



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사학위논문

딥러닝 기반 다발성 골절 CT 영상의 분류

Deep Learning-Based CT Image Classification
of Multiple Bone Fracture



충북대학교 대학원

전기 · 전자 · 정보 · 컴퓨터학부 컴퓨터과학전공

이 상 현

2020 년 2 월

공학석사학위논문

딥러닝 기반 다발성 골절 CT 영상의 분류

Deep Learning-Based CT Image Classification
of Multiple Bone Fracture

지도교수 이 건 명

전기 · 전자 · 정보 · 컴퓨터학부 컴퓨터과학전공

이 상 현

이 논문을 공학석사학위 논문으로 제출함.

2020 년 2 월

본 논문을 이상현의 공학석사학위 논문으로 인정함.

심 사 위 원 장 이 재 성 (인)

심 사 위 원 류 관 희 (인)

심 사 위 원 이 건 명 (인)

충 북 대 학 교 대 학 원

2020 년 2 월

차 례

Abstract	iii
List of tables	v
List of figures	vi
I. 서 론	1
II. 관련 연구	5
2.1 컨볼루션 신경망과 이미지 분류	5
2.2 다중 부류 분류 문제	9
2.3 골절 검출 및 분류	10
III. 다발성 골절 데이터	15
3.1 AO/OTA 골절 및 탈구 분류	15
3.2 CT 영상 데이터 및 레이블	17
3.3 데이터 불충분 및 불균형 문제	19
IV. 제안하는 딥러닝 기반 다발성 골절 다중-부류 분류 방법	23
4.1 다발성 골절 분류 방법의 구조	23
4.2 데이터 전처리 및 데이터 증대	24
4.3 학습/테스트 데이터셋 구성 방법	32
4.4 다발성 골절 분류 모델	34

IV. 실험	40
5.1 교차 검증	40
5.2 제안한 방법의 실험 결과	41
5.3 데이터 오버 샘플링 실험 및 실험 결과	44
V. 결 론	51
참고문헌	53

Deep Learning-Based CT Image Classification of Multiple Bone Fracture*

Lee, Sang Hyun

*School of Electrical Engineering and Computer Science,
Graduate School of Chungbuk National University,
Cheongju, Korea*

Supervised by Professor Lee, Geon Myung

Abstract

The orthopedists use Computed Tomography(CT) to identify the fracture feature of fractured patient such as anatomical location and angle. They determine a proper care and treatment method according to the fracture categories. Fractures often result in multiple fractures and hence a patient may have simultaneously multiple fracture categories. It demands the orthopedists to have well-qualified training experience and careful examination with sufficient amount of time for fracture classification. In some medical practices, some orthopedists see many patients in a limited amount of time or some poorly-trained doctors see patients with fractures. Such situation may result in mistakes in medical diagnosis and subsequent mistreatment on patient

There have been some works for detecting fractures or for classifying fractures into a small number of categories. This thesis addresses how to classify fractures into multiple labels. It first defines fracture classification as a multi-label classification problem. Then it proposes a deep learning-based model for fracture classification. The model is designed to classify a CT image into some out of 20 AO/OTA classes.

The model has been trained for classifying multiple fractures of 149 patients with lower body fractures. Each patient has 9.75 CT images on average which have

been taken while rotating imaging angles gradually. The data set suffers from class imbalance problem, and the number of CT image in the data set is too small for training a deep learning-based model. To expand the data set, the proposed method uses a technique of oversampling along with data augmentation. To balance the class distribution of the data set, an oversampling method is proposed. 85% of the data set is used for model training and the remaining 15% is used for model testing.

The proposed deep learning-based model consists of 49 convolution layers and the following fully connected layer. It has a modified model for ResNet-50 and it is trained from randomly initialized weights. It also contains skip-connections as in ResNet model. At the output layer each node corresponds to a class label and produces a value in range $[0, 1]$. According to the rank of the output node values, the fracture class labels are determined. The threshold value for the class labeling is specified with respect to the difference of adjacent node values in the ranked sequence of the output node values.

In the experiments, we built 5 trained models and evaluate their average performance. The trained model showed 98.70% precision, 96.86% recall, and 95.67% accuracy for the data set which has been constructed with the proposed data augmentation and oversampling method.

* A thesis for the degree of Master in February 2020.

표 목차

Table 3.1	다발성 골절 데이터의 예시	18
Table 3.2	레이블별 데이터 분포	19
Table 3.3	실험 데이터의 레이블 카디널리티와 레이블 밀도	20
Table 4.1	데이터 증대 기준	24
Table 4.2	최종 레이블 결정 예시	39
Table 5.1	기존 딥러닝 모델과 제안된 방법의 실험 결과	42
Table 5.2	이미지의 레이블 개수별 분류 현황	43
Table 5.3	교차 실험동안 한 에폭에 사용되는 학습 데이터 누적 현황	50
Table 5.4	오버 샘플링 실험 결과	51

그림 목차

Figure 1.1	X-ray 영상(좌)과 CT 영상(우)	1
Figure 2.1	컨볼루션 신경망	5
Figure 2.2	컨볼루션 연산	7
Figure 2.3	풀링 연산	7
Figure 3.1	AO/OTA 골절 및 탈구 분류	16
Figure 3.2	환자의 CT 이미지	17
Figure 3.3	골절의 위치 및 형태	21
Figure 4.1	제안하는 다발성 골절 분류 방법의 흐름도	23
Figure 4.2	원본 이미지	25
Figure 4.3	원본 이미지에 크기 변경 적용 예시	25
Figure 4.4	이미지 보간법 예시	26
Figure 4.5	이미지 크기 변경 후 회전 적용 예시	27
Figure 4.6	이미지 회전 후 반전 적용 예시	28
Figure 4.7	이미지 반전 후 이동 적용 예시	29
Figure 4.8	이미지 이동 후 자르기 적용 예시	30
Figure 4.9	이미지 자르기 후 점잡음 적용 예시	30
Figure 4.10	데이터 증대 후 컨볼루션 신경망의 학습을 위한 이미지 입력	32
Figure 4.11	데이터셋 구성 방법	33

Figure 4.12 제안하는 방법의 다발성 골절 분류 모델	35
Figure 4.13 다발성 골절 분류 모델의 출력	37
Figure 4.14 SelcectLabel()의 의사 코드	38
Figure 5.1 교차 검증을 위한 데이터셋 구성 방법	41
Figure 5.2 오버 샘플링	44
Figure 5.3 OverSampling()의 의사코드	46
Figure 5.4 기존 학습데이터와 오버샘플링 된 학습데이터의 각 레이블별 데이터 분포현황	48

I . 서 론

정형외과에서는 골절 환자의 환부를 확인하기 위해 엑스선 촬영(X-ray) 혹은 컴퓨터 단층 촬영(Computer Tomography, CT)을 활용한다. 골절이 단순 형태인 경우 2차원 X-ray 영상만으로 골절의 특징을 파악할 수 있다. 하지만 골절이 복합적으로 발생한 경우 의사는 골절의 상세한 정보를 파악하기 위해 3차원의 CT 영상을 사용한다. 의사는 이러한 CT 영상을 기반으로 골절이 발생한 해부학적 위치, 골절 형태, 골절 각도 등의 정보를 파악하고, 이를 기반으로 골절의 범주를 분류한 뒤, 각 범주에 따라 적절한 치료 방법을 적용한다. 이렇듯 의사가 CT영상을 기반으로 골절의 범주를 분류하는 과정은 각 환자에게 적절한 치료를 적용하기 위해 필수적이다.



[그림 1-1] X-ray 영상(좌)과 CT 영상(우)

보건복지부의 통계에 따르면 2018년을 기준으로 한국의 인구 1,000명 당 전

문의 수는 2.39명에 불과하다. 이는 OECD 국가 35개국 중 최하위이다. 더불어, 의사들은 각 분야의 전문의로 구분된다. 통계에 따르면 정형외과 전문의는 전체 의사 중 약 7.31%에 불과하다. 즉, 평균적으로 약 17.5명의 정형외과 전문의가 10만 명의 정형 관련 내원을 소화해야 하는 것이다. 지역 불균형도 심해 수도권 외 지역은 이 수치의 절반도 되지 않는 경우가 대부분이었으며, 세종의 경우 인구 1,000명당 전문의 수는 1.1명으로 가장 적었다[통계청18]. 이런 상황에서 모든 환자들의 CT영상을 일일이 직접 확인하며 골절의 특징을 식별하고 범주를 분류하는 과정은 정형외과 의사에게 업무부하의 원인이 된다.

실제로 골절이 발생하는 경우 여러 형태의 골절이 동시에 발생하는 다발성 골절인 경우가 많다. 다발성 골절은 한 뼈의 한 곳에서 둘 이상의 형태인 골절이 발생하는 경우, 한 뼈에서 둘 이상의 골절이 동시에 발생 하는 경우, 둘 이상의 뼈에서 동시에 골절이 발생하는 경우 등을 뜻한다. 이렇듯 골절은 수백 개의 범주가 복합적으로 얹혀서 등장하고, 각 특성에 따른 적절한 치료방법이 적용되어야 한다. 따라서 정형외과의에게는 골절을 정확히 식별하고 분류하기 위한 높은 전문성이 요구되며, 숙달된 전문의라 하더라도 적절한 치료법을 결정하기 위해 충분한 시간이 필요하다. 하지만 골절 환자의 경우 응급실을 통해 내원하는 경우가 많다. 동시다발적으로 많은 환자의 처치가 요구되는 응급실에서는 의사들이 환자 개개인에게 충분한 시간을 할애하지 못한다. 이로 인해 의사는 종종 실수나 의료과실을 일으키며 환자는 신속하게 양질의 서비스를 받지 못하게 된다.

본 논문에서는 위와 같은 문제를 해결하고자 CT 영상을 기반으로 딥러닝을 사용해 다발성 골절의 범주를 분류했다. 제안하는 방법은 의사들의 골절 검출 및 분류를 보조하고 더 나아가 자동화하는 것이 목표다. 본 논문에서는 다발성 골절의 범주 분류 과정을 다중 부류 분류(Multi-Label Classification) 문제로 정의했다. 제안하는 방법은 다발성 골절 CT 영상을 입력으로 사용해 골절의 특징을 식별하고, 다발성 골절의 범주를 분류한다.

골절 검출 혹은 분류를 위한 기존 연구들 중 대부분은 단지 의사의 골절 진단을 보조하는 것을 목적으로 한다. 골절의 특징을 기반으로 그 범주를 상세하게 분류하는 것이 아닌 단순히 검출(Detection)하는 것이 목표다. 더 나아가 감지와 더불어 골절의 검출 영역을 시각화하기 위해 열 지도(Heat Map), 경계 상자(Bound Box)를 사용하기도 한다. 이 외에 분류를 목적으로 하는 연구들도 있지만 분류 기준이 실제 정형외과에서 사용되는 골절의 범주가 아니거나, 분류되는 레이블의 개수가 10개미만으로 분류 수준이 상당히 낮다. 골절을 검출 혹은 분류하기 위한 기존 연구들에 대한 내용은 2.3절에서 상세하게 서술한다.

기존 연구들과 비교하여 본 논문에서 제시하는 방법의 차별점은 다음과 같다. 첫 번째로 연구의 목적이 단순히 의사의 골절 진단을 보조하는 것이 아니고, 결과적으로 자동화하는 것을 지향한다. 본 연구에서는 실제 정형외과에서 사용 중인 ‘AO/OTA 골절 및 탈구 분류’를 기반으로 레이블을 작성했다. ‘AO/OTA 골절 및 탈구 분류’는 각 골절을 특성별로 분류하기 위한 기준을

제시한다. 이 내용은 3.1절에서 서술한다. 실험에 사용된 이미지는 해당 골절의 특징에 따라 상세하게 분류되었으며, 실험에서 사용된 범주도 20개로 기존 연구들과 비교해서 최소 2배 이상 많다. 두 번째로 다발성 골절에 대한 분류를 시도했다. 기존 연구들 중에도 다발성 골절을 대상으로 진행된 연구가 있다. 하지만 이는 골절을 분류하는 것이 아닌 단순히 골절이 몇 군데 발생했는지 감지하는 수준이다. 본 논문의 연구는 다발성 골절의 실제 범주를 분류하기 위해 해당 문제를 다중 레이블 분류(Multi-Label Classification)로 정의했고, 이에 따라 하나의 이미지에 대해 최대 3개의 레이블을 동시에 출력한다.

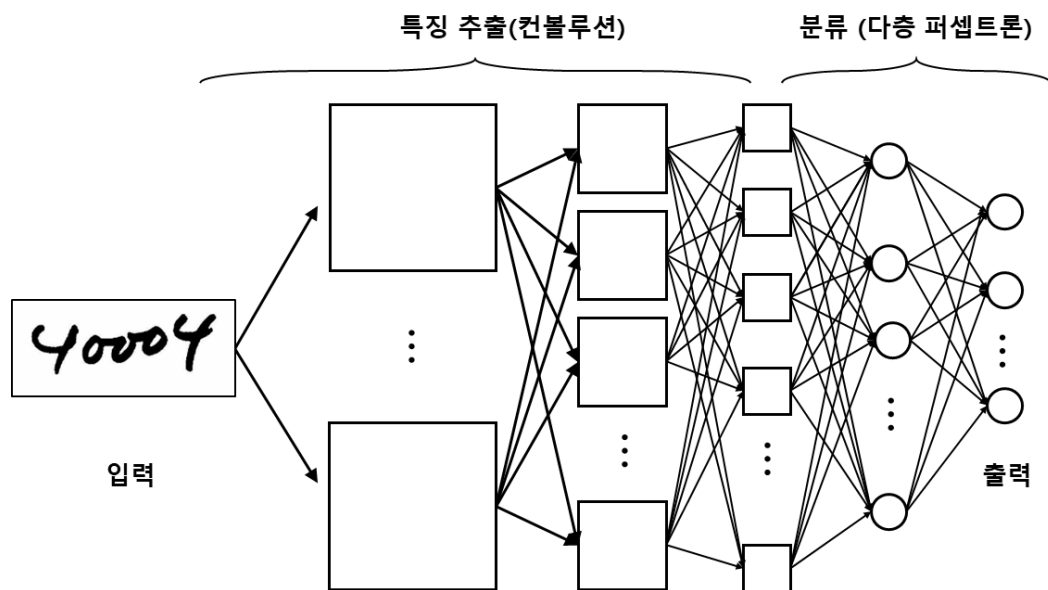
본 논문의 실험은 하반신 골절 환자에 대해서만 진행되었으며, 특히 데이터가 정강이 뼈인 경골(Tibia)와 비골(Fibula)에 집중되어있다. 이는 직접 수집한 데이터를 사용해 실험을 진행했기 때문이다. 이 점이 이 논문의 적용 한계이다. 적용 범위를 신체 전반으로 확대하기 위해서는 데이터 확보와 함께 추가 실험이 필수적이다.

본 논문의 구성은 다음과 같다. 제 2장에서는 본 논문에서 제시하는 방법과 관련된 연구를 살펴본다. 제 3장에서는 다발성 골절 분류 문제의 특성에 대해 설명한다. 제 4장에서는 본 논문에서 제안한 방법에 대해 설명한다. 제 5장에서는 제안한 방법을 활용한 실험과 그 결과를 제시한다. 제 6장에서는 결론을 맺는다.

II. 관련 연구

2.1. 컨볼루션 신경망(CNN)과 이미지 분류

컨볼루션 신경망(Convolution Neural Network, CNN)은 동물의 시각피질 구조를 모방하여 만들어진 딥러닝 신경망 모델이다. LeCun 등이 처음으로 제시 하였으며 필기체 인식에 있어 의미 있는 결과를 도출했다[LeCun89]. 컨볼루션 신경망의 구조는 [그림 2-1]과 같다.



[그림 2-1] 컨볼루션 신경망

[그림 2-1]의 특징 추출 단계에서 수행되는 컨볼루션 연산은 입력된 이미지의 일정 영역의 값들에 대해 가중치를 적용해 하나의 값으로 만든다. 입력되

x_{11}	x_{12}	x_{13}	x_{14}	x_{15}		
x_{21}	x_{22}	x_{23}	x_{24}	x_{25}		
x_{31}	x_{32}	x_{33}	x_{34}	x_{35}		
x_{41}	x_{42}	x_{43}	x_{44}	x_{45}		
x_{51}	x_{52}	x_{53}	x_{54}	x_{55}		

(a)

w_{11}	w_{12}	w_{13}
w_{21}	w_{22}	w_{23}
w_{31}	w_{32}	w_{33}

(b)

y_{11}	y_{12}	y_{13}
y_{21}	y_{22}	y_{23}
y_{31}	y_{32}	y_{33}

(c)

[그림 2-2] 컨볼루션 연산, (a)입력, (b)필터, (c)컨볼루션 결과

는 값이 2차원으로 표현된 데이터라면 컨볼루션 연산에 적용되는 가중치 또한 2차원으로 표현되고 컨볼루션 연산의 결과 또한 2차원으로 얻어진다. 이 때 주어지는 가중치 행렬을 필터라고 한다. 입력이 [그림 2-2]의 (a)처럼 주어지고, 필터가 (b)와 같다면, 컨볼루션 결과는 (c)와 같이 얻어진다. 이 때 y_{11} 을 구하기 위한 연산은 식 (1)과 같다. 또한 입력에 컨볼루션을 적용 할 때는 보통 직전의 컨볼루션 층보다 많은 수의 필터를 사용하여 더 많은 특징 맵 (Feature Map)을 생성한다. 이 때문에 [그림 2-1]을 보면 연산이 진행 될수록 출력력이 많아진다.

$$\begin{aligned}
 y_{11} = & x_{11}w_{11} + x_{12}w_{12} + x_{13}w_{13} + x_{21}w_{21} + x_{22}w_{22} + x_{23}w_{23} \\
 & + x_{31}w_{31} + x_{32}w_{32} + x_{33}w_{33}
 \end{aligned}
 \tag{1}$$

풀링은 컨볼루션 연산을 통해 추출된 특징 맵의 크기를 줄이기 위해 사용된다. 다양한 풀링 연산을 통해 특징들을 결합하거나, 가장 뚜렷한 특징을 선택하는 것이 가능하다. [그림 2-3]의 (a)와 같이 입력이 주어졌을 때 2×2 크기로 최댓값 풀링을 한 결과는 (b), 평균 풀링을 했을 때의 결과는 (c)와 같이 얻어진다. 특징 추출 단계에서는 ‘컨볼루션 - 풀링 - 컨볼루션 - 풀링’ 과 같이 두 연산이 반복적으로 진행된다. 이 때문에 [그림 2-1]을 보면 특징 추출이 진행될수록 특징 맵의 크기가 작아짐을 알 수 있다.

1	6	2	15
3	9	4	7
5	8	6	3
16	5	3	2

(a)

9	15
16	6

(b)

4.75	7
8.5	3.5

(c)

[그림 2-3] 풀링 연산, (a)입력, (b)최댓값 풀링, (c)평균 풀링

컨볼루션 신경망의 분류 단계에서는 앞에서 출력된 특징 맵을 기반으로 분류 또는 회귀를 수행한다. 일반적으로 다층 퍼셉트론(Multi Layer Perceptron)으로 구성되며, 다층 퍼셉트론은 직전 층의 모든 노드로 연결되는 완전 연결(Fully Connected)로 구성된다.

위와 같은 컨볼루션 신경망을 사용해서 이미지 내의 물체를 분류하기 위한 연구들이 진행되었다. 이러한 문제를 다중-클래스 분류(Multi-Class Classification)라고 한다. 다중 클래스 분류 문제는 이미지가 입력되었을 때 전체 레이블 중 확률이 가장 높은 레이블 하나만 선택하여 최종으로 출력한다. 다중 클래스 분류를 위한 대표적인 딥러닝 모델 연구로는 구글넷(GoogleNet)과 레스넷(ResNet) 등이 있다.

구글넷은 인셉션 모듈(Inception Module)을 기반으로 컨볼루션 신경망이 구성됐다. 인셉션 모듈은 직전 층의 처리결과에 1×1 컨볼루션, 3×3 컨볼루션, 5×5 컨볼루션을 동시에 각각 적용해 각 크기에 맞는 특징을 추출한다. 또한 1×1 컨볼루션을 사용해 추출된 특징 맵의 개수를 조정한다[Szegedy15].

레스넷은 특정 규칙의 컨볼루션 블록을 반복적으로 쌓아서 신경망이 구성된다. 레스넷은 이렇게 쌓인 블록으로 인해 다른 컨볼루션 신경망들에 비해 훨씬 많은 컨볼루션 층을 갖지만, 잔차 모듈을 활용해 가중치 소멸 문제를 극복했다. 잔차 모듈은 컨볼루션 층을 거치지 않고 이후의 컨볼루션 층으로 입력을 전달할 수 있는 연결을 말한다. 이 경로를 통한 변환은 컨볼루션 연산을 뛰어넘어 입력 상태를 그대로 유지하기 때문에 컨볼루션 연산이 기대 출력과 입력의 차이를 학습할 수 있다[He16].

2.2. 다중 부류 분류 문제(Multi-Label Classification)

다중 부류(레이블) 분류(Multi-Label Classification)는 하나의 입력을 하나 이상의 레이블로 분류 한다는 점에서 2.1절에서 소개한 다중-클래스 분류(Multi-Class Classification)와는 차이점이 있다. 다중 클래스 분류 문제는 일반적으로 분류 모델의 말단에 소프트맥스 함수를 적용해 각 입력에 대해 최종적으로 하나의 레이블만 출력한다. 반면 다중 레이블 분류에서는 여러 개의 분류기를 동시에 사용하는 등 다양한 방법을 통해 하나의 입력에 대해 여러 개의 레이블을 출력한다.

Tsoumakas 등은 다중 부류 분류 작업을 소개하며 다중 부류 데이터셋의 레이블 특성을 수량화하기 위한 개념으로 레이블 카디널리티(Label Cardinality)와 레이블 밀도(Label Density)를 정의한다. 또한 다중 부류 분류 작업을 평가하기 위한 지표로 정확도(Accuracy), 정밀도(precision), 재현율(Recall), 해밍 손실(Hamming Loss) 정의한다[Tsoumakas07].

Read 등은 다중 레이블 분류를 위해 여러 개의 이진 분류기를 사용했다. 다중 레이블 분류 문제에서 여러 개의 이진 분류기를 사용하는 방법은 레이블간의 상관관계를 모델링할 수 없다. 하지만 높은 예측 성능과 확장성을 보장하는 동시에 낮은 계산 복잡도를 유지할 수 있다[Read11].

Wai 등은 다중 레이블 분류를 위해 하나의 입력 이미지를 여러 개의 창

(Window)으로 분할하는 방법을 사용했다. 제안된 방법은 입력된 이미지에서 BING 혹은 EdgeBox 방법을 통해 물체 검출(Object Detection)을 진행한다. 이 후 검출된 물체들을 기반으로 이미지를 분할하여 창을 생성하고, 창들을 군집화하여 각 군집별로 하나의 창만 선택한다. 이렇게 선택된 창들은 각각 고유한 물체를 포함하게 된다. 컨볼루션 신경망은 이 창들을 각각 입력으로 사용해 각 창별로 하나의 레이블을 출력한 뒤, 출력된 각 레이블들을 종합해서 원본 이미지를 다중 레이블로 분류한다[wai15].

Wang 등은 다중 레이블 분류를 위해 컨볼루션 신경망과 재귀신경망(Recurrent Neural Network, RNN)을 동시에 사용했다. 제안된 방법은 각 레이블의 고유한 임베딩 벡터(Embedding Vector)와 이를 위한 공간이 있다. 컨볼루션 신경망은 이미지에서 각 물체를 식별하여 벡터를 생성하고, 재귀 신경망에서는 주어진 타겟 레이블의 벡터를 출력한다. 이 두 벡터의 합을 통해 해당 물체를 주어진 공간에 임베딩 하고, 재귀 신경망을 통해 각 물체를 순차적으로 임베딩 함으로써 레이블의 동시 발생 종속성을 모델링한다[Wang16].

2.3. 골절 검출 및 분류(Fracture Detection or Classification)

Wang 등은 X-ray 데이터를 사용해 고관절 및 골반의 골절을 검출하기 위한 연구를 진행했다. 이 연구에서는 두 단계를 거쳐 골절의 존재유무를 파악한다. 1단계에서는 원본 X-ray를 DenseNet-121에 입력해 가능성 지도

(Probability Map)를 출력한다. 그 후 가능성 지도를 기반으로 관심영역 (Region of Interest, ROI)을 생성하고, 2단계의 입력으로 사용한다. 2단계에서는 ResNet-18을 사용해 골절의 확률을 계산한다. 이 후 1단계와 2단계의 출력을 기반으로 골절의 존재 유무를 최종 출력한다[Wang19].

Nicolaes 등은 척추 3D CT 이미지를 사용해 골절을 검출하기 위한 연구를 진행했으며, 척추 골절의 검출을 위해 동시에 두개의 컨볼루션 신경망을 사용한다. 하나의 신경망은 원본 이미지를 입력으로 사용하고, 또 다른 하나는 골절부위를 중심으로 분할된 이미지를 입력으로 사용한다. 이미지의 분할은 레이블링 된 학습데이터의 좌표를 기반으로 함께 학습되며, 두개의 신경망은 각각의 컨볼루션 층으로 이루어지고, 이 후 동일한 컨볼루션 및 완전 연결 층으로 연결된다. 이 골절 검출 모델은 최종적으로 골절의 존재 유무를 출력한다 [Nicolaes19].

Rayan 등은 팔꿈치 X-ray를 기반으로 골절을 검출하기 위해 컨볼루션 신경망과 재귀 신경망(RNN)을 함께 사용했다. 최종모델은 CNN-RNN 순서로 이루어져 있으며, 두 단계로 나누어 학습된다. 1단계에서는 X-ray데이터를 흑백-스케일로 전처리 후 입력으로 사용하며, 이 단계에서는 CNN만 학습된다. 2단계에서는 CNN의 말단에 RNN이 추가된다. 이 단계에서는 모델에 같은 골절에 대한 서로 다른 3가지 이미지를 순차적으로 입력하며 학습시킨다. 이 때 CNN은 학습되지 않는다. 모델의 최종 출력은 골절의 존재 유무다[Rayan19].

Tripathi 등은 대퇴골 X-ray 기반으로 SVM을 사용해 골절 검출을 시도했다. 전처리를 통해 이미지 내의 잡음을 제거 후, 캐니 경계선 검출(Canny Edge Detection)을 적용한다. 이후 SVM을 활용해 뼈에 존재하는 비정상적인 경계선을 검출한다[Tripathi17].

골절의 검출과 더불어 골절로 예측되는 위치를 시각화하는 연구도 진행되었다. Lindsy 등은 손목 X-ray를 사용해 골절의 검출 및 골절위치의 시각화를 시도했으며 이를 위해 U-Net을 기반으로 재구성된 모델을 사용했다. 모델은 골절 확률을 계산 하여 골절 여부를 판단하고, 열지도(Heat Map)를 사용해 골절로 예측되는 위치를 시각화한다. [Lindsy18]. Thian 등은 손목 X-ray를 데이터로 사용했으며 R-CNN을 사용해 골절을 검출하고, 골절로 예상되는 위치를 경계상자(Bound Box)로 표시한다[Thian19].

Chung 등은 어깨 X-ray를 기반으로 골절을 검출하여 몇 군데에서 골절이 발생했는지 분류하는 연구를 진행했고, 제안된 방법은 골절이 몇 군데에서 발생했는지를 출력한다. 본 논문에서는 새로운 방법이 제시되지는 않았지만, 정형외과 연구자의 입장에서 데이터가 충분히 누적될 경우 실제 정형외과 의사보다 더 나은 검출성능을 갖출 수 있을 것으로 판단한다. 또한 기존 정형외과 평가와 비교하여 치료 및 치료 결과를 개선할 수 있는지에 대한 추가 연구의 필요성을 언급한다[Chung18].

Kroguе 등은 골반 및 고관절 X-ray 데이터를 기반으로 골절의 분류를 진

행했다. 제안된 방법은 두 단계에 걸쳐 이미지를 입력한다. 첫 단계에서는 ResNet을 활용해 원본 이미지 내에서 골절로 의심되는 부위를 분할(Segmentation)한다. 이 후 분할된 이미지를 Densely Connected CNN에 입력하여 골절의 유형을 분류한다. 제안된 방법의 출력은 두 가지로 ResNet의 경계상자와 Densely Connected CNN의 최종 분류다. 최종 출력 레이블은 6가지로 정상, 골절 유형 3가지, 하드웨어 2가지이다. 하드웨어는 인공관절 혹은 이 전 수술(ORIF, Open Reduction and Internal Fixation)에 사용된 기구를 말한다[Krogue19].

Kazi 등은 대퇴골 X-ray를 사용해 골절을 형태별로 분류하기 위한 연구를 진행했다. 제안된 방법은 원본 이미지와 추출된 관심영역을 함께 입력으로 사용한다. 관심영역 추출 네트워크와 분류 네트워크는 사전학습된 GoogleNet으로 구성되었으며, 학습 단계에서 두 네트워크는 함께 학습된다. 제안된 방법의 출력 레이블은 정상 및 골절 유형([A, B]*[1, 2, 3])으로 총 7가지로 분류된다[Kazi17].

Pranata 등은 발뒤꿈치 CT의 절단면 이미지를 기반으로 골절 분류를 위한 연구를 진행했다. 제안된 방법은 CNN을 통해 생성된 특징 지도(Feature Map)에 케니 경계선 검출을 적용한다. CNN 모델은 ResNet-50 혹은 VGGNet-16이 사용되었다. 제안된 방법의 최종 출력 레이블은 정상 및 골절 유형으로 총 9가지로 분류된다[Pranata19].

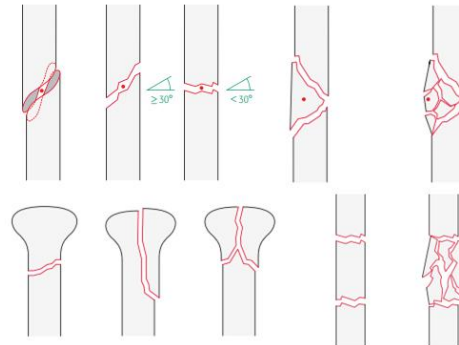
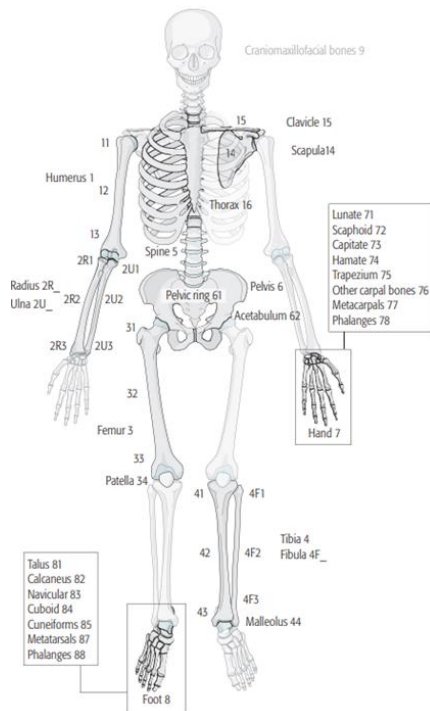
Bayram 등은 대퇴골 X-ray를 기반으로 골절의 범주를 분류하는 연구를 진행했다. 이 연구에서는 원본 X-ray에 뼈와 함께 등장하는 근육 및 기타조직을 제거하기 위해 Niblack Thresholding 기법을 적용했고, 이 과정에서 생성되는 잡음을 제거하기 위해 SVM 기반의 잡음 제거 방법을 제안했다. 제안된 방법은 입자의 둘레, 영역, 각도 등을 기반으로 노이즈와 골절로 생긴 뼈 파편을 구분하고 노이즈를 제거한다. 이 후 노이즈가 제거된 이미지를 기반으로 파편 수, 골절 각도 및 부위 등의 정보를 추출하여 SVM 기반 골절 분류기의 입력으로 사용한다. 분류 모델의 출력 레이블은 9가지([A, B, C] * [1, 2, 3])이다[Bayram16].

Ⅲ. 다발성 골절 데이터

3.1. AO/OTA 골절 및 탈구 분류

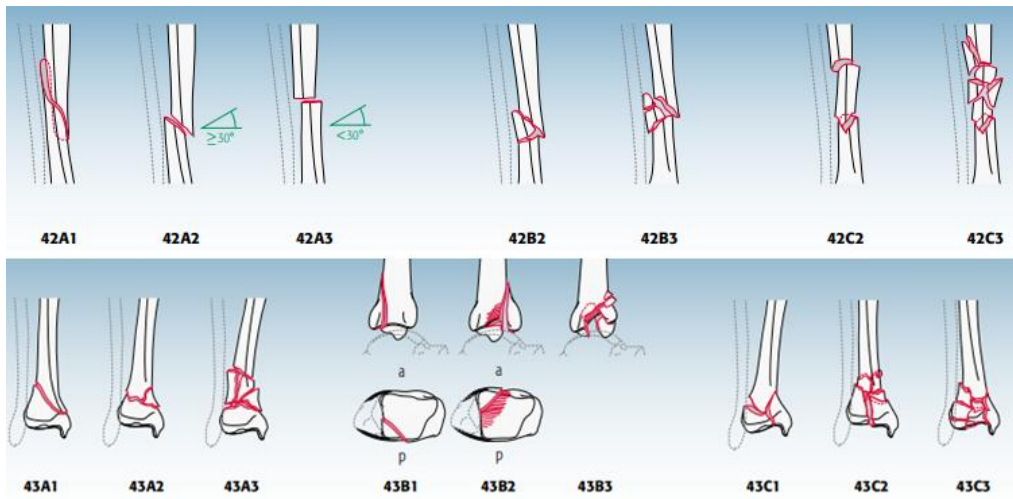
AO재단(AO Foundation)과 정형외과 외상 협회(Orthopaedic Trauma Association, OTA)는 1996년에 골절 및 탈구에 대한 개요서를 최초로 발표했고, 현재까지 지속적인 임상 연구를 통해 오류를 수정하고 새로운 분류를 추가하였다. AO/OTA의 골절 및 탈구 분류 기준은 분류를 위한 표준화된 방법론을 제시하며, 일관된 시스템을 구축하기 위해 반드시 필요하다[AO/OTA18]. AO/OTA 골절 및 탈구 분류는 실제 정형외과에서 골절을 분류하기 위한 기준으로 사용된다. 본 논문의 실험에 사용된 하반신 골절 CT 데이터의 레이블은 AO/OTA 골절 및 탈구 분류를 기반으로 작성되었다.

[그림 3-1]은 AO/OTA 골절 및 탈구 분류의 분류 기준 및 분류 예시를 보여준다. 골절을 분류하는 첫 번째 기준은 (a)와 같이 골절이 발생한 뼈의 종류 및 골절이 발생한 정확한 해부학적 위치이다. 다음으로 골절은 (b)처럼 골절의 각도 혹은 골절의 형태 등을 기준으로 분류 된다. (c)는 각 기준에 따라 분류된 골절 범주의 예시이다. 각 골절의 범주는 기본적으로 4~5자리의 숫자 및 알파벳으로 표기되며, 앞의 2~3자리는 골절의 위치, 그 다음은 골절의 형태 및 특성을 표기한다. 골절의 특징별로 세분화하기 위해 4~5자리의 표기 뒤에 .O 형태로 추가하여 표기하기도 한다. 실험에 사용된 데이터의 레이블은 이 분류를 기준으로 작성되었다. 자세한 사항은 3.2절에서 서술한다.



(b)

(a)



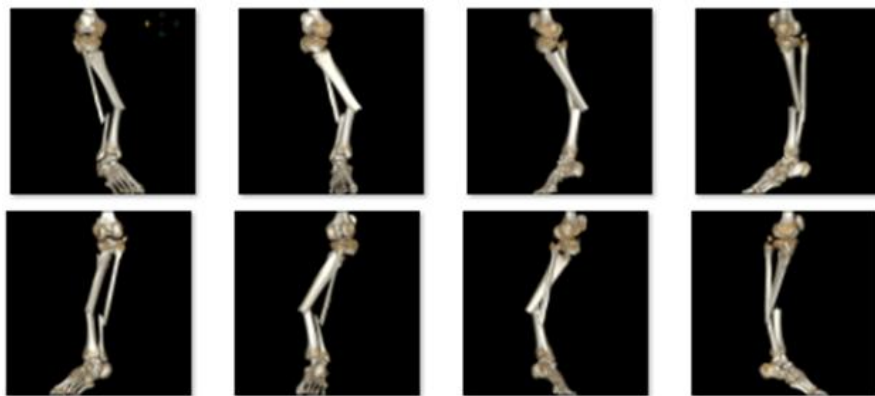
(c)

[그림 3-1] AO/OTA 골절 및 탈구 분류

(a) 골절이 발생한 뼈 및 해부학적 위치에 따른 분류 기준 (b) 골절의 각도, 형태 등에 따른 분류 예시 (c) 각 기준에 따라 세분화 된 골절 범주 예시

3.2 CT 영상 데이터 및 레이블




본 논문의 실험에 사용된 데이터는 직접 수집한 하반신(다리) 골절 CT 영상으로 총 181명의 환자에 대한 데이터가 수집되었다. 하지만 노이즈, 레이블링 등의 문제로 인해 실제 실험에는 149명의 데이터만 사용되었으며, 사용된 이미지 수는 1,453장이다. 했다. 각 환자들은 평균적으로 9.75장의 이미지를 갖고 있으며, 각 환자는 [그림 3-2]와 같이 골절 부위를 일정 각도로 회전시키며 촬영된 CT 이미지를 갖고 있다.



[그림 3-2] 환자의 CT 이미지

실험에 사용된 데이터의 레이블은 3.1절에서 서술한 “AO/OTA 골절 및 탈구 분류”를 기준으로 작성되었다. 이 기준에 따라 다리에 발생한 골절은 (세분화를 제외하고) 총 127개의 범주로 분류된다. 하지만 본 논문에서는 데이터 부족 문제로 인해 수집된 데이터에 실제로 등장하는 20개의 레이블만 사용한다.

수집된 데이터의 환자들은 여러 골절이 동시에 발생한 다발성 골절이다. 따라서 각 환자들은 여러 개의 골절 범주를 동시에 진단 받는다. 이로 인해 각 이미지는 최소 1개, 최대 4개의 레이블을 갖는다. 하지만 전체 환자 181명 중 4개의 레이블을 갖는 환자는 3명이며, 이 환자들은 노이즈 등의 문제로 인해 실제 실험에 사용되지 않았다. 이로 인해 본 논문의 실험에서는 각 이미지별로 최소 1개, 최대 3개의 레이블까지만 분류한다. 실제 데이터의 예시는 [표 3-1]과 같다. 따라서 이 논문에서는 다발성 골절 범주 분류 문제를 다중 레이블 분류 문제로 정의하고, 다발성 골절의 범주를 분류하기 위한 딥러닝 기반의 방법을 제시한다.

환자 번호	이미지	레이블
01147443		41A2 42A3 4F2B
03830112		42B2
04959175		42B2 4F2B

[표 3-1] 다발성 골절 데이터의 예시

3.3 데이터 불충분 및 불균형 문제

이 논문의 실험에서 사용된 데이터는 각 이미지가 1~3개의 레이블을 갖는 다중 레이블 데이터다. 각 이미지에는 20개의 레이블들이 복합적으로 조합되어 등장한다. 다중 레이블 분류 문제는 다중-클래스 분류(Multi-Class Classification, 혹은 단일-레이블 분류) 문제보다 더 많은 데이터를 필요로 한다. 하지만 실험에 가용한 데이터는 149명이고, 결과적으로 1,453장의 이미지를 사용한다. 환자수를 기준으로 각 레이블별 데이터 분포현황은 [표 3-2]와 같다. 10건 이상의 데이터는 파란색, 10건 미만의 데이터는 노란색으로 표기되었다.

※ 단위 : 환자 수

레이블	개수	비율	레이블	개수	비율	레이블	개수	비율
32A3	7	2.18%	42C2	6	1.87%	4F1B	17	5.30%
41A2	1	0.31%	42C3	6	1.87%	4F2A	36	11.21%
42A1	42	13.08%	43A1	1	0.31%	4F2B	63	19.63%
42A2	13	4.05%	43A2	5	1.56%	4F3A	12	3.74%
42A3	10	3.12%	43A3	3	0.93%	4F3B	12	3.74%
42B2	55	17.13%	43C3	2	0.62%	4FB2	2	0.62%
42B3	22	6.85%	4F1A	6	1.87%	총 레이블 20개		

 : 10건 이상 데이터  : 10건 미만 데이터

[표 3-2] 레이블별 데이터 분포

데이터셋의 레이블 카디널리티와 레이블 밀도는 [표 3-3]과 같다. 레이블 카디널리티와 레이블 밀도는 각각 식(2)와 식(3)을 따른다[Tsoumakas07]. 식(2)와 (3)의 $|D|$ 는 데이터의 개수, $|Y|$ 는 각 데이터의 대상 레이블 개수, $|L|$ 은 레이블의 개수다.

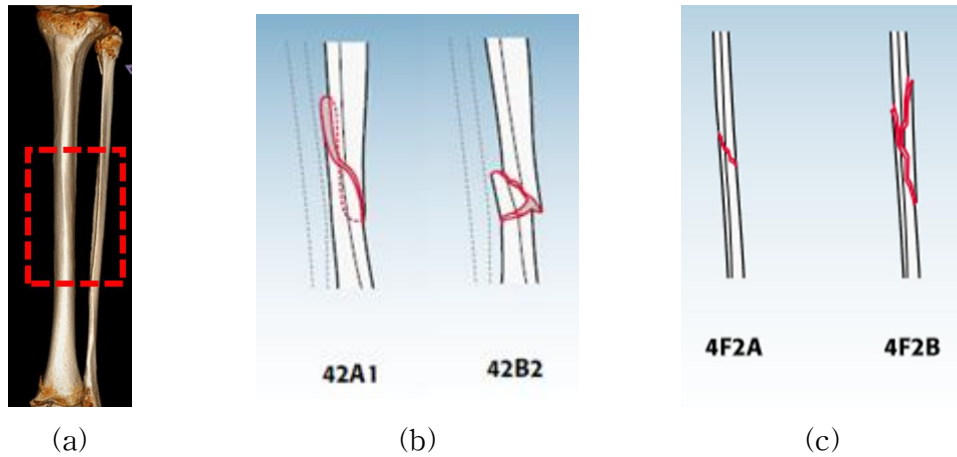
$$\text{레이블 카디널리티 (Label Cardinality)} = \frac{1}{|D|} \sum_{i=1}^{|D|} |Y_i| \quad (2)$$

$$\text{레이블 밀도(Label Density)} = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i|}{|L|} \quad (3)$$

환자 수($ D $)	대상 레이블 수($ Y $)	레이블 카디널리티	레이블 밀도
149	321	2.154	0.107

[표 3-3] 실험 데이터의 레이블 카디널리티와 레이블 밀도

데이터셋에서 가장 많은 비중을 차지하는 레이블은 42A1, 42B2, 4F2A, 4F2B이다. 이 4개 레이블의 비율은 51.06%로 전체 데이터 중 절반 이상을 차지하고, 데이터가 10건 미만인 레이블 10개(32A3, 41A2, 42C2, 42C3, 43A1, 43A2, 43A3, 43C3, 4F1A, 4FB2))는 전체 데이터 중 12.15%밖에 되지 않는다. 이처럼 각 레이블별 데이터 분포가 불균형한 것이 직접 수집한 데이터를 사용한 이 실험의 한계점이다.



[그림 3-3] 골절의 위치 및 형태

(a) 정강이 뼈(경골과 비골) (b) 경골 골절 레이블 (c) 비골 골절 레이블

레이블 분포가 불균형한 이유는 실제로 골절 환자의 대부분은 정강이 뼈인 경골(Tibia)과 비골(Fibula)의 중간 위치가 골절 되는 경우가 많기 때문이다. [그림 3-3]의 (a)는 실제 정강이뼈의 사진이고, (b)와 (c)는 데이터 내에서 가장 많이 등장하는 레이블 4개 중 각각 경골과 비골에 대한 골절 레이블이다. (a)의 굵은 뼈가 경골, 얇은 뼈가 비골이며, 붉은 점선으로 표시된 영역이 (b)와 (c) 골절이 실제로 발생하는 위치이다.

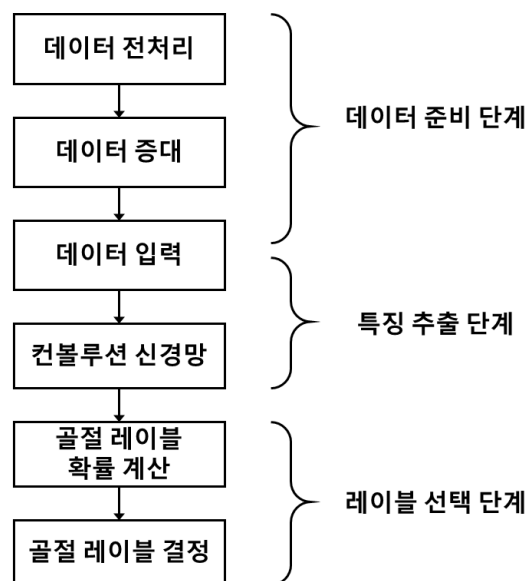
본 논문의 실험에 사용된 데이터셋은 데이터 불충분 문제를 갖고 있다. 이 때문에 학습 데이터셋과 테스트 데이터셋을 구분하는 기준이 환자 단위가 되지 못했고, 데이터셋의 구성에 따라 실험결과에 편차가 존재했다. 따라서 본 논문에서는 실험결과의 유효성 검증을 위해 교차검증을 진행했다. 데이터셋

구성 방법은 4.3절, 교차 검증은 5.1절에서 상세하게 서술한다. 또한 본 논문에서는 데이터 불충분 문제를 극복하기 위해 데이터 증대(Data Augmentation)를 적용했으며 이 내용은 4.2절에서 서술한다. 데이터셋의 불균형 문제를 극복하기 위한 레이블 그룹화 혹은 오버 샘플링을 활용한 추가 실험 결과는 5.3절과 5.4절에서 제시한다.

IV. 제안하는 딥러닝 기반 다발성 골절 다중-부류 분류 방법

4.1 다발성 골절 분류 방법의 구조

본 논문에서 제안하는 다발성 골절 분류 방법은 CT 이미지를 입력으로 각 골절의 특징을 추출 및 식별한다. 그 후 입력 된 이미지에 대해 모든 골절 레이블의 확률을 독립적으로 계산 후 최종적으로 1~3개의 레이블을 결정한다. 제안하는 방법의 전체 과정은 [그림 4-1]과 같다.



[그림 4-1] 제안하는 다발성 골절 분류 방법의 흐름도

4.2 데이터 전처리 및 데이터 증대

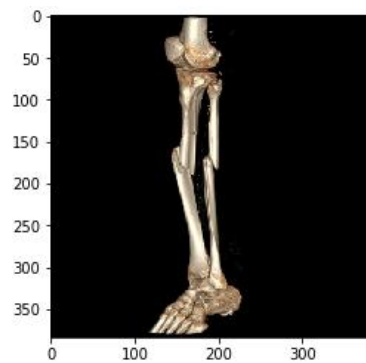
데이터를 모델에 입력하기 위한 데이터 준비과정에서는 데이터 전처리를 먼저 진행한다. CT 기기에서는 CT 이미지를 3D로 출력하기 위해 RGBD 4개 채널로 표현한다. 하지만 실험에 사용된 CT 이미지는 캡처본이기 때문에 깊이(Depth) 차원에 특별한 정보가 표현되지 않았다. 따라서 불필요한 연산을 줄이기 위해 전처리 과정에서 깊이 채널을 제거한다. 또한 CT촬영이 진행되는 기기에 따라 이미지의 해상도가 다르기 때문에 같은 크기로 먼저 조정한다. 그 후 이미지 내의 픽셀들을 0과 1사이의 값으로 선형정규화 해준다. 선형정규화는 식 (4)를 따른다. 식에서 src 는 원본 이미지, dst 는 처리 후 이미지, i 와 j 는 이미지의 y, x 좌표를 의미한다.

$$dst(i, j, src) = \frac{src_{ij} - \min(src)}{\max(src) - \min(src)} \quad (4)$$

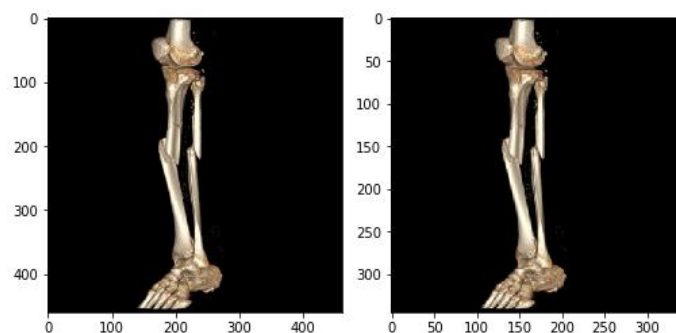
크기 변경(Size Reshape)	80% or 90% or 100% or 110% or 120%
회전(Rotate)	3° or 6° or ... or 30°
반전(Flip)	False or 좌우반전
이동(Translate)	상하 : -8%~8%, 좌우 : -12%~12%
자르기(Crop)	중앙 기반 자르기(Center based Cropping)
잡음(Noise)	점 잡음(Salt and Pepper Noise)

[표 4-1] 데이터 증대 기준

본 논문의 실험에서 사용된 데이터는 불충분 문제를 갖고 있다. 이 문제점을 극복하기 위해 본 논문의 실험에서는 데이터 증대를 사용해 입력 데이터를 변형하여 입력한다. 데이터 증대 기준은 [표 4-1]과 같다. 데이터 증대기준의 적용 방법은 원본 이미지에 [표 4-1]의 각 증대기준을 순차적으로 무작위로 선택하여 적용한다. 예시로 원본 이미지 [그림 4-2]에 데이터 증대를 적용하는 과정을 서술한다. [그림 4-2] 원본 이미지의 크기는 384×384다. 이 이미지에 해상도 변경 기준이 적용 되면 그림[4-3]과 같이 이미지의 크기가 기준 내에서 무작위로 변경된다.

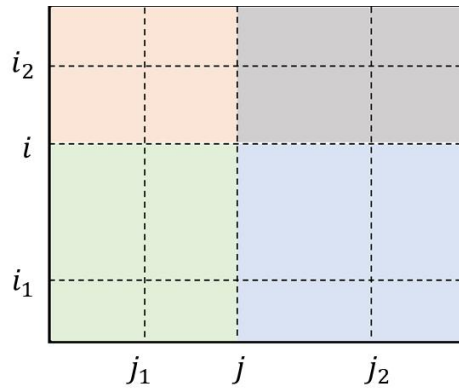


[그림 4-2] 원본 이미지



[그림 4-3] 원본 이미지에 크기 변경 적용 예시

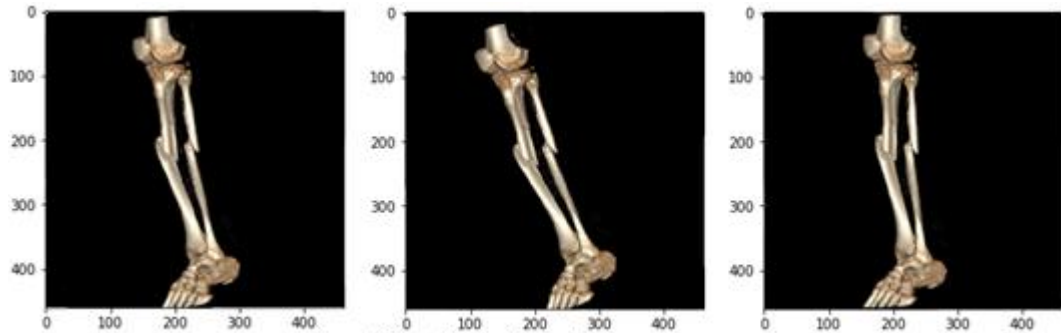
이미지의 크기 변경을 위해 쌍선형 보간법(Bilinear Interpolation)을 사용한다. 쌍선형 보간법을 사용한 이유는 이미지의 크기 변경 과정에서 생기는 에일리어싱(Aliasing)을 최소화하기 위함이다. 쌍선형 보간법은 식 (5)를 따른다. 식의 src 는 원본 이미지, dst 는 변경이 적용된 이미지이며 i 와 j 는 이미지 내 픽셀의 y 좌표와 x 좌표를 나타낸다. 식(4)는 [그림 4-4]의 예시 이미지를 기준으로 작성되었다. dst_{ij} 의 값은 기존 픽셀 중 인접한 픽셀인 $src_{i_1j_1}$, $src_{i_2j_1}$, $src_{i_1j_2}$, $src_{i_2j_2}$ 의 값과 가중치를 기반으로 계산된다. 가중치는 dst_{ij} 와 기존 픽셀과의 거리에 역비례하게 적용한다.



[그림 4-4] 이미지 보간법 예시

$$dst(i, j, i_1, j_1, i_2, j_2, src) = \lfloor \frac{i_2 - i}{i_2 - i_1} \left\{ \frac{(j_2 - j)src_{i_1j_1} + (j - j_1)src_{j_1i_2}}{j_2 - j_1} \right\} + \frac{i - i_1}{i_2 - i_1} \left\{ \frac{(j_2 - j)src_{i_2j_1} + (j - j_1)src_{j_1i_2}}{j_2 - j_1} \right\} \rfloor \quad (5)$$

이미지의 크기 변경 후에는 [그림 4-5]와 같이 기준 내의 무작위 각도로 이미지가 회전된다. 이미지 회전은 식 (6)을 따른다. 회전은 이미지의 중심좌표를 기반으로 적용되며, 회전에 의해 *dst*의 범위를 벗어나는 *src*는 삭제되고, *src*의 범위를 벗어나 *dst*에 생기는 여백은 배경과 같은 픽셀 값으로 채운다.



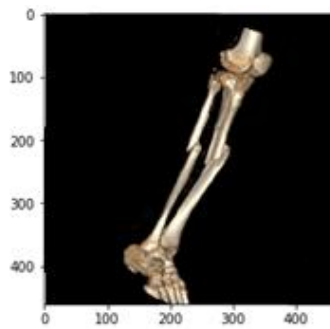
[그림 4-5] 이미지 크기 변경 후 회전 적용 예시

$$dst(i, j, \theta, src) = \begin{cases} src_{\alpha\beta}, & \text{if } (0 \leq \alpha \leq src.rows - 1) \text{ and } (0 \leq \beta \leq src.cols - 1) \\ 0, & \text{else} \end{cases} \quad (6)$$

$$\alpha = \lfloor -j \cdot \sin\theta + i \cdot \cos\theta + (1 - \cos\theta) \cdot \frac{src.rows}{2} + \sin\theta \cdot \frac{src.cols}{2} \rfloor$$

$$\beta = \lfloor j \cdot \cos\theta + i \cdot \sin\theta + (1 - \cos\theta) \cdot \frac{src.cols}{2} + \sin\theta \cdot \frac{src.rows}{2} \rfloor$$

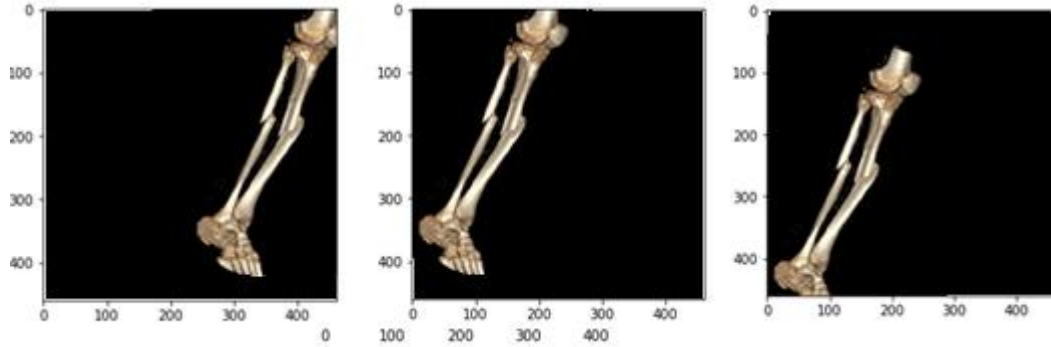
이미지의 회전 후에는 [그림 4-6]처럼 반전이 적용된다. 이미지의 반전 여부는 50% 확률로 결정되며, 상하 반전은 적용되지 않고 좌우반전만 적용된다. 이미지 반전은 식 (7)을 따른다.



[그림 4-6] 이미지 회전 후 반전 적용 예시

$$dst(i, j, src) = src_{i, src.cols-j-1} \quad (7)$$

이미지 반전 후에는 [그림 4-7]과 같이 기준 내의 무작위 범위로 이미지가 상하좌우로 이동된다. 이미지의 이동은 식 (8)을 따른다. dst_{ij} 의 범위를 벗어나는 src_{ij} 는 삭제되며, src_{ij} 의 범위를 벗어나는 dst_{ij} 는 배경과 동일한 값으로 채운다. 이미지의 이동 범위를 상하는 $-8\% \sim 8\%$, 좌우는 $-12\% \sim 12\%$ 로 제한한 이유는 다음과 같다. 기준 이상으로 이동했을 경우 src 에서 어떤 빠인 지 구분하기 위한 특징이 포함된 영역이 과도하게 삭제되고, 골절 부위가 과도하게 제거되어 골절을 분류하기 위한 특징이 손상 될 수 있기 때문이다.

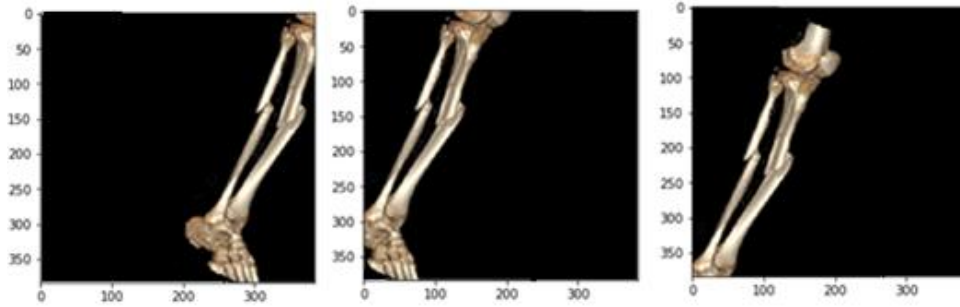


[그림 4-7] 이미지 반전 후 이동 적용 예시

$$dst(i, j, t_i, t_j, src) = \begin{cases} src_{i-t_i, j-t_j}, & \text{if } (0 \leq i-t_i \leq src.rows-1) \text{ and } \\ & (0 \leq j-t_j \leq src.cols-1) \\ 0, & \text{else} \end{cases} \quad (8)$$

$$t_i = \lfloor src.rows \cdot raito_{rows} \rfloor, \quad t_j = \lfloor src.cols \cdot raito_{cols} \rfloor,$$

각 이미지는 크기 변경 단계에서 적용된 기준에 따라 다른 크기를 갖게 된다. 본 논문에서 제안하는 방법의 다발성 골절 분류모델에는 같은 크기의 이미지를 입력해줄 필요가 있다. 따라서 각 이미지에 자르기를 적용하여 [그림 4-8]처럼 모든 이미지가 같은 크기를 갖게 한다. 이미지 자르기는 식 (9)를 따르며 이미지의 중앙을 기반으로 적용된다. 따라서 이미지 자르기 후에는 각 이미지의 가장자리가 제거된다.

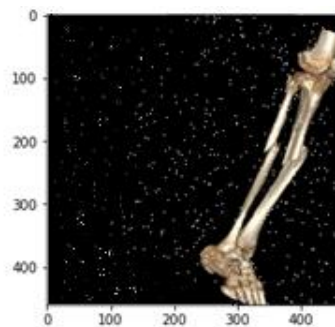


[그림 4-8] 이미지 이동 후 자르기 적용 예시

$$dst(i, j, H, W, src) = src_{i+H, j+W}$$

$$H = \lfloor \frac{src.rows - dst.rows}{2} \rfloor, W = \lfloor \frac{src.cols - dst.cols}{2} \rfloor \quad (9)$$

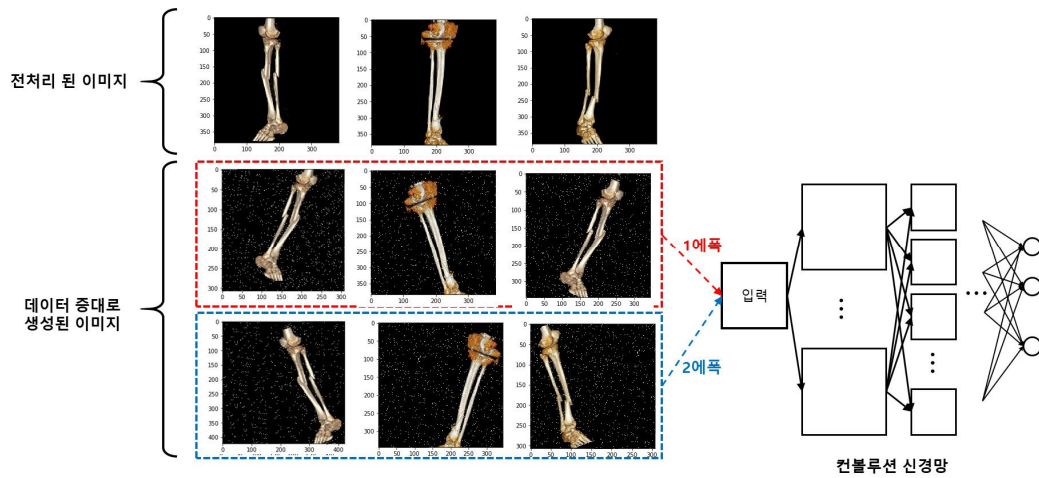
이미지 자르기 후에는 점 잡음(Salt and Pepper Noise)을 적용한다. 적용 방법은 이미지 내에서 무작위로 선택된 픽셀들의 값을 $\min(src)$ 와 $\max(src)$ 로 변경한다. [그림 4-9]는 점 잡음이 적용된 예시를 보여준다.



[그림 4-9] 이미지 자르기 후 점 잡음 적용 예시

[그림 4-2]부터 [그림 4-9]의 이미지는 검은색 배경을 갖는다. 이 검은색 배경은 CT 이미지가 본래 갖고 있는 RGBD 4개의 채널 중 깊이(Depth)채널에 의한 것이다. 하지만 이 깊이 채널은 데이터 증대보다 먼저 진행되는 전처리 단계에서 제거된다. 따라서 실제로 CNN 모델에 입력되는 이미지는 모두 하얀색 배경만 갖게 된다. 본 절에서는 이미지 증대 과정의 가시성을 위해 전처리가 되지 않은 이미지를 기반으로 예시를 보였다.

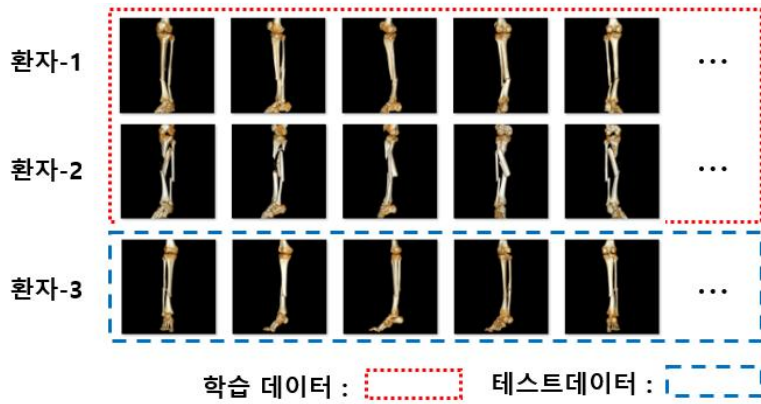
기존 연구들은 데이터 증대를 적용 할 때 사전에 모든 이미지에 데이터 증대를 적용 후 생성된 이미지들을 동시에 모두 학습에 사용한다. 예를 들어, 100장의 이미지를 증대시켜 1,000장의 이미지를 생성했다면 1,000장의 사진을 매 에폭마다 학습에 사용한다. 하지만 본 논문의 실험에서 사용하는 데이터의 원본은 2,000장 미만이다. 이 때문에 본 논문에서 제안하는 방법은 골절 분류 모델의 오버피팅을 방지하고자 [그림 4-10]과 같은 방법으로 증대된 이미지를 사용한다. 예를 들어 이미지가 100장이 있고, 골절 분류 모델을 1,000 에폭동안 학습시킨다고 가정하자. 골절 분류 모델이 학습되는 매 에폭마다 100장의 이미지에 무작위로 데이터 증대를 적용하여 100장의 이미지를 생성하고, 생성된 이미지를 입력하여 골절 분류 모델을 학습시킨다. 즉 기존 방법처럼 데이터 증대로 생성된 이미지들이 모두 한 에폭에 사용하지 않고, [그림 4-10]처럼 각 에폭에 생성된 이미지들은 해당 에폭에만 사용한다. 따라서 제안하는 방법의 다발성 골절 분류 모델은 학습되는 동안 매 에폭마다 서로 다른 이미지를 기반으로 학습된다.



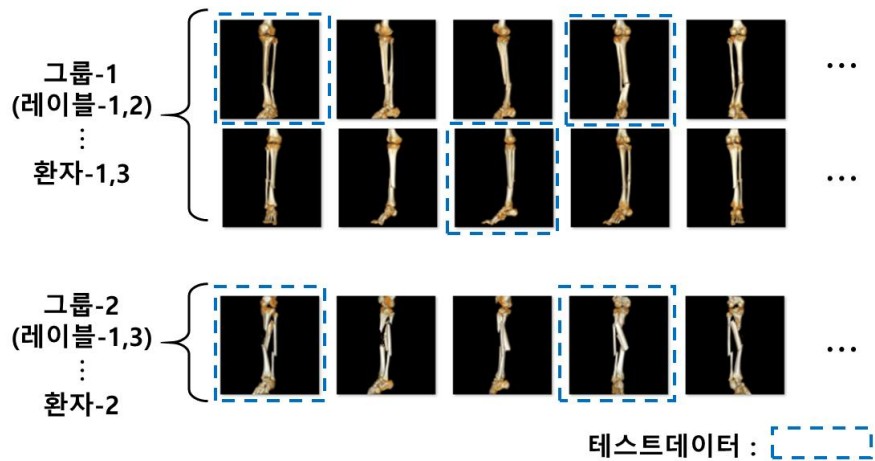
[그림 4-10] 데이터 증대 후 컨볼루션 신경망의 학습을 위한 이미지 입력

4.3. 학습/테스트 데이터셋 구성 방법

본 논문의 실험에서 사용된 데이터는 각 환자의 골절부위 CT 이미지를 일정각도로 회전시키며 캡처한 이미지이다. 따라서 한 명의 환자가 여러 장의 CT 이미지를 갖고 있으며, 각 이미지는 같은 형태의 골절을 서로 다른 시점에서 캡처한 것이다. 따라서 다발성 골절 분류 모델이 각 골절의 시점별 특성을 모두 학습하기 위해서는 한 환자의 이미지를 모두 학습해야한다. 즉, [그림 4-11]의 (a)처럼 학습 데이터셋과 테스트 데이터셋을 구분하는 기준이 환자 단위인 것이 바람직하다.



(a)



(b)

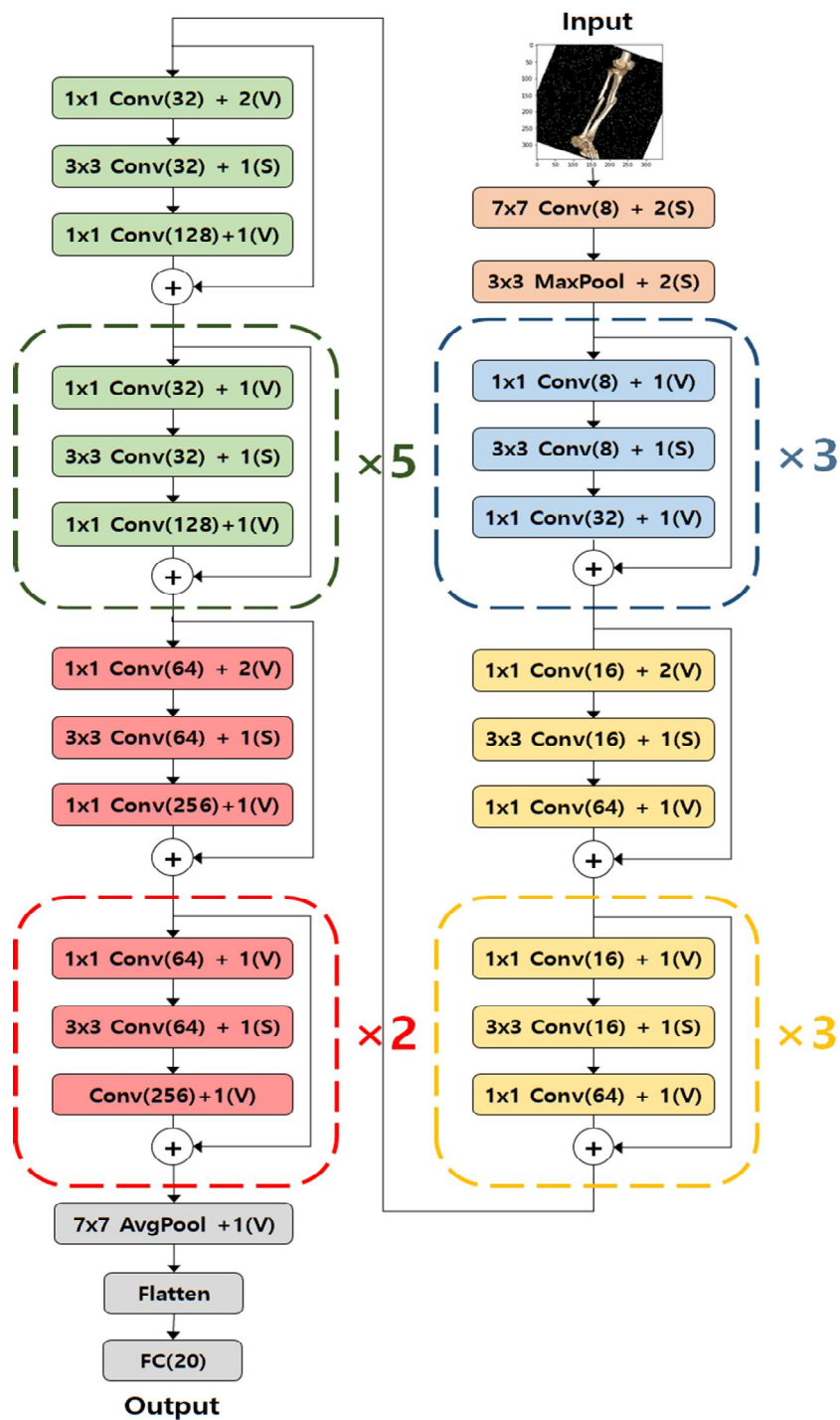
[그림 4-11] 데이터셋 구성 방법

각 레이블들이 복합적으로 조합되어 각 이미지에 존재하는 다중-레이블 분류 문제의 경우 기존 단일-레이블 분류 문제보다도 데이터셋을 구성하는 것이 까다롭다. [표 3-1]을 통해 이 실험에 사용된 데이터는 불충분하고, 각 레이블 별 분포가 불균형함을 알 수 있다. 특히 환자 수가 적은 레이블의 경우 환자 단위로 데이터셋을 구분하는 것이 불가능했다. 이와 같은 문제로 인해 이 실

험에서는 데이터셋을 구분하는 기준이 환자가 되지 않고, [그림 4-11]의 (b)와 같은 방법으로 데이터셋을 구성한다. 같은 레이블 조합을 갖는 환자들을 그룹화하고, 각 그룹에 속한 이미지 중 일정 비율을 무작위로 테스트데이터로 선택한다. [그림 4-11]의 (b)를 기준으로 예를 들면 레이블-1과 레이블-2를 동시에 갖는 환자-1과 환자-3을 그룹-1, 레이블-1과 레이블-3을 갖는 환자-2를 그룹-2로 나눈다. 그 후 그룹-1의 이미지 중 무작위로 테스트데이터를 선택하고, 그룹-2의 이미지 중 무작위로 테스트데이터를 선택한다. 그룹 내에서 테스트데이터를 무작위로 선택하는 이유는 다음과 같다. 일정한 기준에 따라 데이터셋을 추출할 시 특정 각도의 이미지들만 테스트데이터로 선택된다. 이러한 규칙성이 생기지 않도록 해 다발성 골절 분류 모델을 일반화하기 위함이다.

4.4 다발성 골절 분류 모델

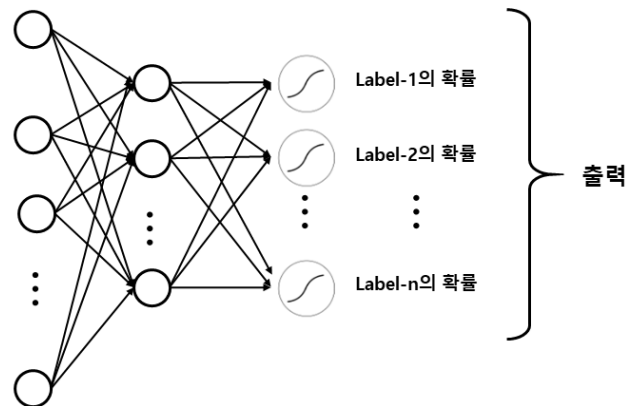
본 논문의 실험에서는 충분한 양의 골절 데이터를 확보하지 못했다. 따라서 CNN 모델이 학습데이터에 과적합 되는 현상을 방지하고자 모델에 사용되는 매개변수의 수를 줄여야 했고, 동시에 CT 이미지에서 효과적으로 특징을 추출할 필요가 있었다. 이를 위해 본 논문에서 제안하는 다발성 골절 분류 모델은 레스넷(Residual Neural Network, ResNet)의 구조를 기반으로 구성되었다. 다발성 골절 분류모델의 구조는 [그림 4-12]와 같다.



[그림 4-12] 제안하는 방법의 다발성 골절 분류 모델

[그림 4-12]의 컨볼루션 블록과 풀링 블록의 ' $x \times y$ '는 필터의 크기를 나타내고, 괄호 안의 숫자는 필터의 개수를 나타낸다. '+z'는 스트라이드(Stride)를 나타내며, 패딩(Padding)여부는 괄호 내의 (S)와 (V)로 각각 패딩 없음(Same)과 패딩 있음(Valid)로 표기한다. [그림4-12] 내의 점선으로 표시된 각 '블록'은 표시된 ' $\times n$ '만큼 반복적으로 존재하기 때문에 축약하여 표시됐다.

제안하는 다발성 골절 분류 모델은 데이터 증대를 통해 생성된 이미지들을 입력으로 사용한다. 입력된 각 골절 CT 이미지는 컨볼루션과 풀링 층을 통해 특징이 추출되고, 완전 연결층(Fully Connected, FC)과 시그모이드 함수를 통해 각 레이블들의 확률이 계산된다. 다발성 골절 분류 모델은 각 '블록' 단위로 잔차 모듈을 활용해 입력과 기대 출력의 차이를 학습하도록 했다. 이를 통해 많은 수의 컨볼루션 층을 사용할 수 있었고, 효과적으로 골절의 특징을 추출한다. 다중-클래스 분류(Multi-Class Classification) 문제에서는 소프트맥스(Softmax) 함수를 주로 사용해 각 레이블들의 확률 총 합이 1.0이 되도록 출력한다. 하지만 본 논문에서 다루는 문제는 다중-레이블 분류(Multi-Label Classification)문제이기 때문에 시그모이드(Sigmoid) 함수를 사용해 각 레이블의 확률을 독립적으로 계산한다. 즉, 제안하는 다발성 골절 분류 모델에서는 [그림 4-13]과 같이 다층 퍼셉트론을 통해 출력되는 정보를 바탕으로 각 노드별로 시그모이드 함수를 적용한다. 이를 통해 출력되는 것은 각 레이블들의 확률이 된다. 예를 들어 다발성 골절 분류 모델의 출력은 레이블-1의 확률은 77.5%, 레이블-2의 확률은 20.3%, ..., 레이블-n의 확률은 95.6%와 같은 형태다.



[그림 4-13] 다발성 골절 분류 모델의 출력

마지막으로 골절 레이블 결정 단계에서는 입력된 이미지에 따라 앞에서 계산된 각 레이블들의 확률을 기반으로 최소 1개, 최대 3개의 레이블을 결정한다. 여기서 최종으로 레이블을 결정하고 출력하기 위해 임계값을 적용한다. 임계값은 다발성 골절 분류 모델의 학습에 영향을 받지 않는 하이퍼-파라미터(Hyper-Parameter)로 데이터의 상태, 모델의 구조, 각 레이블별 데이터 개수 및 분포 등을 기반으로 최적의 값을 찾아야한다. 이 임계값을 기준으로 최종 레이블이 결정되는 과정은 [그림 4-14]의 의사코드와 같다.

Procedure <i>SelectLabel</i> (l, p, t)
<p>입력 : 레이블 후보 $l = \langle l_1, l_2, \dots, l_n \rangle$</p> <p>출력된 레이블 n개의 확률 $p = \langle p_1, p_2, \dots, p_n \rangle$</p> <p>임계값 t ($t < 1.0$)</p> <p>출력 : 최종 출력 레이블 $L = \langle L_1, L_2, \dots, L_m \rangle$ ($m \leq 3$)</p>
<p>begin</p> <ol style="list-style-type: none"> 레이블 확률 p 중 가장 높은 값을 갖는 3개의 레이블 확률 p_{x_m} 식별 $(p_{x_1} > p_{x_2} > p_{x_3})$ $L \leftarrow p_{x_1}$ foreach ($n = 2, 3$) if $p_{x_{n-1}} - p_{x_n} \leq t$ $L \leftarrow p_{x_n}$ else break; return L <p>end</p>

[그림 4-14] *SelectLabel*()의 의사 코드

예를 들어 [표 4-2]와 같이 각 레이블의 확률이 출력 되었고 임계값이 0.05 일 때 CT 이미지-1의 최종 레이블은 레이블-1, 레이블-2, 레이블-3이 되고, CT-이미지-2의 최종 레이블은 레이블-1, 레이블-2가 된다.

※ 임계값 : 0.05

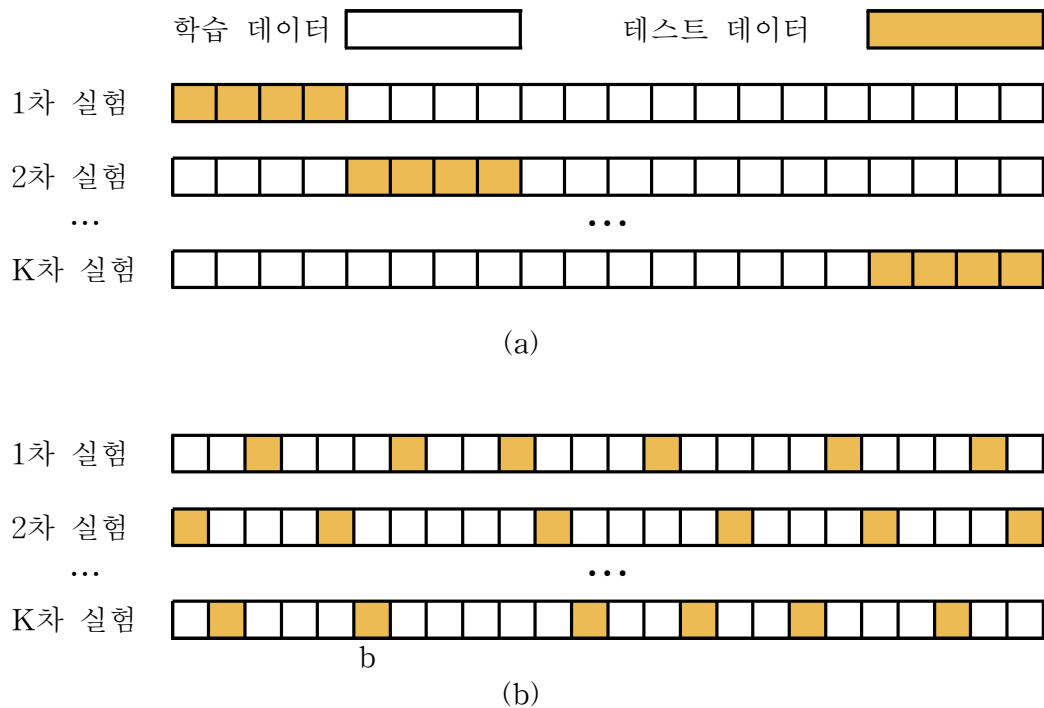
CT 이미지 - 1			CT 이미지-2		
레이블	확률	결정	레이블	확률	결정
레이블-1	0.95	O	레이블-1	0.95	O
레이블-2	0.91	O	레이블-2	0.91	O
레이블-3	0.97	O	레이블-3	0.85	X
...
레이블-n	0.10	X	레이블-n	0.10	X

[표 4-2] 최종 레이블 결정 예시

V. 실험

5.1. 교차 검증

본 논문의 실험에서 사용되는 데이터는 불충분 문제와 함께 각 레이블별 분포가 불균형한 문제를 갖고 있다. 이 때문에 학습 데이터셋과 테스트 데이터셋의 구성에 따라 결과에 편차가 생겼다. 따라서 제시하는 실험 결과의 유효성 검증을 위해 교차 검증을 진행했다. 이를 위해 K-Fold 교차검증(K-Folds Cross Validation)을 응용해 학습 데이터셋과 테스트 데이터셋의 구성을 변경하며 반복 실험하였으며 각 실험 결과의 평균을 최종결과로 사용한다 [Bengio04]. 일반적인 K-Fold 교차 검증의 데이터셋 구성 방법은 [그림 5-1]의 (a)와 같다. 하지만 특정 각도에서 캡처된 이미지들만 테스트 데이터로 추출되는 것을 방지하기 위해 무작위로 테스트 데이터를 추출했으며, 이에 따라 [그림 5-1]의 (b)와 같이 교차 검증을 진행한다.



[그림 5-1] 교차 검증을 위한 데이터셋 구성 방법

5.2. 제안한 방법의 실험 결과

이 절에서는 4장에서 제안한 다발성 골절 분류 방법에 대한 실험 결과를 제시한다. 제안하는 방법의 평가를 위한 지표는 정밀도(Precision), 재현율(Recall), 정확도(Accuracy), 해밍 로스(Hamming Loss)를 사용한다. 각 지표는 식 (10), (11), (12), (13)을 따른다[Tsoumakas07]. 식의 D 는 데이터, Y 는 정답 레이블, Z 는 출력 레이블, L 은 레이블 종류다.

$$\begin{array}{l} \text{정밀도} \\ \text{(Precision)} \end{array} = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Z_i|} \quad (10)$$

$$\begin{array}{l} \text{재현율} \\ \text{(Recall)} \end{array} = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Y_i|} \quad (11)$$

$$\begin{array}{l} \text{정확도} \\ \text{(Accuracy)} \end{array} = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \quad (12)$$

$$\begin{array}{l} \text{해밍 로스} \\ \text{(Hamming Loss)} \end{array} = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \Delta Z_i|}{|L|} \quad (13)$$

실험 결과는 학습 데이터셋과 테스트 데이터셋의 구성에 따라 편차가 생겼다. 이에 따라 5.1절에서 서술한 교차 검증 방법에 따라 5회 반복 실험했으며, 평균값을 최종 실험결과로 제시한다. 실험결과는 [표 5-1]과 같다. GoogleNet 과 ResNet-50을 사용한 분류는 제안한 방법에 비해 정밀도는 최대 6.01%, 재현율은 14.52%, 정확도는 최대 17.82% 낮다. 이전 제안 모델은 앞서 진행된 연구[이상현19]에서 제안한 모델을 해당 데이터셋에 최적화한 결과 값이다. 제안한 방법의 실험 결과는 정밀도 97.62%, 재현율 95.75%, 정확도 93.59%, 해밍 로스 0.0057이다. 제안한 방법은 기존 방법들에 비해 더 높은 분류 성능을 보여준다.

	정밀도	재현율	정확도	해밍 로스
GoogleNet	91.61%	81.23%	75.77%	0.0231
ResNet-50	95.78%	83.57%	80.62%	0.0177
이전 제안 방법 [이상현19]	96.73%	84.33%	81.99%	0.0162
제안 방법	97.62%	95.75%	93.59%	0.0057

[표 5-1] 기존 딥러닝 모델과 제안된 방법의 실험결과

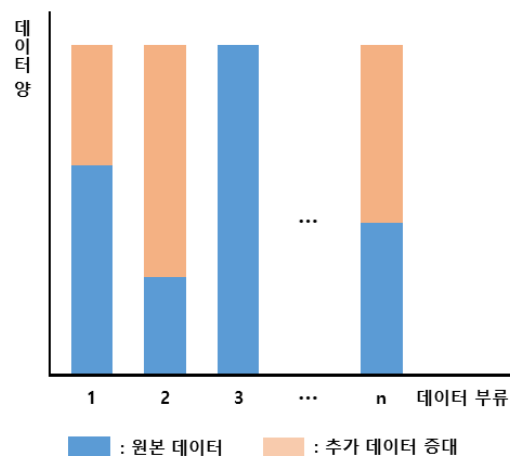
[표 5-2]는 각 이미지의 레이블 개수별로 다발성 골절 분류 모델이 정확하게 분류한 레이블 수를 나타낸다. [표 5-2]의 결과 값은 5회의 교차검증 과정 동안 각 실험들의 결과를 누적한 것이다. 제안하는 다발성 골절 분류 모델은 레이블이 1개 있는 이미지 198장 중 194장을 정확하게 분류했다. 레이블이 2개 있는 이미지 727장에 대해서 정확하게 분류한 레이블이 0개인 경우는 9장, 1개인 경우는 36장, 2개 모두 정확하게 분류한 경우는 682장이다. 레이블이 3개인 이미지 35장에 대해서 정확하게 분류한 레이블이 0개인 경우는 1장, 1개인 경우는 3장, 2개인 경우는 3장, 3장인 경우는 28장이다. [표 5-2]를 통해 이미지에 존재하는 레이블이 많을수록 다발성 골절 분류 모델의 분류 성능이 다소 하락함을 알 수 있다. 레이블이 3개인 경우는 학습을 위한 데이터가 적었다. 따라서 데이터 확보를 통해 분류 성능을 높일 수 있을 것으로 판단된다.

정확하게 분류한 레이블 수 \ 이미지	레이블 1개 이미지	레이블 2개 이미지	레이블 3개 이미지
0개	4장(2.02%)	9장(1.23%)	1장(2.85%)
1개	194장(97.97%)	36장(4.95%)	3장(8.57%)
2개		682장(93.81%)	3장(8.57%)
3개			28개(80.00%)
총 이미지 수	198장	727장	35장

[표 5-2] 이미지의 레이블 개수별 분류 현황

5.3. 데이터 오버 샘플링 실험 및 실험 결과

이 절에서는 데이터 불균형 문제를 극복하기 위해 진행된 추가 실험에 대해 서술하고 그 결과를 제시한다. 5.2절의 실험에서는 데이터 증대를 1:1 비율로 적용하여 골절 분류 모델의 학습에 사용했다. 이 실험에서는 오버 샘플링을 통해 가장 많은 데이터를 갖는 레이블 외의 데이터들을 추가적으로 샘플링해서 각 범주별 데이터 분포의 비율을 맞춘다. 즉, 각 범주별로 기존 데이터 수에 역비례하게 데이터를 오버샘플링 한다. [그림 5-3]은 오버샘플링을 통해 데이터 분포의 비율을 맞추는 예시를 보여준다. 하지만 이 실험에 사용된 데이터는 다중 레이블 데이터이다. 따라서 단일 레이블 데이터에 적용되는 방법과는 다른 방법을 적용한다. 이를 위해 각 레이블의 불균형 비율을 계산하고, 평균 불균형 비율보다 불균형 비율이 큰 레이블들을 대상으로 오버샘플링을 적용한다[Charte15]. 이 때 가장 많은 데이터를 갖는 레이블의 $IRLbl$ 은 1.0이고, 데이터가 적은 레이블일수록 $IRLbl$ 의 값이 커진다. 이 실험에 사용된 데



[그림 5-2] 오버 샘플링

각 레이블의 불균형 비율
(*IRLbl*, *Imbalance Ratio per Label*)

$$IRLbl(y) = \frac{\operatorname{argmax}_{y' \in Y_1} \left\{ \sum_{i=1}^{|D|} h(y', Y_i) \right\}}{\sum_{i=1}^{|D|} h(y, Y_i)}, \quad h(y, Y_i) = \begin{cases} 1, & y \in Y_i \\ 0, & y \notin Y_i \end{cases} \quad (14)$$

$$\begin{aligned} \text{평균 불균형 비율} \\ (MeanIR) \end{aligned} = \frac{1}{|Y|} \sum_{y \in Y_1} (IRLbl(y)) \quad (15)$$

이터를 예로 들면 $IRLbl(32A3)$ 은 15.29, $IRLbl(42B3)$ 은 4.01이다. [그림 5-3]은 이 실험에서 다중 레이블 데이터에 오버샘플링을 적용하기 위해 제안하는 방법을 보여준다. *OverSampling()*의 입력으로는 데이터셋 D , 오버샘플링 횟수 N , 샘플링 시도 횟수 T 가 주어지고, 출력은 오버샘플링된 데이터셋 D' 이다. 첫 번째로 각 레이블의 불균형비율($IRLbl$)과, 평균 불균형 비율($MeanIR$)을 계산한다. 두 번째 단계에서는 오버샘플링을 적용할 이미지를 $minBag$ 에 저장한다. 이 때 저장되는 이미지는 평균 불균형비율보다 불균형 비율이 큰 레이블을 갖는 이미지들이다. 이 후 기존에 충분한 데이터를 갖는 레이블은 *ExclusionLabel*에 저장한다. 이는 해당 레이블을 오버샘플링에서 최대한 배제하기 위함이다. 세 번째 단계에서는 각 $minBag_i$ 에 저장된 이미지 중 무작위 이미지를 선택한다. 이 때 선택된 이미지가 *ExclusionLabel*에 포함된 레이블을 갖고 있다면, 해당 레이블이 없는 이미지가 선택되도록 무작위 선택을 충분한 횟수만큼 다시 진행한다. T 횟수의 시도동안 선택된 이미지가 *ExclusionLabel*에 포함된 레이블을 갖고 있다면 해당 레이블이

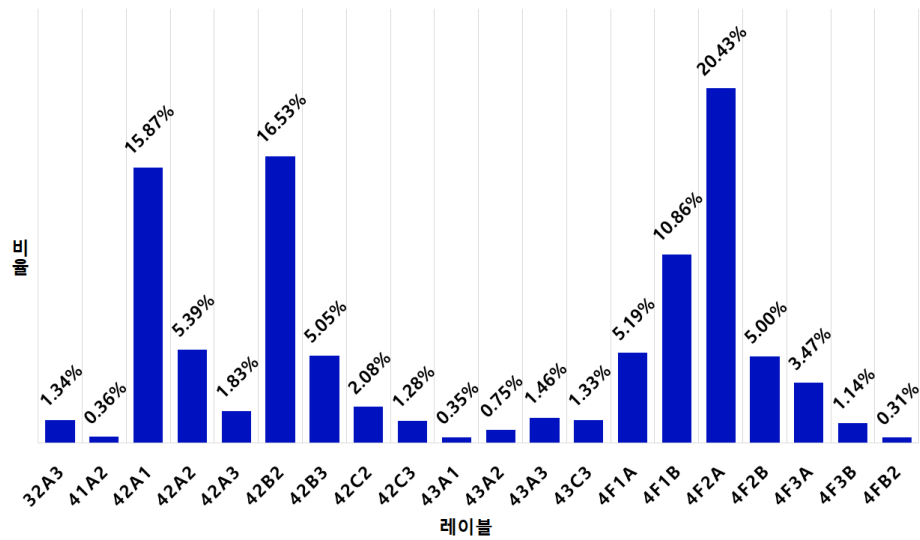
Procedure <i>OverSampling</i> (D, N, T)
<p>입력 : 데이터셋 D, 오버샘플링 횟수 N, 샘플링 시도 횟수 T</p> <p>출력 : 오버샘플링 된 데이터셋 D'</p>
<pre> begin 1. $Image, Label \leftarrow D$에 존재하는 이미지와 레이블 식별 2. for(N times) 3. $IRLbl \leftarrow$ 각 label의 $IRLbl$ 계산 4. $MeanIR \leftarrow MeanIR$ 계산 5. foreach ($label \in Label$) 6. if $IRLbl_{label} > MeanIR$ 7. $minBag_{i++} \leftarrow Bag_{label}$ 8. $ExclusionLabel \leftarrow label_{min_top3(IRLbl)}$ 9. foreach ($minBag_i \in minBag$) 10. $x \leftarrow random(1, minBag_i)$ 11. if ($ExclusionLabel \not\subset label_x$ or $try > T$) 12. $d' += Image_x, label_x$ 13. else 14. $try++$ and go to 9 15. $D' = D + d'$ 16. return D' end </pre>

[그림 5-3] *OverSampling*()의 의사 코드

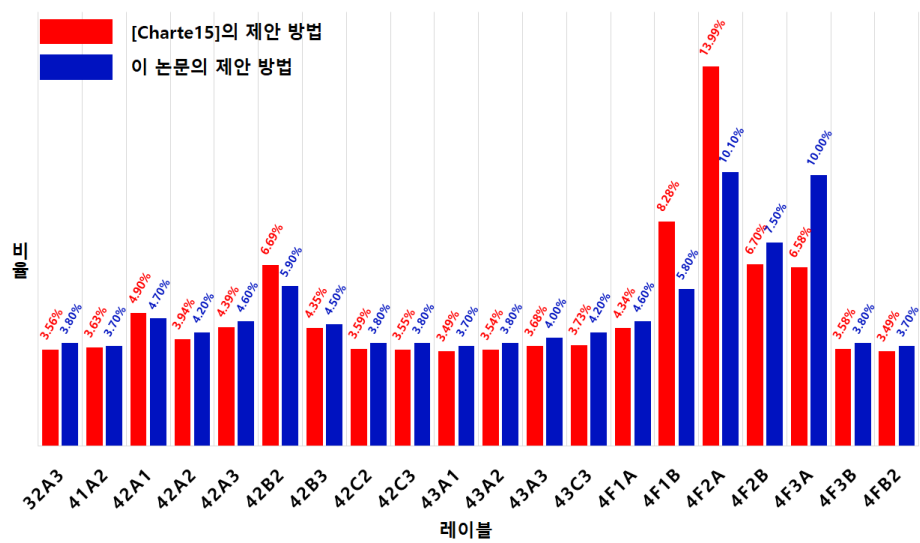
포함되지 않은 이미지가 없는 것으로 판단하고 선택된 이미지를 d' 에 추가하여 오버샘플링을 적용한다. 이 오버샘플링 방법은 학습 데이터셋과 테스트 데이터셋 각각에 적용되며, 그 외의 사항은 5.2절의 실험 방법과 동일하다.

5.2절의 실험에서 사용된 학습데이터의 분포 현황은 [그림 5-4]의 (a)와 같고, 오버샘플링을 적용한 학습데이터의 분포 현황은 [그림 5-4]의 (b)와 같다. (b)의 붉은색 그래프는 [Chart15]에서 제안된 오버샘플링 방법을 적용한 결과이고, 파란색 그래프는 이 논문에서 제안한 방법을 적용한 결과이다. (a)와 (b)의 차이를 통해 오버 샘플링 후 데이터 불균형이 완화되었음을 확인할 수 있고, (b)를 통해 [Chart15]에서 제안된 방법보다 이 논문에서 제안한 방법이 데이터 불균형 완화에 더 효과적임을 확인할 수 있다.

각 레이블별 상세한 현황은 [표 5-3]을 통해 확인할 수 있다. [표 5-3]은 교차검증을 위한 5번의 반복실험 동안 한 에폭에 사용된 학습 데이터의 수를 누적한 것이다. 데이터를 오버샘플링 한 결과 파란색으로 표기된 15개 레이블에 대해 데이터 불균형이 완화되었음을 보여준다. 특히 데이터가 가장 많았던 4개 레이블 중 3개 레이블(42A1, 42B2, 4F2A)은 데이터 비율이 평균적으로 약 10.8% 감소했다. 노란색으로 표기된 3개의 레이블은 불균형도가 다소 악화되었으며, 붉은색으로 표기된 2개의 레이블은 크게 악화되었다. 결과적으로 전체 레이블의 비율을 고려했을 때, 기존 데이터셋의 표준편차는 약 6%, 분산은 약 0.3%였지만 오버샘플링 후에는 표준편차는 약 1.9%, 분산은 약 0.03%로 감소했다. 이를 통해 전체 데이터셋의 불균형이 완화되었음을 확인할 수 있다.



(a)



(b)

[그림 5-4] 기존 학습데이터와 오버샘플링 된 학습데이터의
각 레이블별 데이터 분포 현황 (a) 기존 데이터, (b) 오버샘플링 된 데이터

레이블	기존 실험		오버샘플링 실험		비율 변화
	이미지 수	비율	이미지 수	비율	
32A3	151	1.34%	3,597	3.79%	+2.45%
41A2	40	0.36%	3,538	3.73%	+3.37%
42A1	1,782	15.87%	4,308	4.54%	-11.33%
42A2	605	5.39%	4,031	4.25%	-1.14%
42A3	205	1.83%	4,346	4.58%	+2.75%
42B2	1,856	16.53%	5,408	5.70%	-10.83%
42B3	567	5.05%	4,297	4.53%	-0.52%
42C2	234	2.08%	3,627	3.82%	+1.74%
42C3	144	1.28%	3,584	3.78%	+2.50%
43A1	39	0.35%	3,537	3.73%	+3.38%
43A2	84	0.75%	3,582	3.78%	+3.03%
43A3	164	1.46%	3,937	4.15%	+2.69%
43C3	149	1.33%	3,897	4.11%	+2.78%
4F1A	583	5.19%	4,336	4.57%	+0.62%
4F1B	1,220	10.86%	5,449	5.74%	-5.12%
4F2A	2,294	20.43%	9,533	10.05%	-10.38%
4F2B	561	5.00%	7,215	7.60%	+2.60%
4F3A	390	3.47%	9,506	10.02%	+6.55%
4F3B	128	1.14%	3,626	3.82%	+2.68%
4FB2	35	0.31%	3,526	3.72%	+3.41%
총합	11,231	100%	94,880	100%	

[표 5-3] 교차 실험동안 한 예폭에 사용된 학습 데이터 누적 현황

데이터 오버샘플링을 활용한 실험 결과는 [표 5-4]와 같다. 오버샘플링 실험은 5.2절의 실험보다 정밀도는 1.08%, 재현율은 1.11%, 정확도는 2.08% 높고, 해밍 로스는 0.0027 낮은 결과를 보여주며, [Charte15]에서 제안된 오버샘플링 방법을 적용한 실험과 비교하면 정밀도는 0.31%, 재현율은 0.85%, 정확도는 1.03% 높고, 해밍 로스는 0.0007 낮은 결과를 보여준다.

	정밀도	재현율	정확도	해밍 로스
5.2절 실험	97.62%	95.75%	93.59%	0.0057
오버 샘플링 실험 ([Charte15]방법)	98.39%	96.01%	94.64%	0.0037
오버 샘플링 실험 (제안한 방법)	98.70%	96.86%	95.67%	0.0030

[표 5-4] 오버 샘플링 실험 결과

V. 결 론

본 논문에서는 의사의 업무부하로 인한 의료사고를 방지하고자 골절을 분류하기 위해 딥러닝을 사용하는 방법을 제안했다. 제안하는 방법의 목표는 의사의 골절 진단을 보조하고, 더 나아가 골절 범주 분류 과정을 자동화하는 것이다. 제안하는 방법은 다발성 골절 분류 모델에 다발성 골절 CT 이미지를 입력해 각 이미지를 동시에 여러 개의 레이블로 분류한다.

실험에 사용된 데이터는 하반신 골절 CT 이미지다. 실험에 사용된 환자 수는 149명이며 총 1,453장의 이미지가 사용됐다. 데이터의 레이블은 실제로 정형외과에서 골절을 분류하기 위해 사용 중인 AO/OTA 골절 및 탈구 분류 기준에 따라 작성되었다. 실험에 사용된 CT 이미지는 여러 개의 골절(레이블)이 동시에 등장하는 다발성 골절 데이터이다. 따라서 본 논문에서는 다발성 골절 분류 문제를 다중 레이블 분류(Multi-Label Classification) 문제로 정의했다.

다발성 골절 분류 모델은 ResNet의 잔차 모듈을 기반으로 구성되었다. 다발성 골절 분류 모델은 CT 이미지를 입력으로 사용한다. 다발성 골절 분류 모델은 프론트-엔드에서는 골절의 특징을 추출하고, 백-엔드에서는 다층 퍼셉트론과 시그모이드 함수를 통해 각 골절 레이블의 확률을 독립적으로 계산한다. 이 후 출력된 각 골절 레이블의 확률에 임계값을 적용해 최소 1개, 최대 3개의 다중 레이블을 최종으로 출력한다.

이 실험에 사용된 데이터는 데이터 부족 및 불균형 문제를 포함하고 있다. 이 중 데이터 부족 문제를 극복하기 위해 제안하는 방법에서는 데이터 증대를 적용한다. 이 때 데이터 증대는 매 에폭마다 원본 이미지에 적용되며, 이로 인해 다발성 골절 분류 모델은 매 에폭마다 다른 이미지를 사용해 학습된다. 데이터 불균형 문제를 극복하기 위해 이 논문에서는 데이터를 오버샘플링 하는 방법도 제안했다. 데이터 오버샘플링은 각 레이블의 불균형 비율을 기반으로 적용되며, 결과적으로 제안한 오버샘플링 방법은 기존에 제안되었던 오버샘플링 방법보다도 데이터 불균형 완화에 효과적이었다. 제안한 방법을 활용한 실험결과는 정밀도 98.70%, 재현율 96.86% 정확도 95.67%, 해밍로스 0.0030이다. 이는 제안한 방법이 데이터 불균형 문제를 완화하고 더 높은 분류 성능을 얻을 수 있음을 보여준다.

참고 문헌

- [이건명18] 이건명, 인공지능: 튜링 테스트에서 인공지능까지. 생능출판, 2018
- [이상현19] 이상현, 한찬식, 최승명, 이건명, Classification of Multiple Fractures Using Deep Learning, 한국지능시스템학회 논문지 29 4, pp 285-290, 2019
- [통계청18] 통계청, 의료인력 및 병상수 추이, Available : http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=2772 [Accessed: 11/10/2019]
- [AO/OTA18] AO Foundation and Orthopaedic Trauma Association. Journal of Orthopaedic Trauma 32(1), January 2018
- [Bayram16] F. Bayram, and M. Çakıroğlu, Diffract: Diaphyseal femur fracture classifier system, Biocybernetics and Biomedical Engineering 36(1), pp 157-171, 2016
- [Bengio04] Y. Bengio, and Y. Grandvalet., No Unbiased Estimator of the Variance of K-fold Cross-Validation, Journal of Machine Learning Research 5(Sep), pp 1089-1105, 2004
- [Charte15] F. Charte, A. J. Rivera, M. J. Jesus, and F. Herrera, Addressing imbalance in multilabel classification: Measures and random resampling algorithms, Neurocomputing (163), pp 3-16, 2015

- [Chung18] S. W. Chung, S. S. Han, J. W. Lee, K. S. Oh, N. R. Kim, J. P. Yoon, and Y. M. Noh, Automated detection and classification of the proximal humerus fracture by using deep learning algorithm, *Acta orthopaedica* 89(4), pp 468–473, 2018
- [He16] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 770–778, 2016
- [Kazi17] A. Kazi, S. Albarqouni, A. J. Sanchez, S. Kirchhoff, P. Biberthaler, N. Navab, and D. Mateus, Automatic classification of proximal femur fractures based on attention models, In *International Workshop on Machine Learning in Medical Imaging*, pp. 70–78, 2017
- [Kroguel19] J. D. Kroguel, K. V.Cheng, K. M. Hwang, P. Toogood, E. G. Meinberg, E. J. Geiger, and A. Wright, Automatic Hip Fracture Identification and Functional Subclassification with Deep Learning, *arXiv preprint arXiv:1909.06326*, 2019
- [Lecun89] Y. Lecun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, Backpropagation applied to handwritten zip code recognition, *Neural computation* 1.4, pp. 541–551, 1989.
- [Lindesy18] R. Lindsey, A. Daluiski, S. Chopra, A. Lachapelle, M.

- Mozer, S. Sicular, and H. Potter, Deep neural network improves fracture detection by clinicians, Proceedings of the National Academy of Sciences 115(45), pp 11591–11596, 2018
- [Nicolaes19] J. Nicolaes, S. Raeymaeckers, D. Robben, G. Wilms, D. Vandermeulen, C. Libanati, and M. Debois, Detection of vertebral fractures in CT using 3D Convolutional Neural Network,. arXiv preprint arXiv:1911.01816. 2019
- [Pranata19] Y. D. Pranata, K. C. Wang, J. C. Wang, I. Idram, J. Y. Lai, J. W. Liu, and I. H. Hsieh, Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images, Computer methods and programs in biomedicine 171, pp 27–37, 2019
- [Raghavendra18] U. Raghavendra, N. S. Bhat, A. Gudigar, and U. R. Acharya, Automated system for the detection of thoracolumbar fractures using a CNN architecture, Future Generation Computer Systems 85, pp 184–189, 2018
- [Rayan19] J. C. Rayan, N. Reddy, J. H. Kan, W. Zhang, and A. Annapragada, Binomial classification of pediatric elbow fractures using a deep learning multiview approach emulating radiologist decision making, Radiology: Artificial Intelligence 1(1), e180015, 2019

- [Read11] J. Read, B. Pfahringer, , G. Holmes, and E. Frank, Classifier chains for multi-label classification. *Machine learning*, 85(3) : 333, 2011
- [Shalabi06] L. Al Shalabi, Z. Shaaban, and B. Kasasbeh, Data mining: A preprocessing engine, *Journal of Computer Science*, pp. 735-739, 2006
- [Szegedy15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. nguelov, and A. novich, Going deeper with convolutions, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1-9, 2015
- [Thian19] Y. L. Thian, Y. Li, P. Jagmohan, D. Sia, V. E. Y. Chan, and R. T. Tan, Convolutional neural networks for automated fracture detection and localization on wrist radiographs, *Radiology: Artificial Intelligence* 1(1), e180001, 2019
- [Tripathi17] A. M. Tripathi, A. Upadhyay, A. S. Rajput, A. P. Singh, and B. Kumar, Automatic detection of fracture in femur bones using image processing, In *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp 1-5, 2017
- [Tsoumakas07] G. Tsoumakas, and I. Katakis, Multi-label classification: An overview, *International Journal of Data Warehousing and Mining (IJDWM)* 3(3), pp 1-13, 2007
- [Wai15] Y. Wei, W. Xia, M. Lin, J. Huang, B. Ni, J. Dong, and S.

Yan, HCP: A flexible CNN framework for multi-label image classification, IEEE transactions on pattern analysis and machine intelligence 38(9), pp 1901–1907, 2015

- [Wang16] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, Cnn-rnn: A unified framework for multi-label image classification, In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2285–2294, 2016
- [Wang19] Y. Wang, L. Lu, C. T. Cheng, D. Jin, A. P. Harrison, J. Xiao, and S. Miao, Weakly Supervised Universal Fracture Detection in Pelvic X-Rays, In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp 459–467, 2019