Journal of Korea Multimedia Society Vol. 25, No. 8, August 2022(pp. 1212-1223)
https://doi.org/10.9717/kmms.2022.25.8.1212

다중 스케일 어텐션과 심층 앙상블 기반 동물 피부 병변 분류 기법

곽민호†, 김경태†, 최재영†

Multi-scale Attention and Deep Ensemble-Based Animal Skin Lesions Classification

Min Ho Kwak†, Kyeong Tae Kim†, Jae Young Choi†

ABSTRACT

Skin lesions are common diseases that range from skin rashes to skin cancer, which can lead to death. Note that early diagnosis of skin diseases can be important because early diagnosis of skin diseases considerably can reduce the course of treatment and the harmful effect of the disease. Recently, the development of computer-aided diagnosis (CAD) systems based on artificial intelligence has been actively made for the early diagnosis of skin diseases. In a typical CAD system, the accurate classification of skin lesion types is of great importance for improving the diagnosis performance. Motivated by this, we propose a novel deep ensemble classification with multi-scale attention networks. The proposed deep ensemble networks are jointly trained using a single loss function in an end-to-end manner. In addition, the proposed deep ensemble network is equipped with a multi-scale attention mechanism and segmentation information of the original skin input image, which improves the classification performance. To demonstrate our method, the publicly available human skin disease dataset (HAM 10000) and the private animal skin lesion dataset were used for the evaluation. Experiment results showed that the proposed methods can achieve 97.8% and 81% accuracy on each HAM10000 and animal skin lesion dataset. This research work would be useful for developing a more reliable CAD system which helps doctors early diagnose skin diseases.

Key words: Skin Lesion Diagnosis, Multi-Scale Attention Networks, Deep Ensemble Classification, Joint Learning

1. 서 론

피부 질환은 흔한 질병 중에 하나로써 피부 발진부터 사망에 이를 수도 있는 피부암까지 종류는 다양하다. 매일 대한민국에서는 1,458만 명의 사람들이 피부 질환으로 진료받으며, 매년 10명당 2.8명의 사람들이 간단한 발진부터 치명적일 수 있는 피부암으로 치료받고 있다. 피부 질환은 대부분이 정상 피부와 비교하여 시각적으로 두드러지는 특성이 있으며 이러한 질환 특성을 통해 내리는 빠른 진단은 환자가 받는 고통을 줄이거나 질환으로 인한 후유증을 줄이는 데 큰 역할을 할 수 있다. 이에 따라 피부 질환의

※ Corresponding Author : Jae Young Choi, Address: Engineering Dep. 423, 81, Oedae-ro, Mohyeon-eup, Cheoin-gu, Yongin-si, Gyeonggi-do, 17035, Korea, TEL : +82-10-4304-7422, E-mail : jychoi@hufs.ac.kr
Receipt date : Jul. 28, 2022, Approval date : Aug. 3, 2022
† Pattern Recognition and Machine Intelligence (PMI) Lab., Division of Computer Eng., Hankuk University of Foreign Studies
(E-mail : minos@hufs.ac.kr)
(E-mail : rudxo9251@gmail.com)
※ This research was supported by Hankuk University of Foreign Studies Research Fund.
※ This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2021R1A2C1092322).

조기 진단은 중요한 역할을 할 수 있다.

실제로 일부 피부 질환은 환자에게 치명적일 수 있는 피부암과 외형적 특성이 유사하다. 이러한 특성은 의사들이 진행하는 진단 이후에도, 피부 질환의 정확한 진단을 위해 조직 검사와 같은 추가적인 분석 과정이 있어야 한다. 이러한 과정은 의사의 인력뿐만 아니라 전문 분석 장비와 같은 특수한 장비를 요구로 하는 과정이기 때문에 필연적으로 분석에 많은 시간을 소모하게 된다. 이에 따라 한 번에 많은 분석 요구를 수용하고 인력의 지나친 소모를 막기 위해 피부확대경검사로부터 얻은 영상을 활용하는 컴퓨터 지원 진단 시스템 개발이 활발하게 이루어지고 있다.

현재 개발 중인 컴퓨터 지원 진단 시스템 내부에 활용되는 인공지능 알고리즘 중 가장 많이 활용되는 것은 심층 합성곱 신경망(Deep Convolutional Neural Networks, DCNN) 기술이다. 심층 합성곱 신경망 기반 모델을 활용하여 피부 병변 진단을 위한 분류 기술 개발 뿐만 아니라 영역 분할 같은 연구가 활발하게 수행되고 있다. 그러나 이러한 발전에도 불구하고 현재 딥러닝 모델을 활용하여 피부 질환 영상을 높은 정확도로 분류하는 것은 쉽지 않은 상황이며 다음과 같은 문제점을 해결해야 한다. 첫째 딥러닝 모델 학습을 위해서는 표기된(annotated) 방대한 양의 데이터셋이 필요한데 현재 표기된 피부 질환 영상들의 수가 부족하며 또한 질환별 부류(per class)에 따른 데이터 개수의 불균형이 매우 심각하여 딥러닝 모델을 최적으로 훈련하는 것이 쉽지 않다. 둘째, 피부확대경검사 특성상 환자 피부에 존재하는 털과 피부색, 혈관 및 밝기 또는 대비의 차이와 같은 외부적 요인은 정확한 분류 성능을 얻는 데 장애 요소가 된다. 셋째, 같은 피부 질환 분류에 속하는 영상들 사이에 컬러, 모양 및 텍스처 특징의 가변성이 매우 높다.

본 논문에서는 피부 질환 분류 성능을 향상하기 위해 피부 질환 관심 영역 분할(Region-of-interest segmentation) 정보 활용, 다중 스케일 어텐션(Multi-scale Attention) 기법 및 심층 앙상블 분류(Deep ensemble classification) 기법을 개발하였다. 본 논문에서 제시하는 기술들의 특징 및 장점을 설명하면, 입력 단계에서 원 피부 병변 영상의 영역 분할 부분을 투명도 정보로 전달받아 다중 어텐션 활성화 맵들과 비교하여 피부암 진단의 정확도를 높이는 방법을 제안한다. 또한 기존 개발된 어텐션 모듈을 활용하는 딥러닝 모델들과 달리 여러 개의 어텐션 맵을 사용하고 모든 어텐션 활성화 맵(Activation map)의 평균을 이용하여 피부 병변의 중요 부위를 자동으로 인지 및 집중적으로 학습할 수 있도록 한다. 여러 개의 딥러닝 모델을 하나의 공통 손실 함수로 동시에 학습시켜(Joint training) 딥 모델 앙상블을 구축한 후 테스트 영상 분류 시에 다중개의 분류 결과들을 결합하여 분류 성능을 높이는 딥 앙상블(Deep ensemble) 분류 방법을 활용하였다.

위에서 언급한 제안 기술들을 통해 심층 합성곱 신경망의 분류 성능을 높이고 상대적으로 적은 데이터로 인해 발생하는 불균형(Data imbalance)에 의한 성능 저하를 줄이고 분류 정확도를 높이는 결과를 얻었다. 위 제안 방법들을 적용한 EfficientNet-B0+B1+B2 딥 앙상블 모델의 경우, 사람의 피부 질환 데이터셋인 HAM10000[12]에서 94.4%의 분류 정확도를 얻었다. 이는 다른 최신 연구[13,14,15]와 비교하여 향상된 분류 정확도를 보여주고 있다. 또한 동일한 모델에 대해서 본 연구진이 자체적으로 수집한 동물 피부 병변 데이터셋에서 진행된 실험에서는 분류 정확도 81%를 얻었다.

2. 관련 연구

Krichevsky[1]이 제안한 심층 합성곱 신경망을 사용하는 영상 분류 연구가 다수 진행되면서 피부 질환 영상 분류에도 적용을 위한 연구가 매우 활발하게 진행되고 있다. 그럼에도 불구하고 피부 병변 영상 데이터셋은 대부분이 클래스 간 불균형 문제를 가지고 있으며, 이러한 문제는 피부 병변 자체가 가지는 질환별 부류 간에 낮은 특성 차이로 인해 심층 신경망 모델로부터 높은 분류 성능을 획득하는 데 많은 어려움이 있다. 이런 문제점을 해결하기 위해 아래와 같은 기존 연구가 수행되었다. Masood[2]가 제안한 피부 병변 분류 연구에서는 데이터 보정, 전처리, 영역 분할, 분류별 데이터 불균형 해소와 데이터셋 교차 검증과 같은 방법을 사용하는 피부 병변 분류 프레임 워크를 제안했다. 이러한 제안 방법은 데이터 불균형이 극심한 피부 병변 분류 데이터셋에서 분류 모델이 과적합에 빠지지 않고 안정적으로 성능을 향상하는데 성공했다. Valle[3]의 연구에서는 피부 병변 분류를 위한 딥러닝 모델을 제안했으며 또한 데이터 증강, 입력 영상 정규화, 영역 분할 정보 활용 등을

제안했다. 해당 연구에서는 기타 모델의 구조나 학습 기법보다는 데이터 증강(Data augmentation) 기술이 분류 성능을 향상하는데 있어 가장 효율이 높았다고 주장했다. 그럼에도 불구하고 위 언급된 피부 병변 영상 데이터셋의 태생적 한계인 데이터 불균형 문제는 충분히 해소되지 않았으며 학습 데이터 불균형을 해소하기 위해 Zunair[4]의 연구에서는 데이터의 개수가 적은 피부 질환 부류를 대상으로 생성적 적대 신경망을 이용하여 상대적으로 적은 데이터를 가지는 분류에 대해 임의의 데이터를 생성하여 데이터 불균형을 해소하고 분류 성능을 개선하고자 했다.

피부 질환을 위한 분류 모델로서 주로 심층 합성곱 신경망(Deep Convolutional Neural Network, DCNN)을 활용하여 연구됐으며 다양한 기법을 기반으로 변형된 분류기 모델로서 발전되어왔다. 심층 합성곱 신경망 모델을 기본으로 하는 분류 모델은 AlexNet[5], ResNet[6], EfficientNet[7]이 있으며, 최근 연구에서는 심층 합성곱 신경망 분류기 모델을 기초로 피부 병변 부위의 중요 부분을 스스로 학습하고 이를 분류 결과에 반영하는 어텐션 기법을 추가한 모델[8]과 스스로 생성한 특징 맵에 대해 다른 위치의 요소와 연관성을 비교 및 학습하는 셀프 어텐션 모듈에 관한 연구[9]도 있다. 어텐션을 사용하는 분류 모델의 경우, 피부 병변 영상으로부터 병변 분류에 중요한 부분을 분류 모델이 구분하고 분류기가 목표로 하는 피부 병변으로 분류할 수 있도록 한다. 위와 같이 어텐션을 활용하는 연구 외에도 이와 유사한 역할을 하는 가버 필터와 밀도 기반 공간 클러스터링[13] 활용하여 피부 병변의 이상 감지를 통해 피부 병변을 검출하는 연구 또한 존재하며 주로 실생활에서 촬영된 영상에서 사용이 가능할 정도로 높은 일반화 성능을 지닌다.

위와 같은 연구들 중에서도 EfficientNet은 Tan et al.[7]이 제안한 심층 합성곱 신경망으로 모델 내에 존재하는 대부분의 매개변수에 합성 계수를 적용하여 입력 해상도뿐만 아니라 네트워크 계층(Layers)의 깊이와 넓이까지 조정 가능한 심층 합성곱 신경망이다. EfficientNet-B0가 가장 작은 경량화 모델이며 EfficientNet-B7가 가장 고도화된 모델이다. 본 논문에서는 앙상블 분류 기법을 적용하는 단계에서 네트워크 모델들 사이에 다양성(Model diversity)을 확보하기 위해 EfficientNet 네트워크 계열을 활용하였다.

분류 모델에서 앙상블 기법은 개별로 학습된 분류 모델에 입력 데이터가 들어왔을 때, 일련의 분류기들의 결과를 모아서 결정하는 방법이다. 예측 결과를 모두 모은 후 결합하여 가장 가능성(확률)이 높은 부류(class)를 최종적으로 결정한다. 이때 결정 방식에는 단순 투표(Hard voting)와 가중치 투표(Soft voting) 방법이 있으며[17], 단순 투표(Hard voting)은 각각의 분류 모델이 내린 예측을 모아 가장 많은 득표를 한 분류를 최종적으로 선택하는 방법이다. 가중치 투표(Soft voting)은 각각의 분류 모델에서 내린 예측 신뢰도 확률을 평균하여 최종적으로 분류하는 방법이다. 이 경우, 높은 확률로 예측한 분류기에 가중치가 부여되는 효과를 얻을 수 있으므로 대부분의 분류 문제에서 더 정확한 예측을 보여준다. 앙상블 모델은 서로 다른 구조의 신경망 모델을 사용할 수도 있고, 동일한 신경망 구조에 서로 다른 데이터 증강, 가중치 초기화 및 학습 방법을 적용하여 학습시킬 수도 있다. AlexNet [5], VGGNet [10], DenseNet121 [11]등의 모델을 활용한 여러 모델 중에는 동일한 모델을 서로 다른 초매개변수를 활용하여 성능을 개선한 모델들이 있다.

3. 제안 피부 병변 분류 방법

Fig. 1은 제안 모델의 전체적 설계 구성을 보여준다. Concat.는 채널 간 병합을 의미하며 실선 화살표는 순전파, 점선 화살표는 역전파 진행 과정을 보여준다. 제안 모델은 입력으로 피부 병변 영상과 해당 영상의 영역 분할 정보가 투명도로 변환된 영상을 받는다. 해당 영상은 특징 추출 신경망으로 사용된 EfficientNet을 사용하여 특징 맵으로 변환되며 해당 특징 맵은 아래에서 제안하는 다중 어텐션 모듈을 통해 여러 어텐션 맵을 생성한다. 최종적으로 생성된 어텐션 맵과 특징 맵이 합성되어 심층학습 기반 분류기에 전달되어 피부 병변의 종류를 판단하게 된다.

피부 병변 영상의 특징 맵 추출을 위해 앙상블 심층 신경망 모델을 구성하였고 이를 위해 EfficientNet-B0, B1, B2[7]를 사용했다. 입력으로 사용되는 피부 병변 영상은 영역 분할 투명도 융합 단계를 거쳐 피부 병변 영상과 영역 분할 정보가 하나의 영상에 융합되게 된다. 이후 융합된 영상은 EfficientNet-B0, B1, B2에 전달되어 각각 특징 맵을 얻게 된다. 이후 다중 어텐션 네트워크 모듈들을 통해 어텐션

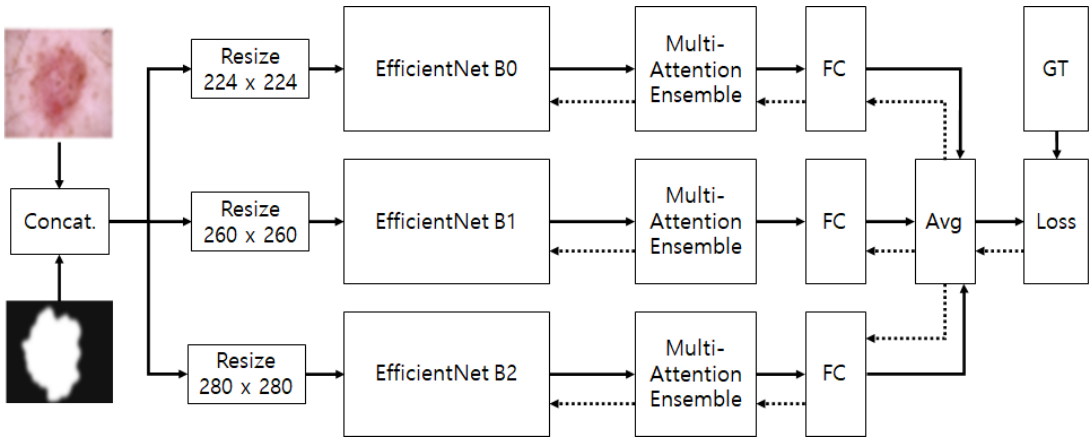

Fig. 1. Overall framework of the proposed skin lesion classification method. FC and GT denotes fully-connected layer and ground-truth, respectively.

맵을 얻고 각각의 특징 맵과 융합되어 최종적으로 앙상블 분류기에서 피부 병변 영상이 속하는 병변으로 분류하게 된다. 분류 결과는 가중치 투표(soft voting)[17]을 사용하며 EfficientNet-B0, B1, B2에서 유래된 결과를 모두 더 하고 평균 내어 가장 큰 값을 가지는 피부 질환 부류를 최종 결과값으로 선택하는 방법을 사용한다. B0, B1, B2 모델의 학습은 개별적으로 이루어지지 않고 하나의 단일 손실 함수로 세 개의 심층 네트워크 모델들을 동시에 학습시키는 (joint training) 방법을 활용한다.

3.1 영역 분할 투명도 융합 단계

제안 모델은 Fig. 2와 같이 입력 영상으로 피부 병변 영상과 해당 영상의 영역 분할 정보를 투명도 정보로 전환 시킨 영상에 병합된 영상을 입력으로 사용한다. 이 과정을 위해서 병변 분할 영상과 해당 영상에 해당하는 영역 분할 정보가 별도로 다른 모델에서 제공되어야 한다. 이때 제공된 영역 분할 정보는 신경망에서 사용되기 전에 피부 병변 영상과 병합되는 과정을 거치며, 이 과정에서 영역 분할 정보는 영상의 투명도 정보로 전환되어 피부 병변 영상과 병합된다. 이는 피부 병변 영상과 피부 병변 영역 분할 정보가 개별적으로 특징 맵을 추출하지 않고 서로의 연관성을 학습하면서 하나의 특징 맵으로 피부 병변 영상과 피부 병변 영역에 관한 결과를 얻을 수 있도록 한다. 또한 EfficientNet이 별도의 영역 분할

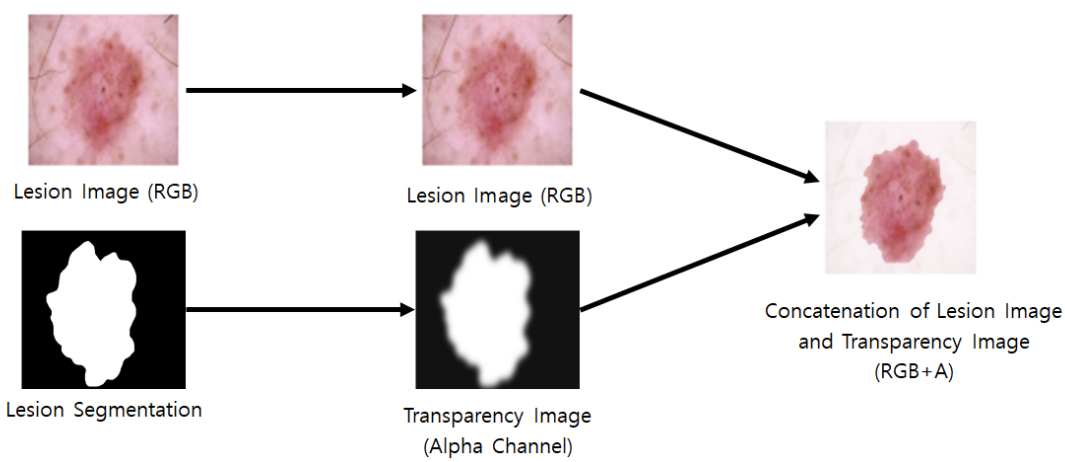

Fig. 2. Fusion of original lesion image and its segmentation information for improved classification.

마스크를 입력으로 받기 위해 모델의 구조를 변경하지 않고도 학습할 수 있도록 하며, 이는 해당 제안 방법이 EfficientNet 뿐만 아니라 다른 특징 추출 모델에도 활용될 수 있는 유연성을 가지게 한다. 결과적으로 투명도 정보와 피부 병변 영상의 병합으로 이루어진 입력 영상은 제안 모델의 다중 어텐션 모듈에서 얻은 어텐션 맵과 비교하여 환부 영역의 중요도를 두 번 판단하고 서로 비교할 수 있도록 하는 역할을 하게 한다.

영역 분할 정보의 형태에 따라 이를 투명도 정보로 전환한다. 집중 영역의 경계가 명확한 하드 어텐션(hard attention)이라면 0과 1의 값에 각각 100%와 0%의 투명도를 부여한다. 이는 영역 분할되어 제거되는 부분을 100% 투명하게, 제거되지 않은 부분은 0%의 투명도를 부여하여 불투명하게 하여 피부 병변 영상에 영역 분할 정보를 부여하게 한다. 집중 영역의 경계가 선형관계를 따르는 소프트 어텐션이라면 0과 1 사이의 값에 반비례하는 100%와 0% 사이의 값으로 부여 할 수 있다. 본 논문에서는 하드 어텐션의 영역 분할을 사용하였기 때문에 모든 투명도 정보는 100% 또는 0%를 사용하였다.

3.2 다중 스케일 어텐션 네트워크

기존 합성곱 신경망 분류 모델에선 정상 피부 영역과 환부 영역을 동일한 중요도로 관찰했다. 하지만 이는 정상 피부 영역이 상대적으로 넓은 영상에서 환부에 대한 특성을 제대로 얻지 못하거나 정상 피부의 특성에 압도되어 신경망이 제대로 학습하지 못하는 경우가 발생할 수 있다. 이러한 경우를 해결하기 위해 본 논문에서는 Fig. 3과 같은 어텐션 모듈을 사용하여 분류 모델이 정상 피부 영역보다 환부 영역에 더 많은 가중치를 두어, 환부 영역에서 획득한 특성에 대해 더 집중하여 학습할 수 있도록 한다. 특히 환부에 비해 정상 피부 영역이 큰 영상에 대해 성능을 안정적으로 유지할 수 있도록 한다. 이때 여러 개의 3차원 특징 맵을 생성하며 이는 특징 맵의 채널간 정보에 대한 정보를 어텐션 맵 생성에 활용하게 하며, 단일 어텐션 맵이 놓칠 수 있는 중요 요소에 대해 상호보완 할 수 있도록 한다. 이후, 0과 1 사이의 값을 가지는 어텐션 맵을 피부 병변 영상의 특징 맵과 곱한다. 낮은 어텐션을 가지는 부분은 0에 가까운 가중치를 가지게 되고, 높은 중요도를 가지는 어텐션 부분은 1에 가까운 가중치를 가지게 되어 정상 피부 영역과 피부 병변 영역이 분류 모델에 미치는 영향력을 조절하고 중요 부위에서 얻은 정보가 모델에 더 많은 영향력을 가질 수 있도록 한다. 다중 어텐션 앙상블 모듈의 출력은 f_{ma}라고 할 때, f_{ma}는 아래와 같이 정의된다.

$$f_{ma} = t\&(t^*(\sum_{n=1}^{K} 3DConv_{Wn}(t))) \quad (1)$$

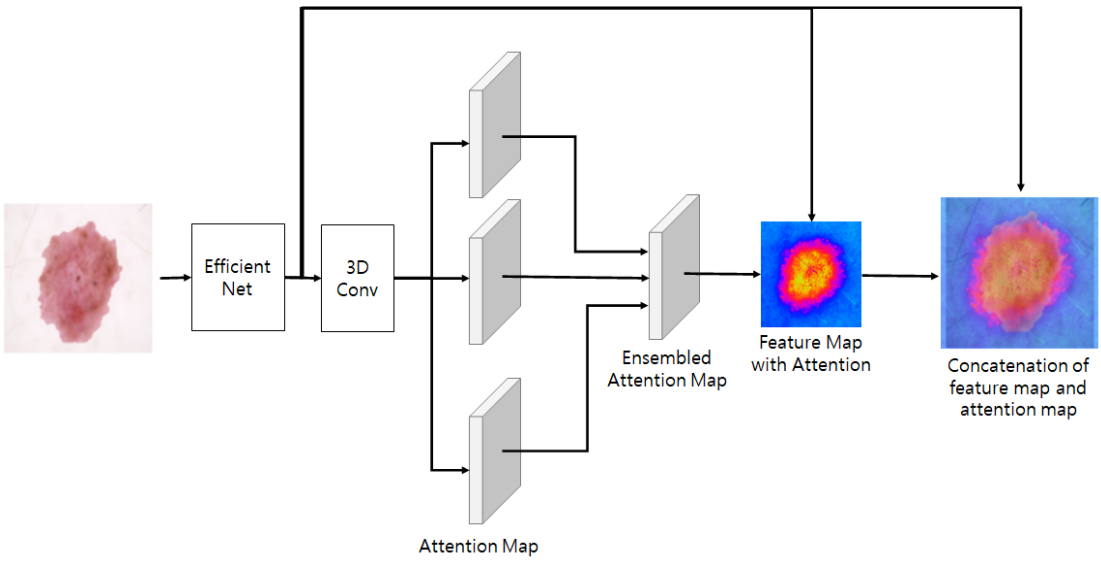

Fig. 3. Architecture of multi-scale attention module.

여기서, t는 EfficientNet으로부터 생성되는 (h, w, d)의 크기를 가지는 특징 맵을 의미한다. h, w, d는 각각의 B0, B1, B2의 계산으로 결정된다. W_k 는 3차원 합성곱 신경망의 가중치에 해당한다. K는 최대 생성할 어텐션 맵의 수이다. &는 채널 간 병합, *는 요소간 곱셈을 의미하며 $3DConv_{Wn}$는 3차원 합성곱 계층의 n번째 필터의 가중치를 사용한 출력에 차원을 한단계으로 낮추며 요소별 덧셈을 수행한 값을 의미한다.

제안 다중 스케일 어텐션 모듈은 3차원 합성곱 계층으로 이루어진다. 특징 추출 신경망에서 얻은 특징 맵을 입력으로 받아 (h, w, d, k)의 형태를 가지는 여러 개의 어텐션 맵을 얻도록 한다. 해당 값들은 활성 함수로 소프트맥수 함수를 사용한다. 이후 모든 어텐션 맵의 차원 수를 낮추면서 모든 값에 요소별 덧셈을 수행한다. 이때 (h, w, d, k)의 형태를 가지는 3차원 합성곱 계층의 출력에서 k개의 차원에 속하는 (h, w, d)에 대해 요소별 덧셈을 진행한다. 이후 (h, w, d)의 형태로 변환 된 어텐션 맵을 소프트맥스 함수를 사용하여 0과 1사이의 값으로 변경하고 EfficientNet으로부터 얻은 특징 맵과 요소 별 곱셈을 실행한다. 어텐션 결과가 곱해진 특징 맵은 어떤 부분의 특징 맵이 중요하고 중요하지 않은지에 대한 지표가 된다. 해당 어텐션 특징 맵은 어텐션이 없는 특징 맵과 병합되어 최종적인 출력을 구성한다(Fig. 4 참조). 이때 3차원 합성곱 계층으로 생성한 어텐션 맵의 수는 지나치게 많으면 연산 요구량이 급격히 증가하기 때문에 배치 크기를 크게 유지할 수 있으면서 EfficientNet의 스케일링을 수치에 비례하는 수치로 선택했다. 따라서 EfficientNet-B0, B1, B2는 각각 8, 8, 16개의 어텐션 맵을 생성하여 앙상블 하는 과정을 거친다.

3.3 조인트 앙상블 분류 기법

본 논문에서는 제안 모델을 다양한 스케일의 EfficientNet 특징 추출 모델에 적용하여 조인트 앙상블 학습 기법을 적용한다. 조인트 앙상블 학습 기법은 기존 앙상블 기법들이 사용하는 방법과 달리 단일 손실 함수를 사용한다. 기존 앙상블 기법은 대표적으로 개별 모델을 각자 학습 단계를 거쳐 추론 단계에서 모든 결과를 합친 후 분류 결과를 얻게 된다. 하지만 조인트 앙상블 학습 기법은 하나의 손실 함수로 모든 모델을 동시에 학습시키게 된다. 이 경우, 각각의 모델은 서로 다른 모델에게 영향력을 미칠 수 있게 된다. 즉, 기존 앙상블 모델은 추론 단계에서만 서로의 결과를 통해 영향력을 미치며 서로 보완할 수 있었던 반면, 조인트 앙상블 학습 기법은 학습 단계에서도 서로 영향을 주어 서로의 학습 단계를 보완할 수 있도록 한다. Fig. 5와 같이 EfficientNet-B0, B1, B2를 사용하여 조인트 앙상블 분류 기법을 사용한다고 할 때, 각 EfficientNet의 출력인 특징맵은 각각의 다중 스케일 어텐션 네트워크에 전달되며 어텐션 맵과 특징맵의 병합 결과를 완전 연결 계층에 전달하여 분류 결과를 얻게 된다. 이때 각각의 EfficientNet에서 얻은 결과를 평균 내어 최종적인 분류를 확정하게 된다. 이후 최종 분류 결과를 정답과 비교하며 아래의 손실 함수 계산에 따라 손실 값을 얻는다. 하나의 손실 함수가 3개의 EfficientNet 모델의 결과로부터 결정되었으므로 역전파 과정에서 각각의 EfficientNet 모델이 손실 함수에 기여한 정도에 따라 가중치를 갱신하는 과정이 이루어 질 수 있도록 한다. 특히, 서로 다른 크기의 입력을 받는 모델을 앙상블 할 때, 입력 크기 및 해상도에 따른 차이점을 모델 간에 공유할 수 있으므로 일반화 성능이 향상 될 수 있도록 한다. Fig. 5에서 묘사한 EfficientNet-B0, B1, B2를 사용하여 3개의 개별 모델을 조인트 앙상블 학습 기법을 활용한다고 할 때, 손실 함수(loss function)는 아래와 같이 표현된다.

$$p_c = p_c^{B0} + p_c^{B1} + p_c^{B2} \tag{2}$$

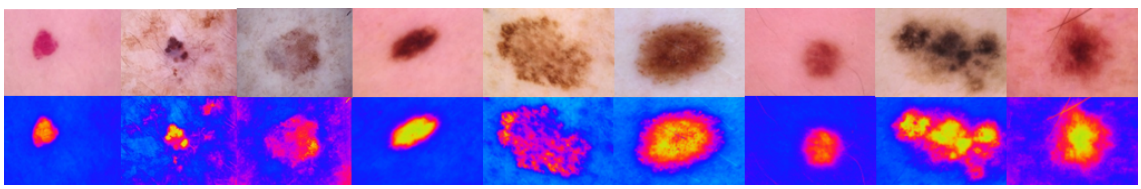

Fig. 4. Examples of original skin lesion images (first row) and their corresponding attention maps (second row).

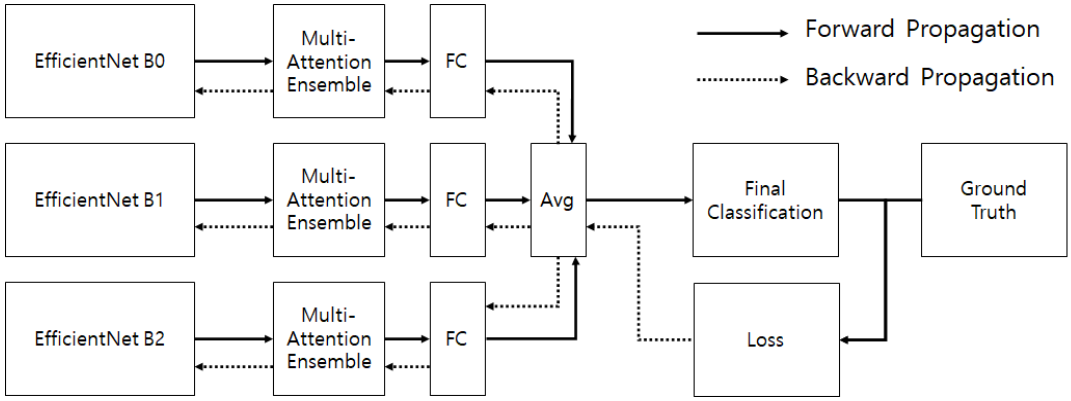

Fig. 5. Architecture of joint ensemble training method.

$$Loss_{IE} = -(W_c^* \sum_{c=1}^{N} y_c \log(p_c)) \quad (3)$$

$$W_c = \frac{img_c}{\sum_{i=1}^{N} Img_i} \quad (4)$$

여기서, N은 총 분류 수, y_c는 분류 c에 대한 지표, p_c^{Bx}은 EfficientNet Bx의 c에 대한 예측 확률을 의미한다. W_c는 불균등한 데이터에 추가하는 가중치를 의미하며 c 분류에 속하는 영상의 수에 반비례하는 값을 손실 함수 계산 시 곱한다. img_c는 c 번째 분류에 속하는 데이터 수이다. 기본적으로 손실 함수는 개별 분류 예측 확률을 모두 더하여 평균값을 사용하며 손실 계산을 위해서 크로스 엔트로피를 사용한다. p_c^{Bx}를 개별 크로스 엔트로피 계산을 통해 손실 함수를 구하고 이를 바탕으로 역전파 과정을 통해 학습하게 되면 해당 손실 함수가 속한 모델만 가중치가 갱신되게 된다. 반면에, 조인트 앙상블 학습 기법의 손실 함수는 학습 과정에서 각각의 모델이 예측한 확률이 최종 $Loss_{IE}$에 기여한 정도에 따라 갱신되며 이는 Fig. 5에서 역전파의 방향과 같이 학습 단계에서 모델 서로가 상호 보완적인 학습이 가능하도록 한다.

4. 실험 결과 및 고찰

4.1 HAM10000 데이터셋

본 논문에서 사람 피부 질환 실험은 HAM10000 [12] 데이터셋을 사용하였다(Fig. 6와 Table 1 참조). HAM10000은 10015개의 피부확대경검사 영상을 포함하고 있는 피부 병변 데이터셋이다. 50%의 영상이 조직 검사를 통해 병명이 확인되었으며 나머지 50%는 임상적 추적 검사, 전문가 의견, IVCM에 의해 결정되었다. 모든 영상은 450 × 600의 크기를 가지고 있다. 영상들은 7개의 분류를 가진다. 분류는 아래와 같다. Melanoma(MEL), Melanocytic Nevi(NV), Basal Cell Carcinoma(BCC), Actinic Keratosis, and

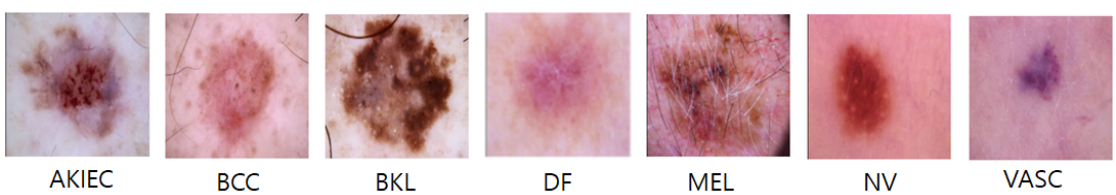

Fig. 6. Examples of HAM10000 Dataset.

Table 1. Details of Animal Skin lesion Dataset.

	Atopic	Bacterial	Etc	Fungal	Normal
Number of images	47	118	27	76	145

Intra-Epithelial Carcinoma(AKIEC), Benign Keratosis(BKL), Dermatofibroma(DF), Vascular lesions (VASC). 실험을 위해서 모든 영상은 360 x 360으로 변환했으며 피부 병변의 영역 분할 정보 또한 동일한 크기로 변환하여 사용하였다.

4.2 동물 피부 병변 데이터셋

본 논문에서는 동물 피부 질환 실험 또한 진행했다. 반려 동물 피부 질환 영상은 ㈜오엔디자인, ㈜주오 및 한강동물병원을 통하여 고양이와 강아지의 피부를 디지털 확대경을 통하여 20배 고배율 영상을 취득하였다(Fig. 7과 Table 2 참조). 취득된 영상은 수의사에 의해 피부 질환의 종류가 분류되었고 이를 기준으로 학습 및 분류를 진행하였다. 모든 영상은 450 × 600의 크기로 변경하였다. 분류는 아토피성, 곰팡이성, 박테리아성, 기타, 정상으로 구성 되어있다. 기타 분류는 피부 질환에 속하지만 아토피, 곰팡이, 박테리아성이 아닌 모든 피부 질환에 속한다.

4.3 실험 환경

병변 영상을 분류하기 위한 딥러닝 모델을 학습하고 성능을 평가하기 위해 HAM10000 데이터셋과 동물 피부 질환 영상 데이터셋에 대해 실험을 구성했다. 특징 추출을 위한 앙상블 모델은 EfficientNet-B0, B1, B2를 사용했으며 단일 모델은 EfficientNet-B4를 사용했다. 모든 모델은 ADAM 최적화 기법을 사용하여 학습한다. 학습률은 1-e4를 시작으로 100회 Epoch를 지날 때 마다 학습률을 0.1배하도록 했다. 배치 크기는 16개로 설정하고 손실 함수는 크로스 엔트로피(Cross Entropy)를 사용하며 데이터 불균형로 인해 발생할 수 있는 특정 분류 학습 편향을 방지하기 위해 모든 분류 기준의 손실 함수 계산은 분류에 속하는 영상 수에 반비례하는 수치의 가중치를 부여했다. 동물 데이터셋에서의 실험 또한 위와 동일한 조건에서 진행되었으나, Epoch의 학습률 감쇄율은 0.1배가 아닌 0.5배로 설정하였다.

딥러닝 모델 구현을 위해 Ubuntu 20.04에서 Python 3.8을 기반으로 Tensorflow 2.5를 사용하여 구현되었다. 모델 학습과 구현을 위해 Intel i7-10700K, Nvidia RTX 3090 1개를 사용한 PC에서 수행되었다. 성능 평가를 위해선 HAM10000 데이터셋과 피부 동물 피부 질환 데이터셋을 학습과 시험용으로 각각 8:2, 6:4로 구성하고 모든 분류 영상에 대한 정확도를 보여주는 Accuracy와 데이터 불균형에 따른 가중치를 적용한 Balanced Accuracy를 성능 지표로 사용하였다. 성능 지표를 위한 시험용 데이터는 K-Fold Validation을 따르며 서로 중복되지 않도록 하여 총 5회 진행되어 계산되었다.

4.4 실험 결과

모델 성능의 정량적 평가를 위해 Accuracy와 Balanced Accuracy 지표를 사용했다. Accuracy는 HAM10000[12]과 동물 피부 병변 데이터셋의 모든

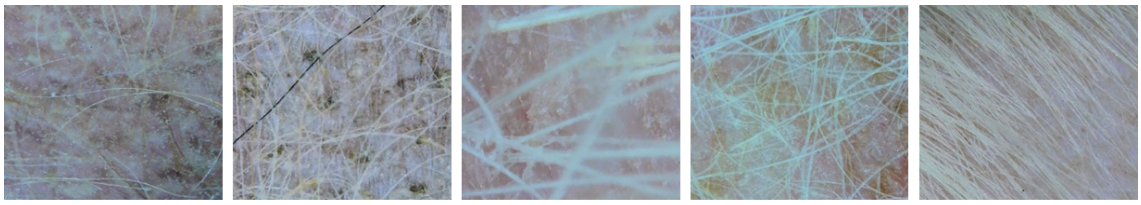

Fig. 7. Examples of animal Skin lesion Dataset.

Table 2. Accuracy and Balanced Accuracy comparison on HAM10000 Dataset.

Model	ACC	Balanced Acc
dermoscopic images and deep CNN models [7]	91.7%	n/a
loss balancing and ensemble efficient nets [14]	92.6%	n/a
Soft-Attention with CNN model [8]	93.7%	91.6%
EfficientNet-B4 Multi-Attention Ensemble Module	94.3%	92.1%

Table 3. Accuracy and Balanced Accuracy comparison on HAM10000 Dataset for demonstrating the effectiveness of using the proposed joint ensemble training method.

Model	ACC	Balanced Acc
Without using joint ensemble method	95.6%	93.8%
Using joint ensemble method	97.8%	94.3%

Table 4. Accuracy and Balanced Accuracy comparison on Animal skin lesion dataset.

Model	ACC	Balanced Acc
InceptionResNetV2[16]	78%	71%
VGG19[15]	77%	73%
EfficientNet[7] B0+B1+B2	80%	74%
Proposed model	81%	78%

영상에 대한 정확도를 보여주며 Balanced Accuracy는 각각 영상의 종류에 속하는 영상 수를 반영한 Accuracy 값을 의미한다. Balanced Accuraccy는 다수에 해당하는 분류에 의해 정확도가 편향되는 것을 반영하고 데이터가 적은 분류에 대해 가중치를 사용하여 상대적으로 많은 영상이 있는 분류에 의해 정확도 지표가 압도되는 것을 막아준다. 본 실험에서는 HAM10000[12]와 동물 피부 병변 데이터셋의 데이터를 활용하여 신경망 모델을 학습하고 비교한다.

Table 3은 제안 방법을 적용하여 사람의 피부 병변 데이터셋인 HAM10000[12]을 학습하고 분류 성능을 검증했을 때의 결과를 보여준다. 제안 모델이 94.4%로 가장 높은 정확도를 보여줬다. 두 번째로 높은 성능을 가지는 최신 방법[8]과 비교하여 1.0%p 차이의 성능을 보여준다. 이는 단일 어텐션 맵을 생성하고 Soft-Attention을 활용하는 최신 피부 병변 분류 기법[8]과 달리 제안 방법의 다중 어텐션 앙상블 모듈의 어텐션 성능이 더 나은 결과를 보임을 알 수 있다. 이는 단일 어텐션 맵을 활용하는 것보다 다중개의 어텐션을 앙상블 하여 활용하는 것이 분류 정확도 향상에 도움이 된다는 것으로 볼 수 있다. 또한 Table 8은 EfficientNet-B4가 배치 크기 1을 사용하면서 입력 영상 10장 분류할 때 걸리는 시간을 보여준다. Table 8에 따르면 제안 다중 어텐션 모듈은 처리 속도에 약 3% 이하의 영향을 주면서 분류 성능은 최대 3.5% 향상 시킬 수 있음을 나타낸다. 이는 처리 속도에 영향이 거의 없으면서 분류 성능은 개선할 수 있음을 보여준다.

Table 4는 제안 방법의 조인트 앙상블 학습 기법의 사용 여부에 따른 성능 지표 비교를 보여준다. 제안 기법을 적용한 모델은 적용하지 않은 모델에 비해 0.5%p 높은 균형 정확도를 보여준다. 정확도 뿐만 아니라 균형 정확도의 값이 상승하였으므로 이는 데이터 불균형에도 불구하고 상대적으로 데이터 수가 적은 분류에 대해 분류 성능이 향상되었다. 이는 모델을 개별 손실 함수에 각각 학습시켜 앙상블 하는 기존의 앙상블 기법보다 하나의 손실 함수에 모델을 동시에 학습시켜 앙상블 하는 제안 기법이 일반화 성능이 뛰어나며 균형 정확도 향상에 도움이 됨을 알 수 있다.

사람의 피부 병변 데이터셋 뿐만 아니라 동물의 피부 병변 데이터에 동일한 모델을 학습 및 검증하였

Table 5. Confusion Matrix of EfficientNet B0+B1+B2 Ensemble model without using proposed joint ensemble method on HAM10000.

Actual / Predicted	AKIEC	BCC	BKL	DF	MEL	NV	VAS
AKIEC	91.3%	0	1.5%	0	0	0	0
BCC	0	88.4%	0	3.3%	0	0	0
BKL	0	7.6%	89.2%	9.9%	0	0.3%	0
DF	4.3%	0	0	86.7%	5.8%	0	0
MEL	4.3%	0	1.5%	0	88.2%	0	0
NV	0	4.0%	7.8%	0	5.8%	99.6%	10%
VAS	0	0	0	0	0	0	90%

을 때의 결과를 Table 5에서 보여준다. Table 5에 의하면 다른 분류 모델과 개별 학습 후 앙상블을 진행한 모델과 달리 제안 기법들을 적용한 모델의 성능이 가장 높은 것을 알 수 있다. HAM10000 데이터셋에 비해 더 큰 데이터 불균형을 가진 데이터셋에서도 제안 방법을 사용하면 데이터 수가 적은 분류에 대한 분류 정확도 및 균형 정확도의 개선이 유효함을 알 수 있다. 또한 Table 6, 7의 Confusion Matrix에 의하면 데이터의 다수를 차지하는 Nv 부분의 정확도가 99.6%에서 99.3%로 감소했으나, 상대적으로 데이터 수가 적었던 BCC, FG, VASC는 평균적으로 1~2%p 더 높은 분류 정확도를 보여주고 있다. 이는 조인트 앙상블 학습 기법을 사용한 경우, 상대적으로 데이터가 적은 분류의 예측 정확도가 개선되는 것을 보여준다. 단일 모델 학습에서는 제안 방법을 적용한 EfficientNet B4는 최대 94.3%의 정확도를 보였으며, 앙상블 기법을 적용한 모델의 경우 최대 97.8%의 검증 정확도를 얻을 수 있다는 결과를 보여준다. 이는 기존 Soft-Attention Improves Skin Cancer Classification Performance [8]에 비하여 Accuracy와 Balanced Accuracy가 약 3.5%p, 2.2%p 향상된 결과이다.

Table 6 및 7은 제안 모델을 HAM10000과 동물 피부 병변 데이터에서 테스트했을 때의 성능을 혼동행렬(Confusion Matrix)로 보여준다. 두 표를 비교하면 상대적으로 제안 모델의 경우, 데이터가 적었던 분류(VASC, MEL, BCC)의 정확도가 높은 것으로 나타난다. 반면 제안 방법을 적용하지 않은 앙상블 모델은 데이터가 적었던 분류의 정확도가 상대적으로 떨어지는 것으로 나타난다. 이와 반대로 제안 모델은 제안 방법을 적용하지 않은 앙상블 모델에 비해 데이터가 많았던 쪽의 분류(NV) 정확도는 미세하게 감소한 것을 볼 수 있다. 이는 데이터 불균형으로 일어나는 특정 분류에 대한 모델의 정확도를 비약적으로 향상시키면서 영상 데이터가 많은 분류의 정확도 감소를 최소화하는 효과를 보여준 것으로 볼 수 있다. 이를 바탕으로 조인트 앙상블 학습 기법이 개별 학습을 통해 얻은 모델을 앙상블 했을 때보다 Balanced Accuracy의 성능이 향상됨을 알 수 있다.

5. 결 론

본 논문에서는 제안한 다중 어텐션 모듈과 투명도 기반 영역 분할 정보를 통한 분류 모델 설계와 조인트 앙상블 학습 기법을 활용하여 피부암 데이터로 구성된 HAM10000에서의 피부암 진단 신경망 모델과 동물 피부 병변에 대한 분류 모델을 구현하였다. EfficientNet의 B0, B1, B2 모델과 B4 모델을 활용하여 제안 방법을 적용하고 앙상블을 사용하여 결과를 얻었다. 단일 어텐션을 사용하는 기존의 어텐션 모듈

Table 6. Confusion Matrix of EfficientNet B0+B1+B2 Ensemble model with the use of proposed joint ensemble methods on HAM10000.

Actual / Predicted	AKIEC	BCC	BKL	DF	MEL	NV	VAS
AKIEC	91.30%	0	0	0	0	0	0
BCC	0	96.2%	0	0	0	0	0
BKL	0	0	93.8%	3.3%	2.9%	0.3%	0
DF	0	0	0	96.7%	0	0	0
MEL	0	0	0	0	94.1%	0.6%	0
NV	8.6%	3.8%	6.1%	0	2.9%	99.3%	0
VAS	0	0	0	0	0	0	100%

Table 7. Inference 10 images latency comparison with/without multi-scale attention ensemble module on batch size 1.

Model	10 Images Latency (ms)
EfficientNet-B4 w/o Multi-scale attention ensemble module	151.1
EfficientNet-B4 w Multi-scale attention ensemble module	156.8

과 달리, 투명도로 전환된 영역 분할 정보와 다중 어텐션을 활용한 기법을 통해 모델이 더 정확한 집중 영역을 학습하고 분류 정확도를 기존 최신 연구와 비교하여 약 1.0%p 향상할 수 있었다. 또한 위 제안 방법을 서로 다른 크기의 3개의 EfficientNet 모델에 적용하고 이를 하나의 손실 함수로 학습하는 조인트 앙상블 기법을 활용하여 불균형이 심한 데이터셋인 HAM10000과 동물 피부 병변 데이터셋에서 상대적으로 적은 수의 영상을 가진 분류의 예측 정확도를 약 1~2%p 향상시킬 수 있었다. 이는 단일 모델 학습에서 제안 방법을 적용하고 앙상블 기법을 적용한 모델의 경우 최신 연구에 비해 약 3.5%p 향상된, 97.8%의 검증 정확도를 얻을 수 있다는 결과를 보여준다. 본 연구 결과는 인공지능을 기반으로 하는 피부 진단 시스템의 정확도를 향상시키는 것으로 의사의 피부 병변 진단에 활용되는 컴퓨터 지원 진단 시스템의 신뢰성을 높이고 대량의 분석 의료를 수용하여 피부 진단을 위한 고급 의료 분석 장비와 의료 인력의 소모를 줄이는데 활용될 수 있을 것으로 예상된다.

REFERENCE

[1] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, Vol. 60, No. 6, pp. 84-90, 2017.

[2] A. Masood and A. Ali Al-Jumaily, "Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms," *International Journal of Bio- medical Imaging*, Article ID 323268, 2013.

[3] F. Perez, C. Vasconcelos, S. Avila, and E. Valle. "Data Augmentation for Skin Lesion Analysis," *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, pp. 303-311, 2018.

[4] H. Zunair and A.B. Hamza, "Melanoma Detection Using Adversarial Training and Deep Transfer Learning," *Physics in Medicine & Biology*, Vol. 65, No. 13, 135005, 2020.

[5] A. Krizhevsky, I. Sutskever, G.E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, 2012.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning For Image Recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[7] M. Tan and Q. Le, "Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks," *International Conference on Machine Learning*, pp. 6105-6114, 2019.

[8] S.K. Datta, M.A. Shaikh, S.N. Srihari, and M. Gao, "Soft Attention Improves Skin Cancer Classification Performance," *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data*, pp. 13-23, 2021.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, and I. Polosukhin, "Attention is All You Need," *Advances in Neural Information Processing Systems*, 30, 2017.

[10] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy, "Going Deeper in Spiking Neural Networks: VGG and Residual Architectures," *Frontiers in Neuroscience*, Vol. 13, Article 95, 2019.

[11] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger, "Densely Connected Convoltional Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700-4708, 2017.

[12] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions," *Scientific Data*, Vol. 5, No. 1, pp. 1-9, 2018.

[13] M.S. Jeon and K.J. Cheoi, "Detection of Abnormal Region of Skin using Gabor Filter and Density-based Spatial Clustering of Applications with Noise," *Journal of Korea*

Multimedia Society, Vol. 32 No. 2, pp. 117-129, 2018.

[14] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin Lesion Classification Using Loss Balancing and Ensembles of Multi-Resolution EfficientNets," *MethodsX*, Vol. 7, 100864, 2020.

[15] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv Preprint*, arXiv:1409.1556, 2014.

[16] C. Szegedy, S. Ioffe, and V. Vanhoucke, and A.A. Alemi, "Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning," *Thirty-first AAAI Conference on Artificial Intelligence*, pp. 4278-4284, 2017.

[17] S. Kumari, D. Kumar, and M. Mittal, "An Ensemble Approach for Classification and Prediction of Diabetes Mellitus Using Soft Voting Classifier," *International Journal of Cognitive Computing in Engineering*, Vol. 2, pp. 40-46, 2021.



곽 민 호

2020년 한국외국어대학교 전자공학과 학사
2020년~현재 한국외국어대학교 컴퓨터공학부 석사과정
관심분야: 딥러닝, 패턴인식, 영상처리



김 경 태

2016년 중원대학교 학사
2017년~2019년 한국외국어대학교 컴퓨터공학부 석사
2019년~현재 한국외국어대학교 컴퓨터공학부 박사과정
관심분야: 머신러닝, 패턴인식, 영상처리



최 재 영

2011년 KAIST 전기및전자공학과 박사
2008년~2012년 토론토대학 연구원
2012년~2013년 펜실베니아대학 연구원
2013년~2014년 삼성전자 책임연구원
2014년~2016년 중원대학교 의료공학과 조교수
2016년~현재 한국외국어대학교 컴퓨터공학부 부교수
관심분야: 딥러닝, 머신러닝, 패턴인식, 영상처리