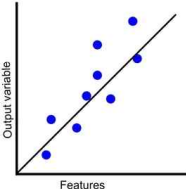
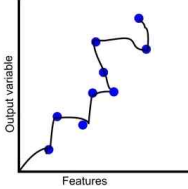


2024 경영경제대학 학술제 참가보고서

제 목	고막 내시경 데이터를 이용한 귀 질환 진단 서비스 (팀명 : CLEAR )				
신청팀 인적사항	성명	학번	생년월일	연락처	E-mail
	강하람	20212049	20021017	010-4636-7634	haramyl@naver.com
	김기남	20191970	20000201	010-2462-8433	asd53214@naver.com
	좌대현	20196783	20010301	010-2619-1099	dhyun0301@naver.com
선정주제	고막 내시경 데이터를 이용한 귀 질환 진단 서비스				
주제 선정 이유 (사회적 현상과 관련지어 서술)	<p>최근 사람들의 이어폰 사용 시간이 증가함에 따라 어지럼증, 이명, 난청과 같은 귀 질환 환자가 꾸준히 늘어나는 추세이다.</p> <p>귀가 간지럽고 불편할 때 전문가의 진단을 받는 데에는 시간과 비용이 많이 소요되어, 간단한 통증에는 그냥 넘어가는 경우가 종종 있다.</p> <p>이러한 상황 속에서 누구나 집에서 간편하게 이용할 수 있는 귀 질환 진단 서비스가 있다면 사람들의 귀 건강에 큰 기여를 할 것이라 판단하였다.</p>				
주제 분석 및 연구	<p><b>1. Image classification (이미지 분류)</b></p> <p>이미지 분류는 컴퓨터 비전 분야에서 가장 대표적인 문제 중 하나로 입력된 이미지가 어떠한 클래스에 속하는지 분류하는 문제이다. 이미지 분류를 수행하기 위해 일반적으로 인공 신경망을 이용한다. 인공 신경망을 지나며 입력 이미지에서 특징을 추출하고 이를 이용하여 이미지를 분류하게 된다.</p> <p><b>1.1 Convolutional Neural Network (CNN, 합성곱 신경망)</b></p> <p>이 이미지 분류에 대표적으로 이용되는 인공 신경망이 바로 CNN이라 불리는 합성곱 신경망이다. 합성곱 신경망은 단순한 선형 퍼셉트론과 같이 연산을 수행하는 것이 아니라 가중치 행렬(필터)을 가지고 도장을 찍듯 연산을 수행해 특징을 추출하는 방법이다. 이 방법은 다층 퍼셉트론 구조보다 파라미터 수가 적지만, 공간적인 구조를 고려하기에 이미지 인식 분야에 뛰어난 성능을 발휘한다.</p> <p><b>1.2 Overfitting (과적합)</b></p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <p>Optimal</p>  </div> <div style="text-align: center;"> <p>Overfitting</p>  </div> </div> <p style="text-align: center;">&lt;그림 1. 과적합&gt;</p> <p>과적합(Overfitting)은 머신러닝에서 모델이 훈련 데이터에만 너무 맞춰져서 일반적인 현상을 잘 설명하지 못하는 현상을 말한다. 과적합은 모델이 너무 복잡하거나 훈련 데이터에 비해 너무 많은 파라미터를 가질 때 발생할 수 있다. 따라서 과적합 방지를 위해서 이미지 분류 문제에서 훈련 데이터의 크기는 무엇보다 중요하다.</p> <p><b>2. 의료 데이터 활용의 어려움</b></p> <p>의료 관련 데이터 활용에는 다양한 어려움이 있다. 첫째는 의료 데이터의 특징에 따른 폐쇄성이다. 의료 데이터는 환자의 신체에 관한 개인정보이기 때문에 환자의 동의를 구하고, 일정 부분 개인정보를 삭제한 뒤에 공급이 된다. 그마저도 구하기 어려운 것이 현실이다. 두 번째는 병원마다 데이터가 표준화 되어있지 않다는 점이다. 병원 데이터의 용어 및 코드가 각각 다르며, 영상 이미지 데이터의 경우에도 촬영된 환경, 장비 등의 차이가 있어 각 병원에서 수집한 데이터가 일관되지 않는다. 우리는 연구를 통해 앞서 말한 의료 데이터의 두 가지 어려움을 해결하고자 했다.</p>				

### 3. 사용 데이터 - Özel Van Akdamar Hospital Eardrum dataset (2020)



<그림 2. 사용 데이터>

Özel Van Akdamar Hospital Eardrum dataset은 터키의 Özel Van Akdamar 병원에서 2018년 10월에서 2019년 6월까지 자원한 입원 환자들로부터 수집한 956개의 고막 내시경 이미지로 구성되어 있다. 정상(Normal) 535개, 급성 중이염(Acute Otitis Media, AOM) 119개, 만성 화농성 중이염(Chronic suppurative Otitis Media, CSOM) 63개, 귀지(Earwax) 140개 등으로 분류되어 각 데이터는 카테고리별로 라벨링되어 있다. 우리 연구의 목표는 급성 중이염과 외이도염에 대해 선제적으로 판별하는 것이기에 데이터를 AOM(급성 중이염), OE(외이도염)를 묶어 Abnormal 클래스로 규정했다. 그렇게 생성한 Abnormal 클래스와 Normal 클래스 간의 이진 분류 문제를 해결하고자 하였다.

### 4. 모델 학습

#### 4-1. 모델 선택

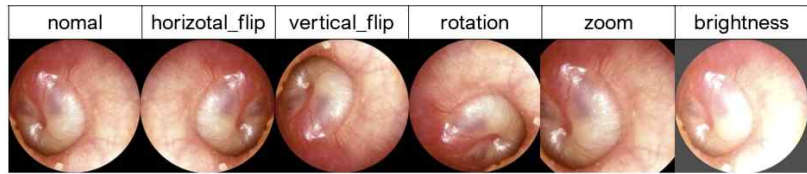
	val accuracy	# of parameter
VIT	0.8208	86,195,456
VGG16	0.7453	21,137,986
Xception	0.8208	21,386,538
CBAM + Residual Block	0.7264	606,080

<그림 3. 모델 성능 비교>

바탕이 되는 모델을 고르기 위해 다양한 사전학습모델의 성능을 비교해 보았다. 이미지의 개수가 적기 때문에 과적합 방지와 완전한 학습을 위해서는 파라미터 수가 작으면서 좋은 성능을 내는 모델을 이용하는 것이 중요했다. 사전학습 모델인 VGG16, Xception 그리고 최근 이미지 분류에서 높은 성능을 보이는 VIT 모델을 학습시켜 평가해 보았다. 추가로 CBAM 블록과 Residual 블록을 조합하여 파라미터가 작은 모델을 설계하여 평가해 보았다. 그 결과, val accuracy가 가장 높으면서 가장 적은 파라미터 수로 구성된 Xception을 baseline 모델로 이용하기로 결정했다.

#### 4-2 데이터 증강 (Data Augmentation)

데이터 증강 기법은 가지고 있는 데이터를 적절하게 변형시켜 새로운 데이터를 생성하는 과정을 말한다. 데이터 증강을 통해 머신러닝 모델의 성능을 향상하고 과적합을 방지할 수 있다. 적용한 증강기법으로는 rotation(회전), horizontal flip(수평 뒤집기), vertical flip(수직 뒤집기), zoom(확대), brightness(명도 조절)가 있다. 증강을 적용하여 모델을 학습시킨 결과 val loss는 0.5326에서 0.2960으로 감소하였고 val accuracy는 0.8208에서 0.8585로 개선되었다.



<그림 4. 데이터 증강>

#### 4-3 Stratified K-fold CV

K-fold Cross Validation(K-fold 교차검증)을 도입하였다. K-fold CV는 데이터를 K개의 폴드로 분할하고 각각의 폴드를 한 번씩 validation set으로 사용하고 나머지 폴드를 train set으로 사용하는 방식이다. 이 방식을 이용하면 적은 데이터에서도 효과적으로 모델을 학습시킬 수 있으며 모델의 분산을 줄이고 데이터 변동에 견고한 모델을 만들 수 있다.

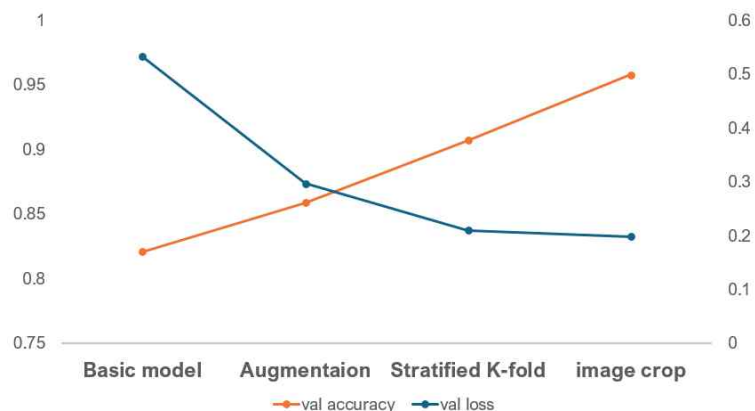
다만 불균형이 존재하는 우리의 데이터 특성을 고려하여 Stratified K-fold를 사용하였다. Stratified K-fold는 각 fold 안의 클래스 비율이 전체 data set과 동일하도록 보장하는 K-fold 교차검증의 한 방법이다. 즉, 특정 클래스가 한곳으로 몰리는 것을 방지하여 불균형한 dataset에서 모델의 성능을 향상할 수 있다. 모델에 Stratified K-fold를 적용하니 val loss가 기존 0.2960에서 0.2093으로 낮아지고 val accuracy는 0.8585에서 0.9068로 향상되는 것을 확인하였다.

#### 4-4 Cost-Sensitive Learning

Cost-sensitive Learning은 클래스 불균형을 해소하기 위한 방법 중 하나로 모델을 학습할 때 클래스별 loss(weight) 값을 다르게 주어 학습하는 방법이다. Loss 설계 시 더 수가 많은 클래스의 loss에 penalty를 주거나, 개수가 더 적은 클래스에 weight를 더 주어 모델의 일반화 성능을 높일 수 있다. 이를 적용하였지만, validation set에 대한 성능이 val accuracy 0.9153, val loss 0.2614로 더 좋아졌다고 보기 어려워 사용하지 않기로 결정하였다.

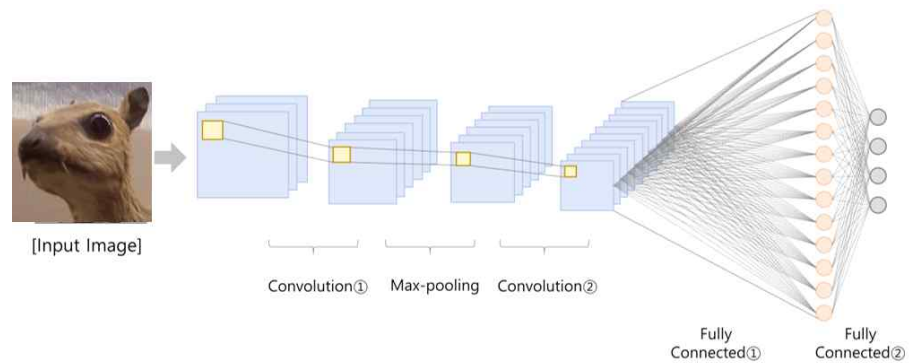
#### 4-5 Center Crop

추가적인 성능향상 방안을 고민하던 중 귀 내시경 사진 이미지의 비어있는 테두리 부분을 줄여 유효한 픽셀값에 더 집중한다면 성능 향상이 될 것으로 가설을 세우고 center crop을 통해 실행하였다. center crop은 이미지의 중심 부분을 기준으로 이미지를 잘라내는 작업이다. 기존 이미지의 0.8 비율만큼을 잘라내어 데이터를 재구성하였다. 그 결과, val loss는 0.2093에서 0.1977로 감소하고 val accuracy는 0.9068에서 0.9576으로 증가하였다. 따라서 이것을 최종 모델로 선정하였다.



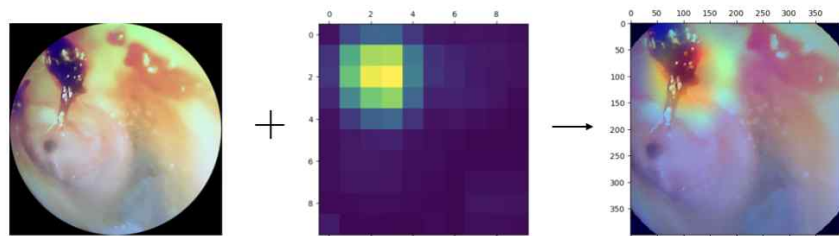
<그림 5. 성능 지표 변화>

#### 4-6 Grad-CAM(Gradient-weighted Class Activation Mapping)



<그림 6. Grad-CAM의 구조>

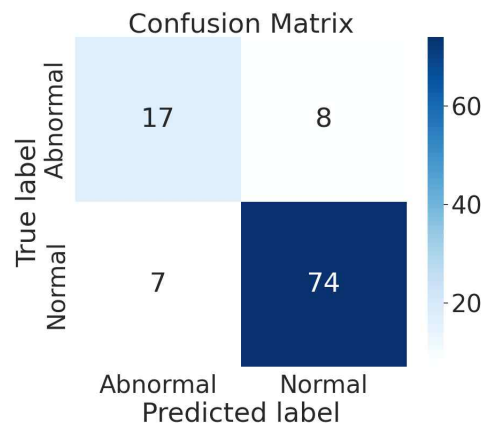
Grad-CAM은 모델의 마지막 컨볼루션 레이어의 특성맵(feature map)에 대한 gradient 정보를 사용하여, 특정 출력 클래스에 대한 각 특성맵의 중요도를 평가하는 방법이다. Grad-CAM을 사용하여 얻은 활성화 맵은 모델이 예측하는 데 있어 가중치를 둔 영역을 heatmap 형태로 보여준다. Grad-CAM을 사용하면 모델이 어느 부분을 보고 특정 클래스로 분류하였는지 시각적으로 확인할 수 있는 것이다. 아래는 Abnormal로 판정된 외이도염 이미지에 대해 Grad-CAM을 출력한 것이다. 염증으로 예상되는 영역에 잘 집중하고 있음을 추론해볼 수 있다.



<그림 7. Grad-CAM 적용 예시>

### 5. 평가

#### 5-1. test set



<그림 8. Confusion Matrix>

	test accuracy	test loss	test precision	test recall	test f1
Final Model	0.8585	0.4062	0.8567	0.8585	0.8575

<그림 9. test set 성능>

		<p>test set에 대해 최종 선정된 모델의 평가 성능을 혼동행렬 및 표로 정리한 것이다. 이에 따르면 이 모델은 약 86%의 정확도를 보인다고 할 수 있다.</p> <p>이때 precision, recall, f1 score는 각 클래스에 속하는 표본 개수로 가중 평균한 weighted avg 값이다.</p> <p><b>5-2. 실제 귀 내시경 촬영 이미지 활용</b></p> <p>우리 서비스의 실현을 위해 우리는 직접 가정용 귀 내시경을 구매해 해당 귀 내시경으로 고막 사진을 촬영한 후 모델에 넣어서 위험도를 판별해 보았다. 그리고 동시에 우리는 이비인후과 전문의에게 해당 사진들에 대한 진단을 의뢰했다. 그 결과는 다음과 같다.</p> <table><tr><td>Picture ID</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr><tr><td>Predict class</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>True class*</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table> <p style="text-align: right;">*True class는 이비인후과 전문의가 직접 판정한 클래스</p> <p style="text-align: center;">&lt;그림 10. 가정용 귀 내시경으로 촬영한 사진 진단&gt;</p> <p>총 10개의 실제 사진 중 7개를 이비인후과 전문의와 같은 클래스로 예측하는 데에 성공했다. 이는 우리의 모델이 우리의 목표처럼 실제 가정용 귀 내시경으로 촬영한 사진에도 실질적으로 적용된다는 것을 보여준다.</p>	Picture ID	1	2	3	4	5	6	7	8	9	10	Predict class	0	1	0	0	1	0	0	1	0	0	True class*	0	0	0	0	0	0	0	0	0	0
Picture ID	1	2	3	4	5	6	7	8	9	10																									
Predict class	0	1	0	0	1	0	0	1	0	0																									
True class*	0	0	0	0	0	0	0	0	0	0																									
현실성 및 지속가능성	주제의 현실 적용 가능성	직접 가정용 귀 내시경을 이용해 촬영한 사진을 모델로 판별 가능하다는 것까지 확인했다는 점에서 추가적인 개선을 거치면 서비스로도 출시할 수 있다는 것을 보여준다. 따라서 주제의 현실 적용 가능성은 매우 높다고 할 수 있다.																																	
	주제의 발전 가능성 및 타당성	사용한 dataset의 양이 너무 적고, 품질도 좋지 않았다. 충분한 양과 좋은 품질의 dataset이 주어진다면 훨씬 높은 정확도의 모델을 도출해 낼 수 있기 때문에 발전 가능성이 매우 높다고 볼 수 있다.																																	
창의성 및 전문성	주제에 대한 분석의 차별성 및 독창성	딥러닝에서 많이 나타나는 클래스 불균형, 데이터 부족 등으로 인한 일반화 능력 부족 문제를 다양한 방법으로 해결했다는 점에서 차별성이 있다. 다양한 증강, Stratified K-fold CV, Cost-sensitive Learning 같은 다양한 기법을 통해 문제를 해결하고 test set에 대한 진단 정확도가 85%로 좋은 성능을 보였다. 또한 학습용 이미지와 다르게 가정에서 귀 사진을 찍을 경우 빛으로 인한 밝기에서 큰 차이가 있을 것으로 생각하여 모델 학습 시 이미지의 밝기를 조절하여 학습하여 가정에서도 성능에 큰 차이가 없도록 하였다.																																	
	의견 도출의 논리 및 과정	데이터 증강 및 Stratified K-fold CV 등을 추가하는 과정에서 새로 추가된 요인 외의 나머지 요인들을 엄밀히 통제하여 성능향상을 이뤄냈다는 점에서 의견 도출의 논리성이 입증된다고 생각한다.																																	
	전공융합 적절성	통계학습개론, 데이터 마이닝 등에서 배운 K-fold CV에 대해 추가적인 탐구를 진행하였고 클래스 불균형 문제에 더 적합한 Stratified K-fold CV를 적용하였다. 이를 통해 데이터 부족 및 클래스 불균형 문제를 동시에 완화할 수 있었다. 추가로 딥러닝 전공 수업에서 배웠던 다양한 이미지 증강 기법, 딥러닝 모델을 활용하여 어떤 환경에서 찍은 사진이든 잘 분류할 수 있게 일반화 능력도 높이고자 하였다.																																	
의의 및 기대효과		<p>귀 불편함을 느끼는 사람들에게 간편하며 자가적인 귀 질환 진단 서비스를 제공할 수 있다는 점에서 의의가 있다.</p> <p>영유아기에 쉽게 노출되는 귀 질환(중이염, 외이도염), 귀가 건강한지 걱정될 때, 이어폰을 많이 사용하여 귀 질환이 의심되는 대학생, 병원에 가야 할지 고민될 때, 의료 서비스 접근성이 낮은 개발도상국, 적절한 의료서비스를 받지 못할 때, 의료 서비스의 가격이 너무 높은 선진국, 병원에 가야 할지 고민될 때, 가정용 귀 내시경과 우리의 서비스를 체온계처럼 구비해두고 간편하고 신속하게 사용한다면 CLean EAR를 유지할 수 있을 것으로 기대한다.</p>																																	