

비소세포폐암 환자의 재발 예측을 위한 흉부 CT 영상 패치 기반 CNN 분류 및 시각화

마세리[○] 안가희[○] 홍헬렌*

서울여자대학교 소프트웨어융합학과

(serie, ghahn, hlhong)@swu.ac.kr

Chest CT Image Patch-Based CNN Classification and Visualization for Predicting Recurrence of Non-Small Cell Lung Cancer Patients

Serie Ma[○] Gahee Ahn[○] Helen Hong*

Seoul Women's University, Department of Software Convergence

요 약

비소세포폐암(NSCLC)은 전체 폐암 중 85%의 높은 비중을 차지하며 사망률(22.7%)이 다른 암에 비해 현저히 높은 암으로 비소세포폐암 환자의 수술 후 예후에 대한 예측은 매우 중요하다. 본 연구에서는 종양을 관심영역으로 갖는 비소세포폐암 환자의 수술 전 흉부 CT 영상 패치의 종류를 종양 관련 정보에 따라 총 다섯 가지로 다양화하고, 이를 입력데이터로 갖는 사전 학습된 ResNet 과 EfficientNet CNN 네트워크를 사용하여 단일 모델과 간접 투표 방식을 이용한 앙상블 모델, 그리고 3 개의 입력 채널을 활용한 앙상블 모델에서의 실험 결과 및 성능을 오분류의 사례와 Grad-CAM 시각화를 통해 비교 분석한다. 실험 결과, 종양 주변부 패치를 학습한 ResNet152 단일 모델과 EfficientNet-b7 단일 모델은 각각 87.93%와 81.03%의 정확도를 보였다. 또한 ResNet152 에서 총 3 개의 입력 채널에 각각 영상 패치, 종양 주변부 패치, 형상 집중 종양 내부 패치를 넣어 앙상블 모델을 구성한 경우에는 정확도 87.93%를, EfficientNet-b7 에서 간접 투표 방식으로 영상 패치와 종양 주변부 패치 학습 모델을 앙상블 한 경우에는 정확도 84.48%를 도출하며 안정적인 성능을 보였다.

Abstract

Non-small cell lung cancer (NSCLC) accounts for a high proportion of 85% among all lung cancer and has a significantly higher mortality rate (22.7%) compared to other cancers. Therefore, it is very important to predict the prognosis after surgery in patients with non-small cell lung cancer. In this study, the types of preoperative chest CT image patches for non-small cell lung cancer patients with tumor as a region of interest are diversified into five types according to tumor-related information, and performance of single classifier model, ensemble classifier model with soft-voting method, and ensemble classifier model using 3 input channels for combination of three different patches using pre-trained ResNet and EfficientNet CNN networks are analyzed through misclassification cases and Grad-CAM visualization. As a result of the experiment, the ResNet152 single model and the EfficientNet-b7 single model trained on the peritumoral patch showed accuracy of 87.93% and 81.03%, respectively. In addition, ResNet152 ensemble model using the image, peritumoral, and shape-focused intratumoral patches which were placed in each input channels showed stable performance with an accuracy of 87.93%. Also, EfficientNet-b7 ensemble classifier model with soft-voting method using the image and peritumoral patches showed accuracy of 84.48%.

키워드: 비소세포폐암, 재발 예측, 딥러닝, 분류, 앙상블 학습, 합성곱 신경망

Keywords: Non-Small Cell Lung Cancer(NSCLC), Recurrence Prediction, Deep Learning, Classification, Ensemble Learning, Convolutional Neural Network(CNN)

[○] These authors contributed equally to this work.

* Corresponding author

*corresponding author: Helen Hong/Seoul Women's University (hlhong@swu.ac.kr)

1. 서론

국립암센터(National Cancer Center)의 통계자료에 따르면 2018 년 한 해 동안 우리나라의 폐암 진단은 총 28,628 건이었으며 전체 암 진단의 11.8%이자 암 발생률 3 위를 차지했다[1]. 또한 폐암의 사망률은 22.7%으로 다른 암에 비해 현저하게 높으며 5 년 상대 생존율은 32.4%로 상당히 낮은 수치를 보인다. 폐암은 폐에 발생하는 악성종양을 말하며 암세포의 크기에 따라 전체 폐암의 15%를 차지하는 소세포폐암(Small Cell Lung Cancer, SCLC)과 나머지 85%를 차지하는 비소세포폐암(Non-Small Cell Lung Cancer, NSCLC)으로 나뉘고, 이 중 종양의 크기가 작지 않은 암을

의미하는 비소세포폐암은 다시 선암(Adenocarcinoma)과 편평상피세포암(Squamous cell carcinoma), 그리고 대세포암(Large cell carcinoma)으로 분류된다. 폐암의 진단은 대부분 흉부 X-선 촬영이나 전산화단층촬영(Computed Tomography, CT)을 통해 이뤄지며 특히 흉부 CT 영상은 폐와 림프절 등을 높은 해상도로 표현할 수 있기 때문에 폐암의 가능성을 파악하기에 용이하게 사용되고 있다. 비소세포폐암은 주로 폐의 중심 기관지나 말단 부위에서 발생하며 암의 진행이 비교적 느리기 때문에 1, 2 기와 3A 기 일부에서 조기 발견할 시에는 종양 자체 및 전이가 예상되는 종양을 둘러싼 림프절까지 모두 제거하는 근치적 절제 수술로 완치를 기대할 수 있다. 하지만 국가암정보센터에 따르면 비소세포폐암의 초기 진단의 55~80%가 전이를 동반하며,

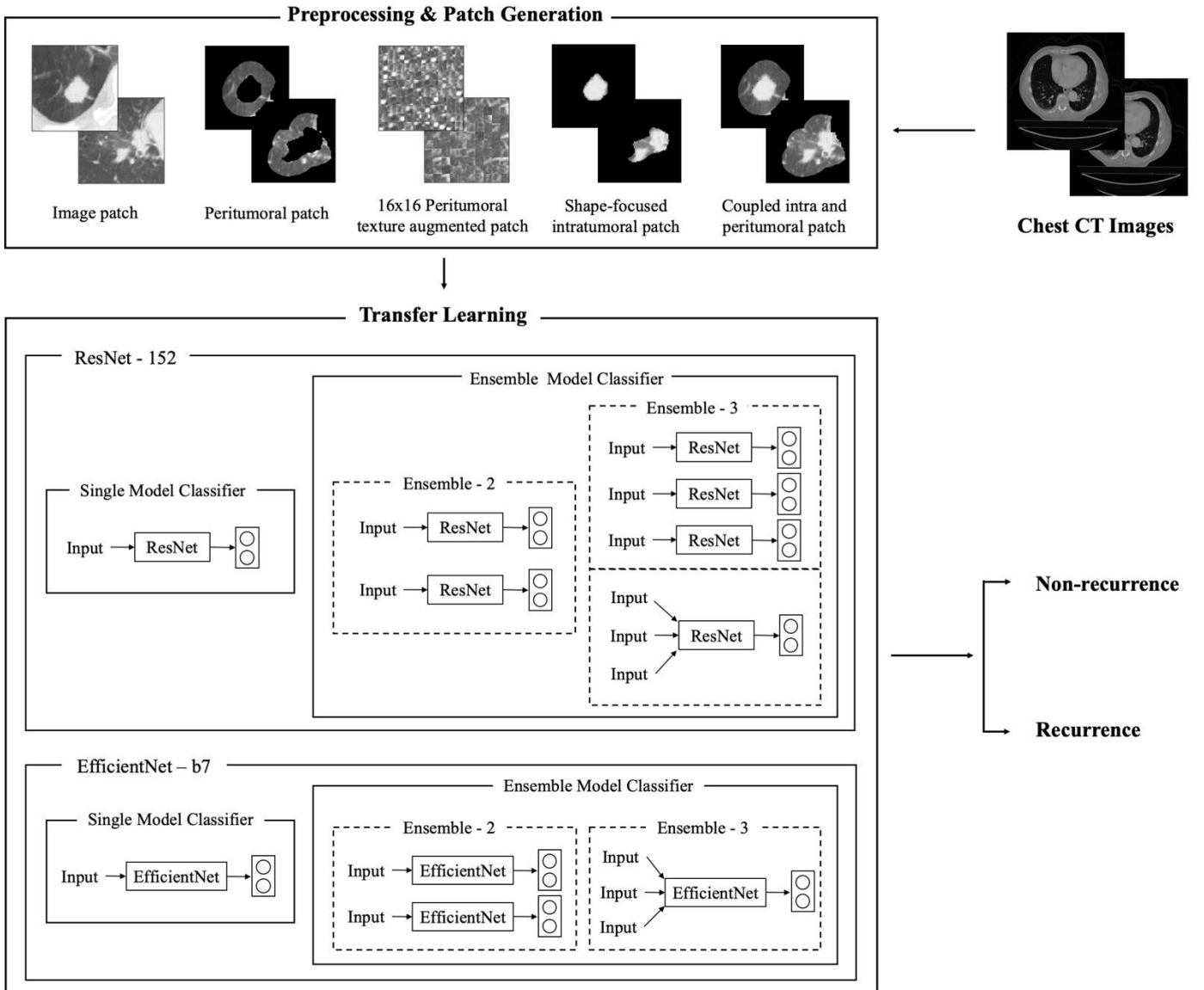


Fig 1: Overview of proposed method. 5 types of tumor related patches are generated after preprocessing for chest CT images. Single and ensemble CNN models are trained in transfer learning method, prediction is performed by classifying patches into two classes.

근치적 절제 수술을 받은 환자의 20~50%에서 재발하는 양상이 나타나기 때문에 비소세포폐암 환자의 수술 후 재발 가능성을 예측하기 위한 연구는 매우 중요하다.

최근에는 비소세포폐암 환자의 예후 예측을 보조하기 위해 CT, PET 등 다양한 의료 영상과 딥러닝 모델을 결합한 인공지능 기반 의료영상 분석 연구들이 활발히 진행되고 있다. 대표적으로는 비소세포폐암 환자의 흉부 CT 영상을 기준 기간 내 생존과 사망으로 이분화하고 변형된 ResNet18 모델이 이미지를 작은 배치 크기로 학습할 수 있도록 함으로써 기존 라디오믹스를 이용한 접근 방식보다 우수한 결과를 얻은 연구와, 폐암 절제술을 받은 비소세포폐암 환자의 수술 전 흉부 CT 영상을 기반으로 입력데이터를 다양화한 후 AlexNet의 전이학습으로 파라미터 고정 층의 범위에 따른 성능 차이를 비교 분석한 연구가 있었다[2,3]. 또한 비소세포폐암을 진단받은 환자의 PET 영상과 임상 정보를 입력데이터로 사용하여 적은 양의 데이터로 생존 시간을 예측하기 위해 VAE(Variational AutoEncoder)와 일반적인 DNN 구조를 결합하여 새로운 네트워크 구조를 제안한 연구들이 진행된 바 있다[4].

본 연구에서는 비소세포폐암 환자의 수술 후 2년 이내의 재발 가능성을 예측함에 있어 유의미하게 사용될 수 있는 종양 정보에 따라 입력데이터로 사용할 흉부 CT 영상 패치를 총 다섯 가지 종류로 다양화한다. 또한 각 종양 패치가 포함하는 종양 관련 정보에 따른 CNN 단일 모델과 간접 투표를 사용한 앙상블 모델, 그리고 3 개의 입력 채널에 서로 다른 패치를 입력한 앙상블 모델에서의 성능을 확인하고, Grad-CAM 시각화를 통해 예측에 영향을 미치는 각 패치의 특성과 우수한 성능을 보인 모델에서의 패치 분류 결과를 분석한다.

2. 제안 방법

Fig 1 은 제안하는 방법의 개요도를 나타낸다. 흉부 CT 영상에 대한 전처리 과정을 거쳐 다양화된 종류의 입력 패치를 생성하고, 이를 전이학습(transfer learning) 방식으로 ResNet152 와 EfficientNet-b7 모델에 각각 학습시킨다. 학습 시에는 단일 모델과 앙상블 모델을 구성하며 두 개의 클래스로 분류한다.

2.1 데이터 전처리

CT 영상은 의료 영상 국제 표준인 DICOM(Digital Imaging Communications in Medicine)에서 입력데이터로 사용하기

위해 PNG(Portable Network Graphics) 포맷으로 변환하여 사용한다. 영상 패치는 종양을 관심영역(Region Of Interest, ROI)으로 가지며 종양을 중심으로 160x160 픽셀 크기로 크롭된 이미지이다. 해당 데이터에 대해 창폭(window width) 1600HU, 창수준(window level) -600HU 로 밝기값을 정규화하고 픽셀 간격 0.54mm 의 공간 정규화를 수행한다. 또한 모델 학습 시 훈련 데이터에 대한 과대적합을 방지하기 위해 훈련 셋에 대해서만 랜덤 수평 뒤집기(RandomHorizontalFlip)와 랜덤 수직 뒤집기(RandomVerticalFlip)를 적용하여 데이터를 증강한다.

2.2 패치 생성

패치의 종류는 총 다섯 가지로 다양화한다. 먼저 종양을 중심으로 주변부 정보를 모두 포함하는 영상 패치(image patch)를 생성한다. 두 번째 패치는 종양 주변부 패치(peritumoral patch)로, 종양을 둘러싸는 주변부의 조직 정보가 암의 침투 및 전이와 관련이 있다는 최근 암 연구와 종양 주변부의 미세환경(microenvironment)이 환자의 생존 및 암의 재발 가능성 등에 영향을 미친다는 연구 결과를 근거로 영상 패치에서 보이는 종양의 경계로부터 외부 15mm 까지의 정보만을 포함하여 생성한다[5,6]. 세 번째 패치는 종양 주변부 패치에서 주변부의 형태나 크기 정보를 제외한 질감에만 집중하여 학습할 수 있도록 하는 종양 주변부 질감 증강 패치(peritumoral texture augmented patch)이며, 다양한 마스크 커널 크기로 생성된 질감 패치들에 대한 단일 모델 실험을 진행하여 가장 우수한 성능을 보인 패치를 한 가지 채택하여 사용하도록 한다. 네 번째 패치는 종양의 형태가 절제 수술 후 예후에 영향을 미친다는 연구 결과를 근거로 종양의 크기와 형태 정보만을 포함하는 형상 집중 종양 내부

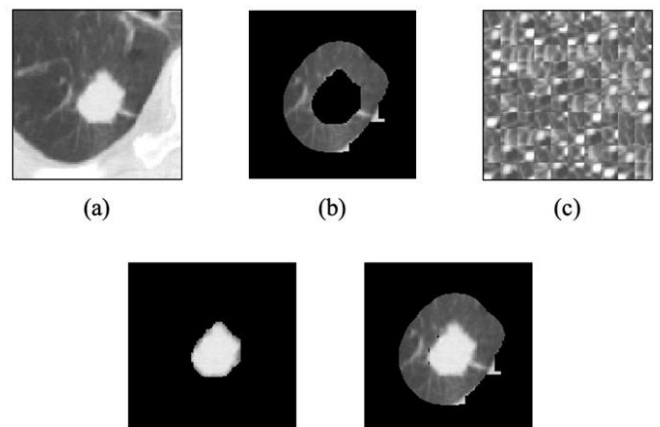


Fig 2: (a) image patch, (b) peritumoral patch, (c) 16x16 peritumoral texture augmented patch, (d) shape-focused intratumoral patch, (e) coupled intra and peritumoral patch.

패치(shape-focused intratumoral patch)이다[7]. 마지막으로 종양 자체와 종양 주변부 15mm 까지의 정보를 포함하는 coupled intratumoral and peritumoral patch 를 생성하여 종양의 형태뿐 만 아니라 주변부 정보까지 포괄적으로 학습할 수 있도록 한다. 모든 패치는 160x160 픽셀 크기로 생성하며 실험에 사용할 CNN 모델인 ResNet152 와 EfficientNet-b7 네트워크에 입력데이터로 활용하기 위해 224x224 크기로 resize 하여 사용한다.

2.3 사용 네트워크

실험에서는 두 개의 CNN 모델을 사용함으로써 결과의 신뢰도를 높이도록 한다. ResNet 은 이미지넷 이미지 인식 대회 ILSVRC 에서 2015 년도에 발표된 모델로, 층이 깊어지면서 일어나는 정보 손실을 잔차블록(residual block)의 출력에 입력을 더함으로써 우회하는 지름길을 통해 해결한다는 특징을 갖는다[8]. 본 실험에서는 ResNet 버전 중 가장 층이 깊은 ResNet152 를 사용한다. 또 다른 모델인 EfficientNet 은 2019 년도에 CVPR 에서 발표된 모델이며 깊이와 너비, 그리고 입력 이미지의 크기 세 가지의 관계에 주목하여 최적의 가중치 매개변수 값을 찾도록 compound scaling 을 제안한 네트워크이다. EfficientNet 은 타겟 성능이 매우 좋아 기존의 합성곱 신경망 모델들과 유사한 정확도를 보이면서도 매개변수와 FLOPS 의 개수를 절감할 수 있다는 특징을 가지며 b0 부터 b7 까지의 다양한 버전을 갖는데, 본 실험에서는 EfficientNet-b7 을 사용한다[9]. 두 가지 네트워크 모두 이미지넷으로 사전 학습되어 가중치가 초기화된 상태의 모델에 본 연구에서 생성한 다섯 가지 패치를 추가로 학습시켜 가중치를 업데이트 하는 전이학습 방식을 사용한다. 각 네트워크 출력층의 소프트맥스(softmax) 함수는 본래 1000 개의 클래스로 분류를 수행하지만, 본 연구에서는 재발(recurrence)과 무재발(non-recurrence) 클래스를 분류하기 위해 출력 클래스의 개수를 두 개로 재정의한다.

모델 학습을 위하여 먼저 다섯 가지 패치를 각각 단독으로 학습한 단일 모델들을 구성한다. 앙상블 모델을 구성하는 방식은 두 가지로 나뉘는데, 먼저 두 개 혹은 세 개의 단일 모델에서 예측한 확률을 더한 뒤 평균을 산출하는 방식으로 최종 예측을 수행하는 간접 투표 방식(soft-voting)을 사용한 앙상블 모델을 구성한다. 이 외에도 ResNet 과 EfficientNet 은 3 개의 입력 채널을 갖는다는 점을 활용하여 입력 채널에 영상 패치와 종양 주변부 패치, 그리고 형상 집중 종양 내부 패치를 입력하는 방식으로 앙상블 모델의 구성 방식을 다양화한다.

2.4 Grad-CAM 시각화

ResNet, EfficientNet 과 같은 딥러닝 모델들은 어떠한 판단 과정을 통해 결과를 도출했는지 알기 어려운 블랙박스 모델이며, 블랙박스 인공지능 모델들은 의사결정의 결과만 알려주고 그 근거를 제공해주지 못한다는 한계를 갖는다. 그러나 특히 의료 분야의 질병 예후 예측 과정에서 인공지능이 의사 등 전문가를 더 효과적으로 지원하기 위해선 딥러닝 모델이 내린 판단의 근거가 사람이 이해할 수 있는 형태로 제시될 수 있어야 한다. Grad-CAM 은 설명 가능한 인공지능(explainable Artificial Intelligence, XAI) 분야에서 모델 판단의 근거를 시각화하기 위해 자주 사용하는 기법으로, CAM(Class Activation Mapping)을 확장하여 고안된 개념이다[10,11,12]. Grad-CAM 에서는 특정 이미지를 재발 환자로 분류했을 때, 해당 영상에서 재발이라는 결과를 도출하는 데에 가장 크게 영향을 준 영역을 확률 값으로 표현하며 그 결과를 영상 위에 히트맵(heat map)으로 시각화한다. 본 연구에서는 각 패치별 학습 시 패치 생성에서 의도한 바와 같이 모델이 중요 영역에 집중하여 예측을 수행하였는지 확인하기 위해 네트워크의 타깃이 되는 층에서의 기울기(gradient) 정보를 사용하여 중요도를 시각화하도록 하며, 타깃이 되는 층은 두 모델 모두 마지막 합성곱층(convolutional layer)으로 지정한다. 특히 중요 영역을 제외한 나머지 영역은 모두 제거한 채 생성된 종양 주변부 패치, 형상 집중 종양 내부 패치, 그리고 coupled intra and peritumoral patch 에서 집중적으로 학습한 부분을 확인하여 학습에 영향을 주는 요인을 분석한다.

3. 실험 방법 및 결과

3.1 실험 데이터

실험 데이터로는 암 연구를 위한 의료 영상 공공 데이터를 제공하는 TCIA(The Cancer Imaging Archive)로부터 얻은 비소세포폐암 환자 100 명의 수술 전 흉부 CT 영상을 사용하였다. 해당 데이터 셋의 구성은 총 73 명의 편평상피세포암 환자와 27 명의 선암 환자로 이뤄지며, 수술 후 재발 기간에 대한 기준을 2 년으로 설정하였을 때 총 32 명의 재발 환자와 68 명의 무재발 환자로 분류되었다. 초기에 DICOM 포맷으로 제공된 데이터의 해상도는 512x512 이며 픽셀 크기는 0.97mm 이다. CT 영상 중 학습에 유용하게 사용할 수 있는 데이터를 엄선하기 위해 종양의 위치 및 형태가 표기된 이진 마스크 영상에서 유클리디안

거리 계산을 사용하여 각 환자마다 종양의 지름이 가장 크게 드러난 영상을 기준으로 상위, 하위, 그리고 기준 단면 영상까지 총 3 장의 2 차원 이미지를 선택하고, 300 장의 이미지 중 종양이 드러나지 않은 11 장을 제외한 총 289 장의 이미지를 학습에 사용하였다. 또한 전체 데이터를 환자 단위에서 8:2 비율로 나누어 훈련 셋과 시험 셋(58 장)을 생성하고 훈련 셋에 다시 동일한 비율을 적용하여 훈련 셋(186 장)과 검증 셋(45 장)으로 분리하였다. 무재발 군과 재발 군 간의 데이터 셋 비율은 약 1 : 2.1로 재발 환자의 비율이 무재발 환자의 2 배 이상의 비중을 차지하였다.

3.2 실험 계획 및 방법

단일 모델 실험에서 학습률(learning rate)과 훈련 조기 종료(early stopping) 조건, 그리고 에폭(epoch) 등의 하이퍼파라미터(hyper parameter)는 각 패치와 네트워크 별로 최적화시킨 값으로 고정한 뒤 진행하였다. 검증 손실이 일정 에폭 간 더 이상 최소값보다 줄어 들지 않는다면 손실이 최소값에 안정적으로 수렴하였다고 판단하고 훈련을 조기에 종료하도록 설정하였다.

종양 주변부 질감 증강 패치로는 각각 종양 주변부를 7x7, 9x9, 11x11, 16x16 픽셀 마스크 크기로 크롭한 패치들 중 EfficientNet-b7 단일 모델에서 가장 우수한 성능을 보인 패치로 종류를 단일화하였다. 또한 ResNet152 모델 실험에는 종양 주변부 질감 증강 패치를 제외한 총 네 가지 종류의 패치를, EfficientNet-b7 모델 실험에는 coupled intra and peritumoral patch 를 제외한 총 네 가지 종류의 패치를 사용하였다.

각 모델에 대한 성능 평가 방법으로는 정확도(accuracy)와 민감도(sensitivity), 특이도(specificity), 양성 예측도(Positive Predictive Value, PPV), 음성 예측도(Negative Predictive Value, NPV), 그리고 ROC(Receiver Operating Characteristic, ROC) 곡선을 기준으로 하였으며, 양성(positive) 클래스와 음성(negative) 클래스는 각각 재발(recurrence)과 무재발(non-recurrence) 클래스로 설정하였다. 전체 실험은 Colab GPU 환경에서 파이썬(python) 3.8.7 버전과 파이토치(pytorch) 1.8.1 버전 라이브러리로 진행되었다.

3.3 실험 결과

Table 1 은 EfficientNet-b7 단일 모델에서 종양 주변부 질감 증강 패치의 마스크 크기 차이에 따른 성능을 나타낸다. 종양 주변부를 크롭한 정사각형의 커널 크기를 16x16 픽셀로 지정한 경우에 정확도 71.42%를 포함한 총 네 가지 지표에서 가장 높은 성능을 보였다. 특히 민감도의 경우 97.36%로 상당히 우수한 값을 보이며 재발 환자에 대한 분류 성능이 뛰어난 것으로 나타났고, PPV 와 NPV 또한 각각 71.15%, 75.0%로 안정적인 예측 성능을 보였다. 다만 질감 패치의 경우 마스크의 커널 크기와 관계없이 모든 경우에 높은 민감도에 대비되는 매우 낮은 특이도를 보이며 재발 환자에 대한 예측 성능은 우수하지만 무재발 환자에 대한 예측도는 매우 떨어지는 것으로 파악되었다. 또한 3x3 나 5x5 등 매우 작은 커널 크기보다 7x7 이상의 커널 크기를 부여했을 때 여러 지표 면에서 유의미하다고 볼 수 있는 성능 및 분류 결과를 확인할 수 있었다. 해당 결과를 토대로 종양 주변부 질감 패치의 종류는 16x16 마스크 크기의 버전으로 단일화하여 이후 EfficientNet-b7 단일 모델과 앙상블 모델 실험을 진행하였다.

Table 2 는 ResNet152 과 EfficientNet-b7 을 사용한 단일 모델과 앙상블 모델에서의 실험 결과를 나타내며 Fig 3 은 ROC 곡선의 분석 결과를 나타낸다. ResNet152 단일 모델 실험 결과, 영상 패치, 종양 주변부 패치, 형상 집중 종양 내부 패치, coupled intra and peritumoral patch 에 대한 정확도는 각각 82.75%, 87.93%, 72.41%, 74.13%로 나타났으며, 87.93%로 가장 높은 정확도를 보인 종양 주변부 패치 학습 모델은 특이도, PPV, NPV 측면에서도 84.21%, 92.1%, 80.0%를 보이며 다른 패치들과 비교해볼 때 가장 우수한 성능을 보였다. 두번째로 우수한 성능을 보인 패치는 영상 패치(정확도 82.75%)로, 민감도 92.3%와 NPV 80.0% 총 두 가지 지표에서 가장 높은 값을 보였다. 형상 집중 종양 내부 패치의 경우 특이도와 NPV 측면에서 각각 31.57%와 66.66%로 매우 낮은 값을 보이며 음성 클래스인 무재발 클래스에 대한 예측 정확도가 가장 떨어지는 것으로 나타났다. Coupled intra and peritumoral patch 에서는 종양 주변부 15mm 에 대한 정보가 포함됨에 따라 형상 집중 내부 패치에서보다 정확도와 특이도, 그리고 PPV 가 향상되었다. 특히 coupled intra and peritumoral patch 에서의 특이도는 형상 집중 내부 패치의 31.57%에 비해 78.94%로 크게 향상되었는데, 종양의 크기나 형태 정보만을

Table 1: Performance comparison of peritumoral texture augmented patches.

CNN Model	Input patch	Accuracy[%]	Sensitivity[%]	Specificity[%]	PPV[%]	NPV[%]
EfficientNet-b7 Single Model	7x7 peritumoral texture augmented patch	67.24	94.87	10.52	68.51	50.0
	9x9 peritumoral texture augmented patch	65.51	82.05	31.57	71.11	46.15
	11x11 peritumoral texture augmented patch	64.91	86.84	21.05	68.75	44.44
	16x16 peritumoral texture augmented patch	71.42	97.36	16.66	71.15	75.0

Table 2: Performance comparison of ResNet152 and EfficientNet-b7 experiment results.

CNN Model	Input patch	Accuracy[%]	Sensitivity[%]	Specificity[%]	PPV[%]	NPV[%]
ResNet152 Single Model	image patch	82.75	92.3	63.15	83.72	80.0
	peritumoral patch	87.93	89.74	84.21	92.1	80.0
	shape-focused intratumoral patch	72.41	92.3	31.57	73.46	66.66
	coupled intra and peritumoral patch	74.13	71.79	78.94	87.5	57.69
ResNet152 Ensemble Model (soft-voting)	image patch & peritumoral patch	84.48	94.87	57.89	82.22	84.61
	shape-focused intratumoral patch & peritumoral patch	72.41	76.92	42.1	73.17	47.05
	shape-focused intratumoral patch & coupled intra and peritumoral patch	68.96	79.48	26.31	68.88	38.46
	image patch & peritumoral patch & shape-focused intratumoral patch	75.86	92.3	47.36	78.26	75.0
ResNet152 Ensemble Model (3 channel input)	image patch & peritumoral patch & shape-focused intratumoral patch	87.93	94.87	73.68	88.09	87.5
EfficientNet-b7 Single Model	image patch	79.31	94.87	47.36	78.72	81.81
	peritumoral patch	81.03	82.05	78.94	88.88	68.18
	shape-focused intratumoral patch	79.31	92.3	52.63	80.0	76.92
	16x16 peritumoral texture augmented patch	71.42	97.36	16.66	71.15	75.0
EfficientNet-b7 Ensemble Model (soft-voting)	image patch & peritumoral patch	84.48	89.74	73.68	87.5	77.77
	shape-focused intratumoral patch & peritumoral patch	81.03	87.17	73.68	87.17	73.68
	image patch & 16x16 peritumoral texture augmented patch	75.0	91.89	10.52	66.66	40.0
EfficientNet-b7 Ensemble Model (3 channel input)	image patch & peritumoral patch & shape-focused intratumoral patch	74.13	71.79	78.94	87.5	57.69

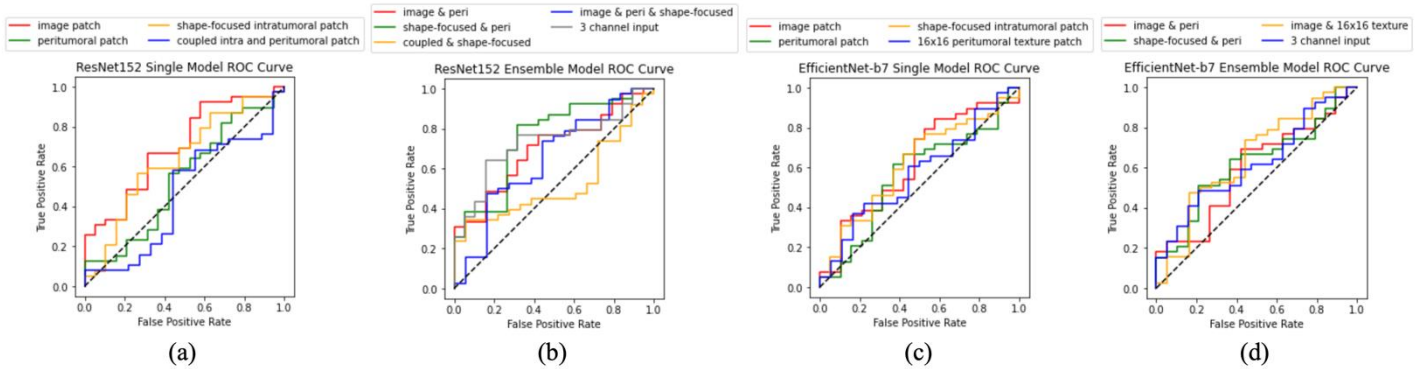


Fig 3: ROC curves from experiment models of (a) ResNet152 single model, (b) ResNet152 ensemble model, (c) EfficientNet-b7 single model, and (d) EfficientNet-b7 ensemble model.

단독으로 사용하는 것 보다는 주변부 정보를 추가하는 것이 전반적인 성능을 보완하는 효과가 있었다.

EfficientNet-b7 단일 모델 실험 결과, 영상 패치, 종양 주변부 패치, 형상 집중 종양 내부 패치에 대한 정확도는 각각 79.31%, 81.03%, 79.31%로 나타났으며, ResNet152 에서와 동일하게 종양 주변부 패치를 학습한 경우에 가장 높은 성능을 보였다. 또한 이 경우에 특이도와 PPV 측면에서도 각각 78.94%와 88.88%로 가장 안정적인 성능을 보였다. 두번째로 유용한 성능을 보인 패치는 NPV 81.81%의 영상 패치와 종양 주변부 질감 증강 패치였다. 특히 16x16 픽셀의 마스크 크기로

생성된 주변부 질감 패치 학습 모델의 경우 민감도 97.36%를 보이며 다른 패치들에서보다 재발 환자에 대한 예측 성능이 가장 우수한 것으로 드러났다.

ResNet 152 앙상블 모델 실험 결과, 간접 투표 방식을 적용한 방식에서 각 각 영상 패치와 종양 주변부 패치를 학습한 모델을 조합하여 예측한 경우 가장 높은 정확도 84.48%가 도출되었고, 정확도를 포함한 나머지 모든 지표들에서도 다른 패치 조합들보다 가장 뛰어난 성능을 보였다. ResNet152 의 각 입력 채널에 영상 패치와 종양 주변부 패치, 그리고 형상 집중 종양 내부 패치를 삽입한

양상블 모델 실험 결과, 간접 투표 방식을 사용하여 동일한 세 종류의 패치들을 조합한 경우에서의 정확도 75.86%보다 더 높은 정확도인 87.93%를 보였다. 또한 이 경우의 민감도와 특이도, PPV, NPV 는 각각 94.87%, 73.68%, 88.09%, 87.5%로 간접 투표 방식의 결과에서보다 훨씬 안정적인 성능을 보였으며 이를 통해 양상블 학습 시 입력 채널을 통한 패치 조합 방식에 따른 성능의 향상을 확인할 수 있었다. 간접 투표를 적용한 EfficientNet-b7 양상블 모델 실험 결과에서도 ResNet152 의 간접 투표 결과와 마찬가지로 영상 패치와 종양 주변부 패치를 조합하여 학습한 경우에 정확도 84.48%를 포함한 총 네 가지 지표에서 가장 우수한 결과를 보였다. 또한 EfficientNet-b7 실험에서 형상 집중 패치를 단독으로 사용하기보다 주변부 패치와 함께 조합한 양상블로 예측을 수행할 때 정확도, 특이도, 그리고 PPV 측면에서 성능이 향상되며 주변부 패치의 유용성을 보였다. 영상 패치와 16x16 종양 주변부 질감 증강 패치를 조합한 양상블 모델에서는 질감 패치를 단독으로 학습했던 단일 모델에서 민감도를 제외한 성능이 전반적으로 낮게 도출되었던 영향으로 무재발 환자인 음성 클래스에 한 예측 성능은 떨어지는 반면, 민감도는 91.89%를 보이며 재발 환자에 대한 예측 성능은 상대적으로 높게 나타났다. 또한 질감 패치를 단일 모델로 학습하는 것보다 영상 패치와 조합하여 양상블을 구성하는 것이 정확도를 소폭 향상시키는 효과를 보였다. 그러나 영상 패치와 주변부 패치를 조합한 ResNet152 과 EfficientNet-b7 의 양상블 모델 모두 특이도 측면에서 다른 지표들에서보다 상당히 낮은 값을 보이며 한계점을 보였는데 그 원인으로는 58 장의 시험 셋 중 동일한 환자에 대한 서로 다른 CT 슬라이스 영상 여러 장이 모두 오분류 되는 경우, 정분류 되는 비율을 크게 저하시키기 때문인 것으로 파악되었다.

3.4 분류 결과 분석 및 Grad-CAM 시각화

실험에서 우수한 성능을 보인 경우들에 대한 혼동행렬(confusion matrix)을 통해 시험 셋에서의 분류 결과를 확인해 보았다. ResNet152 단일 네트워크 학습에서는 종양 주변부 패치를 학습한 모델이, EfficientNet-b7 양상블 모델에서는 영상 패치와 종양 주변부 패치를 조합하여 학습한 모델이 우수한 성능을 보이고 있었다. ResNet152 로 종양 주변부 패치를 학습한 모델은 총 7 장의 이미지를, EfficientNet-b7 로 영상 패치와 주변부 패치를 학습한 양상블 모델에서는 총 10 장의 이미지를 잘못 분류하고 있었다. 두 개 모델에서 중복 오분류 된 경우를 포함하여 False Positive 와 False Negative 의 경우를 분석한 결과, 종양의 크기가 40x50

픽셀 이내로 비교적 작은 경우(2 건)나 종양의 크기가 너무 커서 160x160 픽셀 크기의 패치 내에 전체 종양의 형태가 담기지 않는 경우(3 건), 종양이 두 개 이상의 덩어리로 분리된 경우(4 건)에 오분류 되는 경향이 있는 것으로 나타났다. 또한 종양 내부에 발생한 괴사로 인해 종양의 일부분에 어두운 영역이 생성되거나 종양의 형태를 명확히 알아볼 수 없는 경우(7 건)에도 잘못 분류되고 있었는데, 이는 폐 내부에 발생한 국소적인 폐 침윤으로 인해 기관지나 혈관 등의 경계가 뚜렷하게 그려지는 간유리 음영(ground glass opacity)이 발생함에 따라 폐 내부 일부 영역의 불투명도가 높아지는 현상이 학습 시에 방해 요소가 되었을 것으로

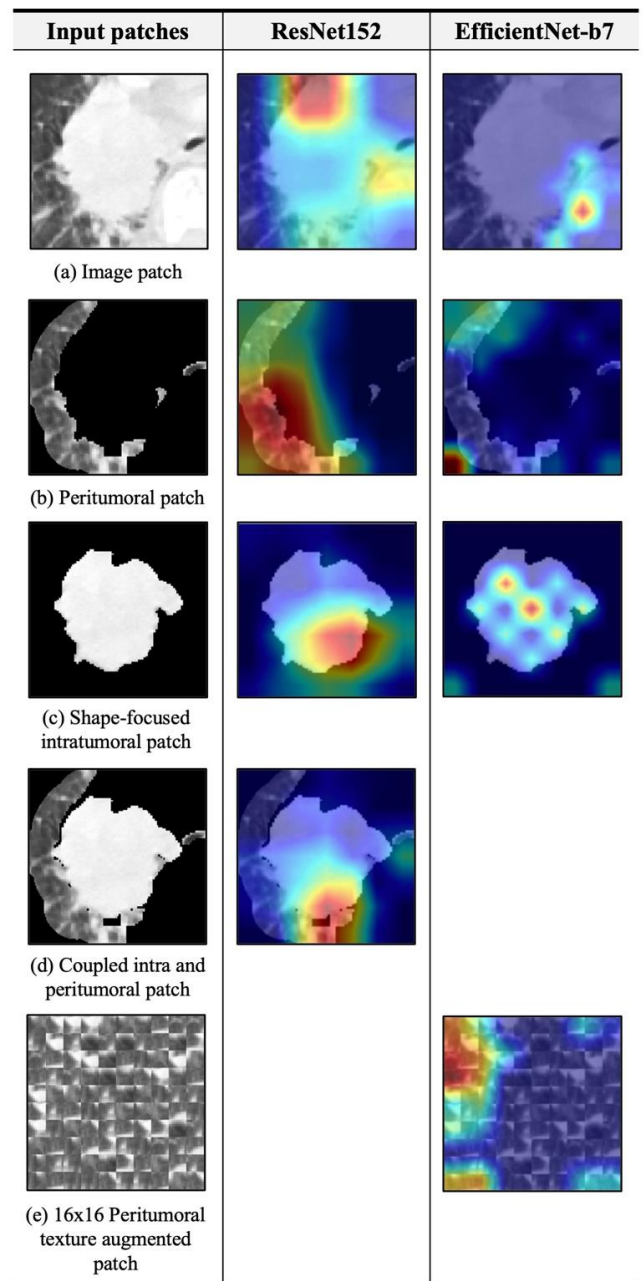


Fig 4: Visualization for each patches using Grad-CAM.

분석된다[13]. 또한 종양의 불규칙한 형태나 크기로 인한 오분류 사례들도 있었는데, 종양의 크기가 큼에 따라 재발할 가능성이 있다고 판단하여 무재발을 재발로 오분류 하거나, 반대로 종양의 크기가 작은 경우 재발하지 않을 것이라고 판단하여 재발을 무재발 클래스로 잘못 분류하는 경향이 나타났다.

Fig 4 는 각 패치에 대해 ResNet152 과 EfficientNet-b7 모델이 각각 집중적으로 학습하는 패치 상의 영역을 Grad-CAM 의 히트맵을 통해 시각화 한 결과이다. Grad-CAM 시각화 결과, 대부분의 영상 패치에서 모델은 종양 내부에만 집중하여 최종적인 예측을 수행하기보다 종양의 경계 부분이나 종양의 주변부에 초점을 맞춰 분류하는 양상을 보였으며, 이 결과로 종양을 둘러싸는 주변부 미세환경이 폐암의 생존 및 재발 등에 영향을 미친다는 최근 연구 결과에 맞게 모델이 판단 시 도움이 되는 유의미한 정보를 적절히 바라보며 학습했음을 알 수 있다. 종양 주변부 패치, coupled intra and peritumoral patch 와 같은 경우, 학습 시 결정적인 요인이 되지 않는다고 판단되는 종양 주변부 15mm 외부의 영역들은 모두 제거한 채 생성되었기 때문에 모델의 판단 시에도 패치 상에서 바라보는 영역이 영상 패치에서보다 훨씬 집중된 모양을 확인할 수 있었다. 형상 집중 종양 내부 패치에서도 종양의 외부는 모두 검정 영역으로 제거되었음에 따라 영상 패치에서보다 모델이 바라보는 영역이 훨씬 집중되었으며, 형상 집중 종양 내부 패치와 coupled intra and peritumoral patch 에서 모델이 검정 픽셀이 아닌 모든 픽셀을 골고루 바라보며 학습하기 보다는 종양의 경계 혹은 주변부의 일정 좁은 영역에 초점을 맞춰 예측을 수행한다는 것이 확인되었다. 시각화 결과 전반에서 ResNet152 가 EfficientNet-b7 보다 상대적으로 더 넓은 영역을 바라보며 분류를 수행하고 있었다.

4. 결론

본 연구에서는 비소세포폐암 환자의 수술 전 흉부 CT 영상을 통해 수술 후 2 년 이내 재발 여부를 예측할 수 있도록 총 다섯 종류의 종양 패치와 두 개의 사전 학습 된 CNN 네트워크를 사용하였다. 또한 각 CNN 모델 별로 단일 모델과 앙상블 모델에서의 실험을 진행하고 그 결과를 다섯 가지 성능 평가 지표로 제시하여 각 패치의 유용성을 분석하였다. 단일 모델과 구성 방식을 다양화 한 앙상블 모델들에서 각각 우수한 성능을 보인 경우에 대한 시험 셋에서의 분류 결과와 특성 중요도를 시각화한 Grad-CAM 을 사용하여 오분류 패치의 사례와 그 원인을 분석하였다. 향후 연구 방향으로

불규칙한 종양의 모양이나 간유리 음영이 포함된 영상 등의 오분류 경우를 더 다양한 사례를 통해 학습하고, Grad-CAM 을 통해 확인한 예측에 영향을 주는 요인을 분석하여 오분류 되는 특성을 더욱 집중적으로 학습할 수 있도록 하는 영상의 전처리 방법 및 최적화 패치를 모색하여 예측 모델의 성능을 향상시킬 수 있을 것으로 보인다. 또한 ResNet 과 EfficientNet 외에도 VGG, DenseNet 등과 같이 깊은 층을 갖는 CNN 네트워크들을 통해 여러 모델들이 동일하게 오분류 하는 패치들의 특성을 파악하여 그에 대한 패치 개선점을 모색해 볼 수 있을 것으로 보이며, 비소세포폐암 환자의 데이터를 추가로 확보하여 더 많은 데이터에 대한 학습을 통해 낮은 특이도를 극복해 볼 수 있을 것으로 기대된다.

감사의 글

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원과(No. 2020R1A2C1102140) 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학지원사업의 연구결과로 수행되었음(2016-0-00022).

References

- [1] Korea Central Cancer Registry, National Cancer Center, "Annual report of cancer statistics in Korea in 2018." Ministry of Health and welfare, 2020.
- [2] C. Haarbuerger, P. Weitz, O. Rippel and D. Merhof, "Image-Based Survival Prediction for Lung Cancer Patients Using CNNs." 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1197-1201, 2019.
- [3] Ye-Sel Lee, A-Hyun Cho and Helen Hong, "Ensemble Learning Based on Tumor Internal and External Imaging Patch to Predict the Recurrence of Non-small Cell Lung Cancer Patients in Chest CT Image." Journal of Korea Multimedia Society, 24(3), 373-381, 2021.
- [4] Thanh-Hung Vo, Guee-Sang Lee, Hyung-Jeong Yang and In-Jae Oh. "Survival Prediction of Lung Cancer Using Small-Size Clinical Data with a Multiple Task Variational Autoencoder." Electronics 10, no. 12: 1396, 2021.
- [5] Tai H Dou, Thibaud P Coroller, Joost J M van Griethuysen, Raymond H Mak and Hugo J W L Aerts, "Peritumoral radiomics features predict distant metastasis in locally advanced NSCLC." PloS one vol. 13,11 e0206108, 2018.
- [6] Hansang Lee, Haeil Lee, Helen Hong, Heejin Bae, Joon Seok Lim and Junmo Kim, "Classification of focal liver lesions in CT images using convolutional neural networks with lesion information augmented patches and synthetic data augmentation." Medical physics, 10.1002/mp.15118, 2021.
- [7] Tetsuro Baba, Hidetaka Uramoto, Masaru Takenaka, Souichi Oka, Yoshiki Shigematsu, Hidehiko Shimokawa, Takeshi Hanagiri and Fumihiko Tanaka, "The tumour shape of lung adenocarcinoma is

related to the postoperative prognosis." Interactive cardiovascular and thoracic surgery vol. 15, 1: 73-6, 2021.

- [8] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.
- [9] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks." 36th Int. Conf. Mach. Learn. ICML, vol. 2019-June, pp. 10691-10700, 2019.
- [10] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh and Dhruv Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE international conference on computer vision, pp. 618-626, 2016.
- [11] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva and Antonio Torralba "Learning deep features for discriminative localization." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2921-2929, 2016.
- [12] Erico Tjoa and Cuntai Guan "A survey on explainable artificial intelligence (XAI): towards medical XAI." pp. 1-21, 2020
- [13] Hualong Yu, Shihe Liu, Chuanyu Zhang, Shaoke Li, Jianan Ren, Jingli Zhang and Wenjian Xu, "Computed tomography and pathology evaluation of lung ground-glass opacity." Experimental and Therapeutic Medicine vol. 16, 5305-5309, 2018.

〈 저 자 소 개 〉

마 세 리

- 2022년 2월 서울여자대학교
소프트웨어융합학과 졸업(학사)
- 관심분야 : 의료영상처리, 컴퓨터비전, 딥러닝
- <https://orcid.org/0000-0003-2893-1969>



안 가 희

- 2022년 2월 서울여자대학교
소프트웨어융합학과 졸업(학사)
- 관심분야 : 의료영상처리, 딥러닝, 인공지능,
영상정합, 영상분할
- <https://orcid.org/0000-0002-5162-3481>



홍 헬 렌

- 1994년 2월 이화여자대학교 전자계산학과
졸업(학사)
- 1996년 2월 이화여자대학교 전자계산학과
졸업(석사)
- 2001년 8월 이화여자대학교 컴퓨터학과
졸업(박사)
- 2001년 9월~2003년 7월 서울대학교
컴퓨터공학부 BK 조교수
- 2006년 3월~현재 서울여자대학교
소프트웨어융합학과 교수
- 관심분야 : 의료 인공지능, 딥러닝, 영상처리 및
분석
- <https://orcid.org/0000-0001-5044-7909>

