KIPS Trans. Comp. and Comm. Sys.
Vol.12, No.3 pp.119~126
ISSN: 2287-5891 (Print), ISSN 2734-049X (Online)

https://doi.org/10.3745/KTCCS.2023.12.3.119

Development and Validation of AI Image Segmentation Model for CT Image-Based Sarcopenia Diagnosis

Lee Chung-Sub[†] · Lim Dong-Wook[††] · Noh Si-Hyeong[†††] · Kim Tae-Hoon[††††] · Ko Yousun[†††††] · Kim Kyung Won[††††††] · Jeong Chang-Won[†††††††]

ABSTRACT

Sarcopenia is not well known enough to be classified as a disease in 2021 in Korea, but it is recognized as a social problem in developed countries that have entered an aging society. The diagnosis of sarcopenia follows the international standard guidelines presented by the European Working Group for Sarcopenia in Older People (EWGSOP) and the d Asian Working Group for Sarcopenia (AWGS). Recently, it is recommended to evaluate muscle function by using physical performance evaluation, walking speed measurement, and standing test in addition to absolute muscle mass as a diagnostic method. As a representative method for measuring muscle mass, the body composition analysis method using DEXA has been formally implemented in clinical practice. In addition, various studies for measuring muscle mass using abdominal images of MRI or CT are being actively conducted. In this paper, we develop an AI image segmentation model based on abdominal images of CT with a relatively short imaging time for the diagnosis of sarcopenia and describe the multicenter validation. We developed an artificial intelligence model using U-Net that can automatically segment muscle, subcutaneous fat, and visceral fat by selecting the L3 region from the CT image. Also, to evaluate the performance of the model, internal verification was performed by calculating the intersection over union (IOU) of the partitioned area, and the results of external verification using data from other hospitals are shown. Based on the verification results, we tried to review and supplement the problems and solutions.

Keywords : Sarcopenia, Image Segmentation, External Validation, U-Net, IOU

CT 영상 기반 근감소증 진단을 위한 AI 영상분할 모델 개발 및 검증

이 충 섭[†] · 임 동 욱[††] · 노 시 형[†††] · 김 태 훈[††††] · 고 유 선[†††††] · 김 경 원[††††††] · 정 창 원[†††††††]

요 약

근감소증은 국내는 2021년 질병으로 분류되었을 만큼 잘 알려져 있지 않지만 고령화사회에 진입한 선진국에서는 사회적 문제로 인식하고 있다. 근감소증 진단은 유럽노인근감소증 진단그룹(EWGSOP)과 아시아근감소증진단그룹(AWGS)에서 제시하는 국제표준지침을 따른다. 최근 진단방법으로 절대적 근육량 이외에 신체수행평가로 보행속도 측정과 일어서기 검사 등을 통하여 근육 기능을 함께 측정할 것을 권고하고 있다. 근육량을 측정하기 위한 대표적인 방법으로 DEXA를 이용한 체성분 분석 방법이 임상에서 정식으로 실시하고 있다. 또한 MRI 또는 CT의 복부 영상을 이용하여 근육량을 측정하는 다양한 연구가 활발하게 진행되고 있다. 따라서 본 논문에서는 근감소증 진단을 위해서 비교적 짧은 촬영시간을 갖는 CT의 복부영상기반으로 AI 영상 분할 모델을 개발하고 다기관 검증한 내용을 기술한다. 우리는 CT 영상 중에 요추의 L3 영역을 분류하여 피하지방, 내장지방, 근육을 자동으로 분할할 수 있는 인공지능 모델을 U-Net 모델을 사용하여 개발하였다. 또한 모델의 성능평가를 위해서 분할영역의 IOU(Intersection over Union)를 계산하여 내부검증을 진행했으며, 타 병원의 데이터를 활용하여 동일한 IOU 방법으로 외부검증을 진행한 결과를 보인다. 검증 결과를 토대로 문제점과 해결방안에 대해서 검증하고 보완하고자 했다.

키워드 : 근감소증, 영상분할, 외부검증, U-Net, IOU

※ 본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술 연구개발사업(HI18C1216), 그리고 한국 연구재단(NRF-2021R1A5A 8029876)(NRF-2020R1I1A1A01074256)의 지원으로 수행함.
※ 이 논문은 2022년 한국정보처리학회 ASK 2022의 우수논문으로 "근감소증 진단을 위한 영상분할 모델 개발 및 외부검증"의 제목으로 발표된 논문을 확장한 것임.
† 정 회 원 : 원광대학교 의료융합연구센터 책임연구원
†† 비 회 원 : 원광대학교 의료융합연구센터 연구원
††† 정 회 원 : 원광대학교 의료융합연구센터 연구원
†††† 정 회 원 : 원광대학교 의료융합연구센터 연구교수
††††† 비 회 원 : 울산대학교 의학과 연구조교수
†††††† 비 회 원 : 서울아산병원 영상의학과 부교수
††††††† 종신회원 : 원광대학교병원 의생명연구원 중점연구센터 교수
Manuscript Received : July 29, 2022
First Revision : September 20, 2022
Second Revision : October 11, 2022
Accepted : October 26, 2022
* Corresponding Author : Jeong Chang-Won(mediblue@wku.ac.kr)

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/ licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서 론

근감소증은 선진국에서는 2016년부터 질병코드를 부여하여 질환으로 분류하였으나 국내는 2021년에 이르러 한국표준질병사인분류에 근감소증(M62.5)을 포함하였다[1]. 근감소증은 노화, 영양 부족, 운동량 감소 등으로 인해 정상적인 근육량과 근력 및 근 기능이 저하되는 질환이다. 20대에 근육량이 최대치가 되고 30대 이후부터 매년 1% 정도의 근육량이 자연스럽게 감소하고 이러한 감소량은 남성은 40대 이후, 여성은 55세 이후부터 더욱 빨라지기 시작하여 70대까지 10년에 8%의 감소가 일어난다고 추정하고 있다. 그 이후로는 급격한 감소가 발생하여 10년 마다 15%까지 발생한다고 알려져 있다[2]. 80대가 되면 20대에 비교하여 전체 근육량의 절반까지 감소하게 된다. 근감소증은 근육량의 평가와 근력의 평가 그리고 신체수행능력의 평가를 통해 진단할 수 있다. 2010년에 유럽노인근감소증 진단그룹(European Working Group on Sarcopenia in Older People, EWGSOP)[3]은 근육량뿐 아니라 악력이나 신체수행능력을 고려해서 평가하는 공통적인 방법을 제시하고 있다[4]. 그러나 아시아인들은 서양인의 기준으로 적용하기에 부족했다. 따라서, 2014년에 아시아 근감소증 진단그룹(Asian Working Group for Sarcopenia, AWGS)이 결성되어 아시아 진단기준을 제시하였고, 근감소증 의심 환자를 진단하는 새로운 알고리즘을 발표하였다[5,6].

우리는 의료영상장비 중 널리 사용하고 있는 CT 의료영상을 기반으로 요추 3번 부위의 피하지방, 내장지방, 근육을 분할하고 정량 분석할 수 있는 애플리케이션을 개발하였다[7]. 그러나 단일 기관의 데이터로 AI 모델을 개발한 경우 범용성을 보장하지 못함을 타기관의 검증용 데이터셋을 통해 확인하였다. 특히, 부족한 데이터를 데이터 증강을 통해 보완하고자 하였으나, 임상에서 나올 수 없는 학습데이터가 생성되어 오히려 나쁜 성능 결과를 보였다. 따라서 본 논문에서는 실제 학습데이터를 확보하고 검증용 데이터도 단일기관이 아닌 한 개 이상의 기관 데이터를 확보하여 평가하고자 하였다. 특히, 타 병원의 복부 CT 데이터 3,019건을 추가하여 외부검증(External Validation)[8-10]을 진행하여 제안한 모델의 정확도와 범용성을 증명하였다.

2. 관련 연구

2.1 근육량 측정 방법

스크리닝을 통해 근육량이나 기능이 감소할 가능성이 높은 것으로 확인되면 다양한 진단 도구를 사용하여 근육량과 근육의 질을 측정한다. 지금까지 알려진 근육량 측정을 위한 진단 도구와 측정 방법을 Table 1에 정리하였다. 진단 도구를 사용한 결과는 개인의 근육량 상태, 근육 질에 대한 정확하고 상세한 정보를 제공한다[11]. 의료영상진단장치인 자기공명영상장치(Magnetic Resonance Imaging, MRI)는 가장 정확하게 근육량을 측정하는 방법이고 이미지의 유형에 따라 다양한 결과를 획득할 수 있다[12]. 그러나 촬영 비용이 비싸기 때문에 근육량 측정만을 위한 방법으로 사용하기에는 어려운 실정이다. 최근 이중에너지 방사선 흡수 계측법(Dual Energy X-ray Absorptiometry, DXA)이 MRI로 측정한 값과 상관관계가 높고 낮은 방사선량과 검사 편의성으로 대표적인 근육량 측정 방법으로 알려져 있다[13]. 또한 컴퓨터단층촬영(Computed Tomography, CT)으로 촬영한 의료영상기반의 측정 방법에 대한 연구가 활발하게 진행되고 있다[14]. 특히 촬영시간이 짧고 질환진단에 가장 많이 사용하고 있는 CT에서 촬영한 의료영상에서 요추 3번 수준의 단면에서 보이는 골격근의 면적 값을 이용한 L3 skeletal muscle index(SMI)가 총 골격근 측정값으로 사용하고 있다[15].

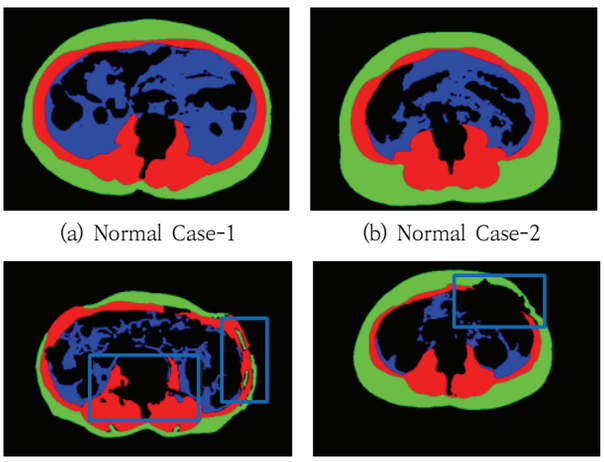
2.2 영상분할을 위한 방법

다양한 질환을 분석하고 진단하기 위해 컴퓨터 단층촬영영상을 기반으로 많은 연구들이 수행되고 있다. 의료영상에서 특정 장기 또는 질환을 영상처리나 인공지능을 이용하여 관심영역을 분할하는 것은 필수적인 단계이다. 이 과정에서 분할이 제대로 이루어지지 못한다면 이후의 정량분석에서 큰 오차가 발생하기 때문에 관심영역을 분할하기 위한 알고리즘

Table 1. Measuring techniques for sarcopenia

Technique	Measurements	Comments
MC	Calf and mid arm circumferences as measures of muscle size	Frequent estimation errors
BIA	Alternating electrical current through body tissue	Loss of accuracy and reliability
DXA	Attenuation of 2 X-ray energies, Total skeletal muscle mass	Fairly accurate and reliable, low radiation exposure
CT	Cross-sectional muscle size quantification	Expensive and radiation exposure
MRI	Cross-sectional muscle size quantification	Expensive and time consuming
SPPB	Lower extremity function	Validated tool for older people
Ultrasound	Analyze key parameters of skeletal muscle structure	Simple, but few validation studies
MB	Valuable information concerning muscle quality	Not widely described as a diagnostic
LM	Biochemical indicators are discussed in Assessing nutritional status and muscle mass	No specific recommendations, references, or cut-off values

* **MC**, muscle circumferences; **CT**, computed tomography; **MRI**, magnetic resonance imaging; **BIA**, bio-electric impedance analysis; **DXA**, dual energy X-ray absorptiometry; **SPPB**, short physical performance battery; **MB**, muscle biopsy; **LM**, laboratory measurements



(a) Normal Case-1 (b) Normal Case-2

(c) Worst Case-1 (b) Worst Case-2

Fig. 1. Various CT Imaging Cases

이 매우 중요하다[16]. 일반적으로 컴퓨터 단층촬영 영상에서는 각 영역별로 HU(Hounsfield Unit)을 갖는다[17]. 그렇기 때문에 보통은 HU 값을 기반으로 단순한 임계처리 방법을 사용하여 영상의 영역 분할이 가능하다. 그러나 근감소증 진단을 위한 영상분할은 근육, 내장지방, 피하지방 부위를 분할하는데 Fig. 1(a), 1(b)와 같이 피하지방, 근육, 내장지방으로 정확하게 분리되는 영상들이 있는가 하면 Fig. 1(c), 1(d)와 같이 안 좋은 케이스의 영상들이 있다. 해당 영상들은 단순한 임계처리 방법을 사용하여 분할하기에는 어려움이 따른다. Fig. 1(a), 1(b)는 명확하게 경계의 구분이 되는 영상이긴 하지만 피하지방과 내장지방 부위가 동일한 HU값을 가지고 있기 때문에 별도의 처리가 필요하다.

Fig. 1(c)의 파란색 박스 부위는 영역별 경계가 명확하지도 않고 근육부위에 지방이 끼어 있는 형태로 영역분할이 어려운 케이스이다. Fig. 1(d)의 파란색 박스 부위는 수술로 인하여 복부 내부에 보조장치를 착용한 케이스로 영역분할에 어려움이 있는 케이스이다. 비록 일부 도메인에 한정되지만 인공지능이 인간 수준의 패턴인식이 가능하다는 사실이 다양한 데이터로부터 입증되었다[18]. 따라서 이런 어려운 케이스를 포함하여 인공지능 기술을 적용하여 영상분할 모델을 개선하고자 하였다. 인공지능 기반 분할 모델은 합성곱 신경망의 등장 이후로 다양하게 제안되었다. 그중 특히 U-Net[19]은 의료영상 분야에서 영상분할을 목적으로 가장 널리 사용되는 구조 중에 하나이다. U-Net은 입력 이미지의 영상 사이즈를 줄여가면서 특징을 찾는 좌측부분과 최종 특징 맵으로부터 원하는 영상 사이즈로 복원해 나가면서 보다 높은 해상도의 분할 결과를 얻기 위한 우측부분으로 구성되어 있다. 따라서 3장에서는 U-Net을 사용하여 근감소증을 위한 인공지능 영상분할 모델에 대해서 설명하고자 한다.

3. 근감소증을 위한 AI 영상분할 모델

3.1 근육량 측정을 위한 AI 영상분할 모델 개발

근감소증 진단을 위한 근육량 측정을 위한 AI 영상분할 모델은 U-Net으로 개발하였다.

Fig. 2는 U-Net 모델의 전체 구조를 보이고 있다. 입력으로 복부 CT 데이터와 라벨링 데이터를 데이터셋으로 학습하여 모델을 생성하였다. 우리는 기존 근감소증 인공지능 모델 개발을 위한 개발 프로세스와 데이터구성, 인공지능 모델에 대한 연구를 논문으로 발표하였다[20]. 이전 연구에서의 분할 모델은 데이터셋은 Train(100), Test(50)으로 학습과 테스트를 진행하였다. 충분한 학습이 되기에는 데이터셋이 부족하기 때문에 영상의 윈도윙을 변경하고 데이터 증강을 통하여 20,000장의 데이터셋을 구성하였고 Train, Validation을 각각 18,000장과 2,000장으로 데이터를 증강하여 학습을 진행하였다.

그러나 Fig. 3과 같이 동일한 영상의 윈도윙을 변경하고 데이터 증강을 적용하여 생성한 영상 분할 결과가 좋지 않은 경우가 발생되었다. 변형된 윈도윙으로 인하여 비슷한 윈도

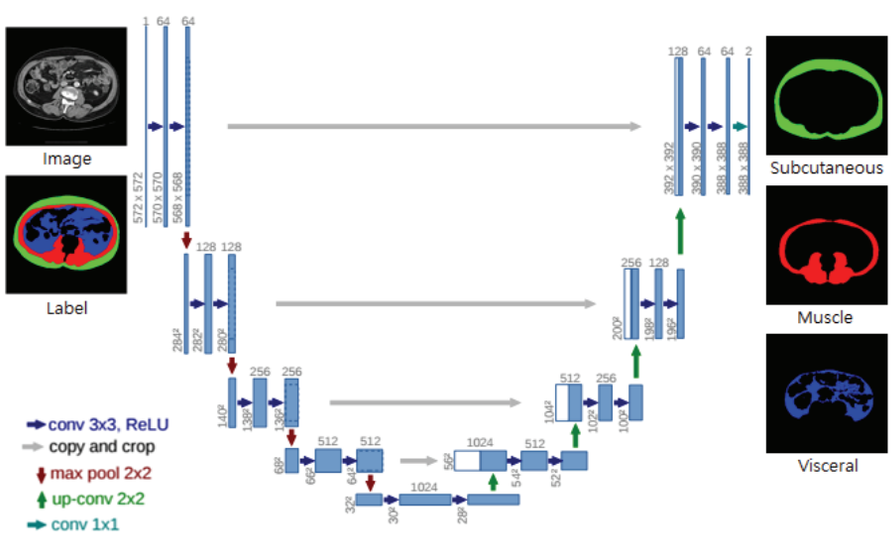


Fig. 2. Overall Structure of U-Net

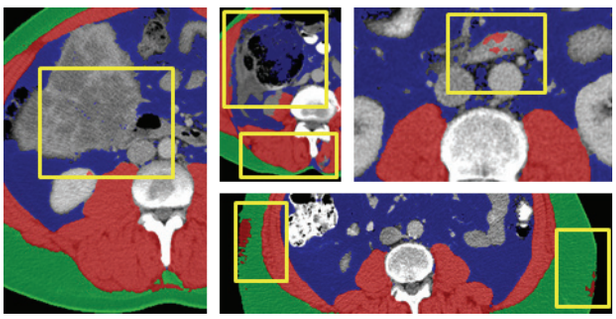


Fig. 3. Segmentation Error

윙을 가진 다른 부분을 분할하는 것을 노란색 박스 부분을 통하여 확인할 수 있었다. 이런 문제로 인하여 우리는 실제 학습 데이터를 확보하여 라벨링을 진행하였고 Train(1480), Test(370)의 데이터셋으로 학습데이터의 부족한 문제점을 해결하기 위해 수행했던 윈도윙과 데이터 증강을 적용하지 않고 실제 데이터만 가지고 학습하였다. 학습을 위한 하이퍼파라미터는 Table 2와 같다. 학습을 위한 하이퍼파라미터에서 특징이 있는 부분은 손실함수이다. 보통 분할 모델에서의 손실함수는 Dice Loss를 사용하지만 Dice Loss 단독으로 사용할 경우 분할이 제대로 안되는 경우가 빈번하게 나왔으며, 성능저하의 결과를 초래하였다.

따라서 Dice Loss[21] 비율 0.5와 Cross-Entropy Loss [22] 비율 0.5를 결합한 Combined Loss를 사용하였다. Dice Loss를 활용하는 이유는 CT 데이터 분할에 있어서 배경인 검은 부분이 사람의 몸체에 비해 많은 부분을 차지하고 있어 클래스 불균형[23,24]에 대한 부분을 Dice Loss를 적용하여 해결하지 않으면 적게 분포하는 클래스에 대해 의도하지 않는 방향으로 학습이 전개될 가능성이 높기 때문이다. Dice Loss와 Cross-Entropy Loss에 가중치를 두고 혼합하여 사용하면 Dice Loss로 클래스 불균형을 보정해 주면서 Cross-Entropy Loss로 Pixel의 정확도를 유지하는 Loss 함수를 사용하였다.

개발된 분할 모델의 학습 결과는 val_loss(0.1730), val_accuracy(0.9713)으로 History Plots은 Fig. 4와 같다. Test 데이터로 개발된 분할 모델의 성능 평가를 해본 결과 Table 3과 같은 IOU값을 얻었다.

분할된 세부분의 영역이 근감소증을 판단하는데 있어서 중요한 지표로 사용되기 때문에 S(피하지방), V(내장지방), M(근육), M+S+V(세부분 영역의 합)으로 분리하여 IOU를 계산하였다. IOU 계산 결과 90%가 넘는 데이터가 전체 데이터 중에 92%가 될 정도로 높은 정확도를 가지는 것을 확인하였다.

3.2 AI 영상분할 모델을 적용한 라벨링 애플리케이션

우리가 제안한 AI 영상분할 모델은 라벨링 애플리케이션에 적용하여 근육, 피하지방, 내장지방 영역을 자동으로 분할하여 환자의 근육량을 측정하는 방법으로 의료영상기반의 근감소증 진단에 중요한 지표로 사용할 수 있도록 하였다.

라벨링 애플리케이션에서는 전체 복부 CT 데이터를 업로드하면 업로드한 Axial 영상을 기반으로 Fig. 5와 같이 Coronal, Sagittal 영상을 재구성하여 보여준다. L3 영역의 판단

Table 2. Hyperparameters for Learning

Name	Description
Input Size	512 x 512
Model	U-Net
Loss function	Dice Loss*(0.5) + Cross Entropy Loss*(0.5)
Batch size	4
Learning rate	2e-4
Epochs	50, Patience=10
Optimizer	RMSProp

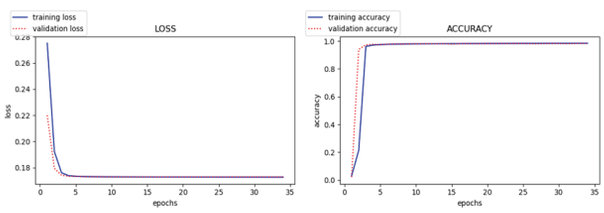
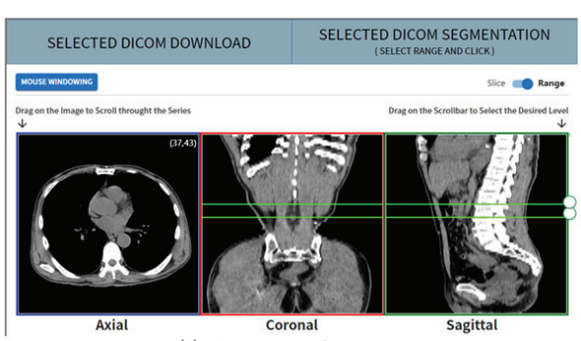


Fig. 4. History plots of Learning results

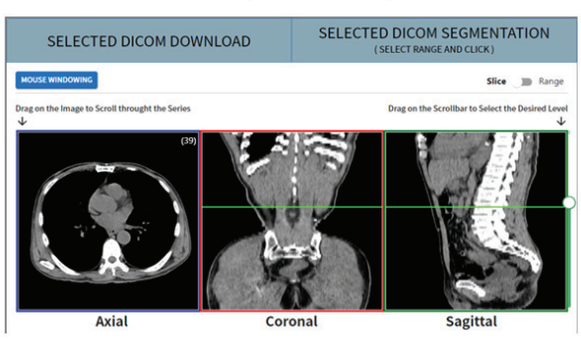
Table 3. IOU Results of Test Dataset

	IOU	M	S	V	M+S+V	Ratio
1	0.9 ~ 1.00%	355	367	291	351	1364/1480 (92%)
2	0.8 ~ 0.89%	14	2	71	18	105/1480 (7%)
3	0.7 ~ 0.79%	1	1	8	1	11/1480 (1%)
4	~ 0.69%	0	0	0	0	0/1480 (0%)

* M+S+V = Muscle + Subcutaneous + Visceral, M = Muscle, S = Subcutaneous, V = Visceral

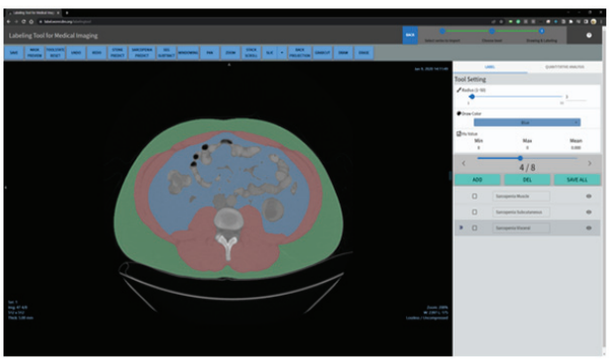


(a) Range : multi image

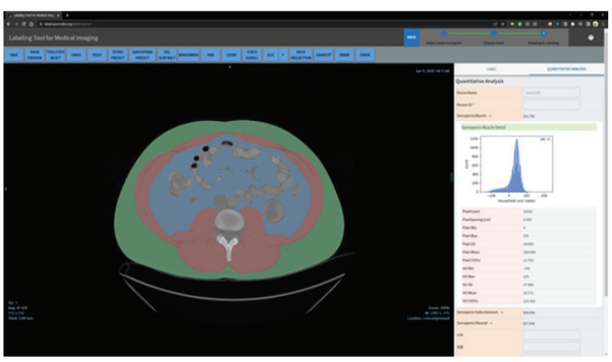


(b) Slice : single image

Fig. 5. L3 Area Selection UI



(a) Image Segmentation



(b) Quantitative Analysis

Fig. 6. Labeling Tool Viewer

은 Axial 영상 만으로는 판단하기 어렵고 재구성한 Sagittal 영상으로 쉽게 판단할 수 있다. 의료영상분석에 대한 연구 목적에 따라 L3 영역 데이터를 한 장만 추출하거나 여러 장을 추출할 수 있도록 개발하였다. 디폴트 세팅은 Fig. 5의 (a)와 같이 L3 영역의 처음과 끝을 범위로 설정하여 여러 장의 이미지 선택이 가능하도록 녹색의 범위 설정라인이 두 개로 시작 이미지와 끝 이미지를 설정할 수 있다. 한 장을 선택할 경우에는 토글 버튼을 Slice로 변경하면 녹색의 범위 설정라인이 하나로 변경되어 대표되는 이미지 하나만 선택할 수 있다. 이렇게 선택된 이미지는 SELECTED DICOM DOWNLOAD 버튼을 클릭하여 다운로드 받을 수 있다. 그리고 SELECTED DICOM SEGMENTATION 버튼으로 라벨링 애플리케이션으로 이동하여 라벨링 작업을 할 수 있다. Fig. 5는 라벨링 애플리케이션 화면을 보인다. Fig. 6의 (a)는 근감소증 인공지능 분할 모델을 사용하여 자동으로 피하지방, 내장지방, 근육 세부분의 영역으로 분할되어 라벨링 데이터가 생성된다. 만약 자동으로 분할된 영역이 잘못된 경우 라벨링 도구를 이용하여 각 라벨을 수정할 수 있다. 이렇게 생성된 라벨링 데이터에 대해 Fig. 6의 (b)와 같이 세부분의 영역에 대한 정량 분석 결과를 확인할 수 있다. 정량분석 결과는 근육, 피하지방, 내장지방에 대한 면적 및 평균값과 표준편차를 정리하였다.

또한 추가적으로 근육면적과 내장지방, 피하지방의 비율을 계산하여 표시하였다. 각 영역에 대한 정량분석 데이터는 추후 근감소증을 진단하는 지표로 사용한다.

4. 검 증(Validation)

우리는 복부 CT 데이터 학습을 통하여 분할 모델을 개발하였고 별도로 수집한 Test 데이터를 통해 검증까지 완료하였다. 그러나 Test 데이터로 사용한 370건만으로는 개발된 인공지능 분할 모델이 실제임상자료(Real World Data, RWD)에 적용하는데 충분한 데이터 셋이라고 보기 어렵다.

특히, 의료 데이터의 특성상 촬영하는 장비와 촬영조건에 따라 다양한 영상이 있을 수 있어 한 병원의 데이터만 가지고 일반화된 성능을 내기 어렵다. 따라서 타 병원의 다양한 데이터로 외부검증이 필요하다. 우리는 한 개이상의 타기관의 복부 CT 데이터를 통해 검증하였다. 해당 데이터로 외부검증 결과의 IOU값은 Table 4와 같다. 비교해 보니 내부검증 결과인 Table 3과 비교하면 정확도가 많이 낮게 나오고 있다. 따라서 IOU 정확도가 낮은 데이터를 실제로 비교한 결과, 타기관 정답 데이터의 근육 표시 부분과 우리가 평가했던 근육 부분이 다른 것을 확인할 수 있었다.

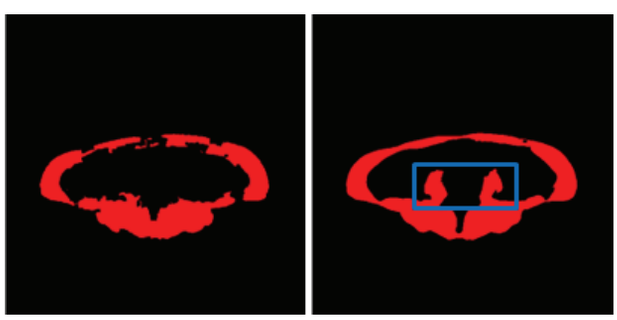
이 결과는 근감소증 연구 목적에 따라 레이블링하는 사람마다 근육이라고 판단하는 영역이 조금씩 차이가 있어 복부 영역의 근육을 라벨링한 결과에서 오는 오차가 발생하였다.

차이가 많이 나는 대표적인 결과의 예로 Fig. 7과 같이 실제 (a) 타 병원 데이터와 (b) 원내 데이터의 라벨링 결과가 차이가 있음을 확인할 수 있었다. (b)의 파란색 박스 부분이 실제 척추를 감싸고 있는 근육 부위이지만 타기관의 라벨링 결과는 해당 부위를 포함하지 않았음을 확인하였다. 이를 통해 공동연구 시에는 라벨링 결과가 서로 통일화할 수 있는 방법을 모색하여 동일한 라벨링 결과가 나왔을 때 보다 좋은 결과를 얻을 수 있다. 이로 인하여 라벨링 데이터의 중요성을 확인할 수 있었다. 앞서 수행한 외부검증 데이터 라벨링 결과가 달라 정확한 비교가 어렵다고 판단되어 동일한 검증용 데이

Table 4. IOU Results of External Validation Dataset

	IOU	M	Ratio
1	0.9 ~ 1.00%	38	38/143 (27%)
2	0.8 ~ 0.89%	61	61/143 (43%)
3	0.7 ~ 0.79%	34	34/143 (24%)
4	~ 0.69%	10	10/143 (6%)

* M = Muscle



(a) other hospital data (b) in-hospital data

Fig. 7. Differences in Labeling Data

Table 5. IOU Results of External Validation Dataset from 4 Hospitals

	n	IOU	MSV	Ratio
A	689	0.9 ~ 1.00%	653	653/689(95%)
		0.8 ~ 0.89%	36	36/689(5%)
		0.7 ~ 0.79%	0	0/689(0%)
		~ 0.69%	0	0/689(0%)
B	576	0.9 ~ 1.00%	558	558/576(97%)
		0.8 ~ 0.89%	18	18/576(3%)
		0.7 ~ 0.79%	0	0/576(0%)
		~ 0.69%	0	0/576(0%)
C	666	0.9 ~ 1.00%	629	629/666(94.4%)
		0.8 ~ 0.89%	36	36/666(5.4%)
		0.7 ~ 0.79%	0	0/666(0%)
		~ 0.69%	1	1/666(0.2%)
D	718	0.9 ~ 1.00%	682	682/718(95%)
		0.8 ~ 0.89%	36	36/718(5%)
		0.7 ~ 0.79%	0	0/718(0%)
		~ 0.69%	0	0/718(0%)

* MSV = Muscle + Subcutaneous + Visceral

Table 6. IOU Results by Percentage of External Validation Dataset

	IOU	MSV	Ratio
1	0.9 ~ 1.00%	2522	2522/2649(95.2%)
2	0.8 ~ 0.89%	126	126/2649(4.7%)
3	0.7 ~ 0.79%	0	0/2649(0%)
4	~ 0.69%	1	1/2649(0.1%)

* MSV = Muscle + Subcutaneous + Visceral

터셋을 타기관에서 개발한 복부근육량 분할 결과[25]와 우리가 개발한 인공지능 이미지 분할 결과로 검증을 진행하였다. 이번 검증에는 4개 기관의 데이터를 수집하고 L3 영역을 추출하여 검증을 진행하였다. 4개 기관의 데이터 현황과 검증 결과의 IOU값은 Table 5와 같다. 우리가 개발한 모듈에서 Segmentation된 결과를 각 기관별 그룹으로 IOU를 계산한 결과이다.

Table 6에서는 전체 데이터의 분포를 보기 쉽도록 정리하였다. 2649건 중에 0.9~1.0%가 2,522건으로 95.2%, 0.8~0.9%가 126건으로 4.7%, 0.6~0.7%가 1건으로 0.1%로 검증 결과 국내의 여러 병원에서 촬영한 데이터를 가지고도 좋은 성능을 보이고 있다. 따라서, 검증결과를 통해 개발한 모델이 정확도와 범용성이 개선되었음을 확인하였다.

5. 결 론

본 논문에서는 근감소증을 위한 AI 영상분할 모델을 개발하였다. 이를 위해 인공지능 모델을 라벨링 애플리케이션에 적용하여 복부 CT 이미지에서 근감소증 진단에 사용하는 요추 L3 영역을 추출하여 다운로드하거나 피하지방, 내장지방, 근육을 분할할 수 있도록 개선하였다. 그리고 내부검증으로 높은 정확도를 확인하였다. 그러나 타기관에서도 사용할 수 있는 일반화된 성능을 갖는 인공지능 모델임을 증명하고자 하였다. 하지만 첫 번째 외부검증에서는 타 병원의 라벨링 데이터와 딥러닝 모델에서 분할하는 라벨링 데이터의 근육 분할 영역의 차이로 인하여 오차가 발생하였다. 두 번째 외부검증에서는 동일한 라벨링데이터로 검증한 딥러닝 모델의 IOU를 측정한 결과, 정확도와 일반화된 성능을 갖는 모델임을 보였다. 향후연구에서는 아직까지도 0.8~0.89% 영역에 있는 데이터가 각 그룹별로 5% 정도 있는 상황이어서 해당 부분의 오차를 줄일 수 있도록 딥러닝 모델을 개선할 계획이다. 또한 더 많은 근감소증 임상데이터를 확보하여 외부검증을 통해 일반화된 성능을 갖는 인공지능 모델로 개선 할 계획이다.

References

[1] J.-Y. Lim, “Consensus update on definition and diagnosis for sarcopenia,” *Geriatric Rehabilitation*, Vol.10, No.2, pp.39-45, 2020.

[2] G. Grimby and B. Saltin, “The ageing muscle,” *Clin Physiol*, Vol.3, No.3, pp.209-18, 1983. doi: 10.1111/j.1475-097x.1983.tb00704 .x. PMID: 6347501.

[3] A. J. Cruz-Jentoft et al., “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.412-23, 2010. doi: 10.1093/ageing/afq034. Epub 2010 Apr 13. PMID: 20392703; PMCID: PMC2886201.

[4] O. Wallengren et al., “Comparison of the 2010 and 2019 diagnostic criteria for sarcopenia by the European Working Group on Sarcopenia in Older People (EWGSOP) in two cohorts of Swedish older adults,” *BMC Geriatr*, Vol.21, No.1, pp.600, 2021. https://doi.org/10.1186/s12877-021-02533-y.

[5] L. K. Chen et al., “Sarcopenia in Asia: Consensus report of the Asian Working Group for Sarcopenia,” *Journal of the American Medical Directors Association*, Vol.15, No.2, pp.95-101, 2014.

[6] L. K. Chen et al., “Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment,” *Journal of the American Medical Directors Association*, Vol.21, No.3, pp.300-307, e2, 2020.

[7] C.-S. Lee et al., “Implementation of medical image labeling web application for machine learning,” *Proceedings of the Korea Information Processing Society Conference*, pp.602-605. 2021, doi:10.3745/PKIPS.Y2021M11A.602.

[8] R. B. Parikh, Z. Obermeyer, and A. S. Navathe, “Regulation of predictive analytics in medicine,” *Science*, Vol.363, No.6429, pp.810-812, 2019. doi:10.1126/science.aaw0029.

[9] C. L Ramspek, K. J. Jager, F. W. Dekker, C. Zoccali, and M. van Diepen, "External validation of prognostic models: What, why, how, when and where," *Clinical Kidney Journal*, Vol.14, Iss.1, pp.49-58, 2021. https://doi.org/10.1093/ckj/sfaa188.

[10] S. E. Bleeker et al., "External validation is necessary in prediction research: A clinical example," *Journal of clinical epidemiology*, Vol.56, No.9, pp.826-832, 2003.

[11] L. L. G. C. Ackermans et al., "Screening, diagnosis and monitoring of sarcopenia: When to use which tool?" *Clinical Nutrition ESPEN*, Vol.48, pp.36-44, 2022.

[12] V. E. Baracos, V. C. Mazurak, and A. S. Bhullar, "Cancer cachexia is defined by an ongoing loss of skeletal muscle mass," *Annals of Palliative Medicine*, Vol.8, No.1, pp.3-12, 2018.

[13] M. T. Imboden, A. M. Swartz, H. W. Finch, M. P. Harber, and L. A. Kaminsky, "Reference standards for lean mass measures using GE dual energy x-ray absorptiometry in Caucasian adults," *PLoS One*, Vol.12, No.4, Article e0176161, 2017.

[14] K. Engelke, O. Museyko, L. Wang, and J.-D. Laredo, "Quantitative analysis of skeletal muscle by computed tomography imaging—state of the art," *Journal of orthopaedic translation*, Vol.15, pp.91-103, 2018.

[15] S. Bamba et al., "Sarcopenia is a predictive factor for intestinal resection in admitted patients with Crohn's disease," *PLoS One*, Vol.12, Vol.6, e0180036, 2017.

[16] R. Hemke, C. G. Buckless, A. Tsao, B. Wang, and M. Torriani, "Deep learning for automated segmentation of pelvic muscles, fat, and bone from CT studies for body composition assessment," *Skeletal Radiology*, Vol.49, No.3, pp.387-395, 2020.

[17] L. Boeri et al., "Hounsfield unit attenuation value can differentiate pyonephrosis from hydronephrosis and predict septic complications in patients with obstructive uropathy," *Scientific Reports*, Vol.10, No.1, pp.18546, 2020. https://doi.org/10.1038/s41598-020-75672-8.

[18] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, Vol.61, pp.85-117, 2015. doi: 10.1016/j.neunet.2014.09.003. Epub 2014 Oct. 13. PMID: 25462637.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *arXiv preprint arXiv:1505.04597*, 2015. 5.

[20] S.-H. Noh et al., "Development of image segmentation model for sarcopenia diagnosis and its application," *Proceedings of the Korea Information Processing Society Conference*, pp.577-79, 2021, doi:10.3745/PKIPS.Y2021M11A.577.

[21] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalized dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Proceeding of Deep Learning in Medical Image Analysis and Multimodal Laerning for Clinical Decision Support*, pp.240-248, 2017.

[22] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *arXiv preprint arXiv:1805.07836*, 2018.

[23] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, Vol.6, No.1, pp.1-5, 2019. https://doi.org/10.1186/s40537-019-0192-5.

[24] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *CoRR abs/1710.05381 arXiv:1710.05381*, 2017. http://arxiv.org/abs/1710.05381.

[25] H. J. Park et al., "Development and validation of a deep learning system for segmentation of abdominal muscle and fat on computed tomography," *Korea Journal of Radiology*, Vol.21, No.1, pp.88-100, 2020. http://doi.org/10.3348/kjr.2019.0470.

이 충 섭

https://orcid.org/0000-0003-1276-4000
e-mail : cslee99@wku.ac.kr
2005년 원광대학교 전기전자 및 정보공학부(학사)
2011년 원광대학교 컴퓨터공학과(석사)
2007년 ~ 2014년 인피니트 헬스케어 과장
2014년 ~ 2018년 레이언스 책임연구원
2018년 ~ 현 재 원광대학교 의료융합연구센터 책임연구원
관심분야 : Medical Imaging, Deep Learning, CDM

임 동 욱

https://orcid.org/0000-0001-6995-2745
e-mail : dw316@wku.ac.kr
2018년 원광대학교 컴퓨터공학과(학사)
2019년 ~ 2020년 전주대학교 정보통신지원실
2020년 ~ 현 재 원광대학교 의료융합연구센터 연구원
관심분야 : Medical Imaging, Deep Learning, CDM, Labeling Tool, Cloud Service

노 시 형

https://orcid.org/0000-0002-1992-6239
e-mail : nosij123@wku.ac.kr
2015년 원광대학교 전기전자 및 정보공학부 (학사)
2017년 원광대학교 컴퓨터공학과(석사)
2017년 ~ 현 재 원광대학교 의료융합연구센터 연구원
관심분야 : Medical Imaging, Machine Learning, Big Data, CDM

김 태 훈

https://orcid.org/0000-0002-2552-0665
e-mail : tae_hoonkim@wku.ac.kr
2010년 전남대학교 의공학협동과정(박사)
2006년 ~ 2009년 전남대학교 의과대학 의과학연구소 연구조교
2009년 ~ 2014년 전남대학교병원 의생명연구원 연구원
2010년 ~ 2013년 전남대학교병원 영상의학연구소 Postdoctoral fellow
2010년 ~ 2014년 전남대학교 산학협력단 의과대학 연구원
2014년 ~ 현 재 원광대학교 의료융합연구센터 연구교수
2020년 ~ 2022년 원광대학교병원 스마트사업팀 교수
관심분야 : 의료영상 분석, 정량분석 소프트웨어, 영상의학 분야 임상 연구, 빅데이터・공통데이터 모델(CDM) 분석 및 활용

고 유 선

https://orcid.org/0000-0002-2181-9555
e-mail : ko.yousun82@gmail.com
2013년 고려대학교 보건대학원 역학 및 보건정보학과(석사)
2019년 서울대학교 융합과학기술대학원 방사선융합의학전공(박사)
2020년 ~ 2022년 서울아산병원 의생명연구소 박사후연구원
2022년 ~ 현 재 울산대학교 의과대학 의학과 연구조교수
관심분야 : Deep learning, AI-based healthcare, Sarcopenia, Clinical Trial, Digital Data Platform for Clinical Trial

김 경 원

https://orcid.org/0000-0002-1532-5970
e-mail : medimash@gmail.com
2002년 서울대학교 의과대학 의학과(학사)
2009년 서울대학교 의과대학원 의학과(석사)
2012년 서울대학교 의과대학원 의학과(박사)
2006년 ~ 2011년 서울대학교병원 영상의학과 전공의/임상강사
2012년 ~ 2013년 다나파버 암센터 하버드 의대
2012년 ~ 현 재 서울아산병원 영상의학과 부교수
2017년 ~ 현 재 서울아산병원 Asan Image Metrics 담당교수
관심분야 : Drug development: Preclinical to clinical research, Preclinical and Clinical Application of MRI, Clinical Trial Design and Operation, Digital Data Platform for Clinical Trial

정 창 원

https://orcid.org/0000-0002-9305-4686
e-mail : mediblue@wku.ac.kr
2003년 원광대학교 컴퓨터공학과(박사)
2004년 ~ 2006년 전북대학교 학술연구교수
2006년 ~ 2013년 원광대학교 전기전자 및 정보공학부 강사
2013년 ~ 현 재 원광대학교 의료융합연구센터 연구교수, 실장
2020년 ~ 현 재 원광대학교병원 의생명연구원 중점연구센터 교수
2020년 ~ 현 재 원광대학교병원 CDM 기관관리자
관심분야 : Bigdata, AI Platform, Medical Data Analysis, Smart healthcare, Common Data Model for Medical information, Radiology_CDM