

CNN 기반 슬관절 골관절염 중증도 판단을 위한 통합 보완된 등급 판정 시스템

(An Integrated and Complementary Evaluation System for
Judging the Severity of Knee Osteoarthritis Using CNN)

윤 예 찬^{1)*}
(YeChan Yoon)

요 약 슬관절 골관절염(OA, Osteoarthritis)은 전 세계적으로 매우 흔한 근골격계 질환이다. 빠르고 정확한 초기 진단이 필요한 슬관절 골관절염의 등급은 현재 분산된 분류 시스템에 따라 다르게 판정되며, 각 분류 시스템마다 기준이 상이하다. 또한 의료진이 X-ray 사진을 직접 보고 판독하기 때문에 의료진의 주관적인 의견에 따라 달라지며 시간이 많이 소요되어 정확한 진단과 명확한 치료 계획 수립에 시간이 지연되고 있다. 따라서 본 연구는 딥러닝 기술인 CNN을 사용하여 슬관절 골관절염 등급 판단 기준이 되는 협착 부분의 길이 측정 알고리즘과 골극의 탐지 및 길이 측정 알고리즘을 따로 설계하였다. 또한 기존 분류 시스템을 통합 보완한 등급 분류 시스템을 만들어 실제 의료진의 판단과 일치하는 결과를 나타내고자 한다. 공개적으로 사용 가능한 OAI (Osteoarthritis Initiative) 데이터를 기반으로 하여, 총 9,786개의 슬관절 방사선 데이터가 본 연구에 사용되었으며, 최종적으로 Accuracy(정확도) 69.8%, F1 score 76.65%를 달성하였다.

핵심주제어: 슬관절 골관절염, X-ray, 협착, 골극, 인공지능, CNN

Abstract Knee osteoarthritis (OA) is a very common musculoskeletal disorder worldwide. The assessment of knee osteoarthritis, which requires a rapid and accurate initial diagnosis, is determined to be different depending on the currently dispersed classification system, and each classification system has different criteria. Also, because the medical staff directly sees and reads the X-ray pictures, it depends on the subjective opinion of the medical staff, and it takes time to establish an accurate diagnosis and a clear treatment plan. Therefore, in this study, we designed the stenosis length measurement algorithm and Osteophyte detection and length measurement algorithm, which are the criteria for determining the knee osteoarthritis grade, separately using CNN, which is a deep learning technique. In addition, we would like to create a grading system that integrates and complements the existing classification system and show results that match the judgments of actual medical staff. Based on publicly available OAI (Osteoarthritis Initiative) data, a total of 9,786 knee osteoarthritis data were used in this study, eventually achieving an Accuracy of 69.8% and an F1 score of 76.65%.

Keywords: Knee Osteoarthritis, X-ray, Joint Space Narrow, Osteophyte, AI, CNN

* Corresponding Author: acg1221@korea.ac.kr
Manuscript received April 19, 2024 / revised May 28, 2024 /

accepted August 15, 2024
1) 고려대학교 산업경영공학과(산업인공지능), 제1저자, 교신저자

1. 서론

1.1 연구 배경

4차 산업 혁명으로 IT 기술이 발전함에 따라 의료 산업에 인공지능 기술이 활발하게 도입되고 있다. 의료 빅데이터로 학습된 인공지능은 질병의 정확한 진단을 도와 환자에게 최적화된 치료를 제공할 뿐만 아니라 불필요한 절차를 대체할 수 있어 의료 분야에서 실효성을 인정받아 꾸준히 연구가 이뤄지고 있다. 2020년 식약처에서 허가한 AI·빅데이터 기반 소프트웨어 의료 기기는 22건으로, 전년 대비 120%로 증가했다. 또한, 2020년 기준 인공지능·빅데이터 기반 독립형 소프트웨어 의료 기기는 세계 시장 규모가 3조 9,739억, 국내 시장 규모가 772.8억인 가운데, 국내 연평균 성장률 44.6%를 보이는 등 전망이 좋다 (Heo, 2022).

슬관절 골관절염(Osteoarthritis; OA)은 진행을 막을 수 없는 질환으로 알려져 빠르고 정확한 초기 진단이 중요하다. 현재 슬관절 골관절염의 등급은 분산된 6개의 분류 시스템(Kellgren-Lawrence, IKDC, Fairbank, Brandt et al. Ahlbäck, Jäger Wirth)에 따라 상이한 기준으로 다르게 판정된다 (Wright, 2014). 이때 각 분류 시스템에 한계가 존재하여 여러 등급 시스템을 적절하게 조합해서 사용하고 있다. 또한, 의료진이 X-ray 영상을 직접 보고 판독하기 때문에 개인마다 판독 결과의 차이가 크고 시간도 많이 소요된다. 이러한 이유로 현재 슬관절 골관절염에 대한 신속하고 정확한 진단이 어려워 명확한 치료 계획 수립이 지연되고 있다. 따라서 본 연구는 기존 시스템의 한계를 보완한 등급 분류 시스템에 기반한 딥러닝 모델을 설계함으로써 의사들의 판독 기준과 일치하는 등급 예측을 제시하여 빠른 치료 계획 수립과 의료의 질을 개선하고자 한다.

1.2 배경 이론

1.2.1 슬관절 골관절염

퇴행성 관절염은 체중을 많이 받는 무릎, 엉

덩이 관절 등과 관계가 깊은 질환으로 ‘골관절염’, ‘골 관절증’이라고도 불린다. 슬관절 골관절염은 무릎의 자연스러운 움직임을 방해하는 근골격계 관절 장애이다. 주요 특징으로 관절 공간 협착, 골극 형성 및 경화증을 포함한다.

퇴행성 관절염의 경우 80세 이상의 인구에서 80% 이상 나타나는 노인성 질환이다. 우리나라 65세 이상의 고령 인구에서 퇴행성 관절염의 유병률은 37.8%이며, 남성은 20.2%, 여성은 50.1%로 여성이 남성과 비교하면 약 3배 높은 유병률을 보인다 (Korean Association of Knee Joints, 2021). 대표적인 노인성 질환이었던 퇴행성 관절염은 점차 20~30대도 피할 수 없는 질환이 되었다. 건강보험심사평가원의 2018년 자료에 따르면 20~30대 퇴행성관절염 환자 수가 185,375명에서 2019년에는 197,246명으로 1만 명 이상 증가했다. 노인의 발병률이 높은 이유는 노화와 연관된 변화가 관절염 발생 위험을 증가시키기 때문이다. 청년층에서 발생하는 원인으로는 유전, 비만, 격렬한 운동이 있다. 다이어트와 건강을 위한 운동이 무릎에 무리를 주는 경우가 많아졌다.

골관절염 진단의 경우, 통증 부위 및 증상을 확인한 후, 간단한 신체검사와 X선 검사를 통해 진단할 수 있다. 그 외에도 골 주사 검사, 관절액 검사, 관절경 진단, 자기 공명영상(MRI)으로도 진단할 수 있다.

골관절염 치료의 경우 보존적 치료와 수술적 치료로 나뉘게 되며, 보존적 치료의 목적은 통증을 줄이고 기능을 향상시키는 것이다. 슬관절 골관절염 초기 단계의 경우 보존적 치료를 시행하게 되며 4가지 방법이 존재한다. 첫 번째는 생활 습관 변경으로 체중을 감소시키고 무릎에 무리가 되는 운동을 피하는 방법이 있다. 두 번째는 운동으로 하지 근력의 강화와 관절 운동 범위를 증가시키고 관절의 유연성을 향상시킨다. 세 번째는 보조기 사용으로 무릎 보호대, 충격 흡수 신발, 지팡이 등의 보조기구가 안정성을 도와주고 기능을 보조해준다. 마지막으로 약물치료는 운동과 물리요법을 보완하기 위해 사용이 된다. 이는 통증을 감소시켜주며 진행을 늦춰주는 역할을 한다.

수술적 치료는 보존적 치료에 반응하지 않는 경우 시행하게 된다. 수술적 방법으로는 근위 경골 절골술을 이용하여 무릎의 정렬을 이동시켜주는 방법과 인공관절 치환술로 손상된 관절을 대체하는 방법이 주로 시행된다. 지속적으로 골관절염을 치료 할 수 있는 방법을 연구 중이며 관절 연골 이식, 줄기세포 치료법의 연구가 진행되고 있다. 다만, 현실적으로 뚜렷한 성과는 아직 보여지고 있지 않다 (MSD Manual, 2021).

슬관절 골관절염의 경우 위와 같이 완치가 어려운 질병으로, 빠르게 진단하여 더이상 악화하지 않게 막는 것이 중요하다.

1.2.2 인공지능 기반 의료영상

인공지능(Artificial intelligence, AI)이란 컴퓨터 공학의 한 분야로 컴퓨터가 인간의 지적 행동을 모방하도록 하는 것이다. 방대한 양의 데이터 학습을 통해 구축된 높은 정확성과 빠른 처리 속도를 가지고 소프트웨어 기반 분야의 전반에 걸쳐 접목되고 있다. 이는 의료 분야에도 적용되며, 특히 AI 의료영상 분석 기술은 의료영상 분야의 발전을 선도할 핵심 기술로 자리 잡고 있다.

AI 의료영상 분석이란 사전 학습된 AI가 의료영상의 특징(feature)을 추출하여 병변을 진단하는 기술이다. 의료영상에서 특징이란 명도(brightness), 대조도(contrast), 공간 주파수(spatial frequency), 균질성(homogeneity), 곡률(curvature), 길이(length) 등 영상의 데이터를 통해 정량적으로 나타낼 수 있는 것을 의미하며 각 병변은 서로를 구분 짓게 하는 고유 특징을 가진다. AI 의료영상 분석 기술은 특징을 추출하는 방법으로 최근 딥러닝(Deep Learning)에 기반을 둔 학습 모델이 주목을 받고 있다 (Trinder et al., 1997; Harman, 2012; Shen et al., 2017; Ravì et al., 2017; Pesapane et al., 2018; Hong et al., 2020).

딥러닝은 인간의 신경계를 모방한 인공 신경망(Artificial Neural Network ; ANN)의 일종으로, 여러 층으로 쌓인 신경망 내부에서 각종 파라미터(parameter)를 갱신하며 학습을 수행하는 기법이다. 딥러닝은 학습 과정에서 데이터의 특

징을 자체적으로 추출하는데, 영상 기기나 질환 부위에 상관없이 일관된 특징 추출 모델을 가진다. 또한, 일관된 특징 추출 알고리즘을 사용하므로 이미 학습된 모델을 재사용하여, 다른 병변에 대해 추가로 학습시키는 것이 학습(Transfer Learning)이 가능하다. 하지만 딥러닝을 사용하기 위해서는 충분한 양의 데이터가 필요하며, 데이터가 부족하면 학습 데이터에 과적합(Overfitting) 될 가능성이 크다. 의료 데이터의 경우 데이터를 구하는 것이 어렵고 정답 데이터를 만들기 어려운 경우가 많아 특히 데이터의 부족이 심각한 분야이다. 또한, 데이터의 불균형이 생길 가능성이 크다. 원천 데이터를 추가로 구하기 어려운 상황에서 데이터의 다양성을 학습시켜 성능을 높이고 싶을 때 사용하는 기법이 데이터 증강(Data Augmentation)이다. 즉, 데이터 증강 기법은 모델 학습에 사용되는 데이터의 다양한 변이를 배우게 함으로써, 학습 모델의 과적합을 방지하기 위해 유용하게 사용된다 (Lee et al., 2018; Kim and Bae, 2020).

딥러닝 기반 의료영상 분석 방법은 X-ray, MRI, 초음파, 내시경, 병리 영상까지 다양한 의료영상에 적용이 되며, 병변 탐지 및 정량화, 병변의 분류에 관한 연구도 활발히 진행되고 있다. 그러나 딥러닝의 한계점도 존재하는데, 판단에 대한 설명력과 결과에 대한 불확실성을 갖는다. 최근에는 한계점을 보완하기 위해 입력 영상의 공간적 정보를 유지하는 모델 구조를 사용하거나, 임상 정보들을 부가 입력하여 해부학적 사전 지식을 활용 할 수 있도록 하는 방법을 사용한다 (Jeong, 2018).

1.3 선행 연구

슬관절 골관절염은 현재 임상 검사와 필요한 경우 방사선 촬영(X-ray) 및 MRI 및 관절경과 같은 영상 기법으로 진단된다. 방사선 영상은 골극 형성, 관절 공간 협착, 연골하 경화증 및 낭종 형성을 포함한 OA의 특징을 탐지할 수 있다.

현재 골관절염의 중증도 평가 시스템은 여러 개 존재한다. 그중에서 Kellgren and Lawrence

(KL) 등급 시스템이 가장 많이 사용되고 있다. KL 등급 시스템은 골관절염의 중증도에 따라 0~4등급으로 분류된다. 슬관절 골관절염 등급은 관절의 공간 폭 변화(Joint Space Width)와 골극 형성(Osteophytes Formation)을 기준으로 분류가 된다. 하지만 KL 등급 시스템도 정확하게 모든 골관절염을 분류하기에는 모호함을 갖고 있다. 현재에도 의료진의 주관성에 의해 판단이 이루어져 등급 기준이 판단하는 사람에 의해 달라진다.

자동 골관절염 감지 및 등급 분류를 위한 현재의 State of the Art(SOTA)는 딥러닝과 같은 기계학습을 기반으로 하며, 특히 CNN 기반 딥러닝 기술을 기반으로 한다. 하지만 딥러닝 모델은 골관절염과 관련된 해부학적인 특징은 고려되지 않는다. 골관절염 평가를 위한 관절 공간 너비의 직접적 측정과 근위 경골 사이의 분리와 같은 사항들이 포함되지 않는다. 또한, CNN 모델은 texture(질감)에 의존적이고 객체의 질감 없이 모양을 인식하는 성능이 좋지 않음을 보인다. 그 외에도 딥러닝 모델의 문제점으로 입력 데이터의 어떤 정보가 결과를 내는데 기여했는지 알 수 없다는 문제점이 존재한다 (Antony et al., 2016; Geirhos et al., 2018; Bayramoglu et al., 2020a).

이 연구에서 Fig. 1과 같이 무릎 관절을 지역화하는 FCN (Fully Convolution Neural Network) 기반 관심 영역 추출(Region of Interest, ROI) 방법을 소개한다. 슬관절 골관절염의 중증도 등급은 관절의 공간 폭의 변화와 무릎 관절의 골극 형성을 감지하여 평가할 수 있다 (Antony et al., 2017).

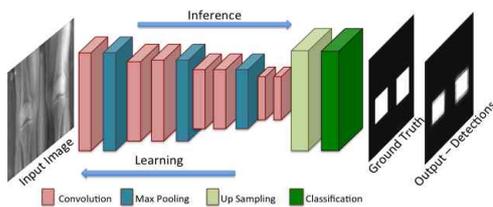


Fig. 1 FCN-Based ROI Extraction Process

무릎 관절을 지역화하는 것은 중증도를 정량화하기 위한 사전 단계이며 큰 데이터 셋일 경우 선호되는 방법이다. 무릎 관절을 감지하는 정확도와 정밀도를 높이기 위해 FCN 기반 모델을 제안하며 CNN 모델의 백본(Backbone)으로 VGG와 ResNet을 사용한다.

OAI(Osteoarthritis Initiative) 데이터 셋과 MOST(Multicenter Osteoarthritis Study) 데이터 셋을 활용하여 훈련과 테스트를 진행한다. 그 결과 Jaccard index(J) 값을 기준으로 0.5 이상의 값을 갖는 정확도는 99.9%이고, J 값이 0.75 이상인 정확도는 91.4%로 자동으로 추출된 무릎 관절이 수동으로 얻은 무릎 관절과 비슷한 결과를 보여주었다 (Bayramoglu et al., 2020b).

본 연구에서도 수동으로 찾는 방법이 아닌 자동으로 무릎 관절을 지역화해주는 FCN 기반의 ROI 모델을 설계하였다.

1.4 연구의 차별성

본 연구는 슬관절 골관절염의 등급을 판정하는 기존 시스템들과 다음과 같은 차별성을 가진다. 첫째로 Kellgren Lawrence(KL) 등급 시스템에 단일 의존한 기존의 연구와 달리 국내 슬관절 골관절염 전문 의료진들과 분류한 등급 체계를 이용한다. 이는 KL 시스템을 기반으로 하되, 의료진들의 임상 경험을 바탕으로 기존의 다른 시스템들의 기준을 적절하게 조합한 것이다. 기존 각 시스템에 존재했던 한계를 보완한 통합 등급 분류 시스템을 사용하여 적합한 최적의 등급을 산출한다.

둘째로 하나의 모델에서 골극, 협착, 등급을 모두 판단하는 기존의 연구와 달리, 골극의 유무, 골극의 폭, 협착 간격을 측정하는 모델을 따로 분리한 후, 이들의 결과를 취합하여 최종 등급을 판정한다. 여러 정보를 하나의 모델 안에서 판단할 경우 개별 정보에 대한 정확도가 떨어질 가능성이 크기 때문이다. 실제로 협착, 골극, 등급을 하나의 모델에서 판정한 논문의 경우 모델의 정확도가 64.9로 좋지 않았다. 따라서 목적이 다른 모델을 분리한 후 개별 모델에 최적화된 튜닝을 따로 작업한다. 이를 통해 개별

모델의 성능을 높이고자 하며, 이는 결국 전체 모델의 성능 향상을 의미한다.

셋째로 최종 등급 판정 시 end to end 방식의 CNN 모델을 통해 판정하지 않고, 알고리즘 논리 구조에 의해 등급이 판정되도록 한다. CNN 모델은 스스로 특징을 추출하여 학습하지만, 이로 인해 나온 결과(등급)는 CNN 모델이 스스로 판정한 것일 뿐, 무엇이 중요한 요소로 판단이 됐는지 알 수 없으며, 실제 의사의 판단과 다를 가능성이 있다. 따라서 기존 등급 시스템을 통합 보완한 등급 분류 기준을 세우고, 의사들의 판독 결과와 일치하는 정확한 등급이 나오도록 등급 시스템에 기반한 알고리즘을 설계한다. 설계에 앞서 인공지능 모델을 통해 예측된 협착/골극 정보를 입력값으로 제공하면 논리 구조에 따라 정확한 최종 등급이 판정된다. 이는 end to end 방식으로의 등급 판정 결과보다 의료진들이 실제 기대하는 결과 값에 부합할 것으로 목표한다.

본 논문의 구성은 다음과 같다. 2장에서는 연구 방법에 대한 설명을, 3장에서는 연구 결과를 확인한다. 마지막으로 4장에서 본 연구가 갖는 의의와 향후 미래 연구 방향에 대한 설명으로 논문을 마친다.

2. 연구내용

2.1 연구 방법

본 연구는 공개적으로 사용 가능한 OAI 데이터를 사용하여 AI 모델 수립 후, 의료 병원에서 사용되는 실제 무릎 X-ray 이미지를 이용하여 새로운 등급 분류 기준으로 의료진의 판독에 도움을 주는 것을 목적으로 한다.

기존에 선행되었던 논문에서는 KL 등급 기준을 이용하여 관절 사이의 협착과 뼈의 변형만 판단 기준으로 사용하여 골관절염 등급을 분류하였다 (Wright, 2014). 하지만 KL 등급 기준이 관절염을 분류하는데 모호한 점이 있고, 뼈의 변형을 인공지능 모델이 잡아내는데 한계점이 있다. 따라서 슬관절 골관절염 전문 의료진들의

경험을 기반으로, 기존 시스템을 통합·보완한 등급 분류 기준을 제안하고, 새로운 기준을 인공지능 모델에 넣어 실제 의료진의 판단과 일치하는 결과를 나타내고자 한다. 골관절염은 초기 진단이 중요한 질병이므로 슬관절 골관절염 중증도 판단 인공지능 모델을 통해 오래 걸리던 초기 진단을 신속하고 정확하게 내려 명확한 치료 계획을 수립하여 환자의 질환 개선에 도움이 되고자 한다.

본 연구에서 사용된 FCN 기반의 ROI(Region of Interest) 추출 모델을 이용하여 필요한 정보가 담긴 부분을 자동으로 추출하고, 새로운 분류 기준에 따라 관절 사이 공간의 길이를 재는 모델과 골극의 유/무와 길이를 측정하는 모델을 통해 등급을 도출해 내고자 한다.

2.1.1 ROI (Region of Interest) 추출

무릎 X-ray 이미지를 보면 슬관절 골관절염 중증도를 판정하는데 불필요한 정보를 포함한 부분이 매우 많다. 슬관절 골관절염의 등급을 판단할 때는 판단 기준이 되는 골극과 협착이 나타나는 일부만 필요하다. 따라서 원본 이미지가 주어졌을 때, 판단 기준에 해당하는 필요한 정보가 담긴 관심 영역(ROI)을 찾는 것이 중요하다. 이는 사람이 직접 수작업으로 추출할 수도 있겠지만 이 경우 추출할 때마다 추출하는 사람의 주관이 개입될 수 있고 인력과 시간이 많이 소요되는 문제가 발생한다. 따라서 원본 데이터가 주어졌을 때 관심 영역을 자동으로 추출해주는 인공지능 모델을 학습시켜 사용하며, FCN 모델을 활용하여 관심 영역을 추출한다.

관심 영역을 추출하기 전, X-ray 이미지의 화질을 높이기 위해 Histogram Equalization 방법을 활용하여 선명도를 높여준다. Histogram Equalization 기법으로 선명도는 높아졌지만 그만큼 이미지에 noise(잡음)가 생길 확률이 높아진다. 잡음은 인공지능 모델의 성능을 낮추는 요인이 되며, 지나친 잡음 제거는 화질을 낮추고 경계를 모호하게 한다. 따라서 경계선은 잘 살리면서 잡음을 제거하는 방법을 적용해야 한다. 이를 위해 양방향 가중치를 이용한 잡음 제거 방법을 사용한다.

본 연구에서는 잡음까지 제거된 수직 분할된 이미지를 사용하여 학습을 진행한다. 수직 분할된 이미지를 학습시키기 위해 관절 부위 라벨링 (labeling)을 진행한다. 인공지능 모델이 스스로 관절의 위치를 찾을 수 있도록 학습시키기 위해 정답을 알려주는 라벨링 과정이 필요하다.

원본 데이터 중 일부인 300개의 방사선 데이터의 라벨링을 진행하였다. 원본 데이터와 라벨링된 데이터를 활용하여 Image Net 기반으로 사전 학습된 VGG16을 사용하여 전이 학습을 진행한다 (Krizhevsky et al., 2012; Ahn, 2017).

모델의 Backbone으로 VGG16 모델을 활용하였고 Fig. 2와 같이 정류 선형 유닛(Rectified Linear Unit ; ReLU)과 맥스 풀링(Max Pooling)을 포함한 5개의 Convolution Layer와 3개의 완전 연결(fully Connected) 계층으로 이루어져 있다. 본 연구에서는 완전 연결 계층 대신 Convolution Layer와 Up Sampling을 쌓아 Fig. 3과 같은 관심 영역을 추출한다.

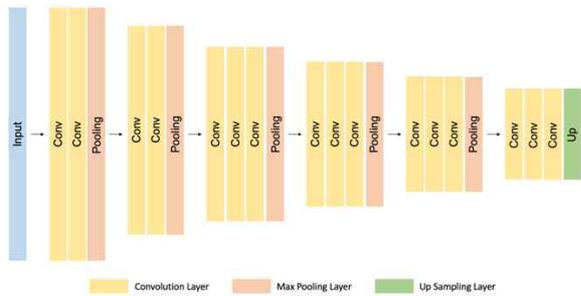


Fig. 2 ROI Model Architecture



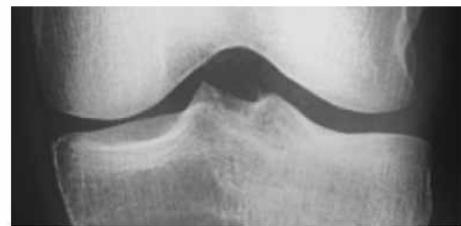
Fig. 3 ROI Extraction Result

2.1.2 이미지 사이즈 규격화

원본 이미지(224x224)에서 관심 영역을 추출할 경우, 추출된 이미지의 크기는 서로 다르게 추출된다. 따라서 이를 동일한 크기로 조정해주

는 작업이 필요하다. 이때 가로/세로 비율을 무시하고 일괄적으로 크기를 통일할 경우 원본 영상에서의 형태학적 구조들이 손실된 채 크기 조정이 발생하게 된다.

형태학적 구조들은 등급 판단하는 기준이 되기 때문에 변환되어서는 안 된다. 따라서 가로/세로의 비율을 유지한 채 동일한 크기로 맞춰야 한다. 본 연구는 형태학적 구조를 그대로 살리며 사이즈를 일괄적으로 맞추기 위해 Padding 기법을 이용하여 Fig. 4와 같이 데이터의 크기를 조정해주었다.



(a) ROI Extraction Image



(b) Image Resize Result

Fig. 4 Image Resizing Using Padding

2.1.3 Mask 이미지 생성 모델

입력된 새로운 데이터에 대해서도 인공지능 모델은 슬관절 뼈의 위치를 찾아내야 하며, 관절의 협착 공간의 길이와 골극의 유/무와 길이를 짚 수 있어야 한다. 새로운 슬관절 방사선 이미지에 대해서도 뼈의 영역을 찾을 수 있도록 Labelme Tool을 이용하여 뼈의 영역을 지정해주는 작업을 진행한다. 라벨링 데이터는 Fig. 5와 같이 위부분은 femur, 아랫부분은 tibia로 명명하여 Fig. 6과 같이 Polygon 방식으로 각 지점들을 지정하여 각 Points의 값과 라벨 이름이 들어간 json 데이터로 저장한다.

텔을 구축하였으며, 이 모델의 구조는 Convolution Layer, ReLU, Max pooling Layer, Up sampling Layer 등으로 구성된다.



Fig. 9 Joint Space Length Measurement Algorithm

2.1.5 골극 형성 판단

골극(Osteophyte)이란 뼈가시라는 뜻으로 뼈 위에 자라는 뼈를 말한다. 골극은 퇴행성 관절염의 대표적인 증상으로 등급 판단 하는데 중요한 요소이다. 우리는 골극을 자동으로 감지할 수 있도록 만들기 위해 골극이 있는 방사선 사진들을 1차로 분류하였다. 골극이 존재하는 것으로 판단되는 745장의 방사선 이미지 중 279장의 이미지를 사용하여, Labelme Tool을 사용하여 골극의 위치를 Bounding Box를 그려 라벨링을 진행한다. 라벨링 된 json 파일은 라벨 이름과 (x_min, y_min), (x_max, y_max) 좌표로 이루어져 있다. 새로운 슬관절 방사선 이미지가 들어왔을 때 뼈의 영역을 찾기 위해 Fig. 10과 같이 Labelme Tool을 사용하여 라벨링 작업을 진행하며, 생성된 Mask 이미지를 통해 뼈의 외곽을 지정한다.



Fig. 10 Osteophytes Labeling Using Labelme

Object Detection은 객체의 분류(classification)와 지역화(localization)를 함께 하는 기법이다. 골극이 형성되었는지 확인하기 위해 Object Detection 모델 중 YOLO_v4 모델을 활용하였다(Zou et al., 2023).

YOLO란 “You Only Look Once”라는 의미로 이미지를 한 번 보는 것만으로 객체의 종류와 위치를 예측해 주는 딥러닝 기반 객체 감지 알고리즘이다. 분류 대상의 객체로 감지하는 대신에, 경계 박스를 다차원적으로 분리하고 클래스 확률(Class Probability)을 적용하여 하나의 회귀 문제로 접근한다 (Lee and Kim, 2020).

YOLO는 1 Stage Detector로 Classification과 Localization을 동시에 수행하는 방법으로 빠르지만, 정확도가 낮다는 특징을 갖고 있으며, Fig. 11은 YOLO 모델의 아키텍처를 설명한다.

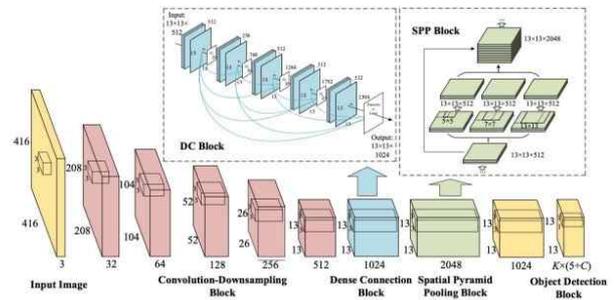


Fig. 11 YOLO_v4 model Architecture

골극 탐지를 위한 YOLO 모델을 학습하기 위해 YOLO 모델에 맞는 라벨링 데이터를 생성해 주어야 한다. YOLO 학습을 위한 데이터 형식은 [Object Class, center x, center y, width, height] 형식으로 txt 파일로 제공되어야 한다. 본 연구에서 사용한 라벨링 데이터의 형식과 다르므로 YOLO 변환이 필요하다. Fig. 12와 같은 방식으로 변경하고 txt 파일로 저장하여 학습을 진행한다.

YOLO 형식으로 변환된 라벨링을 사용하여 YOLO_v4의 tiny 모델 학습을 진행하였다. 총 297장의 데이터와 ‘femur_lateral’ 66개, ‘femur_medial’ 64개, ‘tibia_lateral’ 120개, ‘tibia_medial’ 116개의 각 라벨에 대하여 총 10000 epoch의 학습을 진행하였고, Threshold는 0.2 이상으로 검출되도록

$$x_{center} = \frac{x_{min} + x_{max}}{2}$$

$$y_{center} = \frac{y_{min} + y_{max}}{2}$$

$$width = (x_{max} - x_{min})$$

$$height = (y_{max} - y_{min})$$

$$json\ format = [[x_{min}, y_{min}], [x_{max}, y_{max}]]$$

Fig. 12 Convert Labelme to YOLO Format

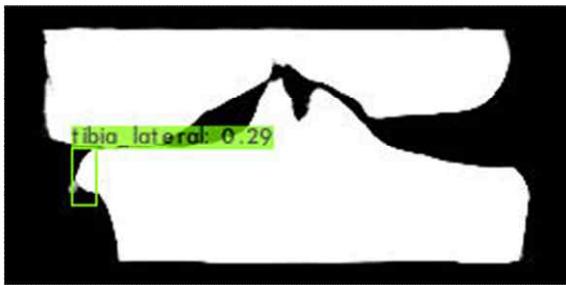


Fig. 13 Osteophytes Detection Result

설정하였다. 그 결과 정확도 72%를 달성하였다. 골극을 감지하기 위해 Labelme Tool을 사용하여 Fig. 13과 같이 골극의 위치를 Bounding Box로 라벨링하고, 이 데이터를 활용하여 객체 감지 모델을 학습을 진행했다.

2.1.6 골극 길이 측정

본 연구에서는 선행 연구와 다르게 새로운 슬관절 골관절염 등급 분류 기준을 활용한다. 중증도를 평가하기 위해서는 골극의 형성 유/무와 형성된 골극의 길이가 필요하다.

앞에서 YOLO 모델을 사용한 골극 형성 탐지 Detection 결과를 활용하여 골극의 길이를 셀 수 있는 알고리즘이다. Detection 된 Bounding Box의 좌표를 이용하여 Box 부분에 포함되는 뼈의 경계선 부분을 구한 후 Box와 경계선의 길이를 구해준다. 만약 경계선의 부분이 Box와 맞닿아 있다면 Box의 Width 길이를 반환해준다. YOLO 모델을 활용하여 Fig. 14와 같이 골극을 탐지하고, 경계 박스를 이용해 객체의 위치를 예측하는 과정을 보여준다.

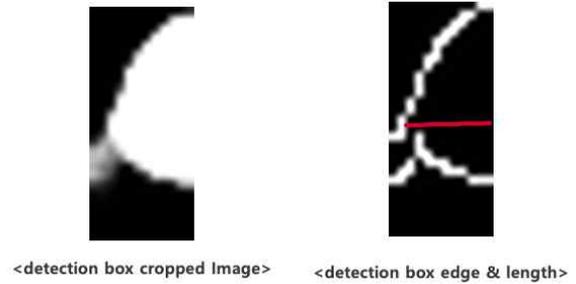


Fig. 14 Osteophytes Length Measure Algorithm

3. 연구 데이터 및 결과

3.1 연구 데이터

본 연구를 위해 무릎 관절 연구를 위한 의료 영상 공공 데이터인 OAI 데이터 셋을 사용하였다. OAI 데이터 셋은 9,786장의 무릎 방사선 이미지 데이터로 구성되어 있다. 정상 슬관절에 해당하는 Grade 0은 3,857장, Grade 1은 1,770장, 관절염으로 판단되는 Grade 2는 2,578장, Grade 3은 1,286장, 중증 골관절염으로 판단되는 Grade 4는 295장으로 이루어져 있다. ROI를 검출해주는 모델링의 경우 학습 데이터 300장, 테스트 데이터 100장을 사용하였고, 나머지 9,386장의 데이터를 예측하여 검출하였다. Mask 이미지를 생성하는 모델링은 학습 데이터 4,097장, 테스트 데이터 1,024장을 사용하였고, 나머지 4,665장의 예측된 Mask 이미지를 생성했다. 마지막으로 골극 감지 모델링은 학습 데이터 236장, 테스트 데이터 43장을 사용하였다.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Fig. 15 Classification Performance Evaluation Method

Table 1 Problems in DIO Processing Systems

Grade	Osteophyte	The Gap Between The Joints
0	x	- 4mm or more - More than 75% compared to the other side
1	- Less than 1mm - More than two, 0~1mm	- 4mm or more - 4mm or more
2	- More than one, 0~1mm x	- 3~4mm - 2~3mm - More than 50% reduction compared to the other side
3	- More than two, 2mm or more - Contour distortion of bone - Subchondral Sclerosis	- 3mm or more
3	- More than two, 1~2mm x	- Less than 3mm - Less than 1.5mm - More than 75% reduction compared to the other side
4	- Secondary changes in contours of Osteophyte and bone - More than two, 0~1mm - More than two, 2mm or more - A marked change in contours of Osteophyte and bone - Subchondral Sclerosis	x - Less than 1.5mm - More than 75% reduction compared to the other side - Less than 3mm

3.2 골관절염 등급 판단 성능평가

본 연구에서 활용한 등급 분류 시스템은 아래의 Table 1과 같다. 등급 분류 시스템을 활용한 슬관절 골관절염의 성능평가 방법으로 정확도 (Accuracy), 정밀도(Precision), 재현율(Recall), F1 Score를 사용한다.

관절염이 존재하지만, 관절염이 존재하지 않다고 판단되는 등급으로 분류되는 것이 더 큰

Table 2 Grade Classification Result

	Precision	Recall	F1-score	Accuracy
Grade 0	0.66	0.68	0.67	0.684
Grade 1	1.0	0.99	0.99	0.988
Grade 2	0.46	0.49	0.47	0.489
Grade 3	0.98	0.69	0.81	0.689
Grade 4	0.80	0.99	0.88	0.986

문제가므로 재현율의 성능이 주요한 지표로 작용한다. 분류 성능평가 공식은 Fig. 15와 같다.

성능평가 지표에 의해 산출된 결과 값은 Table 2와 같다. Grade 0은 Recall(재현율) 68%, F1 score 67%, Grade 1은 Recall 99%, F1 score 99%의 성능이 산출되었다. Grade 2는

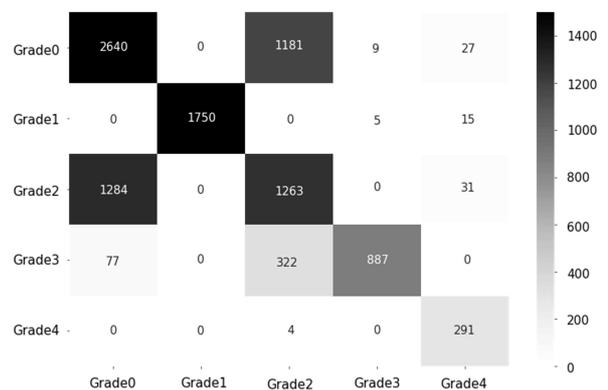


Fig. 16 Confusion Matrix

Recall 59%, F1 score 47%이며, Grade 3은 Recall 69%, F1 score 81%의 성능이 산출되었다. Grade 4는 Recall 99%, F1 score 88%의 성능이 산출되었다. 전체 등급에 대한 정확도 69.8%, F1 score 76.65%를 달성하였다. 위의 결과를 보면 Grade 0, 2의 분류에 어려움이 있음을 알 수 있다.

Fig. 16의 혼동 행렬(Confusion Matrix)에서는 Grade 0과 Grade 2의 분류가 어렵고, 기준에 대한 모호함이 존재함을 알 수 있다. Grade 0과 Grade 2의 분류는 데이터 증강, 앙상블 모델, 세분화된 특징 학습, 전이 학습 그리고 라벨링 개선 등 다양한 접근 방식을 결합하여 모델의 성능을 더 향상시킬 수 있다.

4. 결론 및 한계점

본 논문에서는 빠른 슬관절 골관절염 등급 판단을 위해 무릎 방사선 이미지 데이터를 활용하여 KL 등급 분류 시스템과 전문 의료진들의 판단 기준을 결합한 등급 분류 기준과 분류 방법을 제안하였다. 골관절염 등급을 판단할 때, 관절의 협착 부위와 골극의 형성이 중요한 영향을 주었다. 선행 연구에서는 협착 부분과 골극 형성 판단을 한 번에 모두 측정하는 알고리즘을 선택한 반면, 본 연구에서는 협착 공간의 길이를 측정하는 알고리즘, 골극 형성의 유/무를 판단하는 알고리즘, 골극의 길이를 측정하는 알고리즘을 따로 설계하여 결과를 도출하였다.

본 연구에서는 몇 가지 한계점이 존재한다. 첫 번째로, 새로운 분류 기준을 제시했지만 학습된 라벨링 데이터들은 분류(진단)한 의료진의 기준이 담겨있으며, 모든 Case를 나눌 수 없다는 것이다. 각 등급별 공백이 존재하며, 또한 등급 기준에 포함되지 않는 경우 등급 분류에 어려움이 있다. 공백에 들어있는 데이터는 한 번 더 의료진의 진단이 필요하며, 시간이 더 소요된다. 두 번째로, 골극의 분류(진단)은 의료진마다 판단하는 기준이 천차만별이다. 골극에 대한 정답과 기준을 마련하는 어려움이 있다. 골극 분류(진단)을 하는 사람에 따라 결과가 달라

지기 때문에 인공지능에 학습시키는데도 어려움이 존재한다. 또한 골극이 형성되어있는 데이터(297장)가 적기 때문에 학습 결과가 과적합될 가능성이 컸으며, 성능 또한 좋지 않았다. 마지막으로 정상적으로 찍힌 X-ray 이미지와 비정상적으로 찍힌 방사선 이미지가 섞여 있어 모델의 학습에 방해되는 요소로 작용하여 더 나은 성능을 도출하지 못했다.

향후 연구 방향으로 OAI(Osteoarthritis Initiative) 데이터 셋 확보와 제안한 모델들의 미세조정을 통해 더 나은 성능을 얻고자 하며 등급 분류들 사이의 공백을 줄일 수 있는 기준을 명확히 제시할 수 있도록 연구하고자 한다.

References

- Ahn, H. (2017). A Study on Compression of Connections in Deep Artificial Neural Networks, *Journal of Korea Society of Industrial Information Systems*, 22(5), 17-24. <https://doi.org/10.9723/jksis.2017.22.5.017>
- Antony, J., McGuinness, K., Moran, K. and O'Connor, N. E. (2017). Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity Using Convolutional Neural Networks, *In Machine Learning and Data Mining in Pattern Recognition: 13th International Conference, MLDM 2017, New York, NY, USA, July 15-20, 2017, Proceedings 13* (pp. 376-390). Springer International Publishing. https://doi.org/10.1007/978-3-319-62416-7_27
- Antony, J., McGuinness, K., O'Connor, N. E. and Moran, K. (2016). Quantifying Radiographic Knee Osteoarthritis Severity Using Deep Convolutional Neural Networks, *In 2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 1195-1200). IEEE. <https://doi.org/10.1109/ICPR.2016.7899799>
- Bayramoglu, N., Nieminen, M. T. and Saarakkala, S. (2020a). A Lightweight CNN and Joint

- Shape-Joint Space (JS2) Descriptor for Radiological Osteoarthritis Detection, *In Annual Conference on Medical Image Understanding and Analysis* (pp. 331-345). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-52791-4_26
- Bayramoglu, N., Tiulpin, A., Hirvasniemi, J., Nieminen, M. T. and Saarakkala, S. (2020b). Adaptive Segmentation of Knee Radiographs for Selecting The Optimal ROI in Texture Analysis, *Osteoarthritis and Cartilage*, 28(7), 941-952. <https://doi.org/10.1016/j.joca.2020.03.006>
- Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A. and Brendel, W. (2018). ImageNet-trained CNNs are Biased Towards Texture; Increasing Shape Bias Improves Accuracy and Robustness, *arXiv preprint arXiv:1811.12231*. <https://doi.org/10.48550/arXiv.1811.12231>
- Harman, M. (2012). The Role of Artificial Intelligence in Software Engineering, *In 2012 First International Workshop on Realizing AI Synergies in Software Engineering (RAISE)* (pp. 1-6). IEEE. <https://doi.org/10.1109/RAISE.2012.6227961>
- Heo, Y. S. (2022). AI Medical & Healthcare, *ASTI Market Insight*, 65, 1-9.
- Hong, J.-Y., Park, S. H. and Jung, Y.-J. (2020). Artificial Intelligence Based Medical Imaging: An Overview, *Journal Radiological Science and Technology*, 43(3), 195-208. <https://doi.org/10.17946/JRST.2020.43.3.195>
- Jeong, G. H. (2018). AI Based Medical Image Analysis Technology Trends, *Institute for Information & Communication Technology Planning & Evaluation(IITP)*, Weekly Technology Trends.
- Kim, M. and Bae, H. J. (2020). Data Augmentation Techniques for Deep Learning-Based Medical Image Analyses, *Journal of the Korean Society of Radiology*, 81(6). <https://doi.org/10.3348/jksr.2020.0158>
- Korean Association of Knee Joints, (2021). *Guidebook for Degenerative Arthritis with the Association of Knee Joints*, <https://www.koreaknee.or.kr/pop/file/guidebook.pdf>
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems*, 25.
- Lee, J. H., Kim, B. M. and Shin, Y. S. (2018). Effects of Preprocessing and Feature Extraction on CNN-Based Fire Detection Performance, *Journal of Korea Society of Industrial Information Systems*, 23(4), 41-53. <https://doi.org/10.9723/jksiiis.2018.23.4.041>
- Lee, Y. H. and Kim, Y. S. (2020). Comparison of CNN and YOLO for Object Detection, *Journal of the Semiconductor & Display Technology*, 19(1), 85-92.
- MSD Manual. (2021). *The Korean Orthopaedic Association*, Osteoarthritis (OA), https://www.koa.or.kr/info/index_10_1.php
- Pesapane, F., Codari, M. and Sardanelli, F. (2018). Artificial Intelligence in Medical Imaging: Threat or Opportunity? Radiologists Again at The Forefront of Innovation in Medicine, *European Radiology Experimental*, 2, 1-10. <https://doi.org/10.1186/s41747-018-0061-6>
- Qi, K., Yang, H., Li, C., Liu, Z., Wang, M., Liu, Q. and Wang, S. (2019). X-net: Brain Stroke Lesion Segmentation Based on Depthwise Separable Convolution and Long-Range Dependencies, *In Medical Image Computing and Computer Assisted Intervention - MICCAI 2019: 22nd International Conference*, Shenzhen, China, October 13 - 17, 2019, Proceedings, Part III 22 (pp. 247-255). Springer International Publishing. https://doi.org/10.1007/978-3-030-32248-9_28
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B. and Yang, G. Z.

- (2017). Deep Learning for Health Informatics, *IEEE Journal of Biomedical and Health Informatics*, 21(1), 4-21. <https://doi.org/10.1109/JBHI.2016.2636665>
- Shen, D., Wu, G. and Suk, H. I. (2017). Deep Learning in Medical Image Analysis, *Annual Review of Biomedical Engineering*, 19(1), 221-248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>
- Trinder, J. C., Wang, Y., Sowmya, A. and Palhang, M. (1997). Artificial Intelligence in 3-D Feature Extraction, In *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)* (pp. 257-266). Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-8906-3_25
- Wright, R. W. (2014). Osteoarthritis Classification Scales: Interobserver Reliability and Arthroscopic Correlation, *The Journal of Bone and Joint Surgery*, 96(14), 1145-1151. <https://doi.org/10.2106/JBJS.M.00929>
- Zou, Z., Chen, K., Shi, Z., Guo, Y. and Ye, J. (2023). Object Detection in 20 years: A Survey, *Proceedings of the Institute of Electrical and Electronics Engineers*, 111(3), 257-276. <https://doi.org/10.1109/JPROC.2023.3238524>



윤 예 찬 (YeChan Yoon)

- 고려대학교 산업경영공학과 (산업인공지능)
- (현재) ㈜스칼라데이터 대표
- 관심분야: 머신러닝, 예측모델, 인공지능, 컴퓨터비전