificial intelligence (AI)-based web application that automates the collection of data, classification of the lumbar spine 3 (L3) slices, segmentation of the subcutaneous fat, visceral fat, and muscle areas in the classified L3 slices, and quantitative analysis of the segmented areas.

Methods We developed an automated lumbar spine slice classification model using the CNN (EfficientNetV2) algorithm and an automated domain segmentation model to identify the subcutaneous fat, visceral fat, and muscle areas using the U-NET algorithm. These models were used to identify L3 slices from abdominal computed tomography images and divide the images into the three-segmented domains for sarcopenia diagnosis. Additionally, we developed an algorithm for the calculation of T-Score calculated as (measurement value-Young adult mean)/ (Young adult SD) using the Aggregation Pipeline by MongoDB, with the mean and standard deviation for skeletal muscle area (SMA), SMA/height^2, SMA/weight, and SMA/body mass index (BMI) for both sexes and different age groups.

Results The proposed system demonstrated high accuracy and precision, with an overall accuracy of 97.5% in classifying L3 slices and a segmentation accuracy of 92% for muscle, subcutaneous fat, and visceral fat areas. The T-Score-based analysis provided reliable diagnostic thresholds for sarcopenia, facilitating consistent and accurate assessments. Our diagnostic cutoff points for each index were as follows: SMA (-1.0: 152.55, -2.0: 125.89), SMA/height^2 (-1.0: 38.84, -2.0: 14.50), SMA/weight (-1.0: 2.14, -2.0: 1.89), and SMA/BMI (-1.0: 6.10, -2.0: 5.18) for men; SMA (-1.0: 96.08, -2.0: 76.96), SMA/height^2 (-1.0: 37.20, -2.0: 29.36), SMA/weight (-1.0: 1.80, -2.0: 1.61), and SMA/BMI (-1.0: 4.56, -2.0: 4.01) for women. SMA/BMI best reflected the loss of muscle mass in healthy populations by age, showing a more remarkable decrease in muscle mass in men than in women. The values for men gradually decreased after their 20s,

*Correspondence:
Chul Park
cholssak21@gmail.com
Full list of author information is available at the end of the article

and that for women gradually decreased after their 40s, which progressed to a more dramatic decline in the 70s for both sexes.

Conclusion This AI-based web application addresses the limitations of previous diagnostic techniques by automatically analyzing medical images for the classification, segmentation, and calculation of T-scores. The study findings provide a more reliable and accurate diagnostic technique for sarcopenia that can consequently impact patient treatment and outcomes.

Keywords Sarcopenia, Artificial Intelligence, Tomography, Reference values, Muscle

Background

Sarcopenia can be diagnosed by assessing physical performance, muscle mass, and muscle strength. These diagnostic criteria were first established by the European Working Group on Sarcopenia in Older People in 2010. Subsequently, the Asian Working Group for Sarcopenia was established in 2014, and diagnostic standards specific to Asia were proposed [1–3]. Sarcopenia was assigned a disease code in the United States (ICD-10-CM) in 2016 and in South Korea (KCD-8) in 2021, leading to increased research interest. Hospitals frequently use computed tomography (CT) images for the diagnosis and treatment of this condition and utilize these images for retrospective studies.

Among the several modalities for measuring muscle mass, such as dual-energy X-ray (DXA) and ultrasound, CT has been utilized to quantify muscle mass in areas around various spinal levels. However, CT and muscle mass measurements are time-consuming and resource-intensive. Automated tools, specifically muscle mass quantification tools that utilize artificial intelligence (AI), have the potential to simplify the diagnosis of sarcopenia, making it more accessible and consistent in various clinical settings [4–6]. However, the complexity of the task is evident in the challenge of creating a system that can automate this process and improve diagnostic accuracy by considering the variability in muscle composition across various spine levels.

Despite advances in medical imaging and the recognition of sarcopenia as a critical health issue, a more accurate and efficient diagnosis of this disease is still needed. One of the major issues is that muscle mass can be underestimated or overestimated owing to inconsistencies in diagnostic criteria, especially the reliance on single-slice analysis of CT scans [7]. The lack of standardized cutoff values for different populations and variability in muscle composition across different slices of the lumbar spine further add to this discrepancy. Additionally, despite the potential for diagnostic process automation offered by AI-based technologies, difficulties in guaranteeing accuracy, dependability, and generalizability across various patient populations limit the use of these techniques in clinical practice [8, 9]. To solve these ambiguous problems and enhance patient outcomes, a more thorough and standardized sarcopenia diagnostic strategy that can fully utilize AI is required.

We developed an application that segments and quantitatively analyzes the lumbar spine 3 (L3) region into three areas (subcutaneous fat, visceral fat, and muscle) using abdominal CT images. However, this previous application was not as efficient, as it required physicians to manually input data from the identified L3 region. For AI research, diverse case inputs are necessary for effective AI learning, and by the same token, our study also required the collection of various patient cases to be entered for learning [10]. However, the manual classification of the L3 region in all abdominal CT images is time-consuming and challenging. Therefore, we propose an application that automatically identifies the L3 region in abdominal CT images and segments it into three specific domains (subcutaneous fat, visceral fat, and muscle). This application will also quantitatively analyze each domain and calculate the T-score for diagnosing sarcopenia using data from individuals aged 20–49 years as well as determine the prevalence of sarcopenia using the entire dataset.

Methods

Study population

This study was conducted in accordance with the protocol approved by the Institutional Review Board (IRB) of Wonkang University Hospital (IRB no. WKUH 2023-05-030) and in compliance with Good Clinical Practice. Informed consent was waived because of the observational nature of the study. Before being analyzed, all patient records and data were de-identified and anonymized. Healthy patients who underwent abdominal CT at our hospital's health examination center were recruited, and abdominal muscle mass was measured from the CT findings using an automated AI analysis to predict sarcopenia. The data were collected from 765 patients who visited our health examination center from January 1, 2013 to December 31, 2022. This method was chosen to create a representative cohort of healthy adults, which is essential for developing and validating the AI-based diagnostic tool. Specifically, data were obtained from 421 men and 344 women aged 20–79 years (seven age groups: 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, ≥ 80) who do not

meet the exclusion criteria. Patients who were underweight with a body mass index (BMI) < 18.5 kg/m^2 and patients with a history of diabetes mellitus, kidney disease, or trauma, as reported in the patient questionnaire, were excluded.

Dataset

The abdominal CT scans of the 765 patients were classified. However, due to image boundary uncertainties caused by conditions such as scoliosis and disc compression in patients in their 60s and 70s, we excluded 78 cases, resulting in a final dataset of 687 patients for model development. The classified imaging data were then labeled and divided into training and testing sets in an 8:2 ratio. Supplementary 1 provides detailed information about the distribution of CT images across different spinal segments for both sets. Additionally, we included information about the segmentation model training and testing sets for the L3 region, as shown in Supplementary 2. Initially, abdominal images were categorized into thoracic, lumbar (L1, L2, L3, L4, and L5), and sacral sections for training. Subsequently, imaging data corresponding to the L3 region were separately compiled to train the model to segment muscle mass, subcutaneous fat, and visceral fat.

### AI-based web system for the diagnosis of Sarcopenia

The AI model for sarcopenia diagnosis consisted of two main components: the L3 classification model and the segmentation model for the L3 region. Identifying L3 from abdominal CT scans and locating image data for the corresponding domain is a time-consuming process [11]. To address this issue, we developed an algorithm for automatic segmentation of muscle mass, subcutaneous fat, and visceral fat at the identified L3 level.

The AI models for classification and segmentation were developed and validated using four multicenter validation datasets. The results included T-score calculations and prevalence predictions generated automatically by the model. Finally, a user-friendly web-based platform was developed for end users.

Development of the AI model for the diagnosis of Sarcopenia

An automatic lumbar classification model was developed using EfficientNetV2. Figure 1 shows the overall structure of the lumbar classification model. The model was trained using a dataset comprising seven classes: thoracic, L1, L2, L3, L4, L5, and sacral. To compensate for the limited size of the dataset, augmentation techniques such as rotation, zooming, vertical flipping, and horizontal flipping were applied. The model was trained to classify images using transfer learning by identifying the features of each class using EfficientNetV2 in ImageNet. Supplementary 3 lists the hyperparameters used for training. Although larger datasets typically slow training, the EfficientNetV2 model was designed for faster learning. Compared to EfficientNetV1, it achieves a similar accuracy with a learning
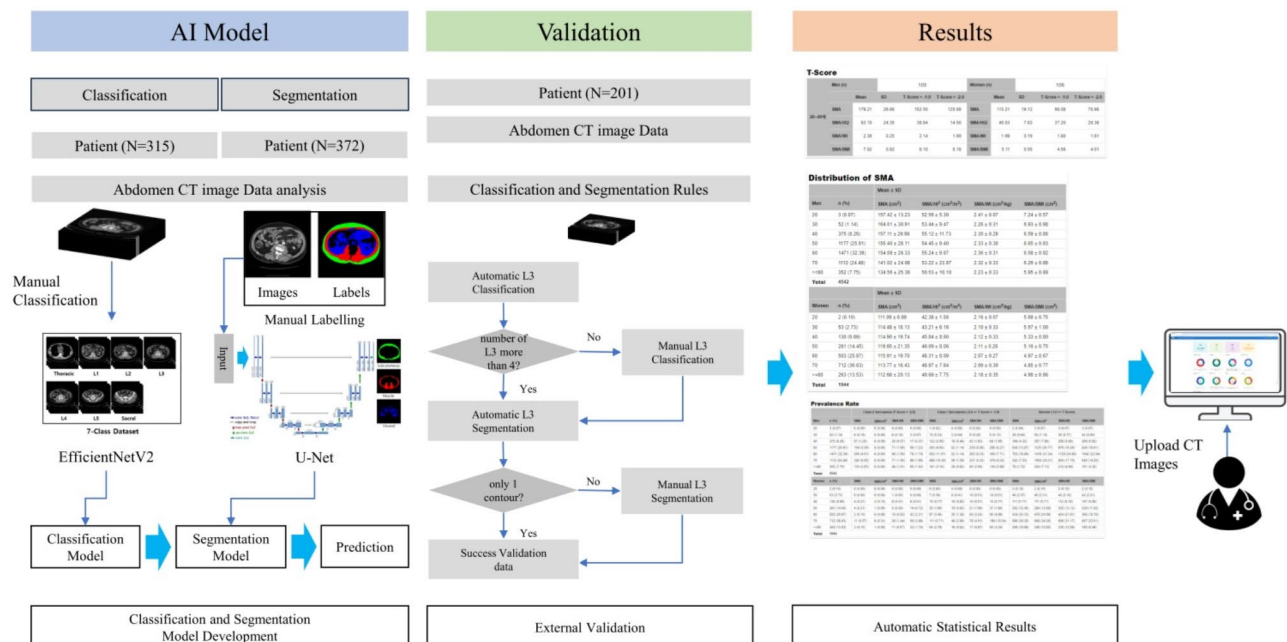
Fig. 1 Study design and study workflow. The AI model used abdominal CT to develop a classification and segmentation model, and a prediction model was developed based on this. The validation was performed by building a dataset not used for learning. Finally, the quantitatively analyzed results were automatically calculated statistically, and the T-Score, age-specific SMA analysis results, and prevalence rates were developed into a user-friendly web-based platform

speed of four times faster and 6.8 times fewer parameters. The model was initially trained using an original CT image of size 512 × 512 pixels. EfficientNetV2 tends to slow down with larger image sizes; however, owing to the limited amount of training data, the original size was used for training. We plan to reduce the image size for training to improve the accuracy once a sufficient dataset is available. The training results of the developed automatic lumbar classification model were as follows: val_loss = 0.1222 and val_accuracy = 0.9757. Historical plots are shown in Supplementary 4. After adjusting the settings for the best model, the model at the 23rd epoch was saved. The performance of the automatic lumbar classification model was evaluated using the test data, and the confusion matrix is presented in Supplementary 5. A well-performing confusion matrix is indicated by higher numbers along the diagonal from top left to bottom right; the matrix shows that the model performed well at the L3 level. As illustrated in Supplementary 5, a few slices along the diagonal were misclassified as adjacent classes; however, misclassifications as distant slices were rare. This indicates that the proposed classification model exhibits excellent performance. Although we obtained the desired results using the classification model, it was equally important to analyze the features extracted by AI to make predictions [12, 13]. Identifying these key influencing factors enhances the predictive reliability of the model and allows for a better interpretation of the underlying reasons for its predictions. Grad-CAM is frequently used for this purpose. Supplementary 6 shows Grad-CAM for the L3 slice, which is significant in diagnosing sarcopenia. Typically, the characteristics of the L3 are determined by its shape. Although each lumbar vertebra varied slightly, it was determined that the model made predictions based on the shape.

Segmentation model for skeletal muscle measurement

We developed an AI image segmentation model to measure muscle mass using U-Net. Supplementary 7 shows the overall structure of the U-Net model. The model was trained using abdominal CT and labeled data. We previously published a study on the development process and data composition for the development of an AI model for sarcopenia [14]. In our previous study, the segmentation model was trained and tested using 100 and 50 datasets. However, this was insufficient for adequate training; therefore, we modified the image windowing and augmented the data to establish a dataset of 20,000 images. We used 18,000 data points for training and 2,000 for validation. However, as shown in Supplementary 8, unsatisfactory segmentation outcomes were observed in the images after applying different windowing values and data augmentation. There were instances in which the modified windowing led to the segmentation of other

areas with similar windowing, as indicated by the yellow-boxed areas. To address these problems, we obtained actual training data and labeled them. We only used a dataset of 1,480 images for training and 370 for testing without applying windowing or data augmentation modifications. Supplementary 9 details the hyperparameters used for training. A notable feature is the loss function. Although dice loss is commonly used in segmentation models, using it alone often leads to segmentation errors, diminishing the model's performance [15, 16]. Thus, we used combined loss consisting of the dice loss and cross-entropy loss at a 1:1 ratio. The rationale for using dice loss is to counter the class imbalances frequently observed in CT data segmentation, where an extensive black background can overshadow the smaller area representing the human body. This imbalance could inadvertently skew the training process for the less-represented class. Integrating dice loss and cross-entropy loss helps correct class imbalances with dice loss and maintains pixel accuracy with cross-entropy loss. The training results of the developed segmentation model were as follows: val_loss = 0.1730 and val_accuracy = 0.9713. Supplementary 10 shows historical plots. The performance of the segmentation model was evaluated using the test dataset, and the intersection over union (IOU) values are shown in Supplementary 11. Because the three segmented areas are important markers for the diagnosis of sarcopenia, we calculated the IOU by dividing them into subcutaneous fat (S), visceral fat (V), and muscle (M), as well as their combined area (M + S + V). The IOU calculations revealed that 92% of the entire dataset had an accuracy of greater than 90%. We developed a segmentation model by training it with abdominal CT data and validated the model using separately collected test data. However, the 370 cases used for testing were not sufficient to enable the application of the AI segmentation model to real-world data. For the validation, we obtained data from four facilities and identified the L3 region from the collected data [17].

Quantitative analysis

We conducted an analysis using the abdominal disease dataset according to the method proposed for the diagnosis of sarcopenia using the proposed system. This involved labeling 1,161 abdominal images and conducting quantitative analyses to measure muscle mass, subcutaneous fat, and visceral fat. Based on these measurements, we calculated the cutoff values of T-score = -0.1 and − 0.2 for a healthy young adult population. Patients showing signs of sarcopenia were identified based on their T-scores. The error range was nearly identical to that of diagnostic markers in a previous study [8]. This error could potentially be attributed to the differences in the number of subjects. To calculate the T-scores, we used

a formula comparing the measured skeletal muscle area (SMA) or skeletal muscle index (SMI) of an individual with the average SMA or SMI of a healthy young population adjusted for height, weight, and BMI. The difference was then divided by the sex-specific standard deviation (SD) of the young adults. The statistical analysis reference values underlying the T-score calculation in our study were based on the reference data presented in the previous study [8]. To ensure accurate and reproducible diagnostic results, the *p*-values and confidence interval values analyzed in the previous study were utilized in the development of an automated AI system conducted in our study.

Results

L3 classification results

The sarcopenia labeling tool developed in this study restructures axial images into sagittal images, allows for manual selection of L3 slices, and enables automatic extraction of the lumbar region using the lumbar classification model. Figure 2 shows the prediction results for the thoracic, L1, L2, L3, L4, L5, and sacral regions for all abdominal CT scan data.

Segmentation results

Figure 3 shows how the sarcopenia AI segmentation model automatically identifies subcutaneous fat, visceral fat, and muscle and produces labeled data. If the segmentation is performed erroneously, a labeling tool can be used to revise the label. The labeled data were then quantitatively analyzed, as shown in Fig. 3. After the labeling is completed, a quantitative analysis of the entire labeled data can be performed by clicking on the quantitative analysis tab. The surface area, mean, and SD are presented for the labeled areas (muscle, subcutaneous fat,

and visceral fat). Additionally, the proportions of muscle, visceral fat, and subcutaneous fat were calculated. Height and body weight were first acquired from medical imaging tags; however, if unavailable, they could be manually entered to automatically calculate BMI. Quantitative data for each domain were used as markers to diagnose sarcopenia. Data were downloaded from Excel for statistical analysis.

Web application for the diagnosis of Sarcopenia

Figure 4 shows the Study List screen of the AI-based web application. Abdominal CT images can be uploaded with the de-identification of personal data and are shown on the screen in the study units. If there is linked patient information, the connection status can be easily verified through an icon in the "Person" row, as shown in Fig. 4. By clicking the "Lumbar" button at the top, the app automatically performs the tasks of classifying the L3 slice from the abdominal CT images uploaded from the server, segmenting the three areas, and calculating the T-score. The "Process" row indicates the progress of these automated tasks. Once all tasks are completed, they are marked as "Analyzed." If an error occurs at any stage, an "error" message pops up, allowing the user to manually identify and resolve the issue. The "Result" button appears at the top-left corner once all workflow tasks are completed for the images. Clicking the "Result" button generates a report with the mean and SD, calculated T-scores, and prevalence rates for each age group and gender. Supplementary 12 shows the SMA, SMA/height^2, SMA/weight, and SMA/BMI, as well as the SD of each factor for men and women aged 20–49 years. With reference to a T-score of -2.0, the SMA/BMI is 5.18 in men and 4.01 in women. Supplementary 13 shows the sarcopenia prevalence rates by age group. Sarcopenia
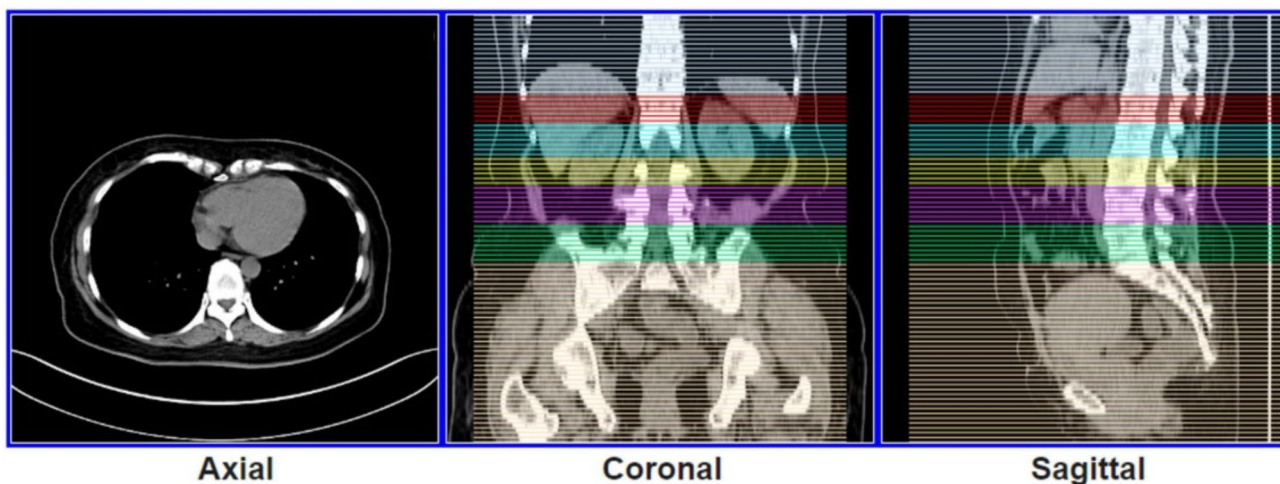
Fig. 2 Visually shows the results of spine-based classification of the uploaded image. Here, muscle mass, subcutaneous fat, and visceral fat are classified and quantified in each image of the nine slices at the L3 level
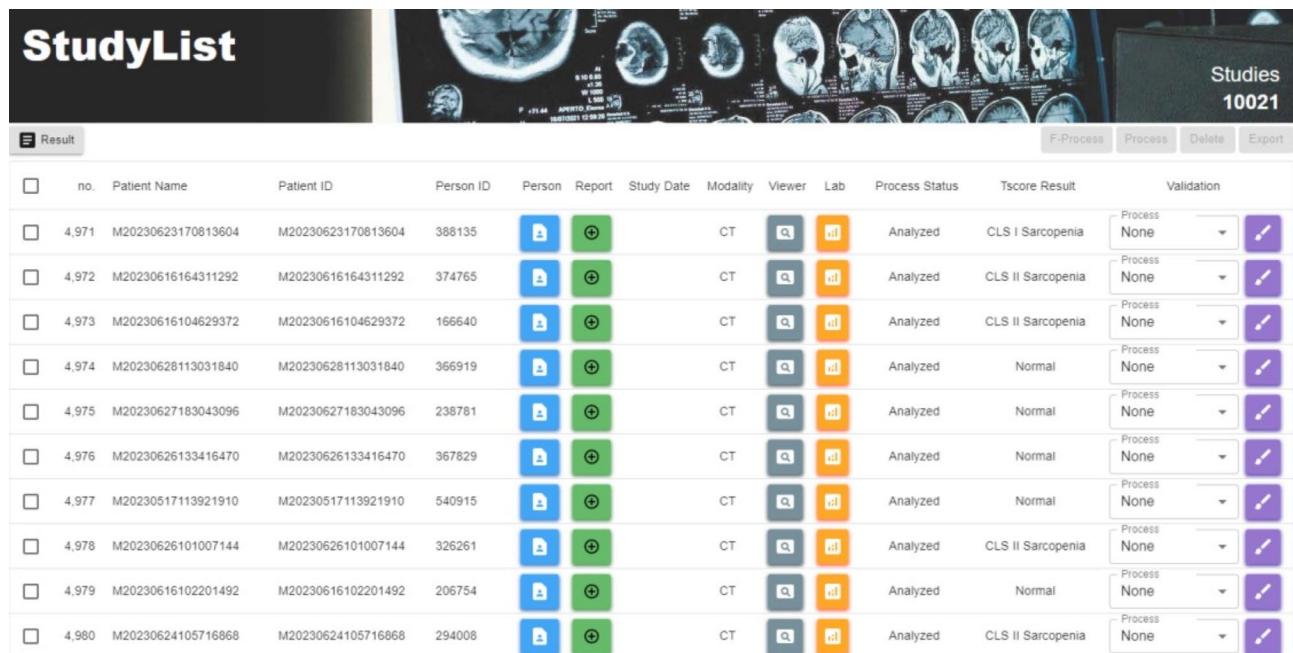
Fig. 3 The image shows the quantified results from the L3 region. It shows the quantified numerical values of muscle mass, subcutaneous fat, visceral fat, age, weight, height, and BMI

prevalence for the collected patient data was determined by defining a T-score of < -2.0 as class II sarcopenia, -2.0 to -0.99 as class I sarcopenia, and ≥ -1.0 as normal. In men, SMA increases until the 40s and subsequently declines. In women, SMA increases until the 30s and subsequently declines. The SMA/height^2 increased until the 40s and subsequently declined in both men and women. The SMA/height ratio decreases with advancing age in women. In men, it increases slightly in their 60s and decreases gradually after that. SMA/BMI peaks in the 20s in both men and women and declines subsequently. The prevalence dramatically spiked in men and women aged ≥ 60 years, and the BMI-adjusted results were particularly significant. The prevalence of class I and class II sarcopenia ranged widely from 14.4 to 24.8%

and 2.3–5.2%, respectively, in men and 13.8–27.2% and 1.0–8.7%, respectively, in women. When the cut-off for sarcopenia was set at a T-score of < -2.0, SMA/BMI led to the highest sarcopenia prevalence (4.2% in men and 8.7% in women), and SMA/height^2 led to the lowest sarcopenia prevalence (2.8% in men and 1.0% in women). Sarcopenia prevalence increased with advancing age when predicted by the SMA and all three indicators, except for those aged 20–29 and 30–39 years, which had small sample sizes.

Discussion

This study presents the validation and implementation of an AI-based web application designed to diagnose sarcopenia using abdominal CT scans. This application

Fig. 4 Study Lists. The Study list manages abdominal CT images uploaded for analysis. Images that are automatically analyzed are displayed as analyzed in the process status. The TscoreResult shows the results of each uploaded image, whether sarcopenia or normal

automated the classification of vertebral regions, muscles, subcutaneous fat, visceral fat, and bone segmentation in CT images. The T-scores were calculated based on the SMA index. We found that among several indices, SMA/BMI best reflected the age-related loss of muscle mass in healthy populations. SMA/BMI peaks in the 20s in both men and women and declines subsequently. In addition, the AI-based system is distinct from previous approaches that usually deal with only a single slice. It can automatically analyze each slice at the L3 level and determine the average value. This method guarantees more consistent and reproducible results, especially in clinical settings where an accurate diagnosis can significantly impact patient treatment and outcomes.

The Foundation for the National Institutes of Health (FNIH) reported that appendicular skeletal muscle mass (ASM)/BMI correlates with muscle weakness and slowness [18]. In line with this, confirmation of sarcopenia prevalence in the Korea National Health and Nutrition Examination Survey (KNHANES) dataset was defined at less than 2 SD of DXA and showed a tendency to underestimate the prevalence of sarcopenia in women [19]. In contrast, we showed the difference in outcome values when DXA was height- or weight-adjusted. In addition, there were differences in muscle mass according to age group. We identified a tendency for the SMI to increase in men and women in their 30s and 40s. SMA/weight and SMA/BMI tended to peak in men and women in their 20s and decrease in their 70s. SMA/weight and SMA/BMI reflected the age-related muscle loss pattern

better than SMI in that the muscle mass peaks in the 20s and decreases with age. In the KNHANES study, ASM/height^2 peaked in men in their 30s and 40s in women, whereas ASM/weight gradually showed a downward trend and peaked in their 20s for both men and women; therefore, it would be reasonable to adjust the weight [19, 20]. Based on these results, we also used BMI as a correction factor to reflect weight.

The results of this study have important implications in clinical practice, especially for improving the practicality and effectiveness of muscle mass assessment. By confirming that the SMA/BMI ratio is the most accurate measure for altering muscle mass with age, this study contributes to our understanding of muscle mass loss. According to earlier research, this is significant because there is a clear correlation between muscle mass and the diagnosis, treatment, and prognosis of sarcopenia. Clinicians can implement prognostic assessments and nutritional support measures to improve clinical outcomes and quality of life during the early detection of sarcopenia [6, 15, 21, 22]. In addition, the time and effort involved in manually analyzing images are decreased by developing AI-based solutions that automate the diagnosis procedure. This reduces the risk of human error, resulting in a more consistent and reliable evaluation. These advances can facilitate faster decision-making in a clinical setting when time and resources are limited. Ultimately, they may be essential for enhancing patient care. These findings are especially pertinent in the context of an aging population where sarcopenia is becoming more common.

Recognizing and managing sarcopenia can lessen the potential for hospital stays, falls, mobility impairments, and patient outcomes.

Another finding of this study was the capability of the AI-based system to perform comprehensive analysis across all L3 slices in abdominal CT scans rather than relying on a single slice, as done in previous studies [23]. Several muscles are distributed, including the psoas, erector spinae, quadratus lumborum, abdominal wall muscles (transversus abdominis, external and internal obliques, and rectus abdominis), and many tissues that are distinguished from muscles, such as visceral fat and subcutaneous fat, are at the L3 level, which has been used in various studies as a landmark for SMA measurements. The fact that fat-free muscles should be used to measure muscle volume or cross-sectional area using CT is also one of the reasons for this [6, 8, 11, 24, 25]. Whether the L3 level of the psoas muscle represents total body muscle mass has not yet been established. Our research is meaningful because we have developed an application that can easily accomplish this goal. This method has the potential to significantly affect clinical practice by providing a more accurate and holistic representation of the patient's muscle mass, leading to more reliable diagnostic outcomes. Additional research is needed on standard or representative survey markers for the spinal level and specific muscles that can define sarcopenia and clinical outcome correlation for each type of disease. The traditional method of analyzing only one slice may overlook the variability in muscle composition across different slices, potentially leading to an underestimation or overestimation of sarcopenia prevalence. The ability of the AI system to analyze multiple slices and average the values addresses this limitation by offering a more consistent and representative measure of SMA. This method reduces the risk of misclassification owing to anatomical variations or technical inconsistencies in image acquisition. By ensuring that the entire L3 region is evaluated, the AI system enhances the accuracy of sarcopenia diagnosis, providing reassurance about the reliability of the diagnostic outcomes.

This study addressed some of the limitations of previous studies. For instance, Kim et al. provided reference data for the SMA measured using a CT scan based on a single L3 slice analysis [8]. Although this method is commonly used for measuring muscle mass, it risks underestimating or overestimating muscle mass owing to the anatomical variability within the L3 region. In contrast, our study overcame these limitations by analyzing all available slices at the L3 level and evaluating muscle mass via volume measurements rather than cross-sectional measurements. Similarly, Ha et al. developed a deep learning system for automatic L3 selection and body composition evaluation, which was a significant advancement in automating the sarcopenia diagnosis process [11]. However, their work primarily focused on developing and verifying deep learning models for slice selection rather than a detailed quantitative analysis of multiple slices. Our study integrates these advancements with a comprehensive quantitative analysis of muscle, subcutaneous fat, and visceral fat for all slices and automates slice selection. By automating both the L3 classification and segmentation processes and applying them across the entire L3 domain, our study introduces a more reproducible sarcopenia diagnostic method that can be used in various patient populations. We have introduced our model's features, including the comprehensive analysis of all L3 slices, integration of automated processes, volume-based muscle mass evaluation, and a user-friendly web platform. These features collectively contribute to a more practical, efficient, and clinically applicable diagnostic tool compared to other AI methods in sarcopenia diagnosis.

This study had some limitations. First, the study population was selected from a health examination center in a single institution, where there may be selection bias, and the study population needs to be representative of the general population, which can limit the generalizability of the study results. A cohort consisting primarily of individuals undergoing routine health examinations may only partially represent the diversity of the general population, particularly with respect to age, ethnicity, and underlying health conditions. We are planning a follow-up study that integrates as much data as possible from multiple centers. It aims to reduce the impact of selection bias by increasing the generalizability of the results across diverse populations. Despite these limitations, the dataset used in this study was an attempt to provide a starting point for evaluating the effectiveness of AI-based sarcopenia diagnosis. By demonstrating the model's applicability within a controlled, healthy population, we sought to lay the groundwork for building a more comprehensive study in the future. Second, the sample size of the reference group used to calculate the T-score, especially for the study population aged 20–49 years old, was relatively small. This may have affected the accuracy of the T-score threshold and the subsequent estimates of sarcopenia prevalence. Larger and more diverse datasets are required to improve the diagnostic accuracy of AI-based systems. To overcome the limitations of the relatively small reference groups for establishing T-score references, we plan to work with other organizations to expand our dataset to include various cohorts. In addition, continuous data collection and integration into AI-based systems can update the T-score calculations in real time. This will enhance the diagnostic accuracy of our AI-based system and ensure our findings are relevant to various populations. Third, although AI-based systems

automate the classification and segmentation of the L3 domain, they did not evaluate the clinical outcomes associated with low muscle mass, such as disability, weakness, or mortality. Evaluating these results is necessary to validate the clinical usefulness of diagnostic tools and understand their impact on patient care [26–28]. Longitudinal studies that track these results can assess the possibility of adopting this technique in clinical practice. This study can improve the tools for understanding the impact of sarcopenia diagnosis on long-term health outcomes and for better predicting and managing risks by tracking clinical outcomes associated with low muscle mass. Finally, although the system showed high accuracy in segmenting the muscle and adipose tissue, this study was limited to abdominal CT scans. The applicability to other examination tools, such as magnetic resonance imaging or DXA, is yet to be investigated. To address this study's focus on abdominal CT scans, we will use cross-modal verification to determine whether it can be effectively used in different imaging modalities and clinical settings. We have identified that integrating AI-based diagnostic tools into clinical workflows may pose logistical and technological challenges. These challenges include ensuring compatibility with various imaging systems and electronic health records (EHRs). Another challenge is the scalability of AI-based diagnosis and management, especially in resource-limited settings. Lastly, maintaining the model's relevance and accuracy over time is crucial, given the rapid evolution of AI technology and clinical guidelines. To address these issues, we plan to conduct future studies that incorporate strategies to reduce algorithmic bias and utilize multi-center datasets to ensure broader generalizability and improved model robustness.

In conclusion, an artificial intelligence web application for the diagnosis of sarcopenia was proposed. The proposed system automatically calculates the artificial intelligence module linkage, quantitative analysis, T-Score calculation, and prevalence by uploading abdominal CT medical images to classify and segment the lumbar 3 spine slices. Based on this, the prevalence by age was calculated. Our proposed system has the advantage of being automatically calculated with continuous data collection, and the diagnosis and prevalence of sarcopenia could be confirmed in patients in various disease groups with the same T-score-based cut-off point value. This result is predicted to offer basic information for understanding and comparing the clinical significance of sarcopenia and exploring its impact on patient outcomes. Our findings suggest that this AI-based tool could enhance the assessment and management of sarcopenia in clinical practice, providing a helpful resource for future research and patient care improvement.

Abbreviations
AI   Artificial intelligence
BMI   Body mass index
CT   Computerized tomography
DXA   Dual-energy X-ray
ICD-10-CM International Classification of Diseases, Tenth Revision, Clinical Modification
KCD-8   Korena Standard Classification of Diseases
KNHNES   Korea National Health and Nutrition Examination Survey
L3   Lumbar spine 3
SD   Standard deviation
SMA   Or skeletal muscle area
SMI   Skeletal muscle index

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12911-025-02900-4.

Supplementary Material 1

Declarations

Ethics approval and consent to participate
This study was conducted in accordance with the protocol approved by the Institutional Review Board (IRB) of Wonkang University Hospital (IRB no. WKUH 2023-05-030), in compliance with Good Clinical Practice, and according to relevant guidelines and regulations, including the Declaration of Helsinki. Informed consent was waived because of the study's observational nature under our center's IRB approval. Before being analyzed, all patient records and data were de-identified and anonymized.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]STSC Center, Wonkwang University, Iksan 54538, South Korea
[2]Smart Team, Wonkwang University Hospital, Iksan 54538, South Korea
[3]Department of Orthopedics, Wonkwang University Hospital, Iksan 54538, South Korea
[4]Division of Pulmonology and Critical Care Medicine, Department of Internal Medicine, Ulsan University Hospital, 25, Daehakbyeongwon-ro, Dong-gu, Ulsan 44033, South Korea

References

1. Chen LK, Liu LK, Woo J, Assantachai P, Auyeung TW, Bahyah KS, et al. Sarcopenia in Asia: Consensus report of the Asian Working Group for Sarcopenia. J Am Med Dir Assoc. 2014;15:95–101.
2. Chen LK, Woo J, Assantachai P, Auyeung TW, Chou MY, Iijima K, et al. Asian Working Group for Sarcopenia: 2019 Consensus update on Sarcopenia diagnosis and treatment. J Am Med Dir Assoc. 2020;21:300–e3072.
3. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, Boirie Y, Cederholm T, Landi F, et al. Sarcopenia: European consensus on definition and diagnosis: report of the European Working Group on Sarcopenia in Older people. Age Ageing. 2010;39:412–23.
4. Bedrikovetski S, Seow W, Kroon HM, Traeger L, Moore JW, Sammour T. Artificial intelligence for body composition and sarcopenia evaluation on computed tomography: a systematic review and meta-analysis. Eur J Radiol. 2022;149:110218.
5. Boutin RD, Yao L, Canter RJ, Lenchik L. Sarcopenia: current concepts and imaging implications. AJR Am J Roentgenol. 2015;205:W255–66.
6. Lortie J, Gage G, Rush B, Heymsfield SB, Szczykutowicz TP, Kuchnia AJ. The effect of computed tomography parameters on Sarcopenia and myosteatosis assessment: a scoping review. J Cachexia Sarcopenia Muscle. 2022;13:2807–19.
7. Fielding RA, Vellas B, Evans WJ, Bhasin S, Morley JE, Newman AB, et al. Sarcopenia: an undiagnosed condition in older adults. Current consensus definition: prevalence, etiology, and consequences. International working group on Sarcopenia. J Am Med Dir Assoc. 2011;12:249–56.
8. Kim EH, Kim KW, Shin Y, Lee J, Ko Y, Kim YJ, et al. Reference data and T-scores of lumbar skeletal muscle area and its skeletal muscle indices measured by CT scan in a healthy Korean population. J Gerontol Biol Sci Med Sci. 2021;76:265–71.
9. Kim S, Kim TH, Jeong CW, Lee C, Noh S, Kim JE, et al. Development of quantification software for evaluating body composition contents and its clinical application in sarcopenic obesity. Sci Rep. 2020;10:10452.
10. Liu R, Gao XY, Wang L. Network meta-analysis of the intervention effects of different exercise measures on Sarcopenia in cancer patients. BMC Public Health. 2024;24:1281.
11. Ha J, Park T, Kim HK, Shin Y, Ko Y, Kim DW, et al. Development of a fully automatic deep learning system for L3 selection and body composition assessment on computed tomography. Sci Rep. 2021;11:21656.
12. Ji D, Oh D, Hyun Y, Kwon OM, Park MJ. How to handle noisy labels for robust learning from uncertainty. Neural Netw. 2021;143:209–17.
13. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. J Big Data. 2019;6:27.
14. Myoung Kee K, Lee C-s, Lim D-w, Kim J-e, Yu Y-j, Kim T-h, et al. Development of cloud-based medical image labeling system and it's quantitative analysis of Sarcopenia. KTCCS. 2022;11:233–40.
15. Nowak S, Theis M, Wichtmann BD, Faron A, Froelich MF, Tollens F, et al. End-to-end automated body composition analyses with integrated quality control for opportunistic assessment of Sarcopenia in CT. Eur Radiol. 2022;32:3142–51.
16. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. Deep learn Med image anal multimodal learn Clin decis support (2017). 2017;2017:240-8.
17. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. 2018;106:249–59.
18. McLean RR, Shardell MD, Alley DE, Cawthon PM, Fragala MS, Harris TB, et al. Criteria for clinically relevant weakness and low lean mass and their longitudinal association with incident mobility impairment and mortality: the foundation for the National Institutes of Health (FNIH) Sarcopenia project. J Gerontol Biol Sci Med Sci. 2014;69:576–83.
19. Moon SS. Low skeletal muscle mass is associated with insulin resistance, diabetes, and metabolic syndrome in the Korean population: the Korea National Health and Nutrition Examination Survey (KNHANES) 2009–2010. Endocr J. 2014;61:61–70.
20. Yoon JL, Cho JJ, Park KM, Noh HM, Park YS. Diagnostic performance of body mass index using the Western Pacific Regional Office of World Health Organization reference standards for body fat percentage. J Korean Med Sci. 2015;30:162–6.
21. Amarasinghe KC, Lopes J, Beraldo J, Kiss N, Bucknell N, Everitt S, et al. A deep learning model to automate skeletal muscle area measurement on computed tomography images. Front Oncol. 2021;11:580806.
22. Zhu M, Zhang L, Wang L, Li D, Zhang J, Yi Z. Robust co-teaching learning with consistency-based noisy label correction for medical image classification. Int J Comput Assist Radiol Surg. 2023;18:675–83.
23. So R, Matsuo T, Sasai H, Eto M, Tsujimoto T, Saotome K, et al. Best single-slice measurement site for estimating visceral adipose tissue volume after weight loss in obese, Japanese men. Nutr Metab (Lond). 2012;9:56.
24. Couderc AL, Liuu E, Boudou-Rouquette P, Poisson J, Frelaut M, Montégut C et al. Pre-therapeutic sarcopenia among cancer patients: an up-to-date meta-analysis of prevalence and predictive value during cancer treatment. Nutrients. 2023;15.
25. Lan Q, Guan X, Lu S, Yuan W, Jiang Z, Lin H, et al. Radiomics in addition to computed tomography-based body composition nomogram may improve the prediction of postoperative complications in gastric cancer patients. Ann Nutr Metab. 2022;78:316–27.
26. Arango-Lopera VE, Arroyo P, Gutiérrez-Robledo LM, Pérez-Zepeda MU, Cesari M. Mortality as an adverse outcome of Sarcopenia. J Nutr Health Aging. 2013;17:259–62.
27. Vashi PG, Gorsuch K, Wan L, Hill D, Block C, Gupta D. Sarcopenia supersedes subjective global assessment as a predictor of survival in colorectal cancer. PLoS ONE. 2019;14:e0218761.
28. Xia L, Zhao R, Wan Q, Wu Y, Zhou Y, Wang Y, et al. Sarcopenia and adverse health-related outcomes: an umbrella review of meta-analyses of observational studies. Cancer Med. 2020;9:7964–78.

Publisher's note