

제 26회 담낭 용종 연구 보고

2021.09.03



실험 세팅

➤ 데이터 구성

[서울대 데이터]

- 177명의 환자 데이터 (전체 데이터)
- Neoplastic : 100명
- Non-neoplastic : 77명

[가톨릭 데이터]

- 360명의 환자 데이터 (전체 데이터)
- Neoplastic : 53명
- Non-neoplastic : 307명

➤ Train & Test Set

[서울대 데이터]

- 2,888장의 데이터셋
- Train Set : 141명(2,256장)
- Test Set : 36명(632장)

[가톨릭 데이터]

- 1,047장의 데이터셋
- Train Set : 288명(838장)
- Test Set : 72명(209장)

➤ Input Image

- Outline, Inline, Polyp, 담낭벽+Polyp 이미지

➤ 모델 구성

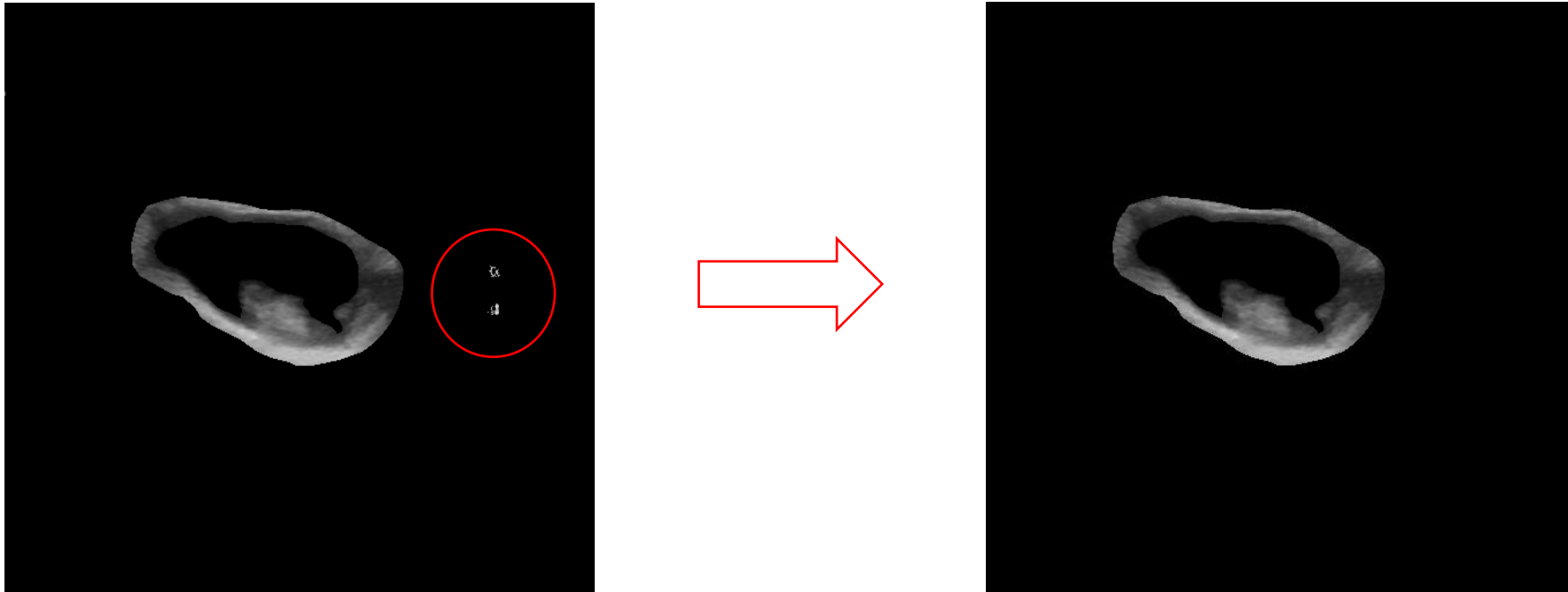
- 갑상선에서 사용한 모델을 담낭 용종 연구에 맞게 조정한 모델
- 데이터 불균형 문제를 해소하기 위한 Focal loss function 적용
- 서울대 데이터를 이용한 가톨릭 데이터 Transfer Learning

➤ 반복 실험 30회

Preprocessing

➤ 입력 이미지에 대한 Noise 제거

- 모델의 입력으로 사용하기 위한 전처리 수행과정에서 노이즈 이미지 발생
- 추가적인 전처리를 통하여 노이즈 이미지 제거



Data Augmentation

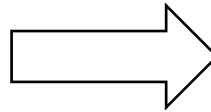
➤ Horizontal Flip

- 가톨릭 데이터에 대한 불균형 문제를 해결하기 위한 data augmentation 적용
- Neoplastic 데이터에 대해 horizontal flip augmentation 적용



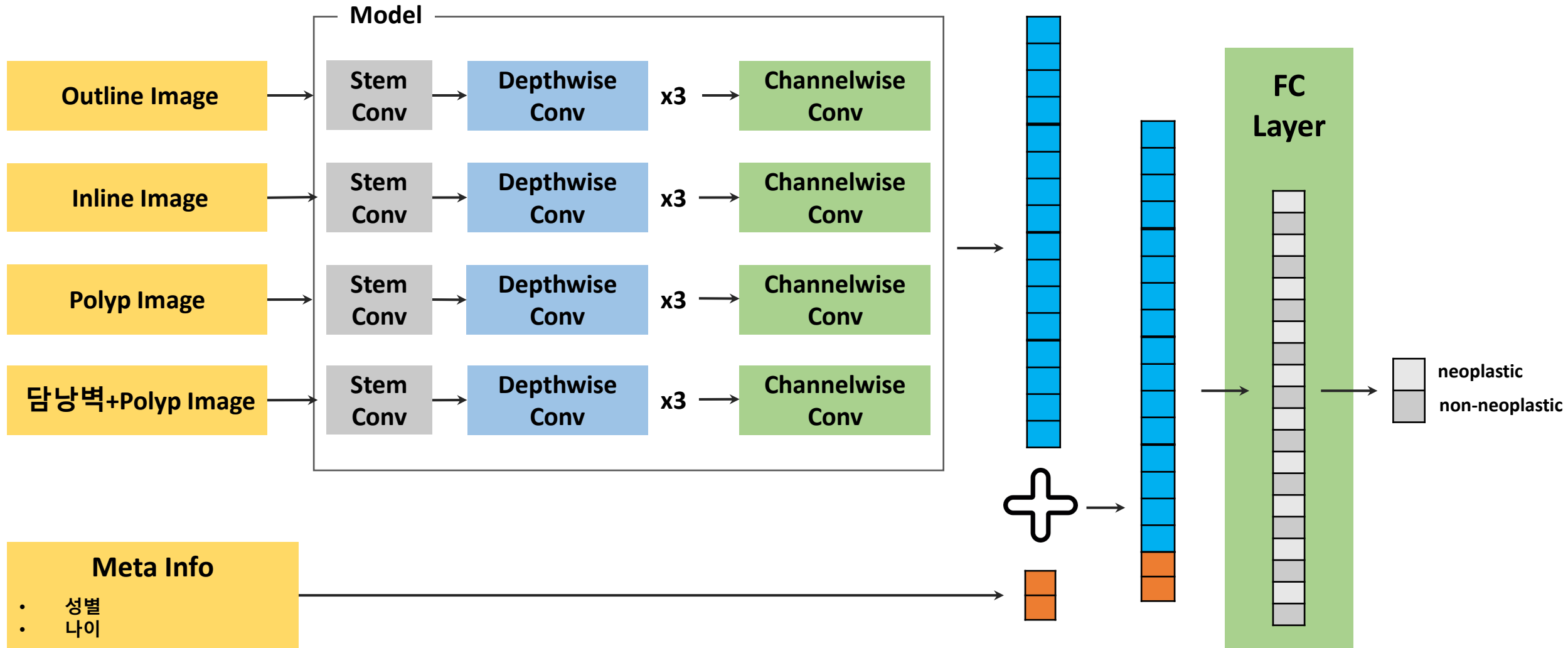
Original Image

Horizontal Flip



Flip Image

모형 파이프 라인



Transfer Learning

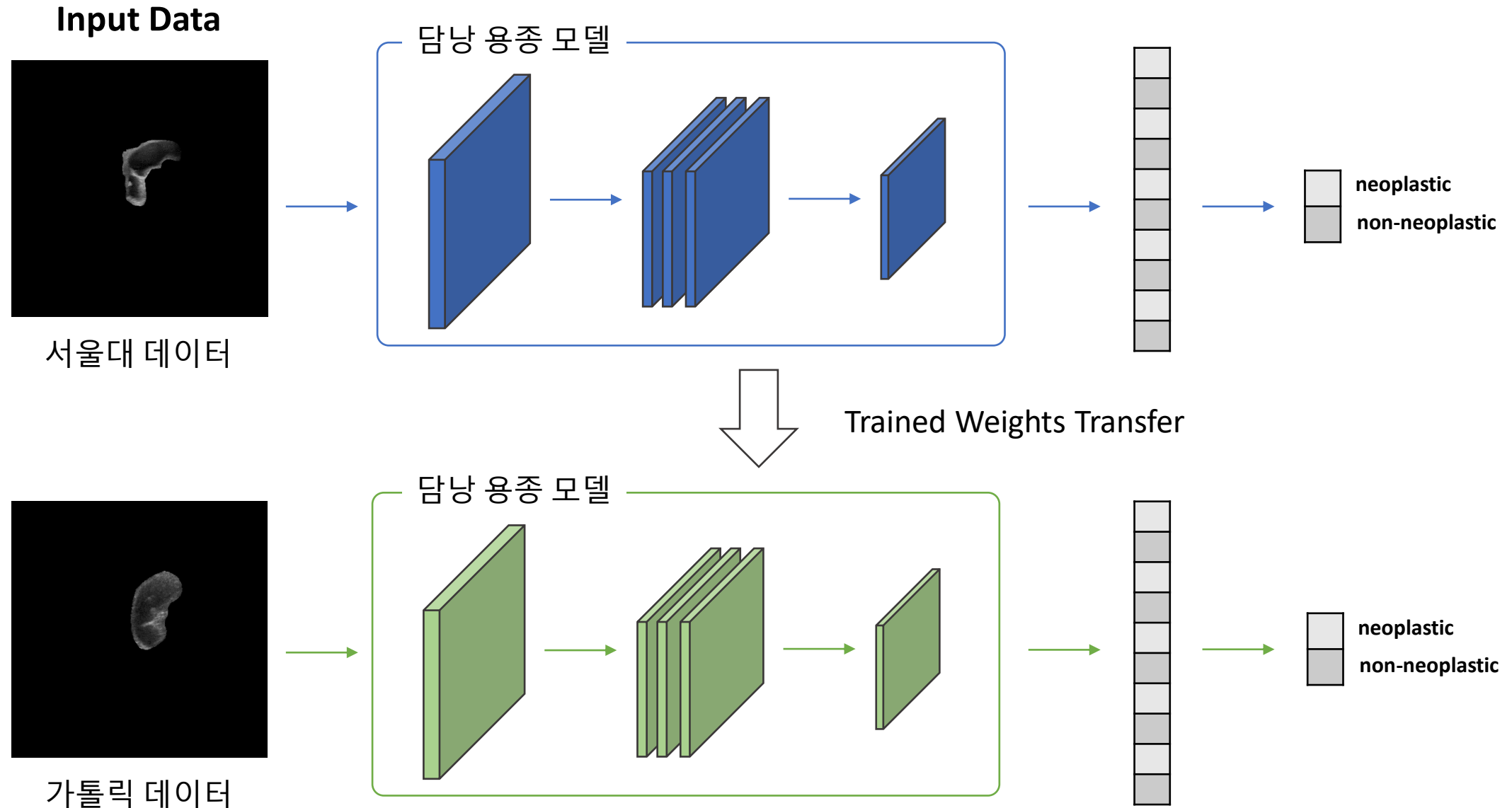
➤ 가톨릭 데이터 Transfer Learning

- 가톨릭 데이터 수 부족에 대한 실험 성능 향상을 위한 Transfer Learning 수행
- 서울대 데이터를 이용하여 학습한 모델의 가중치 값을 fine-tuning을 통해 가톨릭 데이터에 대한 모델 학습
- 가톨릭 데이터 838장을 사용한 Transfer Learning 수행 결과 성능 향상

➤ Experiment Settings

- | | |
|---|--|
| <ul style="list-style-type: none">• Baseline Model(서울대 데이터)• Train Set : 141명(2,256장)• Test Set : 36명(632장)• Loss function : Focal loss• Optimizer : Adam | <ul style="list-style-type: none">• Transfer Learning Model(가톨릭 데이터)• Train Set : 288명(838장)• Test Set : 72명(209장)• Loss function : Focal loss• Optimizer : Adam |
|---|--|

Transfer Learning



Top-3 Vote Method Classification Performance Table

Table . Metrics with 95% confidence interval of five Images (▼/△ indicates that the corresponding method is significantly worse/better than 서울대 데이터 model based on paired t-test at 5% significance level).

Model	Metrics	AUC Mean	95% CI		p-value	
			Lower	Upper		
서울대 데이터	AUC	0.905	0.889	0.922		
	Accuracy	0.832	0.807	0.858		
	Sensitivity	0.822	0.785	0.858		
	Specificity	0.846	0.782	0.909		
GoogleNet	AUC	0.778	0.738	0.819	0.0073	*▼
	Accuracy	0.582	0.559	0.606	0.0002	**▼
	Sensitivity	0.920	0.844	0.996	0.1135	
	Specificity	0.160	0.060	0.261	0.0006	**▼
Inception V3	AUC	0.725	0.697	0.754	0.0007	**▼
	Accuracy	0.587	0.564	0.610	0.0002	**▼
	Sensitivity	0.937	0.899	0.975	0.0192	
	Specificity	0.150	0.067	0.233	0.0003	**▼

AUC, area under the curve; CI, confidence interval; NN, neural network. *: p-value < 0.01, **: p-value < 0.001

Top-3 Vote Method Classification Performance Table

Table . Metrics with 95% confidence interval of five Images (▼/△ indicates that the corresponding method is significantly worse/better than Polyp Image based on paired t-test at 5% significance level).

Input Image	Metrics	AUC Mean	95% CI		p-value	
			Lower	Upper		
가톨릭 데이터	AUC	0.544	0.509	0.579	0.0008	**▼
	Accuracy	0.743	0.689	0.787	0.2907	
	Sensitivity	0.218	0.147	0.289	0.0161	
	Specificity	0.837	0.774	0.900	0.9027	
가톨릭 데이터 Augmentation	AUC	0.608	0.573	0.642	0.0029	*▼
	Accuracy	0.728	0.674	0.781	0.2149	
	Sensitivity	0.324	0.227	0.421	0.1515	
	Specificity	0.801	0.724	0.877	0.4648	
가톨릭 데이터 Transfer Learning	AUC	0.691	0.664	0.717	0.0244	
	Accuracy	0.765	0.730	0.801	0.5935	
	Sensitivity	0.409	0.325	0.494	0.5994	
	Specificity	0.830	0.777	0.882	0.7638	
가톨릭 데이터 Augmentation & Transfer Learning	AUC	0.761	0.739	0.784		
	Accuracy	0.782	0.747	0.817		
	Sensitivity	0.445	0.377	0.513		
	Specificity	0.843	0.794	0.892		

AUC, area under the curve; CI, confidence interval; NN, neural network. *: p-value < 0.01, **: p-value < 0.001

Top-3 Vote Method Classification Performance Table

Table . Metrics with 95% confidence interval of five Images (▼/△ indicates that the corresponding method is significantly worse/better than 서울대 데이터 model based on paired t-test at 5% significance level).

Model	Metrics	AUC Mean	95% CI		p-value
			Lower	Upper	
가톨릭 데이터 Augmentation & Transfer Learning	AUC	0.761	0.739	0.784	
	Accuracy	0.782	0.747	0.817	
	Sensitivity	0.445	0.377	0.513	
	Specificity	0.843	0.794	0.892	
GoogleNet	AUC	0.735	0.709	0.761	0.2553
	Accuracy	0.656	0.575	0.736	0.0672
	Sensitivity	0.521	0.389	0.654	0.4248
	Specificity	0.680	0.564	0.795	0.0871
Inception V3	AUC	0.662	0.616	0.707	0.0272
	Accuracy	0.606	0.568	0.645	0.0042
	Sensitivity	0.380	0.279	0.481	0.4074
	Specificity	0.890	0.815	0.964	0.4159

AUC, area under the curve; CI, confidence interval; NN, neural network. *: p-value < 0.01, **: p-value < 0.001

Experiment Results

➤ 서울대 데이터

- 전처리 과정을 통해 입력 이미지에서 불필요한 부분 제거
- 갑상선 연구에서 사용한 모델을 담낭용종에 맞게 변형한 모델을 사용하여 대표적으로 많이 사용하는 GoogleNet, Inception V3 모델 보다 높은 성능을 보임
- 데이터 불균형 문제를 Focal loss를 이용하여 성능 향상

➤ 가톨릭 데이터

- Data augmentation을 적용하여 데이터 불균형에 대한 문제를 해결함으로써 성능 향상
- 적은 수의 데이터 및 데이터 불균형 문제를 Data augmentation과 서울대 데이터를 이용하여 학습한 모델을 Transfer Learning 통해 가톨릭 데이터에 맞는 모델 학습 및 성능 향상

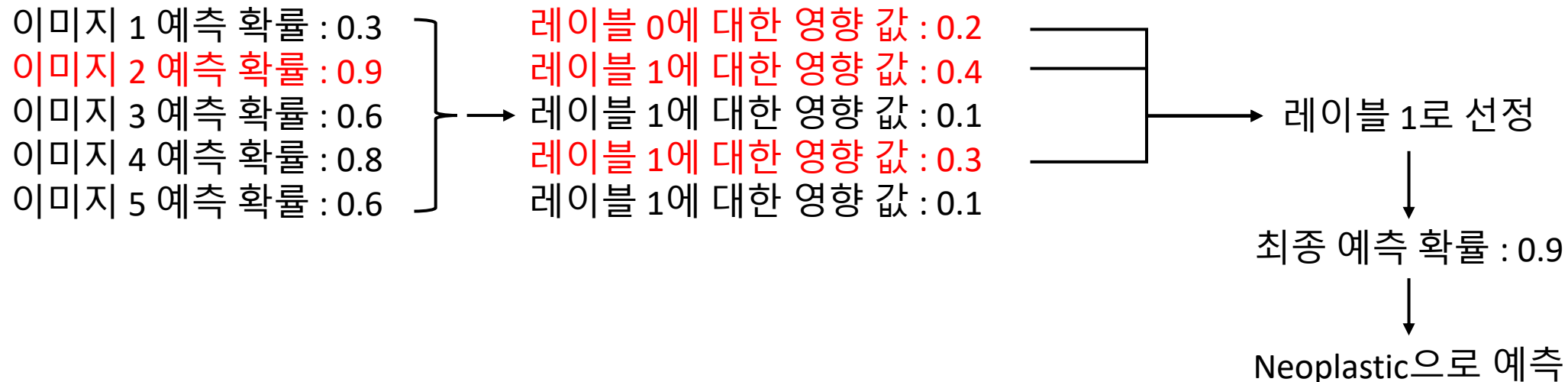
Appendix

➤ Top-k Voting Method

1. 모든 이미지 중 레이블 결정에 대한 영향 값 계산 $f = \begin{cases} |\hat{y} - 0.5| & \text{if } \hat{y} \geq 0.5 \\ |0.5 - \hat{y}| & \text{otherwise.} \end{cases}$
2. 큰 영향을 미치는 이미지 k개를 선택
3. k개의 이미지에서 다수의 레이블에 대한 확률 값 중에서 가장 큰 값을 최종 결과값으로 선정

Ex) Top-3 - 환자 1명에 대해 5장의 이미지로 예측하는 경우

- 0 ~ 1의 확률 값을 가지고 0.5를 기준으로 0에 가까우면 Non-neoplastic, 1에 가까우면 Neoplastic
- 선정 레이블의 확률 값 중 가장 큰 값을 최종 결과값으로 선정



Additional Experimental Results – 서울대 데이터

Table . Metrics with 95% confidence interval of five Images (▼/△ indicates that the corresponding method is significantly worse/better than Top-3 vote method based on paired t-test at 5% significance level).

Vote	Metrics	서울대 데이터				GoogleNet				Inception V3			
		Mean	95% CI		p-value	Mean	95% CI		p-value	Mean	95% CI		p-value
			Lower	Upper			Lower	Upper			Lower	Upper	
Top-1	AUC	0.889	0.872	0.907	0.313	0.778	0.739	0.818	0.932	0.724	0.695	0.754	0.968
	Accuracy	0.829	0.803	0.854	0.868	0.581	0.561	0.600	0.929	0.584	0.563	0.606	0.890
	Sensitivity	0.805	0.766	0.844	0.617	0.925	0.853	0.997	0.938	0.940	0.898	0.982	0.931
	Specificity	0.858	0.797	0.920	0.816	0.150	0.055	0.245	0.903	0.140	0.063	0.216	0.881
Top-3	AUC	0.905	0.889	0.922		0.778	0.738	0.819		0.725	0.697	0.754	
	Accuracy	0.832	0.807	0.858		0.582	0.559	0.606		0.587	0.564	0.610	
	Sensitivity	0.822	0.785	0.858		0.920	0.844	0.996		0.937	0.899	0.975	
	Specificity	0.846	0.782	0.909		0.160	0.060	0.261		0.150	0.067	0.233	
Top-5	AUC	0.907	0.892	0.922	0.903	0.773	0.730	0.816	0.883	0.723	0.694	0.752	0.925
	Accuracy	0.834	0.810	0.859	0.926	0.576	0.553	0.599	0.755	0.582	0.562	0.603	0.806
	Sensitivity	0.837	0.799	0.875	0.640	0.908	0.828	0.988	0.859	0.927	0.883	0.970	0.772
	Specificity	0.831	0.770	0.892	0.787	0.160	0.059	0.262	0.994	0.152	0.070	0.235	0.974
Top-10	AUC	0.909	0.895	0.923	0.783	0.773	0.730	0.815	0.875	0.724	0.695	0.752	0.946
	Accuracy	0.832	0.811	0.854	0.942	0.578	0.555	0.601	0.830	0.585	0.565	0.605	0.914
	Sensitivity	0.842	0.809	0.874	0.517	0.912	0.834	0.990	0.904	0.927	0.883	0.971	0.780
	Specificity	0.821	0.757	0.884	0.654	0.160	0.058	0.262	0.996	0.158	0.073	0.243	0.912

AUC, area under the curve; CI, confidence interval; NN, neural network. *: p-value < 0.01, **: p-value < 0.001

Additional Experimental Results – 서울대 데이터

Table . Metrics with 95% confidence interval of five Images (▼/△ indicates that the corresponding method is significantly worse/better than Top-3 vote method based on paired t-test at 5% significance level).

Vote	Metrics	Outline				Inline				Polyp				담낭벽 + Polyp			
		Mean	95% CI		p-value	Mean	95% CI		p-value	Mean	95% CI		p-value	Mean	95% CI		p-value
			Lower	Upper			Lower	Upper			Lower	Upper			Lower	Upper	
Top-1	AUC	0.866	0.845	0.886	0.768	0.799	0.763	0.836	0.752	0.840	0.808	0.871	0.874	0.823	0.791	0.855	0.885
	Accuracy	0.757	0.723	0.792	0.856	0.705	0.671	0.738	0.765	0.747	0.720	0.775	0.475	0.701	0.666	0.736	0.687
	Sensitivity	0.892	0.860	0.923	0.959	0.912	0.881	0.943	0.911	0.820	0.766	0.874	0.975	0.888	0.847	0.930	0.969
	Specificity	0.590	0.503	0.676	0.844	0.446	0.347	0.544	0.785	0.656	0.568	0.744	0.610	0.467	0.354	0.579	0.797
Top-3	AUC	0.871	0.850	0.892		0.810	0.771	0.848		0.835	0.804	0.867		0.827	0.795	0.859	
	Accuracy	0.752	0.719	0.785		0.696	0.666	0.727		0.730	0.702	0.757		0.689	0.656	0.722	
	Sensitivity	0.893	0.864	0.922		0.915	0.884	0.946		0.820	0.766	0.874		0.887	0.849	0.925	
	Specificity	0.575	0.490	0.660		0.423	0.331	0.515		0.617	0.530	0.703		0.442	0.331	0.552	
Top-5	AUC	0.874	0.854	0.894	0.866	0.815	0.777	0.853	0.872	0.844	0.812	0.876	0.754	0.830	0.781	0.885	0.894
	Accuracy	0.752	0.718	0.785	0.990	0.697	0.666	0.728	0.980	0.738	0.711	0.765	0.727	0.688	0.651	0.718	0.908
	Sensitivity	0.903	0.873	0.934	0.692	0.923	0.896	0.951	0.745	0.837	0.785	0.889	0.714	0.885	0.841	0.929	0.955
	Specificity	0.563	0.477	0.648	0.867	0.415	0.326	0.504	0.919	0.615	0.527	0.702	0.978	0.438	0.321	0.548	0.950
Top-10	AUC	0.845	0.855	0.894	0.758	0.812	0.774	0.850	0.944	0.846	0.815	0.878	0.690	0.830	0.801	0.865	0.857
	Accuracy	0.723	0.715	0.780	0.674	0.689	0.658	0.720	0.784	0.739	0.711	0.767	0.701	0.689	0.655	0.723	0.999
	Sensitivity	0.860	0.874	0.936	0.921	0.928	0.901	0.955	0.613	0.847	0.797	0.897	0.559	0.900	0.861	0.939	0.700
	Specificity	0.503	0.467	0.633	0.594	0.390	0.304	0.475	0.669	0.604	0.517	0.691	0.866	0.425	0.313	0.537	0.863

AUC, area under the curve; CI, confidence interval; NN, neural network. *: p-value < 0.01, **: p-value < 0.001

Additional Experimental Results – 가톨릭 데이터

Table . Metrics with 95% confidence interval of five Images (▼/△ indicates that the corresponding method is significantly worse/better than Top-3 vote method based on paired t-test at 5% significance level).

Vote	Metrics	가톨릭 데이터				가톨릭 데이터 Aug				가톨릭 데이터 TL				가톨릭 데이터 Aug & TL			
		Mean	95% CI		p-value	Mean	95% CI		p-value	Mean	95% CI		p-value	Mean	95% CI		p-value
			Lower	Upper			Lower	Upper			Lower	Upper			Lower	Upper	
Top-1	AUC	0.531	0.499	0.564	0.669	0.611	0.578	0.643	0.917	0.662	0.635	0.688	0.251	0.761	0.737	0.784	0.973
	Accuracy	0.770	0.725	0.816	0.471	0.744	0.689	0.798	0.735	0.796	0.766	0.826	0.316	0.812	0.782	0.842	0.322
	Sensitivity	0.167	0.097	0.237	0.425	0.297	0.201	0.393	0.749	0.345	0.269	0.422	0.385	0.409	0.333	0.485	0.573
	Specificity	0.879	0.815	0.944	0.461	0.824	0.746	0.902	0.730	0.877	0.833	0.921	0.297	0.885	0.841	0.929	0.331
Top-3	AUC	0.544	0.509	0.579		0.608	0.573	0.642		0.691	0.664	0.717		0.761	0.739	0.784	
	Accuracy	0.743	0.689	0.787		0.728	0.674	0.781		0.765	0.730	0.801		0.782	0.747	0.817	
	Sensitivity	0.218	0.147	0.289		0.324	0.227	0.421		0.409	0.325	0.494		0.445	0.377	0.513	
	Specificity	0.837	0.774	0.900		0.801	0.724	0.877		0.830	0.777	0.882		0.843	0.794	0.892	
Top-5	AUC	0.551	0.514	0.588	0.823	0.607	0.571	0.643	0.982	0.694	0.669	0.720	0.871	0.759	0.738	0.780	0.901
	Accuracy	0.736	0.691	0.781	0.928	0.710	0.656	0.765	0.714	0.756	0.720	0.793	0.774	0.776	0.740	0.813	0.855
	Sensitivity	0.248	0.176	0.321	0.631	0.348	0.253	0.444	0.772	0.436	0.355	0.518	0.710	0.464	0.390	0.537	0.762
	Specificity	0.823	0.761	0.886	0.802	0.775	0.698	0.853	0.707	0.814	0.761	0.868	0.740	0.833	0.781	0.885	0.820
Top-10	AUC	0.551	0.514	0.588	0.823	0.607	0.571	0.643	0.982	0.694	0.669	0.720	0.871	0.694	0.669	0.720	0.026
	Accuracy	0.736	0.691	0.781	0.928	0.710	0.656	0.765	0.714	0.756	0.720	0.793	0.774	0.756	0.720	0.793	0.428
	Sensitivity	0.248	0.176	0.321	0.631	0.348	0.253	0.444	0.772	0.436	0.355	0.518	0.710	0.436	0.355	0.518	0.894
	Specificity	0.823	0.761	0.886	0.802	0.775	0.698	0.853	0.707	0.814	0.761	0.868	0.740	0.814	0.761	0.868	0.531

AUC, area under the curve; CI, confidence interval; NN, neural network; Aug, augmentation; TL, Transfer Learning. *: p-value < 0.01, **: p-value < 0.001

Additional Experimental Results – 가톨릭 데이터

Table . Metrics with 95% confidence interval of five Images (▼/△ indicates that the corresponding method is significantly worse/better than Top-3 vote method based on paired t-test at 5% significance level).

Vote	Metrics	가톨릭 데이터 Aug & TL				GoogleNet				Inception V3			
		95% CI				95% CI				95% CI			
		Mean	Lower	Upper	p-value	Mean	Lower	Upper	p-value	Mean	Lower	Upper	p-value
Top-1	AUC	0.761	0.737	0.784	0.973	0.725	0.701	0.750	0.932	0.652	0.608	0.696	0.958
	Accuracy	0.812	0.782	0.842	0.322	0.699	0.628	0.770	0.929	0.606	0.568	0.645	0.897
	Sensitivity	0.409	0.333	0.485	0.573	0.448	0.324	0.573	0.887	0.380	0.277	0.483	0.929
	Specificity	0.885	0.841	0.929	0.331	0.744	0.640	0.848	0.815	0.890	0.812	0.967	0.912
Top-3	AUC	0.761	0.739	0.784		0.735	0.709	0.761		0.662	0.616	0.707	
	Accuracy	0.782	0.747	0.817		0.656	0.575	0.736		0.606	0.568	0.645	
	Sensitivity	0.445	0.377	0.513		0.521	0.389	0.654		0.380	0.279	0.481	
	Specificity	0.843	0.794	0.892		0.680	0.564	0.795		0.890	0.815	0.964	
Top-5	AUC	0.759	0.738	0.780	0.901	0.735	0.709	0.761	0.997	0.660	0.615	0.705	0.812
	Accuracy	0.776	0.740	0.813	0.855	0.656	0.575	0.736	0.987	0.596	0.559	0.633	0.806
	Sensitivity	0.464	0.390	0.537	0.762	0.521	0.389	0.654	0.994	0.368	0.269	0.468	0.772
	Specificity	0.833	0.781	0.885	0.820	0.680	0.564	0.795	0.996	0.881	0.805	0.957	0.864
Top-10	AUC	0.694	0.669	0.720	0.026	0.735	0.709	0.761	0.997	0.666	0.621	0.710	0.825
	Accuracy	0.756	0.720	0.793	0.428	0.656	0.575	0.736	0.987	0.606	0.571	0.642	0.958
	Sensitivity	0.436	0.355	0.518	0.894	0.521	0.389	0.654	0.994	0.387	0.290	0.484	0.780
	Specificity	0.814	0.761	0.868	0.531	0.680	0.564	0.795	0.996	0.881	0.805	0.957	0.912

AUC, area under the curve; CI, confidence interval; NN, neural network; Aug, augmentation; TL, Transfer Learning. *: p-value < 0.01, **: p-value < 0.001

Additional Experimental Results – 가톨릭 데이터

Table . Metrics with 95% confidence interval of five Images (▼/△ indicates that the corresponding method is significantly worse/better than Top-3 vote method based on paired t-test at 5% significance level).

Vote	Metrics	Outline				Inline				Polyp				담낭벽 + Polyp			
		Mean	95% CI		p-value	Mean	95% CI		p-value	Mean	95% CI		p-value	Mean	95% CI		p-value
			Lower	Upper			Lower	Upper			Lower	Upper			Lower	Upper	
Top-1	AUC	0.534	0.505	0.564	0.974	0.576	0.550	0.602	0.899	0.581	0.548	0.613	0.991	0.531	0.499	0.563	0.942
	Accuracy	0.836	0.828	0.843	0.894	0.820	0.809	0.831	0.842	0.828	0.818	0.839	0.875	0.831	0.823	0.839	0.764
	Sensitivity	0.027	0.004	0.050	0.768	0.045	0.010	0.080	0.765	0.085	0.050	0.120	0.821	0.030	0.009	0.052	0.842
	Specificity	0.981	0.971	0.992	0.882	0.960	0.941	0.978	0.785	0.962	0.949	0.976	0.762	0.975	0.964	0.986	0.882
Top-3	AUC	0.532	0.502	0.561		0.575	0.548	0.603		0.581	0.548	0.614		0.532	0.499	0.565	
	Accuracy	0.823	0.808	0.837		0.800	0.783	0.817		0.808	0.791	0.825		0.816	0.802	0.830	
	Sensitivity	0.033	0.013	0.053		0.064	0.023	0.104		0.109	0.069	0.150		0.045	0.014	0.077	
	Specificity	0.965	0.946	0.984		0.933	0.909	0.957		0.934	0.912	0.956		0.955	0.935	0.975	
Top-5	AUC	0.541	0.514	0.568	0.896	0.577	0.547	0.606	0.942	0.585	0.553	0.616	0.921	0.533	0.499	0.567	0.984
	Accuracy	0.822	0.807	0.837	0.942	0.797	0.779	0.815	0.932	0.804	0.787	0.820	0.889	0.813	0.798	0.828	0.942
	Sensitivity	0.052	0.022	0.081	0.766	0.076	0.028	0.123	0.882	0.127	0.085	0.170	0.784	0.048	0.013	0.084	0.862
	Specificity	0.961	0.940	0.981	0.876	0.927	0.901	0.954	0.842	0.926	0.904	0.948	0.792	0.951	0.930	0.973	0.884
Top-10	AUC	0.541	0.514	0.568	0.896	0.577	0.547	0.606	0.942	0.585	0.553	0.616	0.921	0.533	0.499	0.567	0.984
	Accuracy	0.822	0.807	0.837	0.942	0.797	0.779	0.815	0.932	0.804	0.787	0.820	0.889	0.813	0.798	0.828	0.942
	Sensitivity	0.052	0.022	0.081	0.766	0.076	0.028	0.123	0.882	0.127	0.085	0.170	0.784	0.048	0.013	0.084	0.862
	Specificity	0.961	0.940	0.981	0.876	0.927	0.901	0.954	0.842	0.926	0.904	0.948	0.792	0.951	0.930	0.973	0.884

AUC, area under the curve; CI, confidence interval; NN, neural network. *: p-value < 0.01, **: p-value < 0.001