

근감소증 진단을 위한 영상분할 모델 개발 및 외부검증

이충섭¹, 임동욱¹, 김지언¹, 노시형¹, 유영주¹, 김태훈^{1,2}, 정창원^{1,2}

¹원광대학교 의료융합연구센터

²원광대학교병원 정보관리실, 스마트사업팀

e-mail : {cslee99, dw1316, kakasky112, nosij123}@wku.ac.kr,

yeriel.ryu@gmail.com, {tae_hoonkim, mediblue}@wku.ac.kr

Development of Image Segmentation Model for Sarcopenia Diagnosis and Its External Validation

Chung-sub Lee¹, Dong-Wook Lim¹, Ji-Eon Kim¹, Si-Hyeong Noh¹,

Yeong-Ju Yu¹, Tae-Hoon Kim^{1,3}, Chang-Won Jeong^{1,3}

¹Medical Convergence Research Center, Wonkwang University

²Smart Business Team, Information Management of Wonkwang University Hospital

요 약

근감소증은 영양부족, 운동량 감소 그리고 노화 등으로 정상적인 근육의 양과 근력 및 근 기능이 감소하는 질환을 말한다. 근감소증은 보편적으로 유럽 근감소증 실무그룹분석(EWGSOP)에서 정의한 측정 방법을 따른다. 본 논문에서는 근감소증 진단을 위한 영상 분할 모델을 개발하고 외부검증하는 방법에 대해서 제안한다. 우리는 CT 영상에서 L3 영역을 선별하여 자동으로 근육, 피하지방, 내장지방을 분할할 수 있는 인공지능 모델을 U-Net을 사용하여 개발하였다. 또한 모델의 성능을 평가하기 위해서 분할영역의 IOU(Intersection over Union)를 계산하여 내부검증을 진행하였으며, 타 병원의 데이터를 이용하여 같은 방법으로 외부검증을 진행한 결과를 보인다. 검증 결과를 토대로 문제점과 해결방안에 대해서 고찰하고 보완하고자 했다.

1. 서론

근육은 뼈를 지탱해 몸을 움직이게 하고, 신체의 각 기관을 보호하는 중요한 부위이다. 하지만 신체는 나이가 들면서 서서히 근육량이 줄어들고, 코로나 19 장기화로 실내 활동시간이 많아지고 운동량 부족으로 근감소증 발생 위험이 커지고 있다. 근감소증이란 영양 부족, 운동량 감소, 노화 등으로 인해 정상적인 근육의 양과 근력 및 근 기능이 감소하는 질환으로 대개 20~30세에 근육량이 최대가 되었다가 40대 이후부터 근육이 감소하기 시작하며 70대까지 10년에 8%의 감소가 일어난다고 추정하고 있으며, 그 이후로는 더욱 급격한 감소가 발생하여 10년 마다 15%까지 발생한다고 알려져 있다[1]. 그리고 80대 정도가 되면 근육의 절반까지 감소하게 된다. 근감소증은 EWGSOP라는 유럽의 근감소증 실무그룹에서 정의한 방법으로 근육량 측정과 핸드그립 또는 근육성능을 측정하여 진단하는 공통적인 방법이 있다[2]. 특히 의료영상 중 CT에서 3번 요추 수준의 단면에서 보이는 골격근의 면적 값을 이용한 L3 skeletal muscle index(SMI)가 총 골격근 측정을 대신해서 사용되고 있다[3]. 우리는 이전 논문에서 인

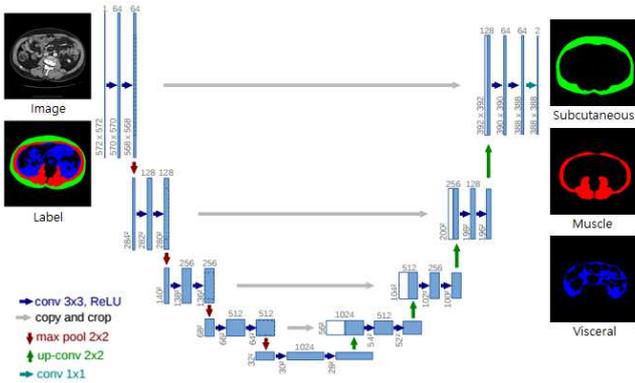
공지능을 이용한 L3 부위의 근육, 내장지방, 피하지방 부위를 분할하고 정량 분석할 수 있는 애플리케이션을 발표하였다[4]. 하지만 단일 병원의 데이터로 모델을 개발한 경우 범용성을 보장하지 못함을 확인하였다. 검증 시 좋은 결과를 얻지 못하였다. 본 논문은 인공지능 모델의 정확도를 4개의 타 병원 데이터를 추가하여 향상시켰으며 외부검증(External Validation)[5,6,7]을 통하여 모델의 정확도를 향상시키고자 하였다.

2. 연구내용

2-1. 근감소증 분할 모델

근감소증 분할 모델은 U-Net[8]으로 개발되었다. (그림 1)은 U-Net의 전체 구조를 보이고 있다. (그림 1)에서 보는 것과 같이 Input으로 복부 CT 데이터와 라벨링 데이터로 구성된 데이터셋으로 학습하여 모델을 생성하였다. 이전 연구의 분할 모델은 데이터셋이 충분하지 못해서 Train:100. Test:50으로 학습과 테스트를 진행하였다. 부족한 데이터셋으로는 충분한 학습이 되지 않기 때문에 Windowing 변경, Augmentation을 통하여 5,000장으로 데이터를

증강하여 학습하도록 하였다.



(그림 1) U-Net의 전체 구조

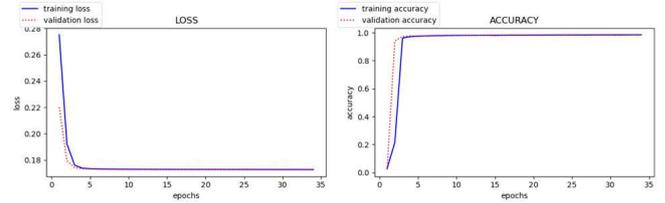
그러나 분할 결과가 좋지 않은 이미지들을 보면 변형된 Windowing으로 인하여 비슷한 Windowing을 가진 다른 부위를 분할하는 것을 확인할 수 있었다. 이런 문제로 인하여 우리는 실제 CT 데이터를 더 라벨링하여 Train:1480, Test:370으로 Windowing과 Augmentation을 적용하지 않고 실제 데이터만 가지고 학습하였다. 학습을 위한 하이퍼파라미터는 아래 <표1>과 같다.

<표 1> 학습을 위한 하이퍼파라미터

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

학습을 위한 하이퍼파라미터에서 특징이 있는 부분은 Loss Function이다. 보통 분할 모델에서의 Loss Function은 Dice Loss를 사용하지만 단독으로 사용했을 경우 분할을 제대로 못하는 케이스가 자주 나왔다. 따라서 Dice Loss[9]와 Cross Entropy Loss[10]를 가중치 0.5대 0.5로 결합한 Combined Loss를 사용 하였다. Dice Loss를 활용하는 이유는 CT 데이터 분할에 있어서 배경인 부분이 다른 부분에 비해 월등히 많아 클래스 불균형(Class Imbalance)[11,12]에 대한 부분을 Dice Loss로 잡아 주지 않으면 적게 분포하는 클래스에 대해 예측하지 않는 방향으로 학습이 전개될 가능성이 높기 때문이다. Cross Entropy Loss와 Dice Loss를 가중치를

두고 혼합하여 사용하면 Cross Entropy Loss로 Pixel의 정확도를 유도하면서도 Dice Loss로 클래스 불균형을 보정해 주는 방향으로 Loss 함수를 사용 하였다. 개발된 분할 모델의 학습 결과는 val_loss:0.1730, val_accuracy: 0.9713으로 History Plots은 (그림 2)와 같다.



(그림 2) 학습 결과의 History Plots

Test 데이터로 분할 모델의 알고리즘 성능 평가를 해본 결과 <표 2>와 같은 IOU(Intersection over Union)값을 얻었다. 본 논문에서는 분할된 세영역의 정확도가 근감소증을 판단하는데 있어서 중요한 지표로 사용되기 때문에 M(근육), S(피하지방), V(내장지방), MSV(세영역)으로 분리하여 IOU를 계산하였다. IOU 계산 결과 90%가 넘는 데이터가 전체 데이터 중에 92%가 될 정도로 높은 정확도를 가지는 것을 확인하였다.

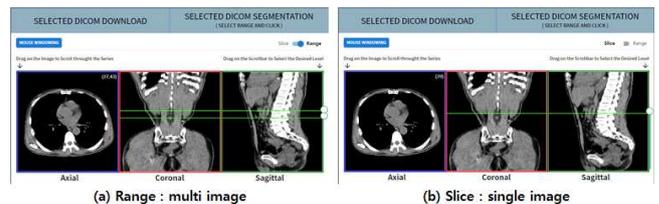
<표 2> Test 데이터셋의 IOU 결과

	IOU	M	S	V	MSV	비율
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%)

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

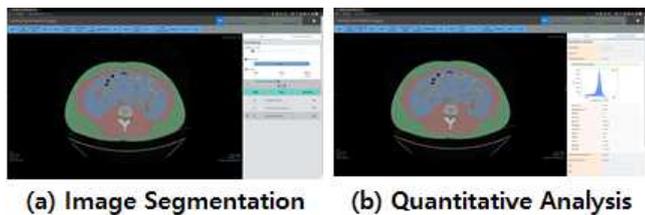
2-2. 근감소증 라벨링 툴

본 논문에서 제안하고 있는 근감소증 모델은 라벨링 툴에 적용하여 근감소증 진단에 중요한 지표로 사용하는 근육, 피하지방, 내장지방 영역을 자동으로 분할할 수 있도록 하였다.



(그림 3) L3 영역 선택 UI

하지만 전체 CT 데이터에서 L3 영역에 해당하는 부위를 선택하는 것 또한 어려운 일이다. 라벨링 툴에서는 전체 CT 데이터를 업로드하면 (그림 3)과 같이 사용자가 L3 영역을 구분하여 판단할 수 있도록 업로드한 Axial 영상을 기반으로 Coronal, Sagittal 영상을 재구성하여 보여준다. 사용자의 연구 목적에 맞도록 L3 영역 데이터를 한 장만 추출하는 사용자가 있고 여러 장을 추출하는 사용자가 있다. 디폴트 세팅은 (그림 3)의 (a)와 같이 여러 장의 이미지를 선택할 수 있도록 가이드라인이 두 개로 이미지의 시작과 끝을 설정할 수 있다. 한 장을 추출할 경우에는 Toggle 버튼을 Slice로 변경하면 가이드라인이 하나로 변경되어 하나만 선택할 수 있다. 여기에서 선택된 이미지는 SELECTED DICOM DOWNLOAD 버튼으로 로컬로 다운로드 받을 수 있다. 그리고 SELECTED DICOM SEGMENTATION 버튼으로 라벨링을 위한 뷰어로 이동하여 라벨링을 할 수 있다. (그림 4)는 라벨링 툴 뷰어 화면을 보인다. (그림 4)의 (a)는 근감소증 분할 모델을 통해서 자동으로 세영역으로 분할되어 라벨링 데이터가 생성되었다. 이렇게 생성된 라벨링 데이터에 대해 (b)와 같이 각 영역에 대한 정량 분석 결과를 확인할 수 있다. 정량분석 데이터는 추후 근감소증을 진단하는 지표로 사용된다.



(그림 4) 라벨링 툴 뷰어

3. 외부검증(External Validation)

우리는 학습을 통하여 분할 모델을 개발하고 별도로 수집한 테스트 데이터를 통해 검증까지 완료하였다. 그러나 테스트 데이터로 사용한 370장만으로는 개발된 분할 모델이 대부분의 데이터에서 일반적인 성능을 낼 수 있다고 평가하기에는 충분한 데이터라고하기 어렵다. 특히, 의료 데이터의 특성상 한 병원의 데이터만 가지고 일반화된 성능을 확신하기 어렵다. 촬영하는 장비와 촬영조건에 따라 다양한 영상이 있을 수 있어 타 병원의 다양한 데이터로 외부검증을 수행해야 한다. 타기관의 데이터는 유병

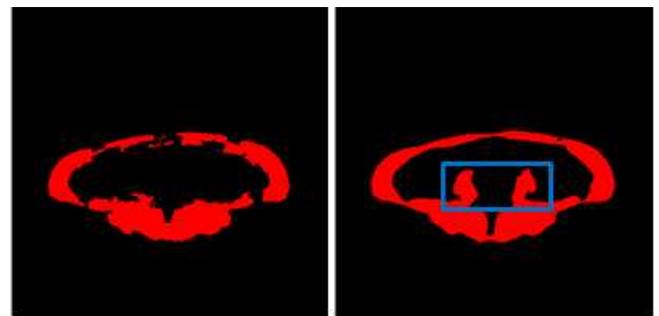
환자로 유방암을 앓고 있는 환자의 복부 CT 데이터이다. 해당 데이터로 외부검증 결과의 IOU값은 <표 3>과 같다.

<표 3> 외부검증 데이터셋의 IOU 결과

	IOU	M	비율
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

외부검증 결과가 내부검증 결과<표 2>와 비교하면 정확도가 많이 낮다. 따라서 IOU 정확도가 낮은 데이터를 실제로 비교해 본 결과 타기관 데이터의 정답 라벨링 데이터의 근육 표시 영역과 우리가 원하는 근육 영역이 다른 것을 확인할 수 있었다. 이는 의사마다 근육이라고 판단하는 영역이 조금씩 차이가 있을 수 있고 근감소증의 연구 목적에 따라 복부 영역의 근육을 측정하는 방식이 다름에서 오는 오차로 볼 수 있었다.



(그림 5) 라벨링 데이터의 차이

차이가 많이 나는 대표적인 결과를 (그림 5)와 같이 실제 (a) 타기관 데이터와 (b) 원내 데이터의 라벨링 결과가 다름을 확인할 수 있다. 이는 특히 (b)의 파란색 박스 부분이 실제 근육 부위이지만 타기관의 라벨링 결과는 포함하지 않았음을 확인하였다.

4. 결론 및 향후연구

본 논문에서는 근감소증을 위한 영상분할 모델을 개발하였고 모델을 라벨링 툴에 적용하여 복부 CT 이미지중 근감소증 진단에 쓰이는 L3 영역을 추출하여 다운로드하거나 근육, 피하지방, 내장지방을 분할할 수 있도록 하였다. 그리고 내부검증을 통하여

높은 정확도를 가지는 것을 확인하였고 외부검증을 통하여 일반화된 성능을 가지는 모델임을 증명하고자 하였다. 하지만 타 병원의 라벨링 데이터와 딥러닝 모델에서 분할하는 라벨링의 근육 표시 영역이 달라 오차가 발생하였다. 향후연구에서는 라벨링의 정확도가 높은 자동분할 모델을 통해 이러한 오차가 발생하지 않도록 라벨링 툴을 개선할 계획이다. 또한 타 병원과 공동연구를 확대해서 근감소증 데이터를 수집하여 외부검증을 늘려서 일반화된 성능을 가지는 모델로 보완할 수 있도록 할 계획이다.

사사표기

본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술 연구개발사업(HI18C1216) 그리고 과학정보통신부의 재원으로 한국연구재단(2021R1A5A8029876)(2020R1I1A1A01074256) 지원에 의하여 이루어진 것임.

참고문헌

[1] Grimby G, Saltin B. The ageing muscle. *Clin Physiol* 1983;3:209-18.

[2] Wallengren, O., Bosaeus, I., Frändin, K. 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* 21, 600 (2021). <https://doi.org/10.1186/s12877-021-02533-y>

[3] Bamba S, Sasaki M, Takaoka A, et al. Sarcopenia is a predictive factor for intestinal resection in admitted patients with Crohn's disease. *PLoS One* 2017;12:e0180036.

[4] Lee, Chung-sub, Dong-Wook Lim, Ji-Eon Kim, Si-Hyeong Noh, Yeong-Ju Yu, Tae-Hoon Kim, and Chang-Won Jeong. "Implementation of Medical Image Labeling Web Application for Machine Learning." *Proceedings of the Korea Information Processing Society Conference*, November 4, 2021, 602 - 5. doi:10.3745/PKIPS.Y2021M11A.602.

[5] Parikh, Ravi B et al. "Regulation of predictive analytics in medicine." *Science (New York, N.Y.)* vol. 363,6429 (2019): 810-812. doi:10.1126/science.aaw0029

[6] Chava L Ramspek, Kitty J Jager, Friedo W Dekker, Carmine Zoccali, Merel van Diepen, External validation of prognostic models: what, why, how, when and where, *Clinical Kidney Journal*, Volume 14, Issue 1, January 2021, Pages 49 - 58, <https://doi.org/10.1093/ckj/sfaa188>

[7] Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;56:826-832

[8] Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint arXiv:1505.04597*, 2015. 5

[9] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, 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 Learning for Clinical Decision Support*, 2017;pp. 240-248.

[10] Zhilu Zhang and Mert R Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *arXiv preprint arXiv:1805.07836*, 2018. 8

[11] Johnson, J.M., Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *J Big Data* 6, 27 (2019). <https://doi.org/10.1186/s40537-019-0192-5>

[12] Mateusz Buda et al. 2017. A systematic study of the class imbalance problem in convolutional neural networks. *CoRR abs/1710.05381* (2017). arXiv:1710.05381 <http://arxiv.org/abs/1710.05381>