



Article

Deep Learning-Based Muscle Segmentation and Quantification of Full-Leg Plain Radiograph for Sarcopenia Screening in Patients Undergoing Total Knee Arthroplasty

Doohyun Hwang^{1,2}, Sungho Ahn^{1,2}, Yong-Beom Park³ , Seong Hwan Kim³, Hyuk-Soo Han^{1,2},
Myung Chul Lee^{1,2} and Du Hyun Ro^{1,2,4,*}

¹ Department of Orthopedic Surgery, Seoul National University College of Medicine, Seoul 03080, Korea; dhhwang@snu.ac.kr (D.H.); tjdgh914758@snu.ac.kr (S.A.); oshawks7@snu.ac.kr (H.-S.H.); leemc@snu.ac.kr (M.C.L.)

² Department of Orthopedic Surgery, Seoul National University Hospital, Seoul 03080, Korea

³ Department of Orthopedic Surgery, Chung-Ang University Hospital, Seoul 06973, Korea; whybe1122@gmail.com (Y.-B.P.); ksh170177@nate.com (S.H.K.)

⁴ Connecteve Co., Ltd., Seoul 03080, Korea

* Correspondence: duhyunro@gmail.com; Tel.: +82-10-6312-4610

Abstract: Sarcopenia, an age-related loss of skeletal muscle mass and function, is correlated with adverse outcomes after some surgeries. Here, we present a deep-learning-based model for automatic muscle segmentation and quantification of full-leg plain radiographs. We illustrated the potential of the model to predict sarcopenia in patients undergoing total knee arthroplasty (TKA). A U-Net-based deep learning model for automatic muscle segmentation was developed, trained and validated on the plain radiographs of 227 healthy volunteers. The radiographs of 403 patients scheduled for primary TKA were reviewed to test the developed model and explore its potential to predict sarcopenia. The proposed deep learning model achieved mean IoU values of 0.959 (95% CI 0.959–0.960) and 0.926 (95% CI 0.920–0.931) in the training set and test set, respectively. The fivefold AUC value of the sarcopenia classification model was 0.988 (95% CI 0.986–0.989). Of seven key predictors included in the model, the predicted muscle volume (PMV) was the most important of these features in the decision process. In the preoperative clinical setting, wherein laboratory tests and radiographic imaging are available, the proposed deep-learning-based model can be used to screen for sarcopenia in patients with knee osteoarthritis undergoing TKA with high sarcopenia screening performance.

Keywords: sarcopenia; deep learning; screening; plain radiograph; segmentation; total knee arthroplasty; osteoarthritis



Citation: Hwang, D.; Ahn, S.; Park, Y.-B.; Kim, S.H.; Han, H.-S.; Lee, M.C.; Ro, D.H. Deep Learning-Based Muscle Segmentation and Quantification of Full-Leg Plain Radiograph for Sarcopenia Screening in Patients Undergoing Total Knee Arthroplasty. *J. Clin. Med.* **2022**, *11*, 3612. <https://doi.org/10.3390/jcm11133612>

Academic Editor: Johannes C. Reichert

Received: 5 June 2022

Accepted: 21 June 2022

Published: 22 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Total knee arthroplasty (TKA) is an effective treatment for end-stage knee osteoarthritis [1,2]. Despite its benefits for relieving pain and restoring mobility, TKA is characterised by several postoperative complications, including the need for blood transfusion, acute kidney injury, prosthetic joint infection and periprosthetic tibiofemoral fractures with varying incidence rates [3–6]. To prevent such adverse outcomes, risk factors for postoperative complications of TKA have been explored [7–9].

Sarcopenia, characterised as age-related loss of skeletal muscle mass and function, has attracted a great deal of interest because of its reported association with an increased likelihood of poor clinical outcomes, including falls, fractures, physical disability and mortality [10]. Recent studies have showed that sarcopenia independently predicts adverse outcomes of various surgical procedures, including TKA [11–14]. Therefore, there is increasing recognition of the importance of preoperative risk stratification by screening orthopaedic surgery patients for sarcopenia [15–18].

Several tools are available for sarcopenia screening. Calf circumference (CC) is an inexpensive and readily available screening method. There is accumulating evidence that it correlates with both muscle mass and sarcopenia [19–21]. However, as the European Working Group on Sarcopenia in Older People (EWGSOP2) does not recommend CC as a screening tool for sarcopenia due to age-related changes in fat deposits and variable skin elasticity [10], its performance has mainly been studied in Asian populations [20,22,23]. The EWGSOP2 and Asian Working Group for Sarcopenia (AWGS) recommended the SARC-F questionnaire as a screening tool in primary healthcare settings [10,24]. However, several studies have reported low-to-moderate sensitivity of this tool to detect sarcopenia, such that a substantial number of potential patients likely go unrecognised [25]. There have been numerous efforts to boost its sensitivity for screening purposes by lowering the cutoffs, adding extra items and combining it with other examinations, but there is as yet no consensus regarding the best tool for screening sarcopenia [26].

As both the EWGSOP2 and AWGS suggest muscle quantity or mass as the confirmatory diagnostic criterion for sarcopenia, a radiological assessment of muscle quantity may represent an alternative method for screening sarcopenia, directly reflecting muscle loss in patients. A radiological assessment by magnetic resonance imaging (MRI) and computed tomography (CT) is considered the gold standard for the noninvasive assessment of muscle mass due to the high accuracy of these modalities. The abdominal skeletal muscle area at the third lumbar level of a single cross-sectional CT image is reported to have correlation with the whole-body skeletal muscle mass, and thus is being deployed to assess sarcopenia [27]. However, the need for time-consuming manual segmentation and potential interobserver variability are major barriers to clinical application [28,29]. In addition, imaging modalities such as CT and MRI are not always available for orthopaedic surgeons in the preoperative setting, hindering their application for sarcopenia screening. Therefore, there is a need for an alternative and more accessible tool to assess muscle quantity in the field of orthopaedics [30].

Here, we propose a novel method of screening for sarcopenia in patients undergoing orthopaedic surgery involving the lower extremities in the preoperative setting. We adopted a convolutional neural network (CNN)-based model for automatic muscle segmentation on full-leg weight-bearing plain radiographs. CNNs are the current state-of-the-art artificial intelligence technique for medical image classification and segmentation. The ability to screen for sarcopenia on full-leg plain radiographs prior to surgical procedures will aid orthopaedic surgeons in terms of risk stratification and patient selection.

The main objective of this study was to present a deep-learning-based muscle segmentation and quantification model using full-leg plain radiographs and illustrate its potential to predict sarcopenia in patients undergoing TKA, which was then confirmed by a bioelectrical impedance analysis (BIA).

2. Materials and Methods

2.1. Study Subjects

This retrospective single-centre study was conducted on two separate cohorts after obtaining institutional review board approval (IRB no. H-2009-181-1161).

For training and validation of the model, healthy volunteers with no history of trauma or prior orthopaedic surgery, enrolled between January 2011 and November 2012, were reviewed (cohort A). Cohort A consisted of two subgroups: a young group consisting of 128 young adults aged 19–35 years (79 females, 61.7%) and an older group consisting of 99 patients aged 60–69 years (51 females, 51.5%). A total of 227 full-leg lower extremity plain radiographs were used as the training set. The muscle segmentation model presented in this study represents a fully automated deep learning system, which was developed, trained and validated using cohort A data.

Patients scheduled for primary TKA to treat degenerative knee arthritis, enrolled between May 2018 and April 2021, were reviewed to test the developed model and explore its potential for predicting sarcopenia (cohort B). Subjects with adequate preoperative

BIA, determined using the InBody S10 device (InBody Co. Ltd., Seoul, Korea), were included in cohort B. A total of 633 patients were initially enrolled. We excluded patients who underwent simultaneous bilateral TKA ($n = 5$), or had inadequate tissue hydration (extracellular water (ECW) ratio > 0.4 ; $n = 175$), no available preoperative plain radiographs (within 8 weeks before scheduled surgery; $n = 4$), low-quality full-leg plain radiographs ($n = 45$) or severe obesity (body mass index (BMI) > 35 kg/m²; $n = 1$) [31–33]. A total of 403 patients (54 males, 349 females) were included in the test set. Data for cohort B were used to evaluate the performance of the developed muscle segmentation model and estimate the muscle quantity of the patients. The machine-learning-based sarcopenia prediction model presented in this study was developed, trained and validated using cohort B data.

2.2. Data Acquisition

All data were collected from the electronic medical records of our institution. Patient baseline characteristics, laboratory data and BIA data were collected. Sarcopenia was determined according to the cutoffs for appendicular skeletal muscle index (SMI) suggested by the AWGS 2019 (<7.0 kg/m² for males, <5.7 kg/m² for females) [33]. SMI is defined as the height-adjusted appendicular skeletal muscle mass and was calculated automatically by the InBody S10 device (InBody Co. Ltd., Seoul, Korea). The baseline characteristics, laboratory data and SMI were compared between the sarcopenic and normal groups (Tables S1 and S2).

Full-leg weight-bearing plain radiographs were acquired using a consistent scanning protocol. The subjects stood barefoot with the feet together in the “stand at attention” position, with the patellae oriented forward. The acquired radiographic images were collected using the picture archiving and communication system (PACS) of our institution.

One of the authors (SA) trained in the segmentation task prepared 227 manually annotated images using Adobe Photoshop (Adobe, Mountain View, CA, USA) using a standardised protocol, to ensure consistency of the data. Muscles of the bilateral thighs and calves were annotated. The upper border of the thigh was determined as the greater trochanter of the femur (lateral) and gluteal fold (medial), and the lower border of the thigh was determined as the lateral epicondyle of the femur (lateral) and medial epicondyle of the femur (medial). The visible parts of the soleus and gastrocnemius muscle on coronal plane radiographs were annotated as the bilateral calf regions. Each ground truth segmentation was reviewed and revised by an orthopaedic surgeon with 14 years of experience (DR) (Figure 1).

2.3. Model

An overview of the study is shown in Figure 2.

2.3.1. Data Preprocessing

All full-leg plain radiographs and ground truth masks were cropped and resized to 240×1200 pixels. Pixel values of grayscale images ranging from 0 to 255 on plain radiographs were normalised (Z-score normalisation) so that each image had a mean pixel value of 0 and standard deviation of 1 [34].

2.3.2. Model Architecture

U-Net shows high performance in medical image segmentation, attributed to its skip connections that allow feature extraction without a significant loss of resolution [35]. In this study, we adopted a U-Net-like architecture to develop an optimal model to segment muscle from full-leg plain radiographs. To increase its feature extraction capability, we added the squeeze-and-excitation (SE) block to the encoding path of the network, which adaptively recalibrated channel-wise feature responses to significantly improve CNN performance, without marked increases in model complexity or computational burden [36]. In addition, we applied group normalisation instead of batch normalisation, which is the

most commonly used normalisation method. Retaining the resolution of the original image was important for segmenting muscle from plain radiographs but had the potential to adversely affect model performance (as a larger image size will inevitably result in a smaller batch size due to the limited memory). Instead, robust normalisation methods such as group normalisation, which performs well even with a small batch size, can be used [37].

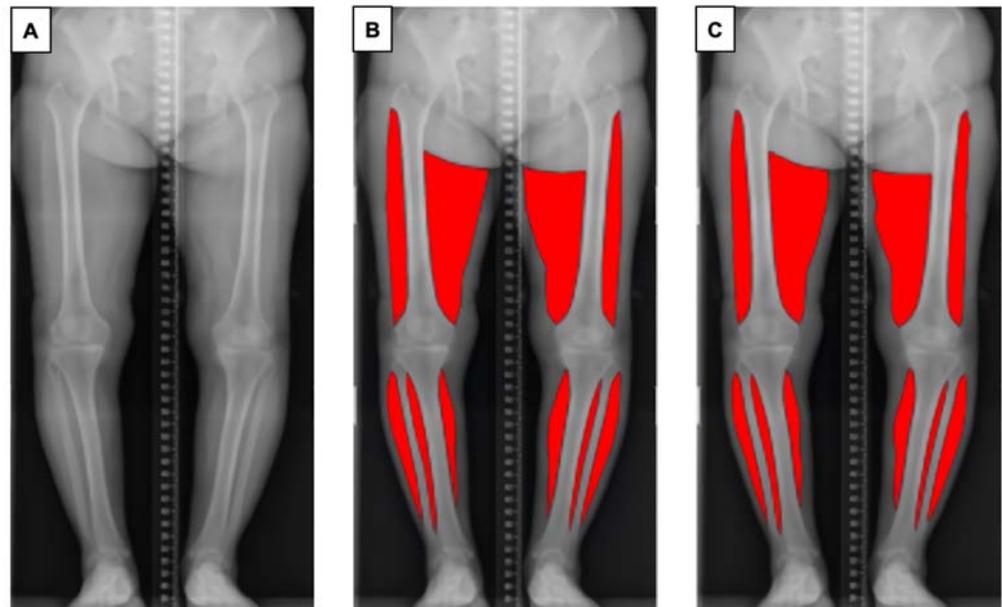


Figure 1. Overview of the pipeline for automatic muscle segmentation of full-leg plain radiographs. (A) Original radiographic image. (B) Ground truth mask of segmented muscle generated by the authors. (C) Segmented muscle predicted by the proposed model.

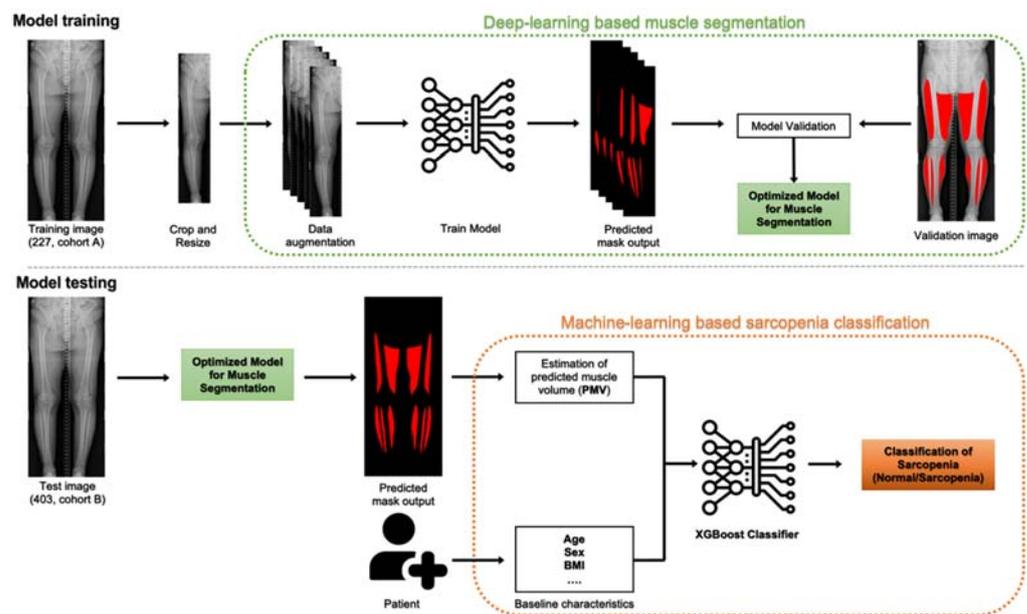


Figure 2. Model overview. Model training, left to right: the model is trained on input images, which undergo preprocessing (crop and resize) and augmentation (enriching the training set) before being fed into the convolutional neural network. The model is trained and validated by 5-fold cross validation. Hyperparameter tuning is conducted to optimise the model for muscle segmentation. Model testing, left to right: the optimised eXtreme Gradient Boosting (XGBoost) classification model is used to estimate predicted muscle volume (PMV) of the patients and, along with baseline characteristics, to classify patients into sarcopenia and normal groups.

2.3.3. Network Training

The Adam optimiser was used to train our model [38]. The Lovász–Hinge loss was used as a loss function and backpropagation was applied to the network weights for training [39]. The training batch size was set to 4. We used a global learning decay strategy that reduced the learning rate by 90% when the loss reached a plateau. The training and model performance assessment were conducted using a 5-fold cross-validation to address potential overfitting due to the small dataset. Data augmentation strategies included rotation (−10, +10), horizontal flip and scaling (0.9, 1.1).

2.3.4. Model Performance Evaluation

The most widely adopted metrics for evaluating models for semantic segmentation are the mean intersection-over-union (IoU) value and dice similarity coefficient (DSC) [40,41]. The performance of the developed model was evaluated using both the training (cohort A) and test sets (cohort B) (Supplementary Table S3).

2.3.5. Muscle Volume Estimation

A total of 403 test images from cohort B were fed into the developed model for muscle segmentation based on full-leg plain radiographs, to estimate the muscle volume. From the output segmentation mask, the predicted muscle area (A) was estimated as the sum of nonzero pixels multiplied by pixel size. We assumed that a single leg was a cylinder to estimate the predicted muscle volume (PMV) as follows: $PMV = \pi a (d_L + d_R)/4$. Here, d_L and d_R are the longest diameters of the mid-thigh region of the left and right extremities, respectively. The mid-thigh muscle area has been reported to be a good predictor of whole-body skeletal muscle mass, as it is highly sensitive to changes therein [10]. The mid-thigh diameter was measured automatically by postprocessing in our model. A Pearson's correlation analysis was performed between the PMV and SMI.

2.3.6. Machine Learning Model for Sarcopenia Prediction

Thirty preoperative variables were initially chosen as candidate predictors based on previous studies [12,14,42–44]. The preoperative variables included patient demographic information, comorbidities, laboratory data and the proposed sarcopenia marker (PMV) (Supplementary Tables S1 and S2). The model for sarcopenia prediction was developed, trained and validated based on PMV and the baseline characteristics of patients of cohort B.

For prediction of sarcopenia, the binary classification (sarcopenia/normal) model of XG Boost was used. The XGBoost model is one of the most commonly used machine learning models for solving both regression and classification problems and has been widely adopted to classify and predict medical events [3,45,46]. The synthetic minority oversampling technique (SMOTE) was utilised to overcome the potential bias arising from class imbalances, by creating synthetic minority class samples (the sarcopenia group in this case) [47,48]. Key features for the XGBoost classification were selected. The comparison of key features between sarcopenia and normal group was conducted. A stratified 5-fold cross-validation of the training dataset was performed to obtain the optimal degree of model complexity. Training accuracy was evaluated from the mean area under curve (AUC) value of the receiver operating characteristic (ROC). Along with the AUC and model accuracy, the sensitivity and specificity of the binary classification were also investigated. To understand the decision-making process, the importance of each feature in the machine learning model was also determined by calculating a “gain”, which refers to the relative contribution of the corresponding feature to the model by taking each feature's contribution for each tree in the model [49].

2.4. Statistical Analysis

Statistical analyses were performed using RStudio for Windows (ver. 1.2.5033; RStudio, Boston, MA, USA). Nominal data are shown as percentages and were analysed by two-sided Pearson's χ^2 test or Fisher's exact test. Continuous data are shown as the mean \pm SD

and were analysed using Student’s *t* test. In all analyses, $p < 0.05$ was taken to indicate statistical significance.

3. Results

The proportion of patients who had undergone primary TKA and had sarcopenia was 8.5% (34/369). There were significant differences between the sarcopenia and nonsarcopenia groups in age, BMI, preoperative total protein and haemoglobin (Hb) levels, skeletal muscle index (SMI) and PMV (Table 1). The patients with sarcopenia were older than those without the condition (74.6 ± 6.5 vs. 70.5 ± 6.5 years, $p = 0.001$), and had a lower BMI (23.9 ± 3.4 vs. 26.7 ± 3.2 kg/m², $p < 0.001$), preoperative total protein level (6.7 ± 0.4 vs. 7.1 ± 0.4 mg/dL, $p < 0.001$), preoperative Hb (12.3 ± 1.2 vs. 13.1 ± 1.8 g/dL, $p = 0.008$), SMI (5.5 ± 0.6 vs. 7.4 ± 1.1 kg/m², $p < 0.001$) and PMV (6972.4 ± 1354.6 vs. 8418.4 ± 1634.8 cm³, $p < 0.001$).

Table 1. Key features comparison of the sarcopenia and nonsarcopenia groups.

Characteristics	Total Population (<i>n</i> = 403)		<i>p</i> -Value
	Yes (<i>n</i> = 34)	No (<i>n</i> = 369)	
Sex (%)			
Female	32 (94.1)	319 (86.4)	0.266
Male	2 (5.9)	50 (13.6)	
Age (SD)	74.6 (6.5)	70.5 (6.5)	<0.001
BMI, kg/m ² (SD)	23.9 (3.4)	26.7 (3.2)	<0.001
Total Protein, mg/dL (SD)	6.7 (0.4)	7.1 (0.4)	<0.001
Albumin, g/dL (SD)	4.1 (0.3)	4.2 (0.4)	0.194
Hemoglobin, g/dL (SD)	12.3 (1.2)	13.1 (1.8)	0.004
Total Bilirubin, mg/dL (SD)	0.6 (0.3)	0.6 (0.2)	0.946
SMI, kg/m ² (SD)	5.5 (0.6)	7.4 (1.1)	<0.001
PMV, cm ³ (SD)	6972.4 (1354.6)	8418.4 (1634.8)	<0.001

Values are shown as the mean ± standard deviation or number (%). Statistical significance was set at $p < 0.05$. BMI, body mass index. SMI, skeletal muscle index. PMV, predictive muscle volume.

The deep learning model developed for the muscle segmentation of full-leg plain radiographs achieved a mean DSC of 0.944 (95% CI 0.936–0.951) and mean IoU value of 0.959 (95% CI 0.959–0.960) for the developmental training cohort A. For the test set images of cohort B, it achieved a mean DSC of 0.913 (95% CI 0.910–0.916) and mean IoU value of 0.926 (95% CI 0.920–0.931).

Pearson’s correlation analysis was performed between PMV and SMI, and the correlation coefficient was 0.654 ($p < 0.001$) (Figure 3).

The stratified fivefold AUC value of the XGBoost model after internal validation with the test set (cohort B) was 0.988 (95% CI 0.986–0.989) (Figure 4). The classification model had an accuracy of 0.945 (95% CI 0.941–0.950), sensitivity of 0.970 (95% CI 0.962–0.978) and specificity of 0.926 (95% CI 0.920–0.931).

Of the 30 preoperative variables, 7 key predictors were selected for the model: age, BMI, total protein, albumin, Hb, bilirubin and PMV. Feature importance ranks, which indicate the relative importance of input features, were calculated to understand the decision-making process of the XGBoost model (Figure 5). PMV was the most important feature in the decision-making process (feature importance: 0.179). Along with the PMV, BMI (0.164), bilirubin (0.158), preoperative Hb (0.132), albumin (0.131), total protein (0.123) and age (0.113) were key features for sarcopenia prediction; PMV, BMI, Hb, total protein and age also showed statistical significance ($p < 0.05$) in a univariate analysis.

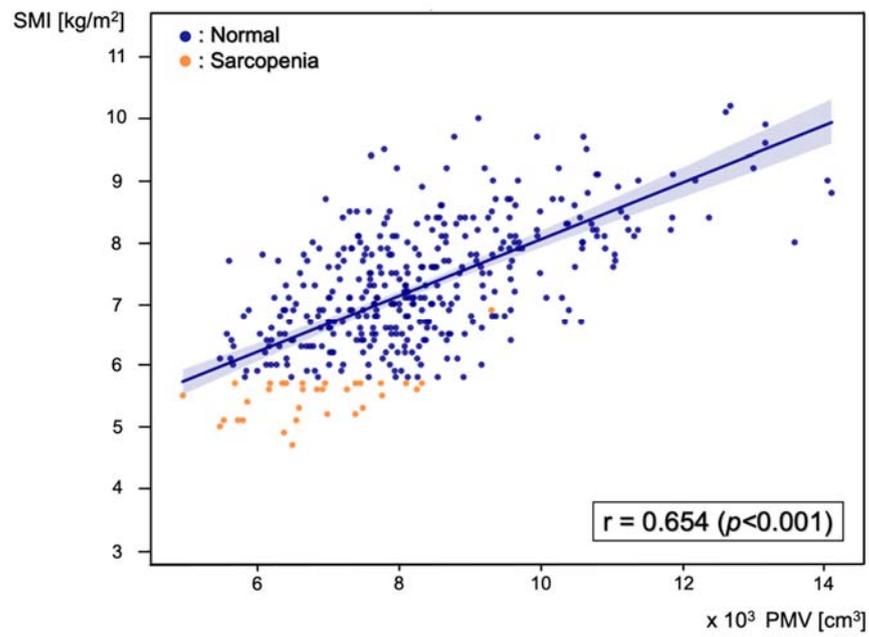


Figure 3. Scatter plot of the correlation between predicted muscle volume (PMV) and skeletal muscle index (SMI), with a Pearson’s correlation coefficient of 0.654 ($p < 0.001$). Blue dots represent patients in the normal group and orange dots represent patients in the sarcopenia group.

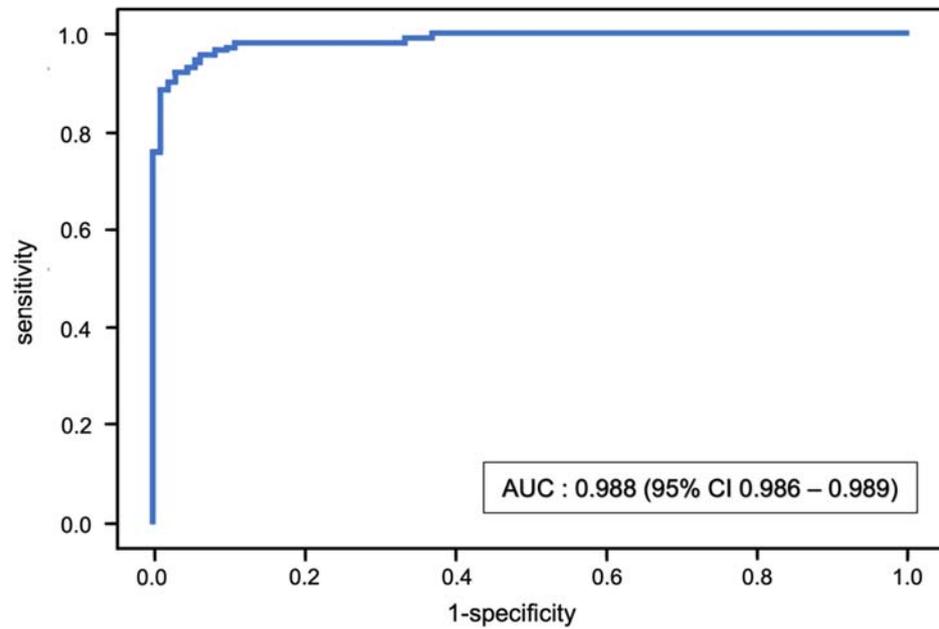


Figure 4. Receiver operating characteristic (ROC) curve and mean area under the curve (AUC) for the binary classification of sarcopenia using the XGBoost model.

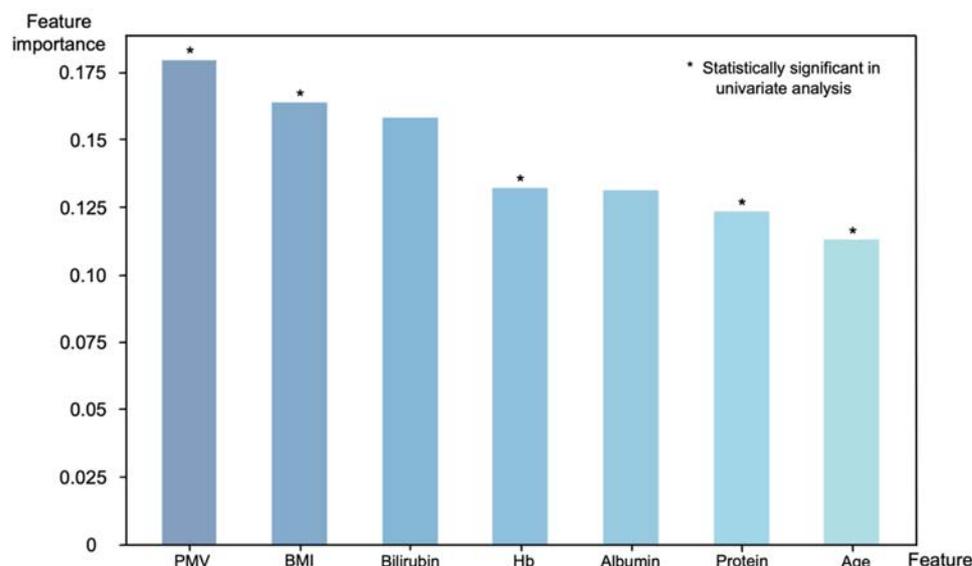


Figure 5. Feature importance with respect to the binary classification of sarcopenia using the XGBoost model. Features that showed statistical significance in a univariate analysis also ranked high in importance in the XGBoost model, with PMV being the most important features.

4. Discussion

The most important findings of this study were that our CNN-based deep learning model showed high performance in terms of automatic muscle segmentation of full-leg plain radiographs, while the XGBoost classification model, which included several important patient features such as PMV, showed high performance for predicting sarcopenia.

The feature importance ranks of our sarcopenia prediction model showed that PMV can serve as a feasible sarcopenia marker. The feature importance score of PMV suggested that it is the most important factor for predicting sarcopenia. Along with PMV, it is notable that several serum markers also contributed to the decision-making process. Total bilirubin has been reported to be positively correlated with SMI, although it was not significantly different between sarcopenia and normal groups [42]. Bilirubin is one of the most active endogenous antioxidant molecules and is therefore thought to have a protective effect against the progression of sarcopenia. Albumin and total protein are considered good markers of nutritional status and were reported to be low in patients with sarcopenia [43,44]. A low Hb level was also reported to be associated with sarcopenia, although the pathophysiology of this relation has not been explored [50]. The selection of serum markers that are correlated with sarcopenia indicates that the decision-making process of the XGBoost classifier was reasonable.

To our knowledge, among all reported sarcopenia screening models, ours showed the highest performance (AUC 0.988, 95% CI 0.986–0.989) [25,33,51,52]. We have developed a pipeline involving both deep-learning- and machine-learning-based models, to take both radiographic images and patient baseline characteristics into consideration (which can be easily obtained in the preoperative clinical setting). Moreover, there have been no previous reports of automatic muscle segmentation or quantification of full-leg plain radiographs for sarcopenia screening. There are studies reporting the potential of chest radiographs to analyse body composition, but not for the purpose of assessing sarcopenia [53]. Although surgeons can use other imaging modalities, such as CT, to investigate the risk of sarcopenia, we evaluated muscle mass on full-leg plain radiographs as they are routinely obtained in TKA patients and are therefore appropriate for screening purposes. Our method could also be applied to other surgical procedures in which full-leg plain radiographs are obtained as part of routine examinations, which would be particularly useful in the field of orthopaedics. Our proposed sarcopenia screening method may not be applicable in a community healthcare setting, where neither laboratory testing nor radio-

graphic imaging is available. In the preoperative setting, on the other hand, sarcopenia can be screened with near-perfect sensitivity and specificity using only elementary laboratory data and full-leg plain radiographic images.

This study had several limitations. Firstly, there were some barriers to a reasonable estimation of the muscle volume on full-leg plain radiographs. Although our deep-learning-based model was generally successful in performing muscle segmentation on full-leg plain radiographs, it tended to slightly overestimate the number of pixels in the muscle layer. Moreover, the model had difficulty in segmenting the medial and the upper margin of the thigh muscle region (gluteal fold).

The small dataset size and homogeneity of the training set may have affected the performance and robustness of the model and given rise to the above issues. Specifically, the difference in patient distribution between the training set (healthy volunteers with no history of osteoarthritis or trauma) and test set (patients with end-stage osteoarthritis undergoing TKA) may have decreased model performance. In addition, several images did not have a distinct gluteal fold and were therefore labelled with reference to the opposite side during the annotation process, to create the ground truth segmentation mask. Such annotation inconsistency may have contributed to the ambiguity in the training process. Therefore, extensive training of the model with additional full-leg plain radiographs from a wide array of patients is required to improve the accuracy of the model.

Secondly, due to fatty infiltration of skeletal muscle in patients with sarcopenic obesity, the estimated muscle mass of patients may not always represent the actual muscle mass [54]. Features such as radio-opacity of the plain radiographs, which may reflect fatty infiltration, will be incorporated in future studies.

For the classification model of sarcopenia, an external validation was not performed, which raises questions regarding its validity when applied to other cohorts. In addition, the predominance of females in our study population may have adversely affected the model feature selection process. As the study population consisted only of patients undergoing TKA, the key features and performance of the model may not generalise to preoperative settings for other surgical procedures. Further studies are required to expand the number of available training sets; the proposed model should also be applied to independent test datasets from other institutions to verify its validity.

As the diagnostic criteria involve not only muscle mass, functional tests of muscle strength and physical performance, such as gait speed, muscle grip strength, the get-up-and-go test and peak expiratory flow, should also be included in the sarcopenia screening process [33]. We expect that incorporation of such tests in future studies will render our approach to screen sarcopenia using full-leg plain radiographs more legitimate.

In a population of 90,438 patients who had undergone primary TKA, Ardeljan et al., reported that 16.7% had sarcopenia. The patients with sarcopenia had longer hospital stays and increased odds of falls, lower extremity fracture, reoperation, 2-year implant-related complication rates, higher surgery costs and higher rates of postoperative blood transfusion and complications within 90 days [17]. As it is an indicator of the risk of adverse events during the postoperative period, cases considered at high-risk of sarcopenia could be confirmed by dual-energy X-ray absorptiometry (DEXA) if necessary, which is the gold standard for quantifying muscle mass [10]. As a modifiable risk factor [55,56], patients suspected of having sarcopenia based on screening can also be managed prior to TKA to improve clinical outcomes. After TKA, they should be treated with more caution and attend more frequent follow-up visits.

5. Conclusions

Here, we presented a novel method for screening sarcopenia in the preoperative clinical setting, using a fully automated deep learning model to automatically segment and quantify the muscle layer on full-leg plain radiographs, and validated its potential to predict sarcopenia in patients undergoing TKA. In the preoperative clinical setting, wherein

laboratory tests and radiographic imaging are available, our model showed high sarcopenia screening performance.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm11133612/s1>, Table S1: Baseline characteristics and sarcopenia markers of sarcopenic and nonsarcopenic patients, Table S2: Laboratory data of sarcopenic and nonsarcopenic patients, Table S3: Performance assessment of the model on Cohort A (training set) and Cohort B (test set).

Author Contributions: Conceptualization, D.H.R. and D.H.; methodology, D.H. and S.A.; validation, Y.-B.P. and S.H.K.; analysis, D.H.; data curation, D.H.; writing—original draft preparation, D.H.; writing—review and editing, H.-S.H., M.C.L. and D.H.R.; visualization, D.H. and S.A.; supervision, D.H.R. and S.H.K.; project administration, D.H.R.; funding acquisition, H.-S.H. and M.C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number HC20C0040030021).

Institutional Review Board Statement: This study was approved by Seoul National University Hospital Institutional Review Board (IRB No. H-2009-181-1161).

Informed Consent Statement: The requirement for informed consent was waived due to the retrospective nature of this study and the use of anonymised patient data by Seoul National University Hospital Institutional Review Board.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chang, M.J.; Kim, S.H.; Kang, Y.G.; Chang, C.B.; Kim, T.K. Activity levels and participation in physical activities by Korean patients following total knee arthroplasty. *BMC Musculoskelet. Disord.* **2014**, *15*, 240. [[CrossRef](#)] [[PubMed](#)]
2. Canovas, F.; Dagneaux, L. Quality of life after total knee arthroplasty. *Orthop. Traumatol. Surg. Res.* **2018**, *104*, S41–S46. [[CrossRef](#)] [[PubMed](#)]
3. Ko, S.; Jo, C.; Chang, C.B.; Lee, Y.S.; Moon, Y.W.; Youm, J.W.; Han, H.S.; Lee, M.C.; Lee, H.; Ro, D.H. A web-based machine-learning algorithm predicting postoperative acute kidney injury after total knee arthroplasty. *Knee Surg. Sports Traumatol. Arthrosc.* **2022**, *30*, 545–554. [[CrossRef](#)] [[PubMed](#)]
4. Jo, C.; Ko, S.; Shin, W.C.; Han, H.S.; Lee, M.C.; Ko, T.; Ro, D.H. Transfusion after total knee arthroplasty can be predicted using the machine learning algorithm. *Knee Surg. Sports Traumatol. Arthrosc.* **2020**, *28*, 1757–1764. [[CrossRef](#)] [[PubMed](#)]
5. Pannu, T.S.; Villa, J.M.; Cohen, E.M.; Hayda, R.A.; Higuera, C.A.; Deren, M.E. Periprosthetic Tibial Fractures After Total Knee Arthroplasty: Early and Long-Term Clinical Outcomes. *J. Arthroplast.* **2021**, *36*, 1429–1436. [[CrossRef](#)]
6. Martinez-Pastor, J.C.; Macule-Beneyto, F.; Suso-Vergara, S. Acute infection in total knee arthroplasty: Diagnosis and treatment. *Open Orthop. J.* **2013**, *7*, 197–204. [[CrossRef](#)]
7. Blanco, J.F.; Diaz, A.; Melchor, F.R.; da Casa, C.; Pescador, D. Risk factors for periprosthetic joint infection after total knee arthroplasty. *Arch. Orthop. Trauma Surg.* **2020**, *140*, 239–245. [[CrossRef](#)]
8. Poultsides, L.A.; Triantafyllopoulos, G.K.; Sakellariou, V.I.; Memtsoudis, S.G.; Sculco, T.P. Infection risk assessment in patients undergoing primary total knee arthroplasty. *Int. Orthop.* **2018**, *42*, 87–94. [[CrossRef](#)]
9. Ebraheim, N.A.; Kelley, L.H.; Liu, X.; Thomas, I.S.; Steiner, R.B.; Liu, J. Periprosthetic Distal Femur Fracture after Total Knee Arthroplasty: A Systematic Review. *Orthop. Surg.* **2015**, *7*, 297–305. [[CrossRef](#)]
10. Cruz-Jentoft, A.J.; Bahat, G.; Bauer, J.; Boirie, Y.; Bruyere, O.; Cederholm, T.; Cooper, C.; Landi, F.; Rolland, Y.; Sayer, A.A.; et al. Sarcopenia: Revised European consensus on definition and diagnosis. *Age Ageing* **2019**, *48*, 16–31. [[CrossRef](#)]
11. Delmonico, M.J.; Beck, D.T. The Current Understanding of Sarcopenia: Emerging Tools and Interventional Possibilities. *Am. J. Lifestyle Med.* **2017**, *11*, 167–181. [[CrossRef](#)] [[PubMed](#)]
12. Jones, A.J.; Campiti, V.J.; Alwani, M.; Novinger, L.J.; Tucker, B.J.; Bonetto, A.; Yesensky, J.A.; Sim, M.W.; Moore, M.G.; Mantravadi, A.V. Sarcopenia is associated with blood transfusions in head and neck cancer free flap surgery. *Laryngoscope Investig. Otolaryngol.* **2021**, *6*, 200–210. [[CrossRef](#)] [[PubMed](#)]
13. Marasco, G.; Dajti, E.; Serenari, M.; Alemanni, L.V.; Ravaioli, F.; Ravaioli, M.; Vestito, A.; Vara, G.; Festi, D.; Golfieri, R.; et al. Sarcopenia Predicts Major Complications after Resection for Primary Hepatocellular Carcinoma in Compensated Cirrhosis. *Cancers* **2022**, *14*, 1935. [[CrossRef](#)] [[PubMed](#)]

14. Bokshan, S.L.; Han, A.L.; DePasse, J.M.; Eltorai, A.E.; Marcaccio, S.E.; Palumbo, M.A.; Daniels, A.H. Effect of Sarcopenia on Postoperative Morbidity and Mortality After Thoracolumbar Spine Surgery. *Orthopedics* **2016**, *39*, e1159–e1164. [[CrossRef](#)] [[PubMed](#)]
15. Babu, J.M.; Kalagara, S.; Durand, W.; Antoci, V.; Deren, M.E.; Cohen, E. Sarcopenia as a Risk Factor for Prosthetic Infection After Total Hip or Knee Arthroplasty. *J. Arthroplast.* **2019**, *34*, 116–122. [[CrossRef](#)] [[PubMed](#)]
16. Lakomkin, N.; Zuckerman, S.L.; Stannard, B.; Montejo, J.; Sussman, E.S.; Virojanapa, J.; Kuzmik, G.; Goz, V.; Hadjipanayis, C.G.; Cheng, J.S. Preoperative Risk Stratification in Spine Tumor Surgery: A Comparison of the Modified Charlson Index, Frailty Index, and ASA Score. *Spine* **2019**, *44*, E782–E787. [[CrossRef](#)]
17. Ardeljan, A.D.; Polisetty, T.S.; Palmer, J.; Vakharia, R.M.; Roche, M.W. Comparative Analysis on the Effects of Sarcopenia following Primary Total Knee Arthroplasty: A Retrospective Matched-Control Analysis. *J. Knee Surg.* **2022**, *35*, 128–134. [[CrossRef](#)]
18. Ho, K.K.; Lau, L.C.; Chau, W.W.; Poon, Q.; Chung, K.Y.; Wong, R.M. End-stage knee osteoarthritis with and without sarcopenia and the effect of knee arthroplasty—A prospective cohort study. *BMC Geriatr.* **2021**, *21*, 2. [[CrossRef](#)]
19. Zeng, Y.; Hu, X.; Xie, L.; Han, Z.; Zuo, Y.; Yang, M. The Prevalence of Sarcopenia in Chinese Elderly Nursing Home Residents: A Comparison of 4 Diagnostic Criteria. *J. Am. Med. Dir. Assoc.* **2018**, *19*, 690–695. [[CrossRef](#)]
20. Nishioka, S.; Yamanouchi, A.; Matsushita, T.; Nishioka, E.; Mori, N.; Taguchi, S. Validity of calf circumference for estimating skeletal muscle mass for Asian patients after stroke. *Nutrition* **2021**, *82*, 111028. [[CrossRef](#)]
21. Chen, C.Y.; Tseng, W.C.; Yang, Y.H.; Chen, C.L.; Lin, L.L.; Chen, F.P.; Wong, A.M.K. Calf Circumference as an Optimal Choice of Four Screening Tools for Sarcopenia Among Ethnic Chinese Older Adults in Assisted Living. *Clin. Interv. Aging* **2020**, *15*, 2415–2422. [[CrossRef](#)] [[PubMed](#)]
22. Zhu, Y.X.; Zhang, Y.; Wang, Y.Y.; Ren, C.X.; Xu, J.; Zhang, X.Y. Low calf circumference is associated with frailty in diabetic adults aged over 80 years. *BMC Geriatr.* **2020**, *20*, 414. [[CrossRef](#)] [[PubMed](#)]
23. Endo, K.; Sato, T.; Kakisaka, K.; Takikawa, Y. Calf and arm circumference as simple markers for screening sarcopenia in patients with chronic liver disease. *Hepatol. Res.* **2021**, *51*, 176–189. [[CrossRef](#)] [[PubMed](#)]
24. Chen, L.K.; Liu, L.K.; Woo, J.; Assantachai, P.; Auyeung, T.W.; Bahyah, K.S.; Chou, M.Y.; Chen, L.Y.; Hsu, P.S.; Krairit, O.; et al. Sarcopenia in Asia: Consensus report of the Asian Working Group for Sarcopenia. *J. Am. Med. Dir. Assoc.* **2014**, *15*, 95–101. [[CrossRef](#)]
25. Bahat, G.; Erdogan, T.; Ilhan, B. SARC-F and other screening tests for sarcopenia. *Curr. Opin. Clin. Nutr. Metab. Care* **2022**, *25*, 37–42. [[CrossRef](#)]
26. Voelker, S.N.; Michalopoulos, N.; Maier, A.B.; Reijnierse, E.M. Reliability and Concurrent Validity of the SARC-F and Its Modified Versions: A Systematic Review and Meta-Analysis. *J. Am. Med. Dir. Assoc.* **2021**, *22*, 1864–1876.e16. [[CrossRef](#)]
27. Gomez-Perez, S.L.; Haus, J.M.; Sheean, P.; Patel, B.; Mar, W.; Chaudhry, V.; McKeever, L.; Braunschweig, C. Measuring Abdominal Circumference and Skeletal Muscle From a Single Cross-Sectional Computed Tomography Image: A Step-by-Step Guide for Clinicians Using National Institutes of Health ImageJ. *J. Parenter. Enteral. Nutr.* **2016**, *40*, 308–318. [[CrossRef](#)]
28. Beaudart, C.; McCloskey, E.; Bruyere, O.; Cesari, M.; Rolland, Y.; Rizzoli, R.; Araujo de Carvalho, I.; Amuthavalli Thiyagarajan, J.; Bautmans, I.; Bertiere, M.C.; et al. Sarcopenia in daily practice: Assessment and management. *BMC Geriatr.* **2016**, *16*, 170. [[CrossRef](#)]
29. Rossi, A.P.; Fantin, F.; Micciolo, R.; Bertocchi, M.; Bertassello, P.; Zanandrea, V.; Zivelonghi, A.; Bissoli, L.; Zamboni, M. Identifying sarcopenia in acute care setting patients. *J. Am. Med. Dir. Assoc.* **2014**, *15*, 303.e7–303.e12. [[CrossRef](#)]
30. Albano, D.; Messina, C.; Vitale, J.; Sconfienza, L.M. Imaging of sarcopenia: Old evidence and new insights. *Eur. Radiol.* **2020**, *30*, 2199–2208. [[CrossRef](#)]
31. Gonzalez, M.C.; Barbosa-Silva, T.G.; Heymsfield, S.B. Bioelectrical impedance analysis in the assessment of sarcopenia. *Curr. Opin. Clin. Nutr. Metab. Care* **2018**, *21*, 366–374. [[CrossRef](#)] [[PubMed](#)]
32. Aleixo, G.F.P.; Shachar, S.S.; Nyrop, K.A.; Muss, H.B.; Battaglini, C.L.; Williams, G.R. Bioelectrical Impedance Analysis for the Assessment of Sarcopenia in Patients with Cancer: A Systematic Review. *Oncologist* **2020**, *25*, 170–182. [[CrossRef](#)] [[PubMed](#)]
33. Chen, L.K.; Woo, J.; Assantachai, P.; Auyeung, T.W.; Chou, M.Y.; Iijima, K.; Jang, H.C.; Kang, L.; Kim, M.; Kim, S.; et al. Asian Working Group for Sarcopenia: 2019 Consensus Update on Sarcopenia Diagnosis and Treatment. *J. Am. Med. Dir. Assoc.* **2020**, *21*, 300–307.e2. [[CrossRef](#)] [[PubMed](#)]
34. Gopal Krishna Patro, S.; Sahu, K.K. Normalization: A Preprocessing Stage. *arXiv* **2015**, arXiv:1503.06462. Available online: <https://ui.adsabs.harvard.edu/abs/2015arXiv150306462G> (accessed on 15 February 2022).
35. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.
36. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *arXiv* **2017**, arXiv:1709.01507. Available online: <https://ui.adsabs.harvard.edu/abs/2017arXiv170901507H> (accessed on 11 January 2022).
37. Wu, Y.; He, K. Group Normalization. *arXiv* **2018**, arXiv:1803.08494. Available online: <https://ui.adsabs.harvard.edu/abs/2018arXiv180308494W> (accessed on 12 January 2022).
38. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980. Available online: <https://ui.adsabs.harvard.edu/abs/2014arXiv1412.6980K> (accessed on 19 June 2022).

39. Yu, J.; Blaschko, M. The Lovász Hinge: A Novel Convex Surrogate for Submodular Losses. *arXiv* **2015**, arXiv:1512.07797. Available online: <https://ui.adsabs.harvard.edu/abs/2015arXiv151207797Y> (accessed on 18 February 2022). [[CrossRef](#)]
40. Csurka, G.; Larlus, D.; Perronnin, F. What is a good evaluation measure for semantic segmentation? In Proceedings of the British Machine Vision Conference 2013, Bristol, UK, 9–13 September 2013.
41. Berman, M.; Rannen Triki, A.; Blaschko, M.B. The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. *arXiv* **2017**, arXiv:1705.08790. Available online: <https://ui.adsabs.harvard.edu/abs/2017arXiv170508790B> (accessed on 19 February 2022).
42. Wang, C.; Jin, C.; Yin, X.; Liu, J.; Liu, J. Relationship between serum bilirubin concentration and sarcopenia in patients with type 2 diabetes: A cross-sectional study. *J. Int. Med. Res.* **2021**, *49*, 1–9. [[CrossRef](#)]
43. Cabrerizo, S.; Cuadras, D.; Gomez-Busto, F.; Artaza-Artabe, I.; Marin-Ciancas, F.; Malafarina, V. Serum albumin and health in older people: Review and meta analysis. *Maturitas* **2015**, *81*, 17–27. [[CrossRef](#)] [[PubMed](#)]
44. Petermann-Rocha, F.; Gray, S.R.; Pell, J.P.; Celis-Morales, C.; Ho, F.K. Biomarkers Profile of People With Sarcopenia: A Cross-sectional Analysis From UK Biobank. *J. Am. Med. Dir. Assoc.* **2020**, *21*, 2017.e1–2017.e9. [[CrossRef](#)] [[PubMed](#)]
45. Fralick, M.; Colak, E.; Mamdani, M. Machine Learning in Medicine. *N. Engl. J. Med.* **2019**, *380*, 2588–2589. [[CrossRef](#)] [[PubMed](#)]
46. Guo, H.; Wang, J.; Ao, W.; He, Y. SGB-ELM: An Advanced Stochastic Gradient Boosting-Based Ensemble Scheme for Extreme Learning Machine. *Comput. Intell. Neurosci.* **2018**, *2018*, 4058403. [[CrossRef](#)] [[PubMed](#)]
47. Wang, S.; Dai, Y.; Shen, J.; Xuan, J. Research on expansion and classification of imbalanced data based on SMOTE algorithm. *Sci. Rep.* **2021**, *11*, 24039. [[CrossRef](#)]
48. Blagus, R.; Lusa, L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinform.* **2013**, *14*, 106. [[CrossRef](#)]
49. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
50. Yoshimura, Y.; Wakabayashi, H.; Nagano, F.; Bise, T.; Shimazu, S.; Shiraishi, A. Low Hemoglobin Levels are Associated with Sarcopenia, Dysphagia, and Adverse Rehabilitation Outcomes After Stroke. *J. Stroke Cerebrovasc. Dis.* **2020**, *29*, 105405. [[CrossRef](#)]
51. Zhou, J.; Li, T.; Chen, X.; Wang, M.; Jiang, W.; Jia, H. Comparison of the Diagnostic Value of SARC-F and Its Three Modified Versions for Screening Sarcopenia in Chinese Community-Dwelling Older Adults. *J. Nutr. Health Aging* **2022**, *26*, 77–83. [[CrossRef](#)]
52. Hanada, M.; Sakamoto, N.; Ishimoto, H.; Kido, T.; Miyamura, T.; Oikawa, M.; Nagura, H.; Takeuchi, R.; Kawazoe, Y.; Sato, S.; et al. A comparative study of the sarcopenia screening in older patients with interstitial lung disease. *BMC Pulm. Med.* **2022**, *22*, 45. [[CrossRef](#)]
53. Ciaffi, J.; Gigliotti, P.E.; Festuccia, G.; Borlandelli, E.; Facchini, G.; Chiaravallotti, A.; Miceli, M.; Meliconi, R.; Ursini, F. Can chest imaging be used to draw information about body mass index and obesity status? *Obes. Res. Clin. Pract.* **2021**, *15*, 187–190. [[CrossRef](#)] [[PubMed](#)]
54. Khanal, P.; Williams, A.G.; He, L.; Stebbings, G.K.; Onambele-Pearson, G.L.; Thomis, M.; Degens, H.; Morse, C.I. Sarcopenia, Obesity, and Sarcopenic Obesity: Relationship with Skeletal Muscle Phenotypes and Single Nucleotide Polymorphisms. *J. Clin. Med.* **2021**, *10*, 4933. [[CrossRef](#)] [[PubMed](#)]
55. Morley, J.E. Sarcopenia: Diagnosis and treatment. *J. Nutr. Health Aging* **2008**, *12*, 452–456. [[CrossRef](#)] [[PubMed](#)]
56. Morley, J.E.; Argiles, J.M.; Evans, W.J.; Bhasin, S.; Cella, D.; Deutz, N.E.; Doehner, W.; Fearon, K.C.; Ferrucci, L.; Hellerstein, M.K.; et al. Nutritional recommendations for the management of sarcopenia. *J. Am. Med. Dir. Assoc.* **2010**, *11*, 391–396. [[CrossRef](#)]