

## Efficient Osteoporosis Prediction Using A Pair of Ensemble Models

Se-Heon Choi\*, Dong-Hwan Hwang\*\*, Do-Hyeon Kim\*, So-Hyeon Bak\*\*\*, Yoon Kim\*

\*Student, Dept. of Computer Science and Engineering, Kangwon National University, Chuncheon, Korea

\*\*Researcher, Department of Research and Development, ZIOVISION Co. Ltd, Chuncheon, Korea

\*Student, Dept. of Computer Science and Engineering, Kangwon National University, Chuncheon, Korea

\*\*\*Professor, Dept. of Radiology, Kangwon National University School of Medicine, Chuncheon, Korea

\*Professor, Dept. of Computer Science and Engineering, Kangwon National University, Chuncheon, Korea

### [Abstract]

In this paper, we propose a prediction model for osteopenia and osteoporosis based on a convolutional neural network(CNN) using computed tomography(CT) images. In a single CT image, CNN had a limitation in utilizing important local features for diagnosis. So we propose a compound model which has two identical structures. As an input, two different texture images are used, which are converted from a single normalized CT image. The two networks train different information by using dissimilarity loss function. As a result, our model trains various features in a single CT image which includes important local features, then we ensemble them to improve the accuracy of predicting osteopenia and osteoporosis. In experiment results, our method shows an accuracy of 77.11% and the feature visualize of this model is confirmed by using Grad-CAM.

▶ **Key words:** CT, Osteopenia, Osteoporosis, Dissimilarity loss function, Feature map, CNN

### [요 약]

본 논문에서는 컴퓨터 단층촬영(CT) 이미지를 이용한 합성곱 신경망(CNN)을 기반의 골감소증 및 골다공증 예측 모델을 제안한다. 기존의 CNN은 단일 CT 이미지에서 예측에 중요한 지역정보를 활용하지 못하다는 문제가 있다. 본 논문에서 이를 해결하고자 CT 이미지를 정규화하여 질감 정보가 다른 두 개의 이미지로 변환하고, 해당 이미지를 활용한 한 쌍의 신경망 네트워크를 제안한다. 동일한 구조를 가진 네트워크 각각의 신경망은 질감 정보가 다른 이미지를 입력으로 사용하고 비유사성 손실함수를 통해 다른 정보를 학습한다. 최종적으로 제안 모델은 중요한 지역정보를 포함한 단일 CT 이미지의 다양한 특징 정보를 학습하며, 이를 앙상블하여 골감소증 및 골다공증 예측 정확도를 높인다. 실험 결과를 통해 제안 모델의 정확도 77.11%를 확인할 수 있으며 Grad-CAM을 이용하여 모델이 바라보는 특징을 확인할 수 있다.

▶ **주제어:** 컴퓨터 단층촬영, 골감소증, 골다공증, 비유사성 손실함수, 특징 정보, CNN

• First Author: Se-Heon Choi, Corresponding Author: Yoon Kim

\*Se-Heon Choi (shchoi2220@kangwon.ac.kr), Dept. of Computer Science and Engineering, Kangwon National University

\*\*Dong-Hwan Hwang (donghwan1994@ziovision.co.kr), Department of Research and Development, ZIOVISION Co. Ltd

\*Do-Hyeon Kim (abc3698@kangwon.ac.kr), Dept. of Computer Science and Engineering, Kangwon National University

\*\*\*So-Hyeon Bak (arsngm05@naver.com), Dept. of Radiology, Kangwon National University School of Medicine

\*Yoon Kim (yoozni@kangwon.ac.kr), Dept. of Computer Science and Engineering, Kangwon National University

• Received: 2021. 10. 28, Revised: 2021. 12. 02, Accepted: 2021. 12. 02.

## I. Introduction

골다공증은 뼈의 칼슘 질량이 감소하고 강도가 약해져서 골절의 가능성이 높아지는 골격계 질환이다[1]. 골다공증은 고관절 골절의 주요 원인이 되며, 골절환자의 절반 이상이 진단을 받은 적이 없을 정도로 사전 인지가 어려운 질병이다[2,3]. 고령화 시대로 진행됨에 따라 골다공증 환자가 증가하는 추세이며, 이에 따른 경제적 비용, 부상률, 사망률도 증가할 것으로 예상되고 있다[4]. 골다공증 치료는 초기 단계에서의 치료가 효과적이고 이를 위해서는 조기 진단과 치료를 통한 예방이 중요하다[5].

최근 딥러닝 기술이 발전함에 따라 의료 영상 분야에서도 관련 연구가 진행되고 있다[6, 7]. 이에 골다공증 진단 분야에서도 Convolutional Neural Network(CNN)를 이용한 진단 예측 방법이 제안되어왔다[8, 9, 10]. 하지만 CNN 모델에서 특징을 학습 시에 데이터에 따라 특징을 편향적으로 학습하는 문제가 있다[11]. 이는 단일 CT 영상을 이용한 CNN 진단은 특정 정보에 편향되어 골다공증 진단에 중요한 지역 정보보다 다른 지역의 정보를 더 중요하게 판단하는 문제점이 있다. 본 논문에서는 이러한 CNN의 문제를 극복하고 골다공증의 판단 성능을 높이기 위해 동일한 구조를 가진 두 개의 CNN 모델을 통해 각기 다른 정보를 학습하여 특정 정보에 편향되지 않는 네트워크를 제안한다. 제안 방법에서는 먼저 전처리를 통해 픽셀 도메인이 다른 정보가 다른 두 이미지를 생성한다. 그 다음에 해당 이미지들을 제안하는 한 쌍의 신경망에 학습 데이터로 사용하여 제안 신경망이 서로 다른 특징 정보를 학습하도록 한다. 본 논문에서는 실험을 위해 정상, 골감소증, 골다공증에 대해 각각 훈련 데이터셋 1200장, 테스트셋 300장을 사용하였다.

본 논문은 CT 데이터를 이용한 골감소증 및 골다공증 예측에서 다음과 같은 기여를 한다.

- 단일 알고리즘을 이용해 정상, 골감소증, 골다공증을 예측한다.
- CT 이미지를 이용한 예측에서 비유사성 손실함수를 활용하여 각기 다른 특징을 학습하는 신경망을 한 쌍으로 결합하고, 서로 다른 특징을 학습하도록 하여 예측 성능이 개선된 네트워크를 제안한다.

본 논문의 구성은 다음과 같다. 2장에서 골다공증 예측에 관한 이전연구를 소개한다. 3장에서 CT 이미지의 전처리 및 제안하는 네트워크 구조와 손실 함수를 소개한다. 4장에서는 다양한 조건에서의 실험을 통하여 제안 모델의 성능을 입증하고 마지막 5장에서 결론을 도출한다.

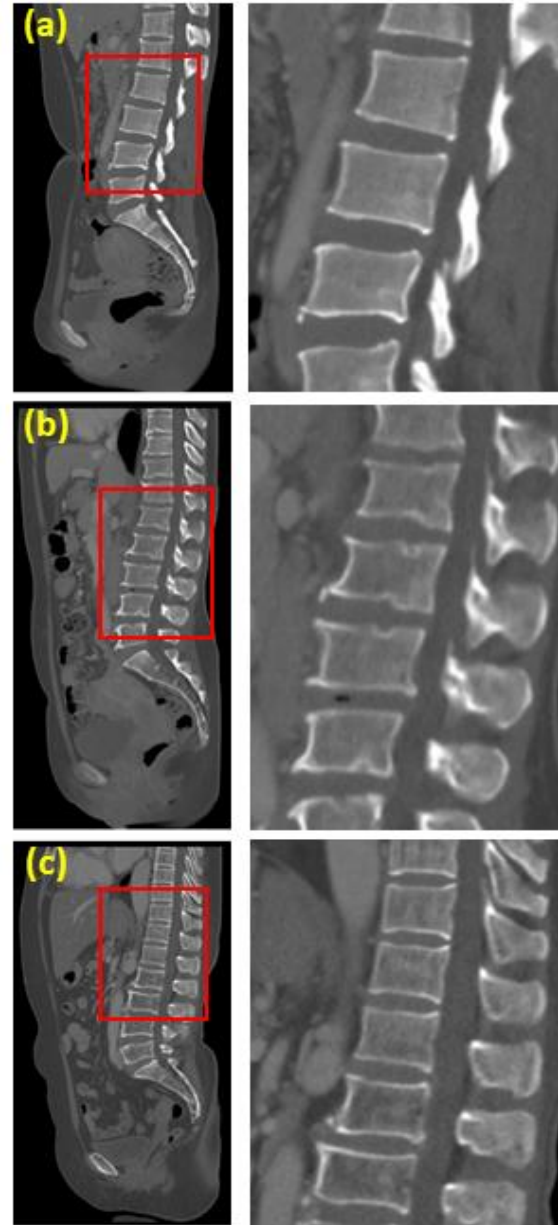


Fig. 1. CT data - (a) Normal, (b) Osteopenia, (c) Osteoporosis

## II. Related Works

Fig 1은 각각 정상, 골감소증, 골다공증 환자의 CT 이미지를 보여준다. 정상 경우에는 직사각형의 형태와 균일한 질감 정보를 가지고 있으며, 골감소증과 골다공증 경우에는 HU(Hounsfield Unit)의 감소로 인한 질감 정보에 차이가 있다. 따라서 골감소증 및 골다공증의 판단은 CT 이미지에서 HU의 감소에 따른 질감 차이로 구분한다.

## 1. Osteoporosis Prediction with Deep Learning

최근 딥러닝의 발전으로 의료분야에서도 딥러닝을 이용한 많은 시도가 진행되고 있다. X선, CT 데이터를 활용한 딥러닝 모델의 성능이 발전됨에 따라 골다공증 예측 또한 여러 연구가 시도되고 있다. Yasaka 등[8]은 CT 이미지를 간단한 CNN 모델을 이용해 골밀도를 예측하는 모델을 제안하였으며, 해당 논문을 통해 CNN을 이용한 골밀도 예측에 대한 가능성을 보여주었다. Pan 등[9]은 저선량 흉부전신단층촬영 이미지를 통해 골밀도를 자동으로 측정하는 딥러닝 기반 시스템을 제안하였다. 저선량 흉부전신단층촬영 이미지를 샘플링 한 3D 패치를 입력으로 사용하여 같은 크기의 세분화 마스크 4종을 출력하는 3D U-net 기반의 모델을 활용하였다. Yamamoto 등[10]은 고관절 방사선 이미지를 이용해 골다공증을 예측하기 위해 CNN을 사용하였으며, 추가적인 임상 정보를 사용하여 예측 진단의 정확성을 조사하였다. 또한, ResNet, GoogleNet 등의 CNN 모델을 이용하여 골다공증 예측 성능을 비교평가하였고 임상 정보를 앙상블 하여 성능을 개선하였다. 본 논문에서는 이전연구에서 사용된 CNN의 문제를 개선하기 위해 질감 정보가 다른 CT 이미지를 이용하여 특정 정보에 편향되지 않고 다양한 특징을 학습하는 모델을 제안한다. 제안 모델의 성능을 검증하기 위해서 안정성과 성능이 우수한 CNN 모델인 ResNet, GoogleNet 등을 사용하여 비교를 진행한다.

## 2. Biased towards specific feature in CNNs

사람은 시각적 판단에 있어 형태와 질감 등을 복합적으로 판단하여 결정한다. 하지만 CNN 모델도 사물이 가지는 형태와 질감의 특징을 학습하고 판단하지만 특정 정보에 편향되는 현상을 보인다. R Geirhos 등[11]은 이미지넷으로 사전학습된 CNN이 코끼리 질감을 가진 고양이 이미지가 코끼리로 예측된다는 결과를 통해 CNN이 질감 편향적으로 학습한다는 것을 증명하였다. 이는 이미지넷 데이터셋으로 학습된 CNN은 질감에 정보가 편향된다는 것을 의미하며, 학습 이미지의 질감 정보를 지속적으로 변형하면서 학습할 경우에는 형태의 편향을 증가되어 정확성이 향상될 수 있음을 발표하였다. MA Islam 등[12]의 논문에서는 네트워크가 학습 초기 대부분의 형태 정보를 학습하고 이 정보가 CNN의 마지막 몇몇 계층에 인코딩되어 있음을 실험을 통해 증명하였다.

본 논문에서 학습에 사용된 CT 이미지는 골격과 장기 정보를 포함하고 있으며, 해당 정보의 편차가 적어 효과적인 특징 학습이 어렵다. 이를 극복하고자 본 논문에서는

입력 데이터로 사용되는 CT 이미지를 정규화하여 질감 정보가 다른 두 개의 이미지로 변환하고, 두 이미지를 각각 다른 신경망을 이용해 학습하는 모델을 제안한다. 제안 모델은 동일한 구조를 가지는 한 쌍의 신경망으로 구성하고 비유사성 손실함수를 이용하여 각각의 신경망이 서로 다른 특징을 학습하도록 한다. 이를 통해 입력 이미지의 질감 차이를 바탕으로 각각의 신경망이 다른 특징을 학습하며, 비유사 손실함수를 통해 학습하는 특징의 다양성을 확보한다. 이를 통해서 두 개의 신경망이 특정 정보의 편향되는 것이 아닌 골다공증 예측을 위한 정확한 특징 정보를 학습하도록 유도한다. 실험 결과를 통해 모델의 성능 결과를 확인하고, Grad-CAM을 통해 시각화하여 제안 모델이 바라보는 특징을 확인하였다.

## III. The Proposed Scheme

### 1. Pre-processing

본 연구에서 CT 영상을 입력으로 사용하고 CT 영상의 한 픽셀은 Hounsfield Unit(HU)으로 이루어져 있으며 범위는  $[-1024, 3071]$ 이다. 일반적인 CNN의 입력으로 사용되는 이미지의 범위는  $[0, 255]$ 로 CNN에서 CT 영상을 사용하기 위해서는 HU를 일반 이미지의 픽셀 범위로 변환이 필요하다. 본 논문에서는 CT 영상에서 HU의 특정 범위를 지정하고 이를 정규화하여 회색조 이미지의 픽셀 범위  $p \in \mathbb{R}^{H \times W}$ 로 변환한다. CT 영상의 픽셀  $p_{hw}$ 에 대한 계산식은 아래와 같다.

$$p_{hw} = 255 \times \frac{HU_{hw} - HU_{\min}}{HU_{\max} - HU_{\min}}, \quad (1)$$

$$h \in H, w \in W,$$

$HU_{\min}$ 과  $HU_{\max}$ 은 각각 지정한 HU 범위에서의 최소값과 최대값을 의미한다. 본 논문에서는  $HU_{\max} = 1024$ 로 고정하고  $HU_{\min} \in \{-450, -128\}$ 을 통해 질감 정보가 다른 두 이미지를 생성하기 위해서  $x^s(HU_{\min} = -128)$ ,  $x^t(HU_{\min} = -450)$ 를 사용한다.

### 2. Proposed Neural Network

본 논문은 질감 정보가 서로 다른 두 이미지  $x^s$ ,  $x^t$ 를 각각 학습하는 한 쌍의 CNN 신경망( $G_s$ ,  $G_t$ )과 신경망들의 특징 정보를 적응적으로 결합하여 정상, 골감소증, 골다공증을 분류하는 완전 결합 층  $F$ 로 구성된 신경망으로 구성된다. 제안하는 신경망의 구조는 Figure 2와 같다. 한

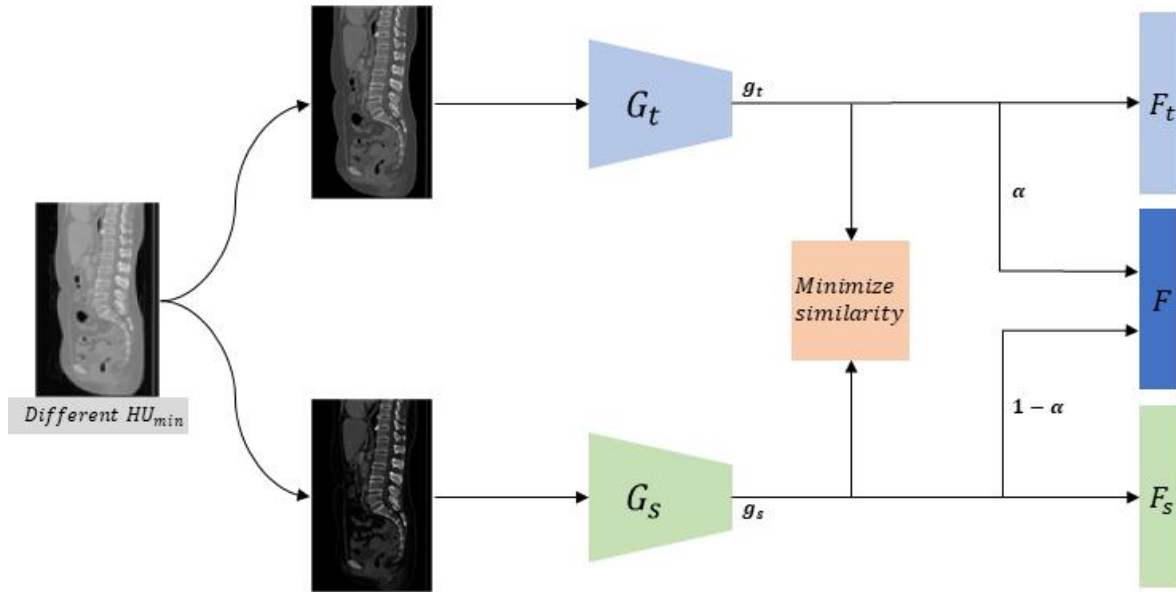


Fig. 2. Network Architecture

쌍의 신경망( $G_s$ ,  $G_t$ )은 ResNet18[13]+ACM[14]으로 각각 완전 결합 층 갖고 있으며  $F$ 에서는 정상, 골감소증, 골다공증을 분류한다. 이때  $F$ 는 CT 이미지의 각 클래스 확률을 가진다.

제안 신경망에서 완전 결합 층  $F$ 는  $G_s$ 과  $G_t$ 의 대표 특징 정보  $g_s$ ,  $g_t \in \mathbb{R}^{C \times 1 \times 1}$ 을 적음적으로 결합한 정보  $g \in \mathbb{R}^{C \times 1 \times 1}$ 을 입력으로 사용한다.  $g_s$ 와  $g_t$ 는 각각  $G_s$ 와  $G_t$ 의 완전 결합 층의 입력이 되는 대표 특징 정보로 채널별 평균 정보를 갖는 벡터이다.  $F$ 는  $g_s$ 와  $g_t$ 를 가중치 합 연산하여 결합한 특징 정보  $g$ 를 입력으로 사용한다. 특징 정보  $g_s$ 와  $g_t$ 의 결합 과정에는  $\alpha$ 라는 가중치를 이용하여  $g_s$ 와  $g_t$ 의 중요도를 조절한다. 중요도 조절은  $g_s$ 에  $\alpha$ 를 곱하고  $g_t$ 에  $1 - \alpha$ 를 곱하여 서로 더하는 연산으로 이루어진다. 이때,  $\alpha$ 는 학습 가능한 파라미터로 학습을 통해 최적의 값을 학습한다.  $\alpha$ 를 이용한 가중합 연산방법에 대한 실험은 4.6에서 자세히 다룬다.  $g$ 의 계산식은 아래와 같다.

$$g = \alpha g_s + (1 - \alpha)g_t. \quad (2)$$

### 3. Loss function

배치 단위  $B$ 에서 입력 데이터  $\{x_i^s\}_{i=1}^B$ 와  $\{x_i^t\}_{i=1}^B$ 가 있고 이들에 대한 공통된 정답 값  $\{y_i\}_{i=1}^B$ 를 가진다. 본 논문에서는 분류 작업을 위한 손실함수와  $G_s$ 와  $G_t$ 가 서로 다른 특징을 학습하기 위해서 비유사성 손실함수를 통합한 손실함수를 사용한다. 제안 신경망에서  $G_s$ 와  $G_t$ ,  $F$ 의 분류 작업 손실함수 수식은 다음과 같다.

$$\begin{aligned} L_{cls} &= L_{cls}^s + L_{cls}^t + L_{cls}^f, \\ L_{cls}^s &= \frac{1}{B} \sum_{i=1}^B (L_{ce}(G_s(x_i^s), y_i)), \\ L_{cls}^t &= \frac{1}{B} \sum_{i=1}^B (L_{ce}(G_t(x_i^t), y_i)), \\ L_{cls}^f &= \frac{1}{B} \sum_{i=1}^B (L_{ce}(F(g), y_i)), \end{aligned} \quad (3)$$

식 (3)의  $L_{ce}$ 는 교차 엔트로피 손실함수이다. 분류 작업 손실함수는 그 값을 최소화하여 신경망들의 분류 작업을 진행한다.

두 신경망이 서로 다른 관점에서 분류 작업을 수행하기 위해  $G_s$ 와  $G_t$ 는 질감 정보가 다른 입력  $x^s$ 와  $x^t$ 을 통해 학습한다. 본 논문에서는 두 신경망이 다른 정보를 학습하기 위해 비유사성 손실함수  $L_{dis}$ 를 제안한다.  $L_{dis}$ 은 두 신경망의 특징 정보의 유사성이 적도록 학습하여 두 신경망의 서로 다른 특징 정보를 추출하도록 한다. 비유사성 손실함수  $L_{dis}$ 에 대한 계산식은 다음과 같다.

$$L_{dis} = \frac{g_s^T g_t}{\|g_s\|_2 \|g_t\|_2}, \quad (4)$$

비유사성 손실함수  $L_{dis}$ 는  $g_s$ 와  $g_t$ 간의 코사인 유사도를 이용하여 계산한다. 학습 과정에서는 비유사성 손실함수를 최소화하여  $G_s$ 와  $G_t$ 의 파라미터들을 최적화한다.  $L_{dis}$ 를 통해서 두 신경망이 분류 작업 손실함수  $L_{cls}$ 의 최적화 과정에서 비슷한 특징 정보를 학습하지 않도록 조절한다.

제안 신경망은 분류 작업 손실함수  $L_{cls}$ 와 비 유사성 손

실함수  $L_{dis}$ 을 합한 손실함수를 학습하며 계산식은 식 (5)와 같다.

$$L = L_{cls} + \lambda L_{dis}, \quad (5)$$

식 (5)에서  $\lambda$ 는 비유사성 손실함수  $L_{dis}$ 의 영향을 조절하는 매개변수이다. 비유사성 손실함수의 유무에 따른 실험은 4.3장에서 자세히 다룬다.

## IV. Experiment

### 1. Implementation details

본 논문은 실험을 위한 이미지는 정상, 골감소증, 골다공증 각각 1500장으로 이루어지며 총 4500장의 CT 영상을 사용한다. 실험을 위해 80%는 학습으로 사용하고 나머지 20%는 검증 데이터로 사용하였다. 제안 신경망의  $G_s$ 와  $G_t$ 는 실험적으로 성능이 좋은 ResNet18+ACM을 사용했다. 학습을 위한 하이퍼파라미터로는 배치 크기 128, 학습률 0.001,  $\lambda$ 는 0.1을 사용하였고  $\alpha$ 는 0.5로 초기화하여 학습을 진행하였다. 최적화를 위해 Adam optimizer와 Cosine Annealing 학습률 스케줄러를 사용하였다. 모든 실험은 이미지넷에 대하여 사전 학습된 신경망으로 이루어졌다. ViT[16]를 제외한 모든 신경망은 60 epoch으로 학습을 진행하였으며, ViT는 임베딩의 학습이 오래 걸리다는 특징 때문에 500 epoch 학습하였다. 실험 환경은 아래의 Table 1과 같다. 하드웨어는 CPU Xeon Silver 4114, GPU는 RTX 3090, RAM은 256GB를 사용하였으며, Ubuntu Server 18.04 운영체제에서 파이토치 1.7.1 버전 딥러닝 프레임워크를 사용하였다.

Table 1. Experiment Environment

OS	Ubuntu Server 18.04
CPU	Xeon Silver 4114
GPU	RTX 3090
RAM	256GB
Deep learning framework	Pytorch 1.7.1

### 2. Ablation study

Table 2는 서로 다른 질감 정보를 갖는 이미지  $x^s$ 와  $x^t$ 에 대한 ResNet18+ACM의 성능과 제안 방법에서 비유사성 손실함수를 생략한 방법(Ours- $L_{dis}$ ), 제안 방법과의 성능 비교이다. ResNet18+ACM은  $x^s$ 데이터에서 64%의 성능을 보이며  $x^t$ 에 대해서 68.44%의 성능을 보였다. 비

유사성 손실함수를 생략한 방법의 경우에는 74%의 성능으로 두 ResNet18+ACM 방법보다 좋은 성능을 보이나 비유사성 손실함수를 포함한 제안 방법보다는 3.11% 낮은 성능을 보였다. 이와 같은 실험 결과를 통하여 제안한 비유사성 손실함수가 모델의 성능 향상에 도움이 되는 것을 확인할 수 있다.

Table 2. Ablation study

Method	Accuracy
ResNet18+ACM( $x^s$ )	64.00%
ResNet18+ACM( $x^t$ )	68.44%
Ours- $L_{dis}$	74.00%
<b>Ours</b>	<b>77.11%</b>

### 3. Ensemble perspective

Table 3은  $G_s$ 와  $G_t$ 의 입력 이미지를 동일하게 했을 때와 다르게 했을 때에 대한 성능 비교이다. 비유사성 손실함수의 목적을 고려했을 때, 성능 비교를 위해 비유사성 손실함수는 생략하여 실험하였다. 질감 정보가 상대적으로 적은  $x^s$ 을  $G_s$ 와  $G_t$ 의 입력으로 사용한 경우에는 실험 결과에서 가장 낮은 성능을 보였다. 반대로  $x^s$ 와  $x^t$ 를 모두 사용한 제안 방법이 가장 높은 성능을 보였다. 실험 결과를 통해  $x^s$ 와  $x^t$ 을 입력으로 두 관점에서 학습한 신경망이 효과가 있음을 확인하였다.

Table 3. Ensemble perspective

Method	Input	Accuracy
Single perspective	$(x^s, x^s)$	63.89%
	$(x^t, x^t)$	66.67%
Two perspective	$(x^s, x^t)$	74.00%

### 4. Analysis on $\alpha$ ratio rules

제안 방법은 앙상블을 이용한 방법으로  $G_s$ 와  $G_t$ 의 특징 정보를 적절히 결합하는 것이 중요하다. Table 4는 두 신경망의 특징 정보를 결합하는 연산에서  $\alpha$ 를 특정한 값으로 고정했을 때와 학습 가능한 파라미터(Ours)로 변경한 경우에 대한 성능 비교이다.  $\alpha$ 를 고정적으로 한 경우에는 실험적으로 0.5일 때 가장 높은 성능을 보였다.  $\alpha$ 가 0.1로 설정 했을 때보다 0.22%p 높은 정확도를 보여주고 있다. 실험 결과에서  $\alpha$ 가 0.9인 경우가 가장 낮은 성능을 보였다.  $\alpha$ 를 학습 가능한 파라미터로 설정한 경우에 77.11%의 정확도가 나왔으며 해당 결과는 실험 방법 중 가장 좋은 성능이다.

Table 4. Ensemble method

$\alpha$	Accuracy
0.1	74.00%
0.3	73.00%
0.5	74.22%
0.7	73.56%
0.9	71.22%
<b>Ours</b>	<b>77.11%</b>

## 5. Ensemble method

Fig 3은  $G_s(S)$ 와  $G_t(T)$ 에 대한 클래스별 정확도 성능 비교이다.  $G_s$ 와  $G_t$ 는 골다공증 클래스에서는 비슷한 정확도를 보이나, 정상과 골감소증 클래스에서는 성능차를 보였다.  $G_s$ 는 정상 클래스에서 가장 낮은 정확도를 보였고  $G_t$ 은 정상 클래스에 대하여 가장 높은 정확도를 보였다. 위 실험 결과를 통해  $G_s$ 와  $G_t$ 가 클래스별 정확도 관점에서 서로 다른 특징을 보이는 것을 확인하였다.

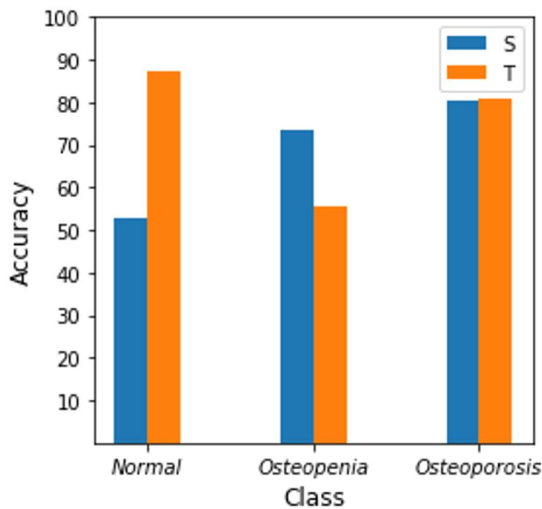


Fig. 3. Class-wise accuracy(%)

## 6. Comparisons with vision algorithm

본 논문에서는 제안 모델의 성능을 비교하기 위해서 일반적으로 성능이 검증된 학습 모델을 선택하여 제안 모델과 비교를 진행하였다. Table 5는 이미지넷 데이터셋에서 우수한 성능을 보인 방법들과 제안 모델의 정상, 골감소증, 골다공증 분류 평균 정확도이다. 성능 비교에 사용된 비교 신경망들의 입력 데이터로  $x^t$ 를 사용했다. 실험 결과를 통해 기존의 모델들 중 EfficientNet-b0 모델이 가장 높은 성능을 보이며, ViT-16×16은 가장 낮은 성능인 것을 확인하였다. 실험 결과는 제안 모델이 가장 좋은 성능

이 77.11%를 보이며, 해당 정확도는 비교 모델에서 가장 좋은 성능을 보인 EfficientNet-b0 보다 5.00%p 높은 정확도이다.

Table 5. Performance with vision algorithm

Network	Accuracy
ResNet18[13]	66.67%
ResNet18+ACM[13,14]	68.44%
Xception[17]	69.22%
Inception-v3[18]	68.77%
EfficientNet-b0[16]	72.11%
ViT-16×16[15]	55.67%
<b>Ours</b>	<b>77.11%</b>

## 7. Visualization with Grad-CAM

Fig 4는 Grad-CAM[19]을 적용하여 검증 데이터셋 CT 이미지에서 해당 클래스에 대하여 네트워크들의 공간적 중요도를 시각화한 것이다. ResNet18+ACM과 비교하였을 때 제안 신경망들은 척추 위치에 더 집중하여 클래스를 예측하는 것을 볼 수 있다. 특히 골다공증 클래스에서 제안 신경망들과 비교 시에 ResNet18+ACM은 척추 이외의 부분을 높은 중요도로 학습한 것을 확인 할 수 있다. 또한, 제안 신경망의 두 신경망들은 이미지에서 보는 지역에 차이가 있는 것을 볼 수 있다. 해당 실험을 통해 제안 방법이 신경망의 특징 개선에 효과적이며 두 신경망은 본 논문의 의도대로 학습되었음을 알 수 있다.

## V. Conclusions

본 논문에서는 골감소증 및 골다공증을 진단하기 위한 딥러닝 기반의 예측 모델을 제안하였다. 제안 모델에서는 CT 이미지를 서로 다른 두 가지 도메인의 영상으로 변환하고, 한 쌍의 신경망을 이용하여 다양한 특징 정보를 앙상블한 모델을 구성하였다. 두 신경망이 서로 다른 특징 정보를 학습하도록 하는 비유사성 손실함수를 제안하였으며 실험을 통해 최적의 모델을 확인하고 한 쌍의 신경망 네트워크가 서로 다른 특징을 학습함에 따라 성능 개선 효과가 있음을 확인하였다. 성능평가에서 제안 모델이 비교 모델들보다 높은 정확도 77.11%의 성능을 보이며, Grad-CAM 결과를 통해 진단을 위한 특징 영역을 효과적으로 학습함을 검증하였다. 추가적으로 향후 연구에서는 CNN 이외에 GCN(Graph Convolutional Network) 등을 이용한 방법 등에 대해서도 연구할 필요성이 있다.



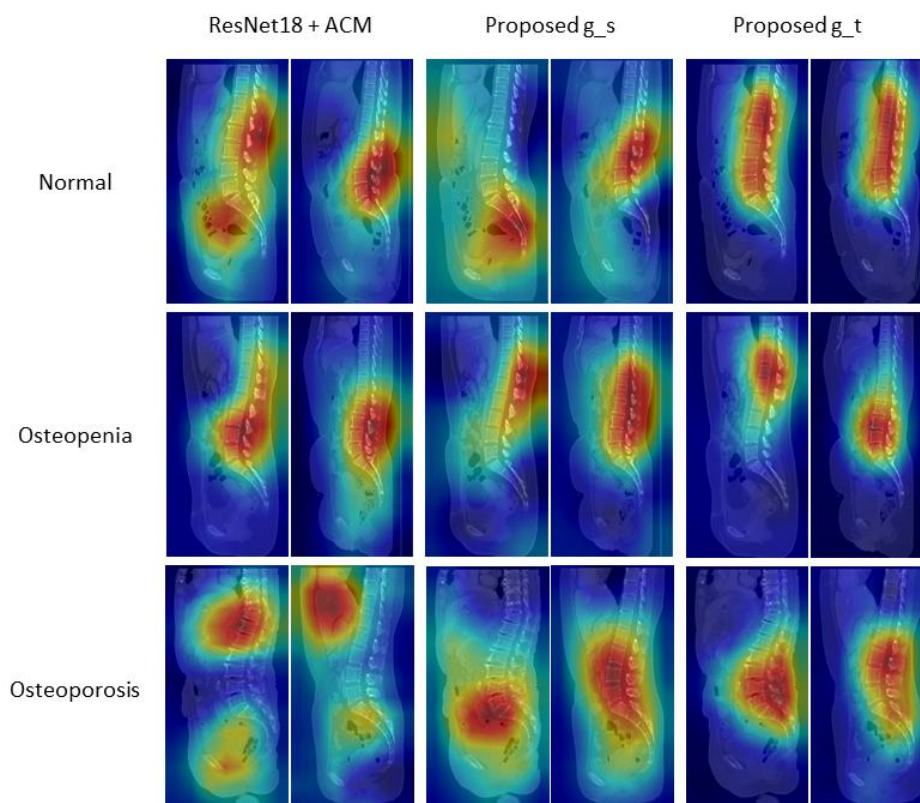


Fig. 4. Experiment result – Grad-CAM

## REFERENCES

- [1] Wani IM, Arora S. Computer-aided diagnosis systems for osteoporosis detection: a comprehensive survey. *Med Biol Eng Comput.* 2020 Sep;58(9):1873-1917. doi: 10.1007/s11517-020-02171-3. Epub 2020 Jun 25. PMID: 32583141.
- [2] Sözen T, Özışık L, Başaran NÇ. An overview and management of osteoporosis. *Eur J Rheumatol.* 2017 Mar;4(1):46-56. doi: 10.5152/eurjrheum.2016.048. Epub 2016 Dec 30. PMID: 28293453; PMCID: PMC5335887.
- [3] Smith AD. Screening of Bone Density at CT: An Overlooked Opportunity. *Radiology.* 2019 May;291(2):368-369. doi: 10.1148/radiol.2019190434. Epub 2019 Mar 26. PMID: 30917293.
- [4] Wu J, Qu Y, Wang K, Chen Y. Healthcare Resource Utilization and Direct Medical Costs for Patients With Osteoporotic Fractures in China. *Value Health Reg Issues.* 2019 May;18:106-111. doi: 10.1016/j.vhri.2018.11.008. Epub 2019 Mar 23. PMID: 30909083.
- [5] Mitchell PJ. Fracture Liaison Services: the UK experience. *Osteoporos Int.* 2011 Aug;22 Suppl 3:487-94. doi: 10.1007/s00198-011-1702-2. Epub 2011 Aug 17. PMID: 21847771.
- [6] Lee, JaeWon, DoHyeon Kim, and Yoon Kim. "Online Video Synopsis via Multiple Object Detection.", *Journal of the Korea Society of Computer and Information*, Vol. 24, No. 8, pp. 19-28, Aug. 2019. DOI: <https://doi.org/10.9708/jksci.2019.24.08.019>
- [7] D.-H. Hwang, G.-S. Moon, and Y. Kim, "SKU-Net: Improved U-Net using Selective Kernel Convolution for Retinal Vessel Segmentation," *Journal of the Korea Society of Computer and Information*, vol. 26, no. 4, pp. 29-37, Apr. 2021.
- [8] Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O. Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network. *Eur Radiol.* 2020 Jun;30(6):3549-3557. doi: 10.1007/s00330-020-06677-0. Epub 2020 Feb 14. PMID: 32060712.
- [9] Pan Y, Shi D, Wang H, Chen T, Cui D, Cheng X, Lu Y. Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening. *Eur Radiol.* 2020 Jul;30(7):4107-4116. doi: 10.1007/s00330-020-06679-y. Epub 2020 Feb 19. PMID: 32072260; PMCID: PMC7305250.
- [10] Ferizi U, Honig S, Chang G. Artificial intelligence, osteoporosis and fragility fractures. *Curr Opin Rheumatol.* 2019 Jul;31(4):368-375. doi: 10.1097/BOR.0000000000000607. PMID: 31045948; PMCID: PMC7282383.
- [11] GEIRHOS, Robert, et al. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*, 2018.
- [12] ISLAM, Md Amirul, et al. Shape or texture: Understanding discriminative features in cnns. *arXiv preprint arXiv:2101.11*

604, 2021.

- [13] HE, Kaiming, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.
- [14] Kim M., Park J., Na S., Park C.M., Yoo D. (2020) Learning Visual Context by Comparison. In: Vedaldi A., Bischof H., Brox T., Frahm JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science, vol 12350. Springer, Cham. [https://doi.org/10.1007/978-3-030-58558-7\\_34](https://doi.org/10.1007/978-3-030-58558-7_34)
- [15] DOSOVITSKIY, Alexey, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
- [16] TAN, Mingxing; LE, Quoc. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. PMLR, 2019. p. 6105-6114.
- [17] CHOLLET, François. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 1251-1258.
- [18] SZEGEDY, Christian, et al. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 2818-2826.
- [19] SELVARAJU, Ramprasaath R., et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 618-626.

## Authors



Se-Heon Choi received the B.S. degrees in Computer Science and Engineering from Kangwon National University, Korea, in 2018. He is currently a M.S. Student in the department of Computer Science and

Engineering at Kangwon National University. His research interests are in the areas of computer vision, machine learning and medical image processing .



Dong-Hwan Hwang received the B.S. degrees and an M.S. degree in Computer Science and Engineering from Kangwon National University, Korea, in 2019 and 2021. He is currently researcher in ziovision.

His research interests are in the areas of machine learning and computer vision.



Do-Hyeon Kim received a B.S. degree and an M.S. degree in 2016. He is now a Ph.D. candidate in Department of Computer and Communications Engineering at Kangwon National University.

His research interests are in the areas of machine learning, and computer vision



So-Hyeon Bak received B.S., M.S. and Ph.D. degrees in Biological Science and Medicine from Konkuk University in 2004, 2009 and 2017, respectively. Dr. Bak is currently assistant professor in Department of

Radiology, Kangwon National University school of medicine. She is interested in computed tomography, magnetic resonance imaging, quantitative analysis, and artificial intelligence.



Yoon Kim received a B.S. degree in 1993, an M.S. degree in 1995, and a Ph.D. degree in 2003, in electronic engineering with the Department of Electronic Engineering from Korea University In 2004, he joined the

Department of Computer Science and Engineering, Kangwon National University, where he is currently a professor. From 1995 to 1999, he was with the LG-Philips LCD Co., where he was involved in research and development on digital image equipment. His research interests are in the areas of machine learning, multimedia communications, and computer vision.