

2020년도  
학사학위논문

# 딥러닝 기반 귀내시경 영상을 이용한 중이 병증 진단

Deep learning-based middle ear disease  
diagnosis with otoendoscopy images

2020년 11월 9일

순천향대학교 공과대학  
컴퓨터공학과

신지환

딥러닝 기반 귀내시경 영상을 이용한  
중이 병증 진단

Deep learning-based middle ear disease  
diagnosis with otoendoscopy images

지도교수 남 윤 영

이 논문을 공학사학위 논문으로 제출함

2020년 11월 9일

순천향대학교 공과대학  
컴퓨터공학과

신지환

신 지 환 의 공학사학위논문을 인준함

2020년 11월 9일

심 사 위 원 천 인 국 인

심 사 위 원 남 윤 영 인

순천향대학교 공과대학

컴퓨터공학과

## 초 록

본 논문은 딥러닝(CNN)을 통해 중이 이미지 데이터 셋을 기반으로 중이 질환 분류 방법 그리고 이를 제공할 수 있는 웹서비스에 대해 소개한다. 연구에 사용된 이미지는 순천향대학교 천안병원 이비인후과에서 제공받았다. 본 연구는 순천향대학교 천안병원 연구윤리심의위원회(Institutional review board, 이하 IRB)의 심의를 거쳐 대상자의 의무기록을 후향적으로 조사하였다(IRB No. SCHCA 2020-02-022). 카테고리는 Normal, Traumatic Perforation, OME, Congenital Cholesteatoma, COM, AOM 여섯 개로 구분하였다. 데이터 셋을 Mask R-CNN을 통해 ROI(Region of Interest)를 추출한 후 Shift, Zoom, Rotate 등 여러 기법으로 증량시켜 데이터 셋을 재구성하였고 이를 5-Fold로 훈련을 시켰다. 모델은 EfficientNetB0과 Inception-V3를 배깅을 통하여 구성하였다. 그리고 테스트 셋으로 예측을 하였을 때 정확도는 97.29%였다. 원본 데이터 셋과 비교하였을 때 약 3.5%정도 우월하였다. 그리고 이를 Python Flask를 사용하여 웹으로 구성하여 이미지를 예측하는데 걸리는 시간이 약 3초 이내인 사이트를 구축하였다.

키워드 : CNN, EfficientNet-B0, Inception-V3, MASK R-CNN, 종이염, 분류,  
웹

## ABSTRACT

This paper introduces the mechanism classifying the type of disease based on image data sets produced by CNN. The images used in the study is given by Soonchunhyang Cheonan hospital otorhinolaryngology. The study studied subject's medical record retrospectively after being reviewed by IRB of Soonchunhyang Cheonan hospital. The categories were classified by six classes: Normal, Traumatic Perforation, OME, Congenital Cholesteatoma, COM, AOM. After applying Mask R-CNN to detect ROI(Region of Interest), the data set was reconstructed by using the mechanism such as Shift, Zoom and Rotate of augmentation and trained the data set by using 5-Fold. The model was constructed by using EfficientNetB0 and Inception-V3. And the accuracy of the test prediction was 97.29% proceeded by test set. The data set was superior to the original data set by 3.5%. and therefore, it took approximately a second to construct the web site by using Python Flask to predict images.

keyword : CNN, EfficientNet-B0, Inception-V3, MASK R-CNN, Otitis Media, Classification, Web

# 차 례

제 1 장 서 론.....	1
제 2 장 선행연구 및 관련 기술 분석.....	3
2.1 종이 병증 분류.....	3
2.2 관련 기술 사례.....	5
제 3 장 시스템 설계.....	9
3.1 개발 환경.....	9
3.2 데이터 셋.....	10
3.3 전처리.....	11
3.3.1 Mask R-CNN.....	11
3.3.2 데이터 증량.....	14
3.4 분류.....	15
3.4.1 분류 모델 Architecture.....	16
3.5 Web Server.....	17
3.5.1 UI(User Interface).....	16
제 4 장 실험 및 결과.....	21
4.1 분류 결과.....	21
4.2 실험 결과 비교.....	23
4.3 웹 실행 속도 실험.....	24
제 5 장 결론 및 향후 과제.....	25
참고문헌.....	26
감사의 글.....	29

## 그 림 차 례

[그림 1] Deep Learning for Detection of Diabetic Eye Disease, 2016-11-295	
[그림 2] 뷰노 .....	6
[그림 3] K-Fold 교차 검증 .....	8
[그림 4] 고막 이미지 .....	10
[그림 5] VGG Image Annotaor .....	11
[그림 6] Mask R-CNN architecture .....	11
[그림 7] Mask R-CNN 결과 .....	12
[그림 8] 고막 영역 추출 결과 이미지 .....	13
[그림 9] 고막이 검출 되지 않은 사진 .....	13
[그림 10] 데이터, 모델 구성도 .....	15
[그림 11] 웹 프로우 차트 .....	17
[그림 12] 초기 화면 .....	18
[그림 13] 회원가입 화면 .....	18
[그림 14] 이미지 업로드 화면 .....	19
[그림 15] 대기 화면 .....	19
[그림 16] 결과 화면 .....	20
[그림 17] 결과 그래프 .....	23

## 표 차 례

[표 1] 개발 환경 사양.....	9
[표 2] Mask R-CNN을 통한 고막 검출 결과.....	12
[표 3] 데이터 셋 크기.....	14
[표 4] EfficientNet-B0 Baseline Network.....	16
[표 5] Inception-V3 Baseline Network.....	16
[표 6] 원본 데이터 셋을 통해 얻은 실험 결과.....	21
[표 7] 고막 영역만 저장하여 재구성한 데이터 셋을 통해 얻은 실험 결과.....	21
[표 8] Confusion Matrix.....	22
[표 9] Result of Precision, Recall, F1-score.....	22
[표 01] 처리 속도 결과.....	24



## 제 1 장 서 론

중이염은 3세 이하에서 3명 중 2명 비율로 1회 이상 앓게 되고, 3명 중 1명의 비율로 3회 이상 앓는 매우 흔한 질환이다. 국내에서 급성중이염에 대한 발병률은 정확하지 않으나 외국의 보고에 의하면 생후 1세까지 62%, 생후 3세까지 83%가 최소 1회 이상 걸린다고 하였다. 국내에서는 중이염의 발병률에 대한 전국 규모의 연구에서 15세 미만 대상군에서 급성중이염은 0.08%, 삼출성중이염은 1.22%의 유병율이 보고된 바 있다 (Kim et al., 1993). 2008년 국민건강보험 심사평가원의 통계자료[1]에 따르면, 중이염은 10세 미만의 환자들이 의사를 찾는 빈도에서 10위, 그리고 병,의원을 찾는 빈도에서 6위를 차지한다. 유소아 중이염은 다른 상기도감염과는 달리 전문적인 의학적 지식과 더불어 적절히 치료되지 못하였을 때 발생할 수 있는 합병증 및 후유증 동반될 수 있다고 한다.

본 논문에서는 유소아에게 많이 발병하는 중이염과 그로 인한 합병증과 후유증을 방지하기 위해 조기 진단을 위한 딥러닝 기반 웹 서비스를 소개하려 한다. 본 연구는 순천향대학교 천안병원 이비인후과와 협력하여 연구하였고 본 연구는 순천향대학교 천안병원 연구윤리심의위원회(Institutional review board, 이하 IRB)의 심의를 거쳐 대상자의 의무기록을 후향적으로 조사하였다(IRB No. SCHCA 2020-02-022). 병증은 Normal, AOM, COM, Congenital cholesteatoma, OME, Traumatic Perforation 총 6개의 카테고리로 나누었고 4808장의 데이터 셋을 가지고 연구를 하였다. 또한 단일 CNN 모델이 아닌 앙상블 모델과 전처리 과정으로 고막 영역을 추출하는 MASK R-CNN 모델을 더하여 정확도를 향상하

였다. 이를 통해 상대적으로 작은 데이터 셋에 전처리 과정이 필요하다는 점과 앙상블 모델의 정확도 향상에 대해 제안 하고 진단에 소요 시간이 3초 이내인 웹 서비스를 소개한다.

본 논문의 구성은 2장에서는 관련 연구 및 기술 분석, 3장에서는 시스템 설계 및 딥러닝을 통한 분류 모델에 대한 설명 그리고 연구 방법, 웹서비스에 대해 소개 하고 4장에서는 결론을 보여준다.

## 제 2 장 선행연구 및 관련 기술 분석

### 2.1 중이 병증 분류

중이염은 증상에 따라 외상으로 인한 천공(Traumatic perforation), 삼출성 중이염(Otitis media with effusion, OME), 선천성 중이 진주종(Congenital cholesteatoma, CC), 만성 중이염(Congenital cholesteatoma, COM), 급성 중이염(acute otitis media)으로 구분할 수 있다[2].

외상으로 인한 천공은 고통, 출혈, 청력 저하, 이명, 어지러움을 유발 할 수 있는 중이염의 일종으로써, 중이로 어떠한 물건이 삽입되거나, 폭발, 뺨을 때리는 행동, 머리 외상, 갑작스러운 감압, 기압 장애 등으로 인해 발생한다. 중이 내부의 기관이 보이는 천공, 이 천공으로 인해 발생하는 출혈을 검이경을 통해 진단한다[3].

삼출성 중이염은 박테리아나 바이러스에 의한 감염으로 인해 발생한 중이염을 의미하며, 삼출물을 동반한다는 특징이 있다. OME는 주로 아주 어린 영유아에게 빈번하게 발생하며 삼출물로 인해 고막이 팽창되거나, 삼출물로 인해 고막의 압력이 증가하여 고막에 구멍이 나고, 이 구멍을 통해 삼출물이 흘러나오는 증상 등이 발생한다[4].

선천성 중이 진주종은 귀가 없거나, 형성되지 않거나, 출산 시에 불완전하게 개발된 것을 의미하는데, 고막과 외이도가 형성되지 않아서 청력 장애를 겪는 경우, 외이가 형성되지 않은 경우, 귓바퀴가 형성되지 않은 경우 등이 존재한다. 본

논문에서는 고막의 형성이 덜 된 경우만을 고려했다[5].

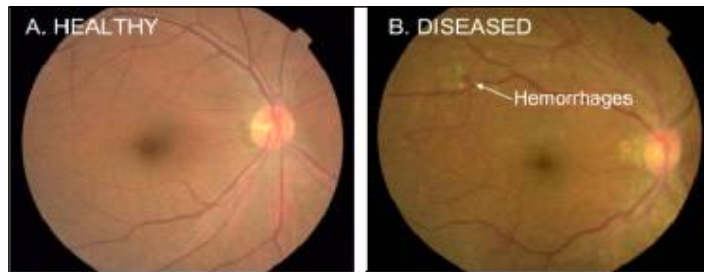
COM은 화농성 중이염, Traumatic perforation이 6주 이상 지속되었을 때를 의미한다. 이 종류의 중이염은 고통을 동반하지 않은 이루와 전도성 청력 손상을 포함하며, 용종, 담지종 등 다른 감염을 포함한다. 이 경우, 이루의 누출이 관찰되고 진단하기 위해 CT와 MRI를 이용한 진단이 요구된다[6].

AOM은 고막에 갑작스럽게 염증이 발생한 경우를 의미하는데, 빠른 주기의 고통, 발열, 액체가 발생한다. 영유아의 경우 울음, 과민성 반응, 수면 부족, 간호를 꺼리는 반응이 나타난다. 이 경우, 고막이 적색, 노란색으로 변화하게 되며 이를 검이경을 통해 진단한다. 성인의 경우 시술 및 항생제를 통해 치료하나 영유아의 경우 항생제만을 이용하여 치료한다[7].

## 2.2 관련 기술 사례

### 1) 구글 안저 사진 연구

세계적인 의학 저널인 JAMA, Ophthalmology, Nature 등에서 DR의 진단을 위해 딥러닝을 활용한 수많은 연구 결과들이 발표되었다[8]-[10],[13]. DR을 판독하기 위해 딥러닝을 적용한 연구 결과를 의학 전문 저널에 최초로 발표한 사례는 2016년 구글이 JAMA에 게재한 연구 논문이다[8]. 논문에서 제시한 성능은 AUC 0.99 수준으로, 전문 안과 의사의 실력과 버금가는 수준이었다.



[그림 1] Deep Learning for Detection of Diabetic Eye Disease, 2016-11-29

구글은 심층신경망의 하나인 InceptionV3에 12만 여 장의 안저영상을 입력하여 기계학습을 하였고, 학습한 모델은 EyePACS-1[10] 데이터 8,788장과 Messidor-2[12] 데이터 1,745장을 가지고 성능을 테스트하였다. 구글이 발표한 논문에서는 영상의 품질과 레이블링 품질에 대해서 상세한 자료를 제공하고 있다. 그러나, EyePACS-1 데이터베이스에서는 학습과 테스트 모두에서 데이터 표본을 뽑은 것으로 나와 있는데, 이때 상호 중복이 없이 표본을 뽑았다는 언급이 없는 것이 아쉬운 점이다. 또한, 테스트 데이터의 구성에서 정상과 질병의 분포 편차가 크다는 점도 성능평가 결과의 해석에서 고려할 사항으로 보인다.

## 2) 의료용 인공지능 진단 소프트웨어

뷰노[13]는 의료 분야에 인공지능(AI) 기술을 적용하여, 다양한 의료 데이터를 분석하고 질환 진단을 보조하는 의료 인공지능 솔루션 개발 기업이며 흉부 X-ray 영상 및 흉부 CT 영상 판독 보조, 안저 질환 판독 보조를 위한 의료 인공지능 솔루션 개발 및 상용화를 통해 인공지능 의료 영역을 넓혀가고 있다. 아래 [그림 2]는 뷰노의 실제 적용 사례 사진이다.



[그림 2] 뷰노

### 3) Flask

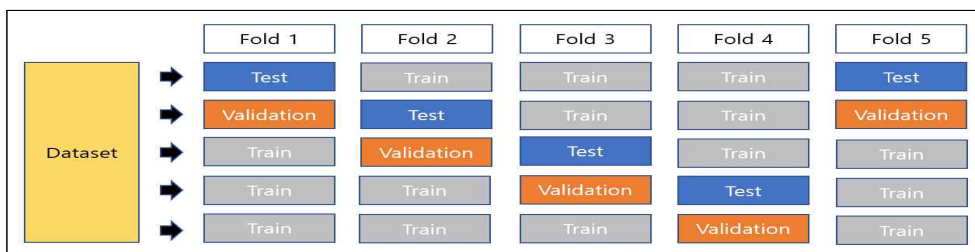
플라스크(Flask)[14]는 파이썬으로 작성된 마이크로 웹 프레임워크의 하나로, Werkzeug 툴킷과 Jinja2 템플릿 엔진에 기반을 둔다. BSD 라이선스이다. 특별한 도구나 라이브러리가 필요 없기 때문에 마이크로 프레임워크라 부른다. 데이터베이스 추상화 계층, 양식 유효성 확인, 기타 기존의 서드파티 라이브러리가 공통 기능을 제공하는 구성 요소가 없다. 그러나 플라스크는 플라스 자체에서 구현된 것처럼 애플리케이션 기능을 추가할 수 있는 확장 기능을 지원한다. 확장 기능은 객체 관계 매핑, 양식 유효성 확인, 업로드 관리, 다양한 개방형 인증 기술, 여러 공통 프레임워크 관련 도구들을 위해 존재한다.

본 논문에서는 웹 상에서 덤러닝 서버에 Python 코드를 실행시키기 위해 Flask를 사용하였다.

#### 4) K-Fold 교차 검증

K-Fold 교차 검증(Cross validation)[15]이란 통계학에서 모델을 평가하는 한 가지 방법이다. 소위 Held-Out Validation이라 불리는 전체 데이터의 일부를 Validation set으로 사용해 모델 성능을 평가하는 것의 문제는 데이터 셋의 크기가 작은 경우 테스트 셋에 대한 성능 평가의 신뢰성이 떨어지게 된다. 테스트 셋을 어떻게 잡느냐에 따라 성능이 다르면, 우연의 효과로 인해 모델 평가 지표에 편향이 생기게 된다.

이를 해결하기 위해 K-Fold 교차 검증은 모든 데이터가 최소 한 번은 테스트 셋으로 쓰이도록 한다. 아래의 그림을 보면, 데이터를 5개로 쪼개 매번 테스트 셋을 바꿔나가는 것을 볼 수 있다. 첫 번째 Iteration에서는 BCDE를 Train 셋으로, A를 Test 셋으로 설정한 후, 성능을 평가한다. 두 번째 Iteration에서는 ACDE를 Train 셋으로, B를 Test 셋으로 하여 성능을 평가한다. 그러면 총 5개의 성능 평가지표가 생기게 되는데, 보통 이 값들을 평균을 내어 모델의 성능을 평가하게 된다. 이 때, 데이터를 몇 개로 나누었는지가 K-겹 교차검증의 K가 된다. 본 논문에서는 5-Fold로 구성하였다.



[그림 3] K-Fold 교차 검증



## 제 3 장 시스템 설계

### 3.1 개발 환경

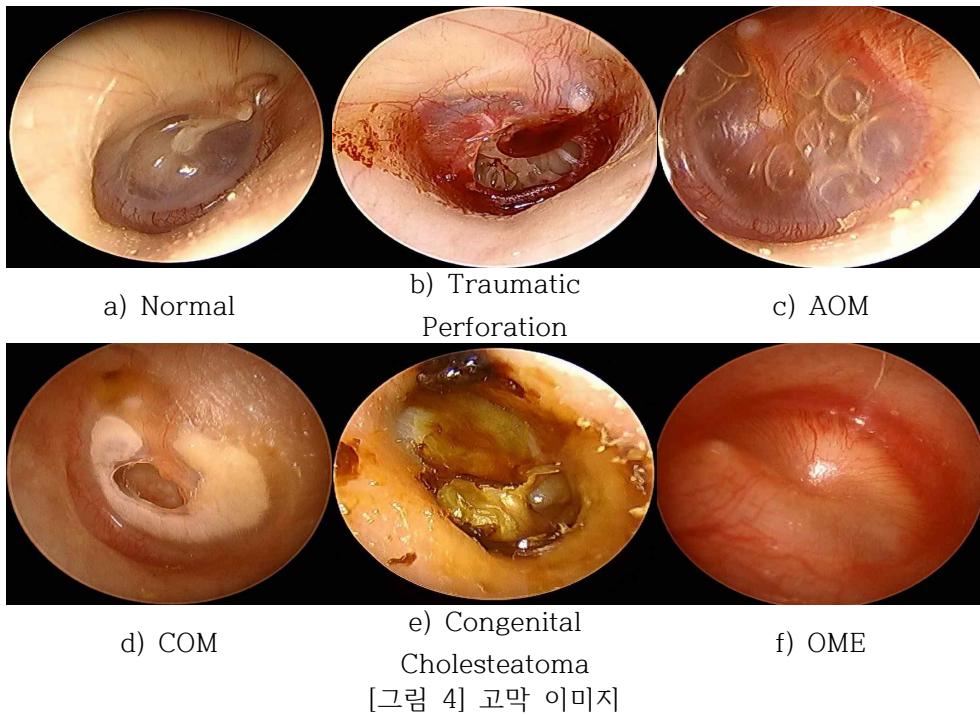
아래 [표 1]와 같은 사양으로 연구를 했다. Epoch는 1000으로 설정하였고 Early Stopping 기법을 통해 Validation Loss가 Epoch 20을 진행하는 동안 계속 증가하는 시점에 훈련을 중단토록 하였고, VGG19, Resnet50, Resnet101, Resnet152, InceptionV3, EfficientNetB0 모델에 적용 하여 비교하였으며, 소요 시간은 총합 70시간이었다.

[표 1] 개발 환경 사양

부품	사양
OS	Window 10 Education 64bit
CPU	Intel(R) Xeon(R) Silver 4114 CPU @ 2.2GHz (20EA)
RAM	192GB
GPU	NVIDIA TITAN RTX (3EA)

### 3.2 데이터 셋

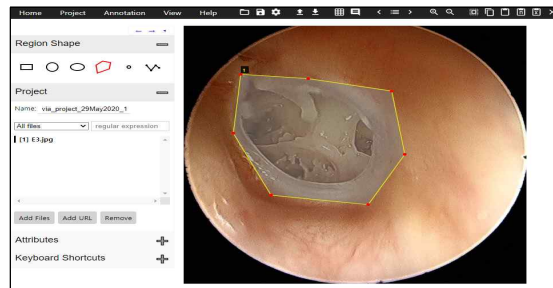
순천향대학교 천안병원과 협력하여 연구 하였고 병원이 제공한 총 이미지 개수는 4808장이었다. Normal, Traumatic Perforation, AOM, COM, Congenital Cholesteatoma, OME 총 6개의 카테고리를 갖는다. 아래 [그림 4]는 각 라벨에 대한 고막 이미지이다.



### 3.3 전처리

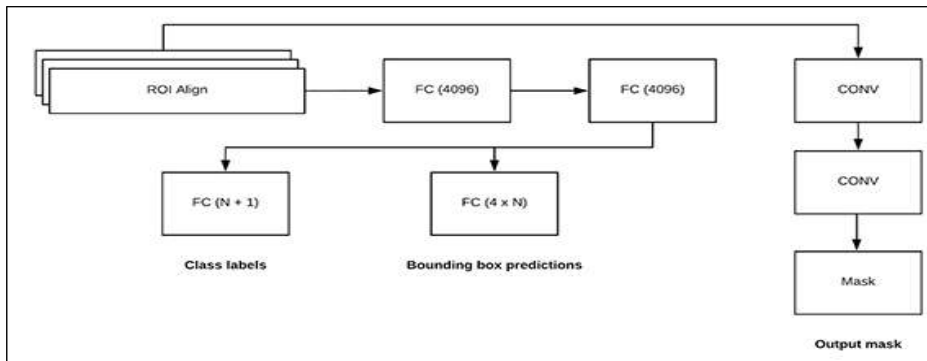
#### 3.3.1 Mask R-CNN[16]

고막 이미지에 귀지나 혹은 다른 이물질로 인해 고막이 가려져 있는 경우 딥러닝 시 정확도 저하로 이어졌다. 데이터 셋의 수가 적기 때문에 최대한 양질의 이미지를 얻는 것을 목표로 하였고 Mask R-CNN을 통해 이미지에서 고막 부분을 추출하여 시각화하였다. 훈련 전 아래 [그림 5]와 같이 VGG Image Annotator(VIA) Tool[17]을 사용하여 주석처리를 하였다.



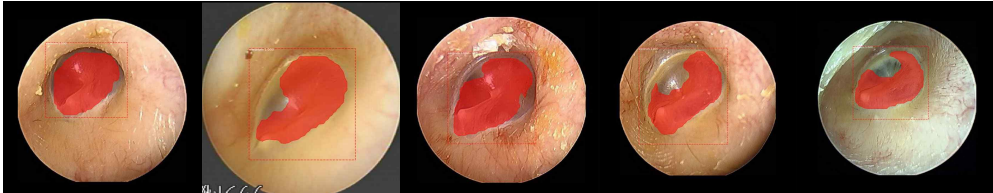
[그림 5] VGG Image Annotaor

그리고 Mask R-CNN의 모델 구성을 아래 [그림 6]으로 표현하였다.



[그림 6] Mask R-CNN architecture

Train Set은 309장, Validation Set은 101장으로 훈련을 하였고 아래 [그림 6]는 결과이다. Validation Set에서 고막 Detection 정확도는 99%를 얻었다.



[그림 7] Mask R-CNN 결과

그리고 테스트 셋으로 순천향대학교 천안병원에서 제공 받은 4808장의 데이터 셋에 적용하였다. 이 때 84.63%의 검출률을 나타냈으며 아래 [표 2]는 그 결과이다. 낮은 검출률의 이유는 Mask R-CNN의 훈련과 검증 셋을 정상 진단을 받은 고막 영상 데이터를 Annotation하였기 때문이다. 문제점으로는 상대적으로 정상 진단을 받은 영상 데이터 셋보다 병증 데이터 셋이 영상 품질이 낮았다. 여기서 품질이란 이미지 중 고막 영역의 밝기가 어둡거나, 화질이 낮아 고막이라고 육안으로도 확인할 수 없거나, 귀지가 고막을 가린 경우를 말한다. 테스트 셋에 적용하여 얻은 4254장의 고막 영역만 저장하여 데이터 셋을 재구성 하였다.

[표 2] Mask R-CNN을 통한 고막 검출 결과

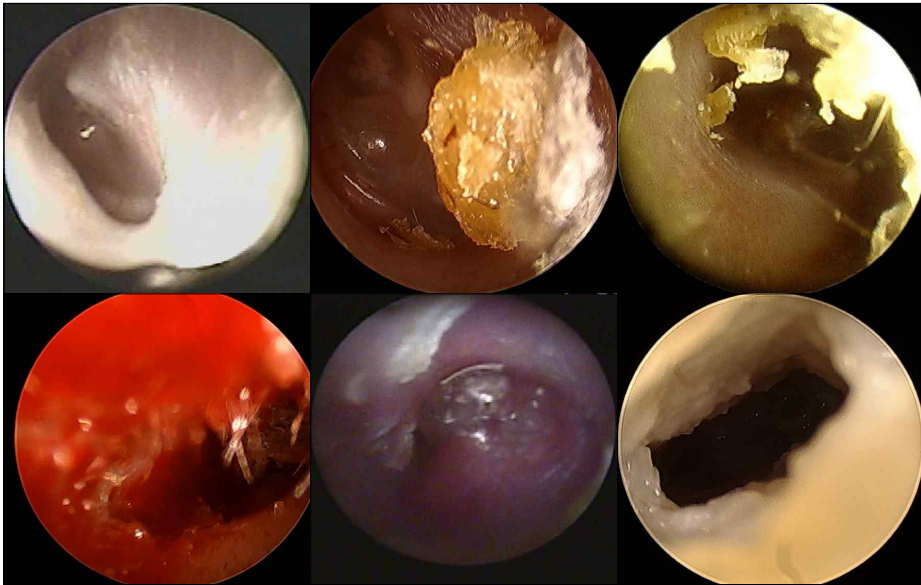
	원본(장)	고막 추출(장)
Normal	2,040	1,971
Traumatic Perforation	356	295
AOM	804	654
COM	824	698
Congenital Cholesteatoma	384	312
OME	400	324
합계	4,808	4,254

아래 [그림 7]는 Mask R-CNN을 통해 검출한 결과를 시각화 한 것이다.



[그림 8] 고막 영역 추출 결과 이미지

Mask R-CNN을 통해 화질이 좋지 않고 귀지로 고막이 가려진 경우에 필터링 효과를 얻을 수 있다. 아래 [그림 8]은 Mask R-CNN을 통해 검출이 안 된 고막 사진들이다. 이를 통해 자동으로 데이터 셋의 정제를 하였고 그 결과 획기적으로 빠른 속도로 노이즈 부분을 제거하고, ROI(Region of Interest)만을 추출할 수 있었다.



[그림 9] 고막이 검출 되지 않은 사진

### 3.3.2 데이터 증량

앞서 데이터 셋의 정제, 관심영역 추출 그리고 훈련 효과를 높이기 위해 Mask R-CNN 모델을 통해 고막 영역을 추출하였다. 그리고 Python 모듈인 OpenCV를 이용하여 CLAHE(Contrast Limited Adaptive Histogram Equalization) 알고리즘을 적용하여 데이터의 노이즈를 제거하고 이미지 크기를 384, 384로 변경하였다. 그리고 RGB 값을 255로 나누어 주었다. 마지막으로 회전, 확대, 상하반전, 밝기 조절 등 여러 기법을 사용하여 데이터 증량하였다. 데이터 증량은 과적합을 방지 하는 데에 목적을 두었다.

[표 3] 데이터 셋 크기

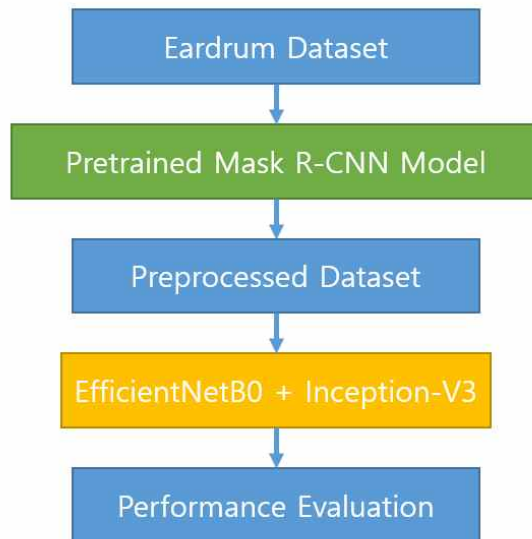
	원본 증량(장)		고막 추출 증량(장)	
Normal	2,040	5,765	1,971	5,695
Traumatic Perforation	356	5,765	295	5,695
AOM	804	5,765	654	5,695
COM	824	5,765	698	5,695
Congenital Cholesteatoma	384	5,765	312	5,695
OME	400	5,765	324	5,695
합계	4,808	34,590	4,254	34,170

위 [표 3]은 데이터 셋의 크기를 나타내는 표이다. 그리고 훈련 셋, 검증 셋, 테스트 셋의 비율은 3:1:1로 나누었고 K-Fold 교차검증을 위해 5개의 Fold로 데이터를 나누었다. 더 많은 Fold로 구성해도 되지만, 소요 시간과 확률적으로 5-Fold가 가장 적절하였다.

### 3.4 분류

분류를 위해 Tensorflow, Keras(Python) 모듈을 사용하였고 Adadelta를 Optimizer로 사용하였으며 Early Stopping 기법을 통해 Validation Loss가 훈련을 진행하면서 20회 동안 감소하지 않을 때 멈추도록 설정하였다. 이는 과적합을 방지 하고 정확도를 높게 얻기 위함이다.

EfficientNet-B0, Inception-V3, 그리고 제안하는 앙상블 모델을 이용하여 원본 사진 데이터 셋과 Mask R-CNN 모델을 통해 고막 영역을 추출하여 재구성한 데이터 셋 두 가지로 나누어 실험을 하였다. 아래 [그림 9]는 데이터처리 흐름도이다.



[그림 10] 데이터처리 흐름도

### 3.4.1 분류 모델 Architecture

본 논문에서 제안하는 모델은 EfficientNet-B0와 Inception-V3 모델을 배경 [18]을 통해 모델을 구성했고 아래 [표 4], [표 5]는 각 모델의 아키텍처다.

[표 4] EfficientNet-B0 Baseline Network[19]

Stage	Operator	Resolution	Channels	Layers
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv 1x1 & Pooling & FC	7×7	1280	1

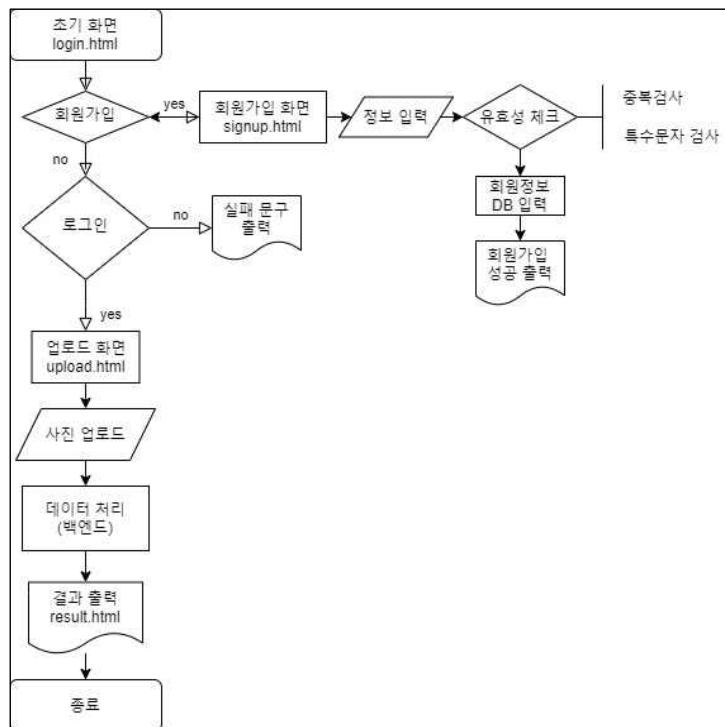
[표 5] Inception-V3 Baseline Network[20]

Type	Patch size/Stride or Remarks	Input Size
Conv	3×3/2	299 × 299 × 3
Conv	3×3/1	149 × 149 × 32
Conv	3×3/1	149 × 149 × 32
Pool	3×3/2	149 × 149 × 64
Conv	3×3/1	73 × 73 × 64
Conv	3×3/2	71 × 71 × 80
Conv	3×3/1	35 × 35 × 192
3 Inception	[20]레퍼런스 참조	35 × 35 × 288
5 Inception		17 × 17 × 768
2 Inception		8 × 8 × 1280
Pool	3×3/1	8 × 8 × 2048
Linear	Logits	1 × 1 × 2048
Softmax	Classifier	1 × 1 × 1000



### 3.5 Web Server

웹 서버에 사용된 모듈은 Flask이며 이를 통해 별도의 소켓 사용 없이 Python 내에서 딥러닝 결과를 프론트 엔드로 쉽게 넘길 수 있었다. 그리고 모듈 자체 메소드를 통해 라우팅을 하였고, 포트포워딩을 통해 외부에서 웹에 접근을 가능하게 하였고 아래 [그림 10]은 웹 페이지의 플로우 차트다. 총 5개의 HTML 문서로 구성되어 있다.



[그림 11] 웹 프로우 차트

### 3.5.1 UI(User Interface)

#### 1) 초기 화면

아래 [그림 11]는 초기화면이며 로그인과 회원가입을 할 수 있다. 로그인 시 서버에서 DB를 분석하여 아이디와 패스워드의 유효성을 검사한다.



The image shows the initial screen of the '중이염 진단 사이트' (Otitis Media Diagnosis Site). It features a logo of a hand holding a stethoscope in the top left corner. The title '중이염 진단 사이트' is prominently displayed in the top center. Below the title, there is a 'LOGIN' section with two input fields: 'Email : Enter Your Email' and 'PW : Enter Your Password'. To the right of these fields is a 'LOGIN' button. Below the input fields, there is a link that says 'You Don't Have Your Account?' followed by a 'SIGN UP' button.

[그림 12] 초기 화면

#### 2) 회원 가입 화면

아래 [그림 12]는 회원가입 화면이며 이메일 주소와 패스워드의 중복 검사 후 조건을 만족 시 DB에 저장시킨다.



The image shows the 'Sign Up' screen of the '중이염 진단 사이트' (Otitis Media Diagnosis Site). It features the same logo and title as the initial screen. Below the title, there is a 'SIGN UP' section with three input fields: 'Email : Enter Your Email', 'PW : Enter Your Password', and 'PW 확인 : Check Your Password'. To the right of these fields are two buttons: 'SIGN UP' and 'BACK'.

[그림 13] 회원가입 화면

### 3) 업로드 화면

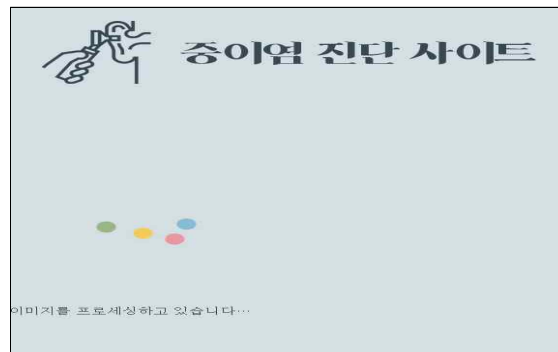
아래 [그림 13]은 병증 사진을 업로드하는 단계이며, 제출 버튼 클릭 시 서버에서 받은 이미지로 테스트를 시작시킨다. 이 과정에서 업로드하는 이미지의 용량이 클수록 소요시간은 증가한다.



[그림 14] 이미지 업로드 화면

### 4) 처리 화면

아래 [그림 14]는 딥러닝에 소요되는 시간 동안 로드되는 HTML 문서이다. Tensorflow-GPU 버전을 사용 시 약 8초의 GPU 할당 시간이 걸렸으나, CPU 버전으로 테스트를 하였을 때는 약 0.8초의 시간이 걸렸다.



[그림 15] 대기 화면

## 5) 결과 화면

아래 [그림]15 는 결과를 출력하는 화면이다. 서버에서 결과가 도출되면 그 결과를 웹 상에 출력시키는 단계이다. 3) 업로드에서 5) 결과 화면까지 업로드하는 시간과 약 0.8초 정도의 테스트 시간이 걸린다. 이미지의 용량이 적을수록 반응 속도는 증가한다.



[그림 16] 결과 화면

## 제 4 장 실험 및 결과

### 4.1 분류 실험 결과

원본 데이터 셋과 앞서 Mask R-CNN을 통해 정제한 데이터 셋 두 가지의 데이터 셋으로 실험을 진행하였다. 분류를 위해 사용한 모델은 VGG19, Resnet50, Resnet101, Resnet152, Inception-V3, EfficientNet-B0이다. 모든 모델의 Optimizer는 AdaDelta였고 손실함수로는 Sparse Categorical Crossentropy[21], 훈련율은 0.13으로 설정하여 훈련을 진행하였다. 그리고 아래 두 표는 실험 결과를 나타낸다.

[표 6] 원본 데이터 셋을 통해 얻은 실험 결과

	KF0	KF1	KF2	KF3	KF4	Params
VGG19	84.35%	85.21%	85.39%	86.01%	84.56%	1304.0M
Resnet50	90.53%	90.29%	90.28%	89.33%	89.12%	90.3M
Resnet101	88.99%	89.37%	89.72%	89.11%	88.45%	163.0M
Resnet152	89.71%	88.43%	88.49%	89.11%	89.86%	223.0M
Inception-V3	89.41%	90.27%	90.69%	89.89%	89.43%	91.8M
EfficientNet-B0	92.87%	90.83%	92.49%	91.14%	90.72%	20.9M
Proposal Model	93.23%	92.46%	93.41%	93.86%	92.56%	112.7M

[표 7] 고막 영역만 저장하여 재구성한 데이터 셋을 통해 얻은 실험 결과

	KF0	KF1	KF2	KF3	KF4	Params
VGG19	85.13%	85.43%	86.43%	86.23%	85.01%	1304.0M
Resnet50	93.27%	92.59%	93.23%	92.48%	91.61%	90.3M
Resnet101	91.93%	90.16%	89.89%	90.81%	91.12%	163.0M
Resnet152	90.12%	89.55%	89.49%	89.18%	88.57%	223.0M
Inception-V3	94.01%	94.99%	95.18%	94.50%	93.68%	91.8M
EfficientNet-B0	95.13%	95.76%	95.93%	96.11%	95.10%	20.9M
Proposal Model	96.86%	96.03%	96.77%	97.29%	97.01%	112.7M

그리고 가장 높은 정확도를 얻은 결과인 고막 영역만 저장하여 재구성한 데이터 셋에 제안하는 모델로 훈련하여 얻은 가중치에서 테스트 셋을 통해 97.29%의 정확도를 얻었고 아래 [표 8]는 Confusion Matrix를 나타내며 [표 9]는 그에 따른 Precision, Recall, F1-Score를 나타내는 표이다.

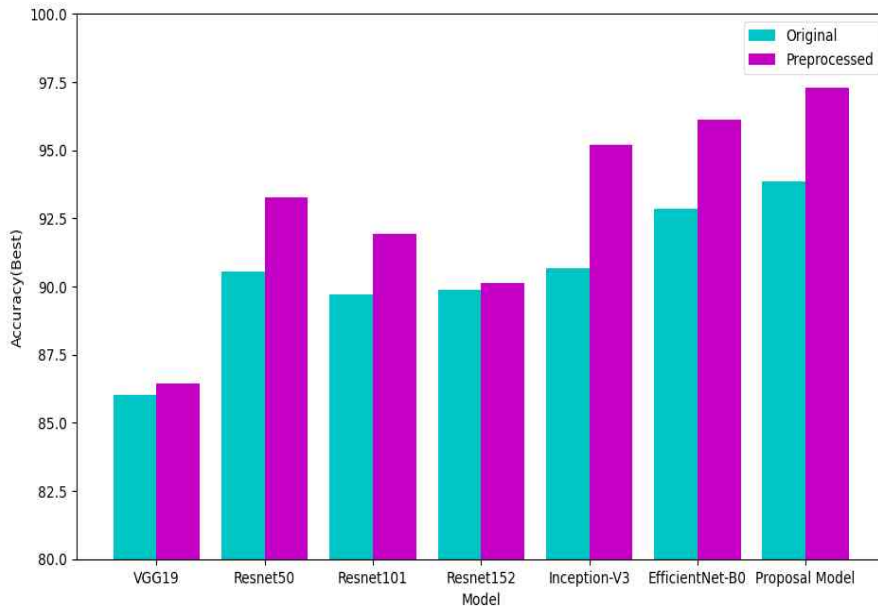
[표 8] Confusion Matrix

		Predict					
		Normal	TP	AOM	COM	C_C	OME
Actual	Normal	1126	0	6	0	1	6
	TP	0	1088	0	51	0	0
	AOM	46	0	1082	7	2	2
	COM	0	0	0	1139	0	0
	C_C	0	0	0	64	1075	0
	OME	0	0	0	0	0	1139

[표 9] Result of Precision, Recall, F1-score

	Precision(%)	Recall(%)	F1-Score(%)
NR	96.075	98.859	97
TP	100	95.522	98
AOM	99.449	94.996	97
COM	90.325	100	95
CC	99.722	94.381	97
OME	99.303	100	99

## 4.2 비교



[그림 17] 결과 그래프

위 [그림 16]는 5개의 Fold중 가장 높은 정확도를 가지고 그래프로 나타낸 것이다. 이를 통해 위 실험 결과 단일 모델보다 앙상블 모델의 정확도가 높다는 점, 그리고 데이터 셋이 한정적일 경우 데이터를 가공하여 재구성하여 분류를 하였을 때 정확도가 향상되는 것을 알 수 있었다.

### 4.3 웹 실행 속도 실험

실험은 이미지를 업로드하는 시점부터 결과가 도출되는 시간까지를 실험하였다. 아래 [표 10]은 처리 속도 실험 결과를 나타내는 표이다. 실험 방식은 전송된 파일 확인, 전송된 파일 이름 확인, 전송된 파일을 지정된 디렉토리에 이동, 전송된 파일 중복 확인, 순천향대학교 ICT 딥러닝 서버로 전송, 예측, 결과값 수신 총 7 단계로 나누어 모든 소요 시간을 합하여 측정하였다. 횟수는 20회씩 이미지의 크기는 1MB~5MB였다. 예측하는 시간과 이미지가 업로드, 결과 송수신 되는 시간을 합쳐서 결과가 나온다. 이미지 용량이 큰 경우 결과가 나오는 시간이 길어진다는 단점이 있다.

[표 10] 처리 속도 실험 결과

	1MB	2MB	3MB	4MB	5MB
평균 (sec)	2.116	2.231	2.551	2.436	2.590



## 제 5 장 결론 및 향후 과제

4차 산업혁명에 맞춰 빠르게 진화하는 현실에서 의료 시스템도 마찬가지로 스마트한 방법으로 진화하고 있으며 본 논문에서는 이러한 현실에 맞춰 딥러닝을 이용한 중이 병증 진단 모델을 제시하였다. 단일 모델이 아닌 두 개의 모델을 배경 방법을 통해 모델을 재구성하였고, 원본 이미지에서 관심 영역을 추출하여 데이터셋을 재구성하여 훈련을 시켰고 이를 통해 정확도 향상을 이루었고 가장 높은 정확도는 97.23%였다. 그리고 웹 서버를 구축하여 3초 이내로 중이 병증을 진단할 수 있었다. 하지만 한계점으로 이미지를 업로드하는 시간이 더해져 고화질의 이미지를 업로드 할 시 시간이 오래 걸리는 단점을 지니고 있다. 이는 내시경과 같은 하드웨어 단에서 처리를 하여 용량이 너무 크지 않도록 설계해야 할 것이다.

병을 진단하는 시스템이기 때문에 더 정확, 정밀한 결과를 도출해야 한다. 그렇기에 병원과의 협력을 통해 더 많은 데이터를 바탕으로 훈련과 검증을 반복한다면 이를 해결 할 수 있을 것이다. 또한 Mask R-CNN을 더욱 정밀하게 튜닝하거나 데이터 셋을 추가하여 정확도를 높이고, 귀지 또한 Detection 하여 귀지가 너무 커서 고막을 가리는 것을 판별하는 알고리즘을 더하여 사용자로 하여금 더 정확한 진단을 얻을 수 있도록 인간 친화적인 프로그램으로 개발해 나아가야 할 것이다. 그리고 추후 의료 전문가의 개입으로 체계적인 임상실험을 통하여 성능과 효율이 입증되고 이를 통해 의료기기와 해당 기술을 접목한다면 긴급한 상황에서 병을 진단할 수 있는 전문적인 의료기기로 발전할 수 있을 것이다.

## 참 고 문 헌

- [1] 국제 데이터 주식 회사 2014, [www.idc.com/](http://www.idc.com/)
- [2]<https://www.msdmanuals.com/en-kr/home/ear,-nose,-and-throat-disorders/biology-of-the-ears,-nose,-and-throat/ears>
- [3]<https://www.msdmanuals.com/en-kr/professional/ear,-nose,-and-throat-disorders/middle-ear-and-tympanic-membrane-disorders/traumatic-perforation-of-the-tympanic-membrane>
- [4]<https://www.msdmanuals.com/en-kr/professional/ear,-nose,-and-throat-disorders/middle-ear-and-tympanic-membrane-disorders/otitis-media-acute>
- [5]<https://www.msdmanuals.com/en-kr/professional/pediatrics/congenital-craniofacial-and-musculoskeletal-abnormalities/congenital-ear-abnormalities>
- [6]<https://www.msdmanuals.com/en-kr/professional/ear,-nose,-and-throat-disorders/middle-ear-and-tympanic-membrane-disorders/otitis-media-chronic>
- [7]<https://medicalguidelines.msf.org/viewport/CG/english/acute-otitis-media-aom-16689234.html>
- [8] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016.
- [9] Krause, Jonathan, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. Ophthalmology, 2018.

- [10] Abramoff, Michael D., et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digital Medicine*, 2018, 1.1: 39.
- [10] EyePACS, <http://www.eyepacs.com/>, access date: 2018. 10. 21
- [11] Van Der Heijden, Amber A., et al. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta ophthalmologica*, 2018, 96.1: 63-68.
- [12] Messidor-2 Decenciere et al.. Feedback on a publicly distributed database: the Messidor database. *Image Analysis & Stereology*, V.33, N.3, pp.231-234, aug. 2014. ISSN 1854-5165.
- [13] vuno, <https://www.vuno.co/static/pdf/BoneAge.pdf>
- [14] <https://flask.palletsprojects.com/en/1.1.x/>
- [15] April 2010IEEE, *Transactions on Pattern Analysis and Machine Intelligence* 32(3):569 - 575
- [16] Mask R-CNN Kaiming He Georgia Gkioxari Piotr Dollar Ross Girshick Facebook AI Research (FAIR), <https://arxiv.org/pdf/1703.06870.pdf>
- [17] VGG Image Annotator, <http://www.robots.ox.ac.uk/~vgg/software/via/>
- [18] Bagging, Boosting and Ensemble Methods-Peter Bühlmann January 2001, [https://www.researchgate.net/publication/45130375\\_Bagging\\_Boosting\\_and\\_Ensemble\\_Methods](https://www.researchgate.net/publication/45130375_Bagging_Boosting_and_Ensemble_Methods)

- [19] EfficientNet, <https://arxiv.org/abs/1905.11946>
- [20] Inception-V3, Rethinking the Inception Architecture for Computer Vision,  
[https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2016/papers/Szegedy\\_Rethinking\\_the\\_Inception\\_CVPR\\_2016\\_paper.pdf](https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.pdf)
- [21] Sparse Categorical Crossentropy,  
<https://cwiki.apache.org/confluence/display/MXNET/Multi-hot+Sparse+Categorical+Cross-entropy>

## 감사의 글

전역 후 2018년 1학기에 복학하고 바로 남윤영 교수님을 찾아가서 연구실에 들어가게 해달라고 말씀을 드렸습니다. 흔쾌히 받아주셨고 많은 경험을 해보고 싶어서 이것저것 다 도전해봤습니다. 교수님의 피드백, 세미나 그리고 연구에 필요한 기자재 등 적극적인 도움과 많은 가르침을 받았습니다. 제일 기억에 남는 프로젝트는 ‘부정맥 진단 체중계’입니다. 이 프로젝트로 코엑스에 가서 발표도 해보고 부천병원에 가서 환자분들에게 시험해보기도 하고 이 프로젝트로 논문을 써서 해외 학회에 참석도 해보고 외부 디자이너분과 같이 설계하면서 의견도 내보는 등 평범한 대학생들이 해볼 수 없는 경험을 하게 해주셨습니다. 그리고 제가 가장 뜻깊게 생각하는 가르침은 저에게 ‘겸손’이라는 미덕을 가르쳐 주신 것입니다. 세미나 때 교수님은 가끔 개발자, 연구원으로써의 소양을 말씀해주시곤 했는데 그 말들을 듣고 저는 하나하나 와닿아서 점차 발전할 수 있었던 것 같습니다. 그리고 글로 표현할 수 없는 많은 것을 배웠습니다.

졸업하고 나서도 많이 생각날 것 같은 친구는 자취, 연구실, 뭐든 같이 흔쾌히 해준 준수입니다. 뭐든 끝까지 하려 하고 파고드는 모습이 멋있는 친구인데 옆에서 보고 배울 점이 많았었습니다. 지금은 휴학하였지만 그래도 저는 제일 소중하게 생각한 친구입니다. 그리고 같이 연구실 생활하며 잊지 못할 추억과 경험을 같이 한 제 동기들 모두 성공하였으면 좋겠습니다. 그리고 저의 동네 친구 김상진씨 남은 시간 고생하시고 포기하지 말고 끝까지 열심히 했으면 좋겠고 요새 많이 바쁘고 취업 준비 때문에 힘든데 옆에서 다독여주고 묵묵히 기다리는 지은이한테 고맙고 인사드리면 항상 밝게 인사받아주시고 안부 물어봐 주시는 모든 교수님들과 고생하시는 청소아주머님 모두에게 감사하고 기억하겠습니다. Good Bye 순천향.