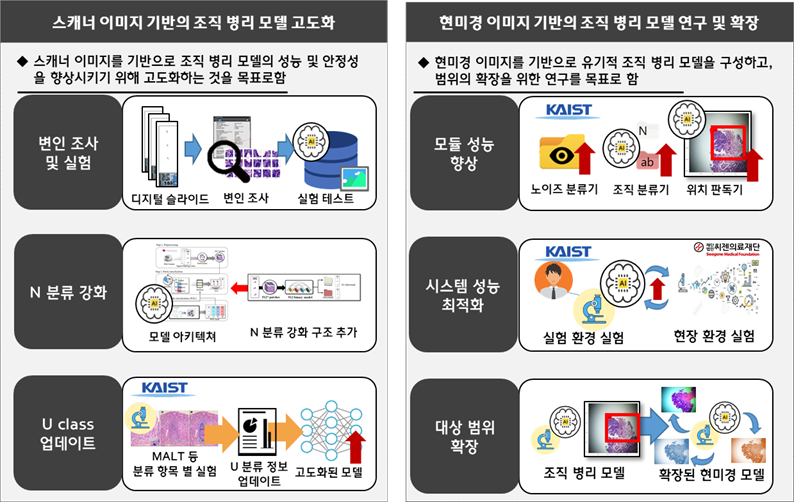
1.2 연구 목표 및 범위

* **스캐너 이미지 기반의 조직 병리 자동 판독 모델 고도화**
* 씨젠의료재단의 조직 병리 시스템 지원을 위해서, 딥러닝 기술을 기반으로 환자의 위, 대장 부위의 이상을 자동으로 판단하는 비전 검사 시스템 모델 연구가 이전 연구과제를 통해 진행되었음.
* **[목표]** 구축된 업무 체계와 인공 지능 모델을 기반으로 스캐너 이미지 기반의 조직 **병리 자동 판독 모델의 성능과 안정성 향상을 위해서 모델의 고도화**를 목표로 하며,
* **[방법]** 
  1. 이를 위해 슬라이드 이미지의 추가 수집 및 전처리를 통해서 높은 수준의 인공지능 학습이 가능하도록 **데이터베이스를 업데이트**할 계획임.
  2. 또한, 고도화를 위한 다양한 전략들을 수립하고 실험을 통해서 선택 가능한 전략들을 정의하여, 유효성 있는 **모델 고도화**를 진행할 예정임.
  3. 마지막으로 유효성이 확인된 **고도화 전략들을 모델에 적용**하는 것을 목표로 함
* 스캐너 모델의 고도화를 위하여 다음 **3가지 접근법을** 계획하였으며, 각 접근법의 구체적인 내용은 ‘3. 연구 개발 내용’에서 서술함 [주요 골자들 넣기]
* **1) 변인 조사 및 성능 일관성 확보:** 스캔 데이터에 따른 성능 차이를 확인했고, 이에 대한 원인을 조사하고 실험할 계획
* **2) 정상 슬라이드 분류 성능 강화를 위한 아키텍쳐 구현**: 정상 슬라이드가 다수 발생하는 환경에서도 성능을 유지하기 위해, 정상 슬라이드를 사전에 분류하는 레이어를 추가할 계획
* **3) Uncategorized 그룹 업데이트:** 분류 난이도가 높은 Uncategorized 그룹의 사례를 조사하고, 실험을 통해 모델을 업데이트하여, Uncategorized 그룹에 대한 분류 정확도를 향상할 계획
* 

**2. 연구 개발 범위**

**2.1 연구 개발 범위 - 1차년도**

* 1차년도는 스캐너 이미지 기반 모델의 고도화 파트와, 현미경 이미지 기반 모델 개발 파트를 포함함

**1) 스캐너 이미지 기반의 조직 병리 자동 판독 모델 고도화 [전체 프로젝트를 묶는 하나의 제목]**

* 스캐너 이미지 기반 조직 병리 자동 판독 모델의 고도화를 위해서 변인 조사 및 성능의 일관성을 확보하고, 정상 슬라이드 분류 성능을 확보하고자 함.
* 또한, U 그룹의 업데이트 이전 3 class 분류 모델을 우선적으로 개발하고, U 그룹의 업데이트 이후 4 class 분류 모델의 개발을 목표로 함

**2.2 연구 개발 결과 요약 [주요 프로젝트 별 결과 요약 ]**

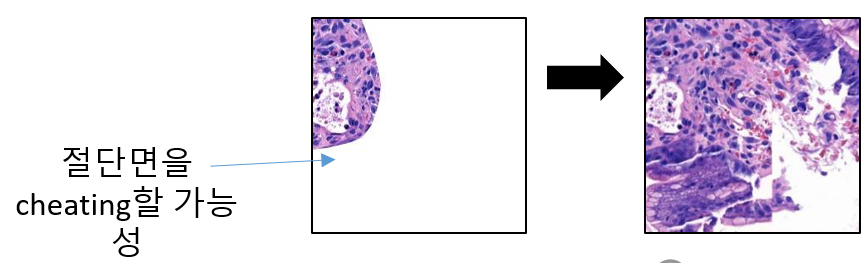
* [표 1]은 제안되었던 당해 연도 연구 계획에 대한 연구 개발 달성 내용을 요약한 표임
* 스캐너 이미지 기반의 조직 병리 자동 판독 모델 고도화 [각 프로젝트 별 한줄 요약 제시]
* 모델의 학습 및 예측에 영향을 줄 수 있는 변인들을 조사하고 실험 했으며, 이를 통해 이미지 수집 조건을 정의하여, 일관성 있는 모델을 생성하는 지침을 정의하고자 했음
* 정상 슬라이드 분류 성능을 확보하기 위해서, 모델이 구분을 잘하지 못하는 사례들에 대해서 조사하고, 해당 사례들에 대한 슬라이드를 수집하여, 모델을 업데이트함
* 기존 3 class 분류 모델의 활용 범위 확장을 위해서, 4번째 그룹인 U에 대한 범위를 다시 정의 하고, 데이터를 확보하고자 하였으며, 일부 모델에 대해서 4 class가 분류 가능하도록 업데이트함

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 일련 번호 | 범위 | 연구 개발 내용 | 연구 개발 결과 | 비고 |
| 1 | 스캐너 | 스캐너 변인 조사 및 성능 일관성 확보 | 변인 조사 및 성능 일관성 확보 | 달성 |
| 2 | 정상 슬라이드 분류 성능 강화를 위한 아키텍쳐 구현 | 정상 슬라이드 분류 성능 확보 | 달성 |
| 3 | U 그룹 업데이트 | NET 그룹 업데이트 | 일부 달성 |
| 4 | 현미경 | 구성 모듈 성능 향상 | 구성 모듈 고도화 | 달성 |
| 5 | 모델 설치 및 현장 테스트 | 전체 모델 프레임 워크 생성 및 테스트 | 일부 달성 |
| 6 | Malignant 디텍션 고도화 | Segmentation 모델 개발 및 적용 | 달성 |

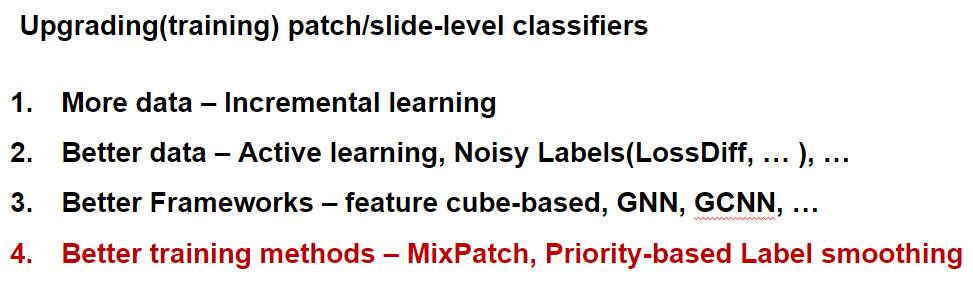
[표 작성 필요 ]

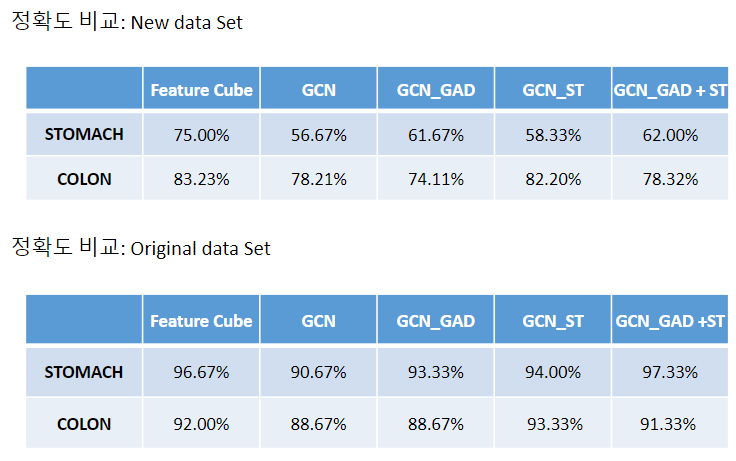
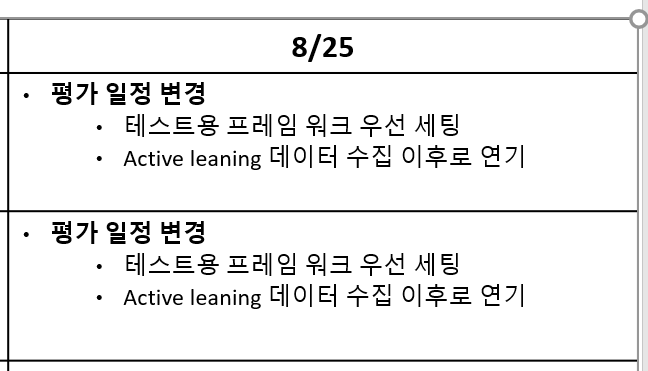
**질의 사항**

* **[Patch] Active learning vs Noise filtering 문제는 데이터 전처리 딴에서 해결이 된 건가?** 
  + **LossDiff 로 필터링 된 후, AL 추가될 데이터는 noise 발생 가능성이 현저히 낮음 : 사실상 전문의에 의한 전수 검사**
  + **=> 이 때부터 Low confidence 기반으로 selection 하기로 결정한 건가?**
  + **또는 Confidence score 와 learning loss는 서로 관련성이 없는 지표로 병행해도 된다라고 결론이 난 것인가?**
* **인공적인 절단면 / ‘절단면을 cheating할 가능성’의 정확한 의미가 무엇인가?**



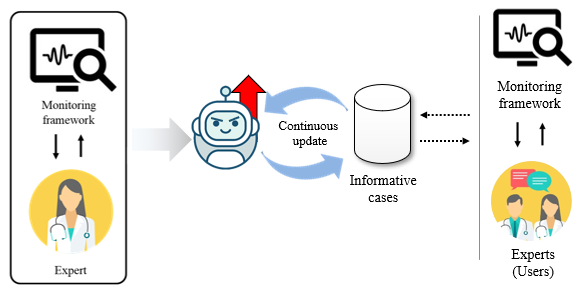
* **Patch classifier의 사전 평가 시스템 도입 건의 추후 방향성**
  + **[배경]Patch classifier의 정량적 성능은 버전에 따라 큰 차이를 보이지 않음. 실제로 과거, 가장 높은 수치를 보였던 lossdiff 모델의 경우, 실 사용에서는 좋지 않은 평가를 받은 적이 몇 차례 있음.**
* **WSI Classification Framework Upgrade : better training part – Priority based label smoothing은 따로 추진하고 있는 사람이 있나?**

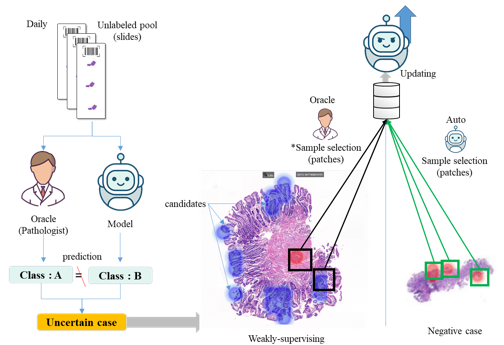


* **현재 GCN / Feature cube 모두 새로운 Dataset에서의 성능 악화로 인해서 Active learning 모듈 도입 전까지 연기한 건가?** 
  + 
  + 

**+ 아래 진행 상황 별 궁금한 것들**

**Active learning 전체 Framework**





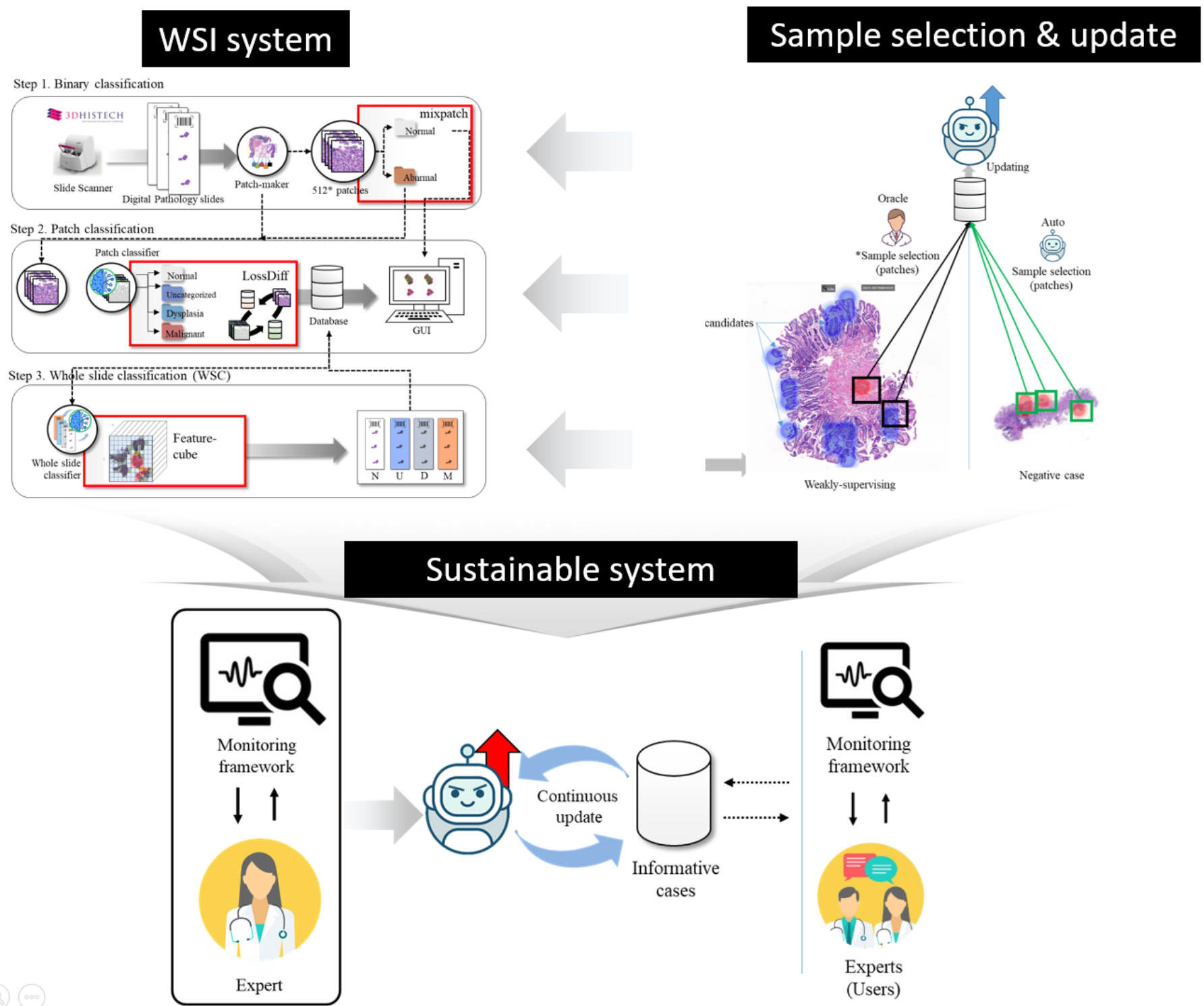
**3. 해당 연도 추진 현황**

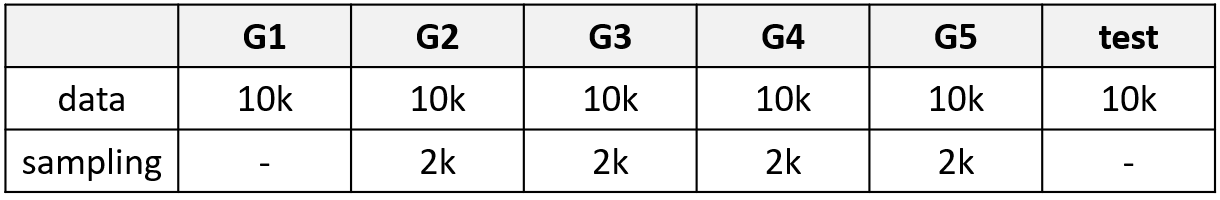
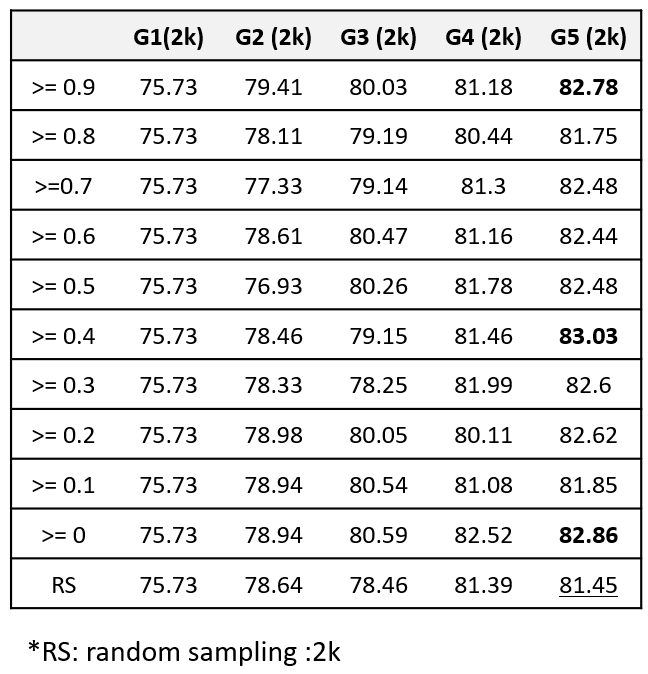
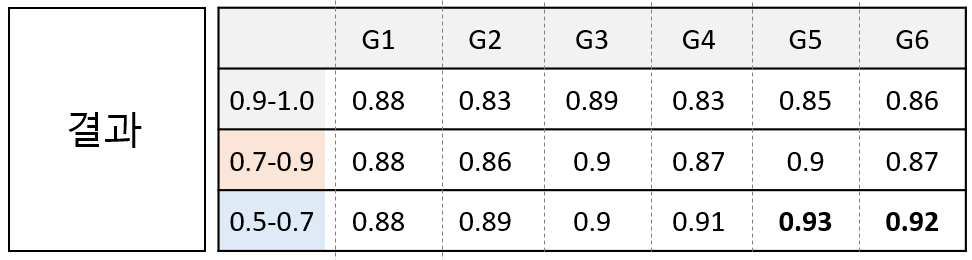
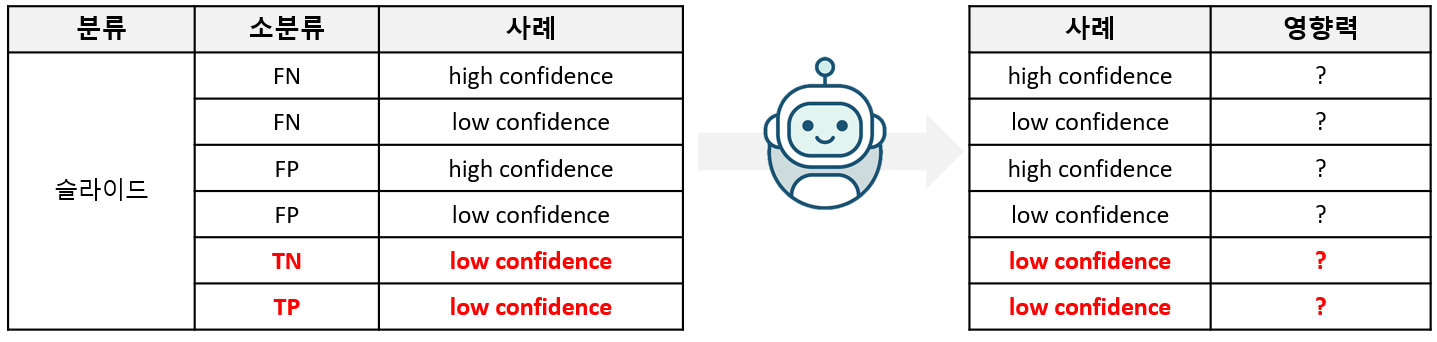
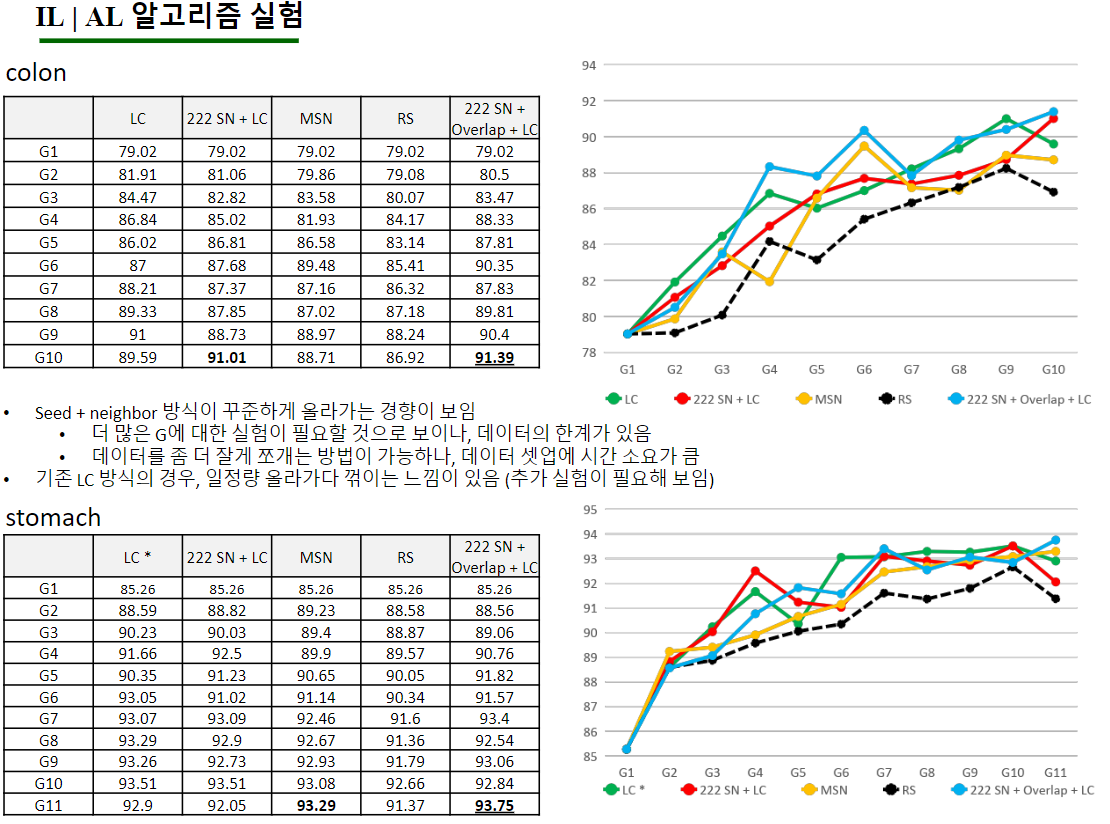
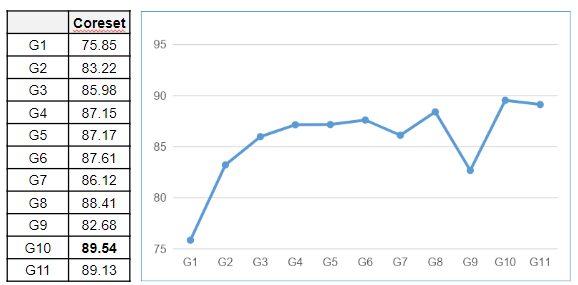
**3.1. 스캐너 이미지 기반의 조직 병리 자동 판독 모델 고도화**

**3.1.1. Incremental learning 시스템 구축**

* **[목표 / 요구 사항] 스캐너의 지속적인 업데이트를 위한 효율적 incremental learning 시스템 구축**

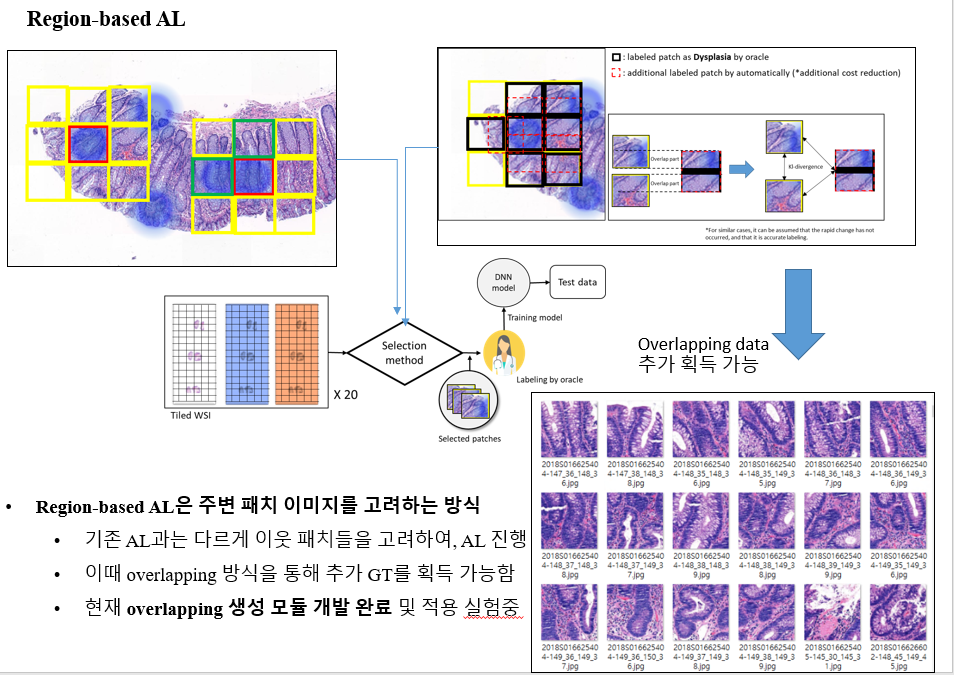
**+ 목적 부분이 좀더 강조가 되어야 할 것 같은데**

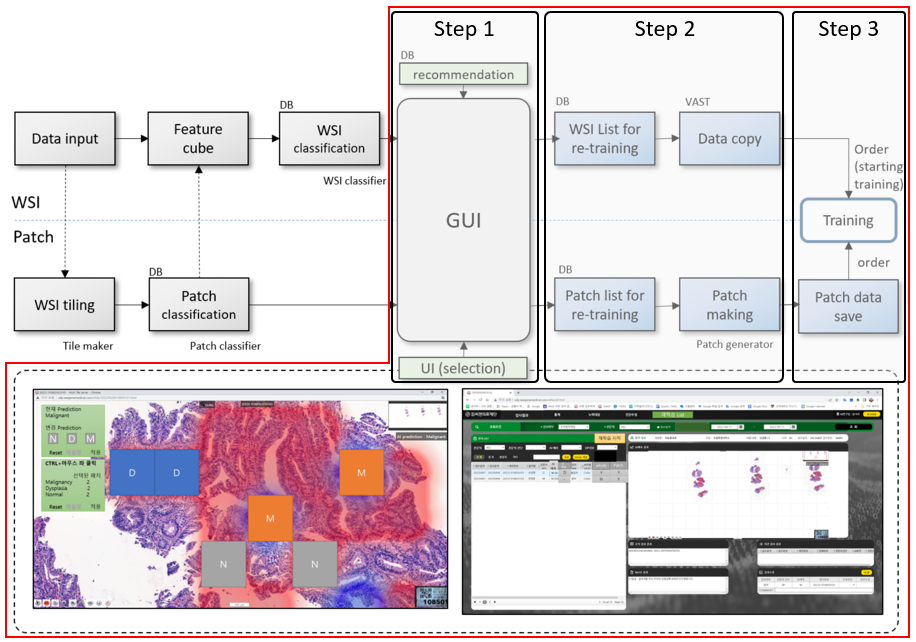


* + 시스템이 변화하는 환경적 요인에 대한 유연한 대응성 획득
  + 업데이트를 위한 Annotation 부담 최소화 및 학습 효율 극대화
* **1) Patch / WSI 별 Selection strategy 설정**
  + - **배경 : 서로 다른 환경으로 서로 다른 전략이 필요함.** 
      * Patch : 일반적인 Active learning 시나리오와 동일함. 라벨 확인 어려움. 노이즈와 혼동될 가능성 존재.
        + Noisy label scenario : 불확실성이 높은 샘플을 노이즈로 보고 학습에서 제외
        + Active learning : 불확실성이 높은 샘플을 선택하여 학습
      * WSI : 일반적인 Active learning 시나리오와 다름. 라벨 확인이 쉬운 확인
  + **결론** 
    - **WSI : GT가 False N / D / M Class 별**
    - **Patch :**
  + **[조치 1] Patch Strategy 방향성 점검 : 낮은 Confidence 샘플 Selection Policy가 일부 성능 개선에 긍정적인 요소임을 확인.** 
    - 점검 방식 : 공개 데이터(Cifar 10) 활용 – Confidence 값에 따른 영향에 대한 사전 실험 진행
      * 총 60K의 데이터를 5개의 그룹(G)와 Test 데이터로 구분. 이후 각 그룹에서 Confidence 기반으로 Sampling 진행.
      * 
    - 
    - 0, 4, 9 구간에서 가장 높은 성능을 냄. Random sampling을 했을 때, 성능이 가장 낮은 것으로 보아 특정 Policy를 선정하는 것이 성능에 영향을 주는 것으로 식별함.
  + Active learning이 잘 작동하지 않은 이유 [Region base 가 추가된 이유/배경]
    - 1) 어려운 예시만 선택하면 대표적이지 않은 feature 들이 학습되어 모델의 품질이 저하됨
      * > 기존 방식은 모델이 확실한 레이블링을 강제로 추가하는 방식을 사용함
    - 2) 단일 데이터 지점 선택에 사용됨
    - 3) 소프트 맥스 아웃풋은 불확실성을 잘 나타내지 않는다.
      * 기존 over-confident 이슈, calibration 방법을 활용함.
    - => 모르는 것만 가르치는 건 의미가 없어 보임. 그렇다고 해서 Random sampling 방식으로 다양하게 가르치는 것은 더욱 효과가 없음.
      * Q. 잘 모르는 패치가 있을 때 정말 주변 패치들을 잘 알면서 자신있게 말을 하는 걸까?
      * => 주변 정보를 추가하여 모델의 정확도를 올릴 수 있지 않을까?
  + **[조치 2] WSI Strategy 방향성 점검 : 전문의의 예측과 차이가 있는 사례 중 낮은 Confidence 기반 샘플 Selection Policy가 모델 성능 개선에 긍정적인 요소임을 확인** 
    - [추가 문제] Over-confidence 이슈 확인 => Calibration 방안(Soft labeling) 적용
    - 3 구간 중심의 실험 진행 : 데이터의 수가 목표 샘플수를 충족하지 못하더라도, 구간 정보를 더 중요시 하여 샘플링
      * 각 구간별 경향성 분석 및 충분한 데이터 수 확보를 위해 3구간 진행
      * Confidence Score가 낮을 때 가장 높은 성능 확인. 낮은 Confidence 기준으로 선발하는 것으로 결정
      * 
    - **~~[조치 2] WSI 업데이트 전략 분석 : 예측 차이 [Q. 이거 연구 결과가 있나?]~~**
      * ~~‘모델-전문가 예측’ 차이에 따른 샘플의 학습 효과 영향 분석~~
      * ~~각 사례에 대한 학습 시 모델의 변화 형태 관찰~~
      * ~~~~
  + [문제] Confidence 값을 기준으로 뚜렷한 경향성 확인 이 어려움
    - 테스트 데이터 재구축(경계선 학습 이슈 우려) / 새로운 추천 방안 개발(Region – based )
      * 경계선 학습 이슈 : Loss diff로 필터링 진행(?)
  + **[결과]** (Patch) Region-based 형태로 Sample selection 진행 / WSI 날짜 기준 각 클래스 별로 가장 낮은 Confidence 값을 가진 20개 샘플 선택
* **[조치] 2) Active learning 아키텍쳐 업데이트** 
  + AL 알고리즘 성능 점검 [최종적으로 무언가를 쓰자는 결정이 있었나?]
    - 
      * **Seed + Neighbor 방식이 꾸준히 성능 향상하는 경향이 보임**
      * **Overlap / Seed 등 다양한 요소 성능 점검. Hyperparameter test 느낌.** 
        + **Overlapping 방식의 문제점 : 1) Context 정보를 포함하지 못하는 데이터 생성 가능, 2) 어려운 케이스의 데이터 생성 => Loss diff를 통한 필터링 고려중**
      * **Region base AL이 다른 기존 알고리즘에 비해서 0.4% 정도의 좋은 결과를 보여줌. 추가적인 실험이 필요해보임.**
      * **Coreset 방식은 다른 방식에 비해 성능이 떨어질 뿐더러, 연산 시간도 매우 오래 걸렸다.**
      * 

**Region base AL : 모듈**

* **주변 패치 이미지를 고려하는 방식** 
  + **현재 overlapping 생성 모듈 개발 완료 및 적용 완료**



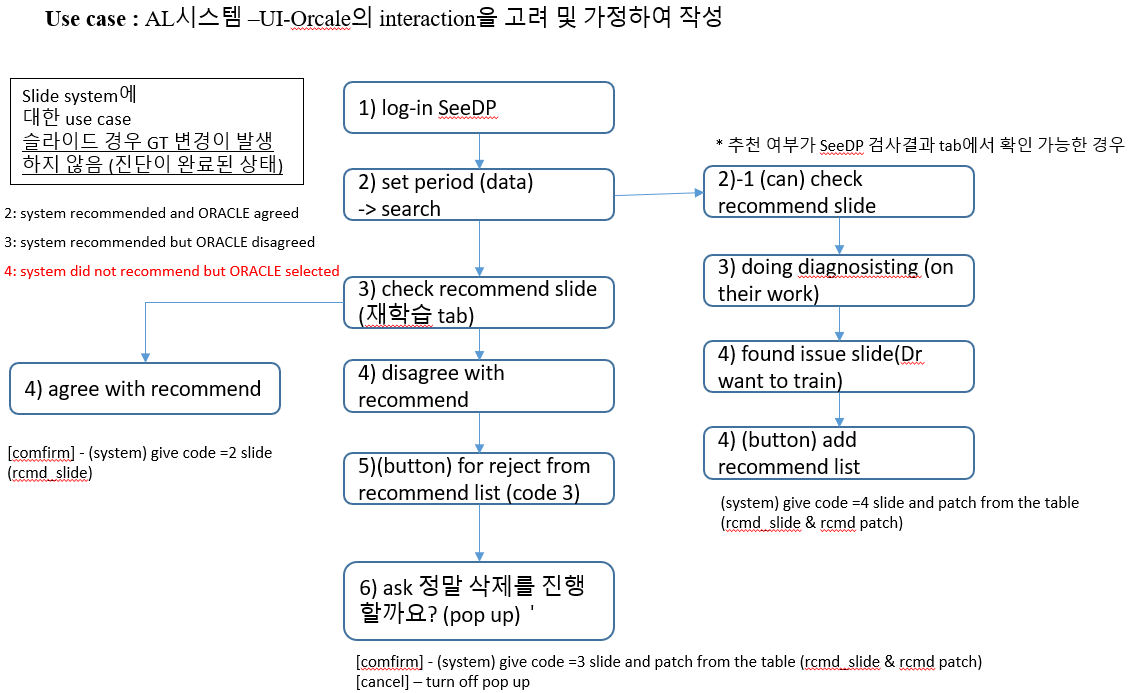
* 2) 시스템 Workflow 및 DB 구조 설계
  + **배경 : QC 목적으로 사용되고 있다.**
* **왜 전문가가 개입하게 되었는가에 대해서 배경에 대해서 알고 싶다.** 
  + **대부분의 Active leanring 연구의 경우 주어진 데이터셋에서 중요한 데이터를 선별하는 Pool-based sampling 방식에 초점을 잡는다. 하지만 현 프로젝트는 계속해서 새로운 데이터가 추가되며, 스캐너의 지속적인 업데이트로 인해 새로운 변인이 추가되는 등 데이터셋이 가변성을 띈다는 큰 차이점을 가진다.**
  + **이런 환경 속에서 지속가능한 판독 시스템을 구축하기 위해서는 지속적으로 Clean한 labeled 데이터를 확보할 수단이 필요하다.**
  + **Clean labeled 데이터를 확보할 방안은 전문의 진단 뿐이다. 모델은 새로운 변인이 발생할 경우, 예측이 부정확할 수 있다. 예측이 틀린 사례를 모델의 학습에 활용할 경우 confiremation bias를 유발할 수 있다. 즉, Clean label 데이터 확보를 위해 전문의의 검수를 필요로 한다.**
* **전문가의 개입을 포함한 전반적인 시스템 Workflow을 구상할 필요가 있다. 그 중 하나가 GUI 딴이다. 이를 위한 창구로써 GUI 딴을 구상할 필요가 생겨졌다.** 
  + **또한 전문의의 선택이 추후 Training 파트에서의 하나의 조건으로 고려되기 위해서 전반적인 DB 설계를 필요로 한다.**
* **Framework**
* 

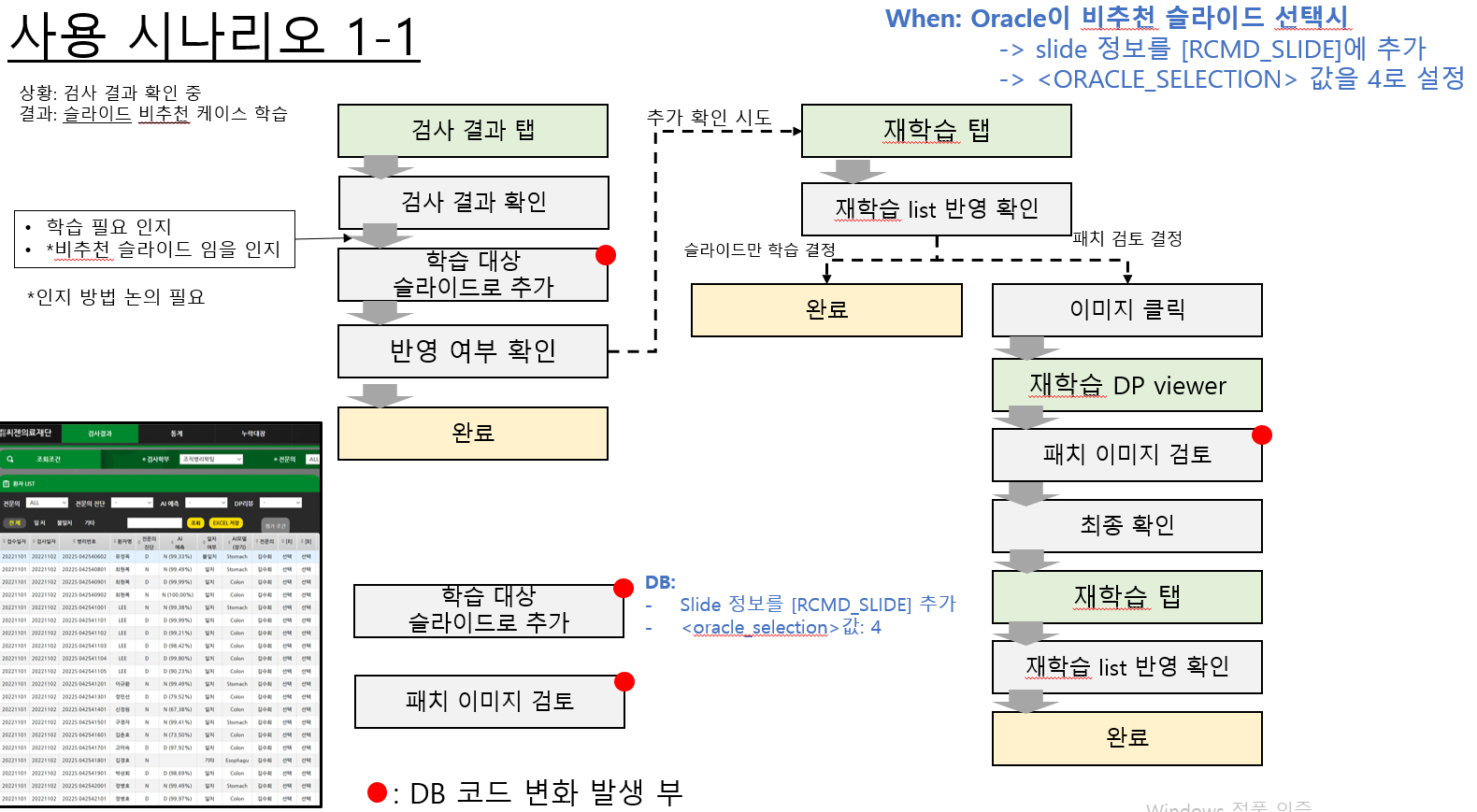
**[시스템 관점 Work Flow]**

* **그림 0의 빨간 영역이 AL 시스템이 추가된 영역으로, 크게 (1) 재학습 데이터 선정, 2) 재학습 데이터 생성, 3) 재학습 데이터 학습 영역으로 나눠짐.** 
  + **(1)은 씨젠 홈페이지 재학습 탭에서 표현된다. 앞서 정립한 WSI 추천 전략에 따라, 하루 당 각 20개의 N/D/M Class WSI을 [재학습 추천 list]에 추가한다.**
  + **또한 전문의의 선택이 추후 Training 파트에서의 하나의 조건으로 고려되기 위해서 전반적인 DB 설계를 필요로 한다.**
* **Framework**

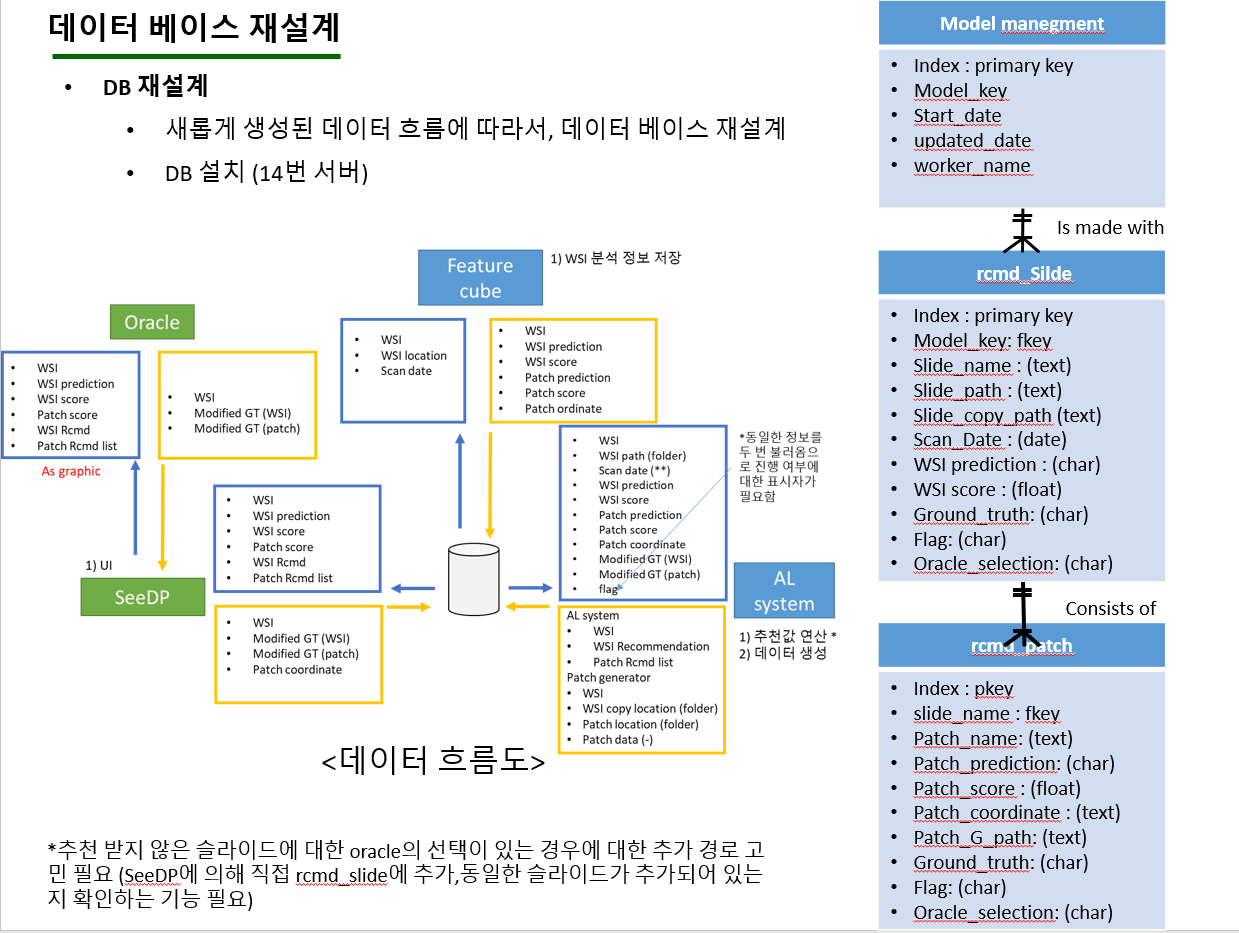
GUI 관련하여 전문의 관점에서의 접근 방식

* 이번 단락에서 설명할 내용
* User case 설명하기



=> 이렇게 까지 구체적인 사용 시나리오는 몰라도 된다. 이건 실제 유저 관점인 거고

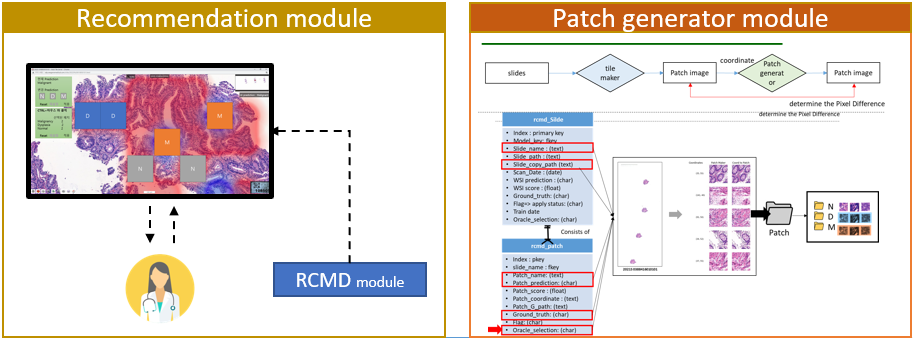
- 전반적인 Work Flow를 참고할 수 있을 정도면 된다.



* 전반적인 DB 설계 구상을 앞 딴에서 설명해줘야 양의 비중이 맞겠다.
* 최신화된 DB 구조를 기반으로 그림 그리기

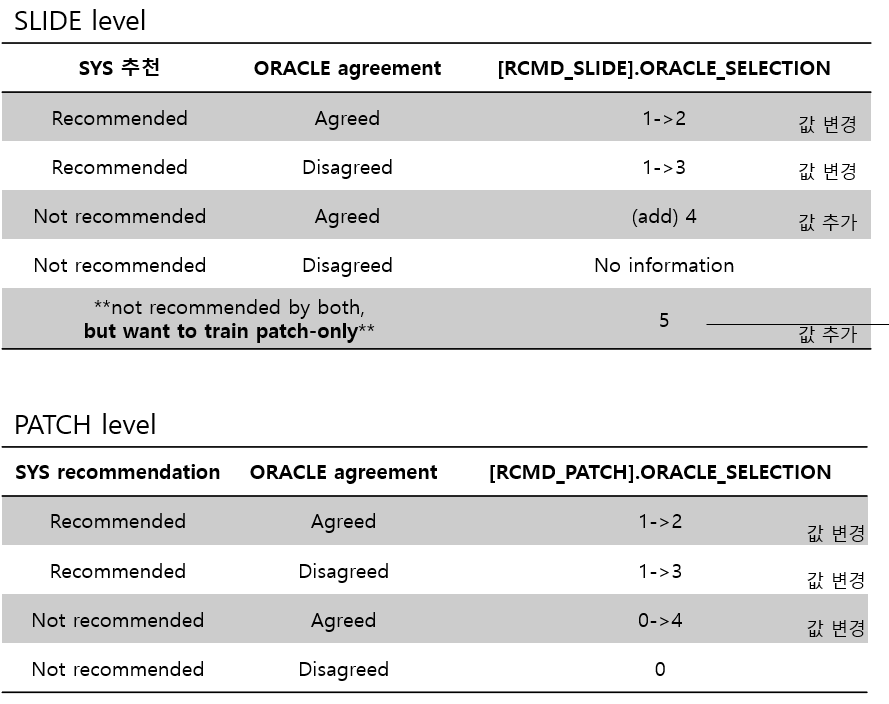
1. **추천 및 훈련 파트 개발 : PPT에 있는 내용 넣기**

**추천 파트**

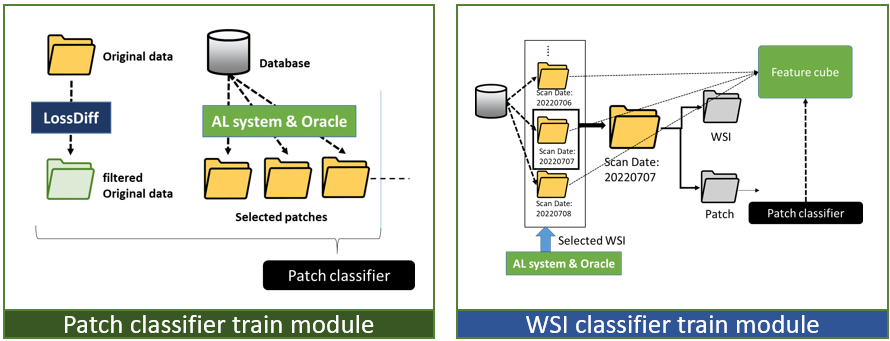


* + **WSI 추천 모듈 작동방안 [위치 최종확인하기]**
    - **위치 : server#266 >> Computer/vast/active\_learning/slide\_recommendation**
  + **Patch 추천 모듈 작동방안** 
    - **위치 : server#226 >> Vast/active\_learning/patch\_recommendation**
    - **작동 방식 : Terminal> go to file location > “python3 main.py”**
    - **Result** 
      * **패치 정보가 [recommend\_patch] 테이블에 복사**
      * **Confidence score가 Threshold 값보다 낮은 patch의 경우 oracle\_selection의 값을 1로, 해당 patch의 주변에 있는 8칸(상하좌우)에 대해선 2의 값으로 변경**
  + **작동 방식** 
    - **1단락에서 설명한 것처럼 confidence score을 기반으로 하여 Ranking 순으로 진행한다.**

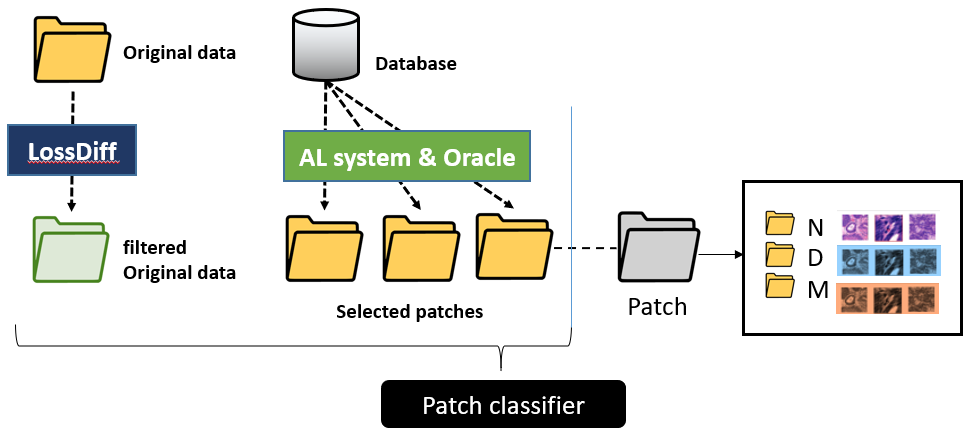
DB 반영 방식에 대해서 서술할 것

* 아래 표 느낌처럼 변경할 것
* 
  + - **Patch generator 설치 위치** 
      * **작동 방안**
      * **Result 설명하기**

**훈련 파트**

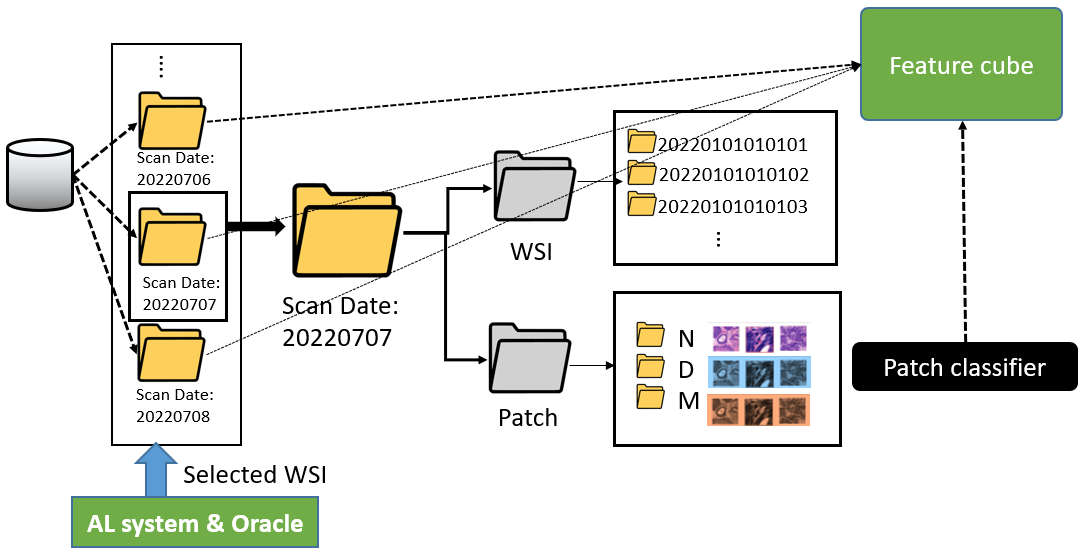


* **Patch classifier train module**



**훈련 대상 서술하기**

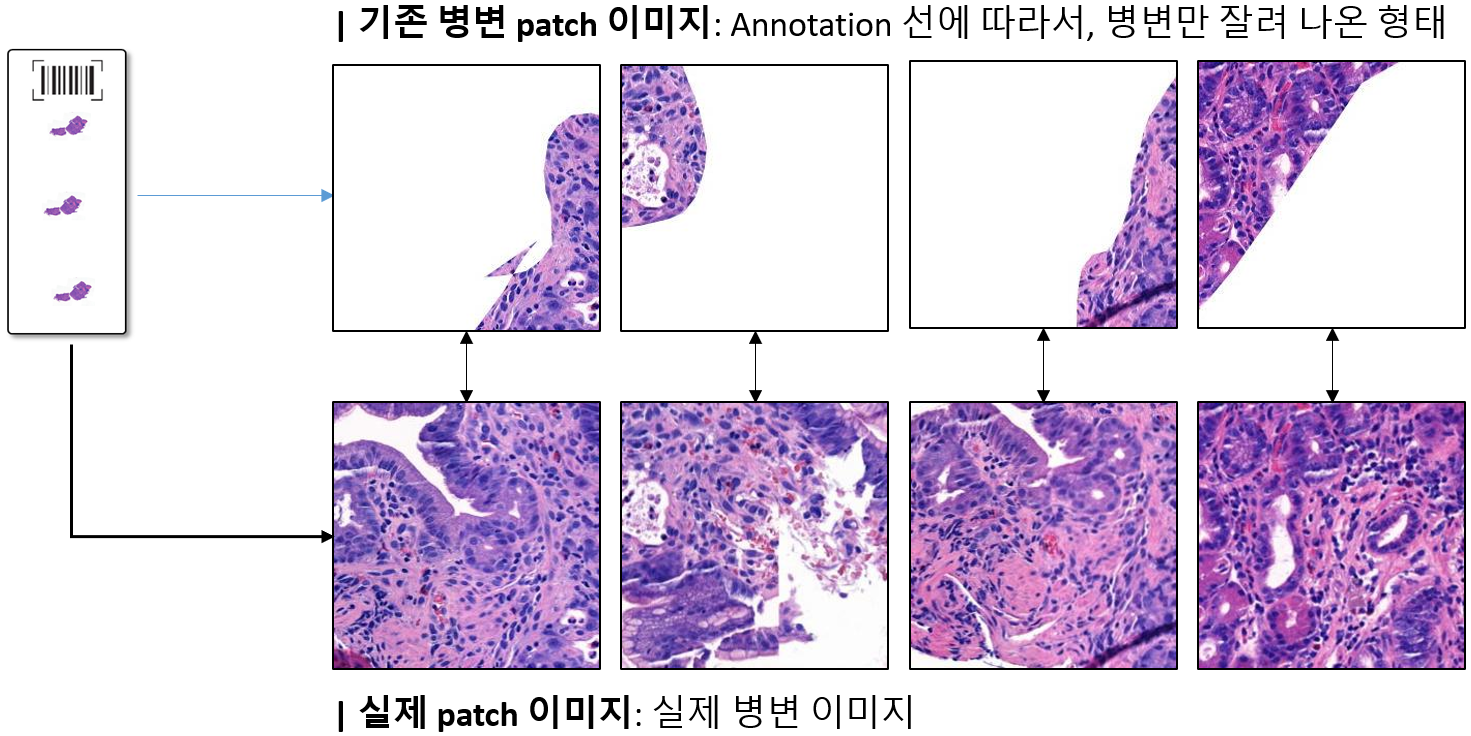
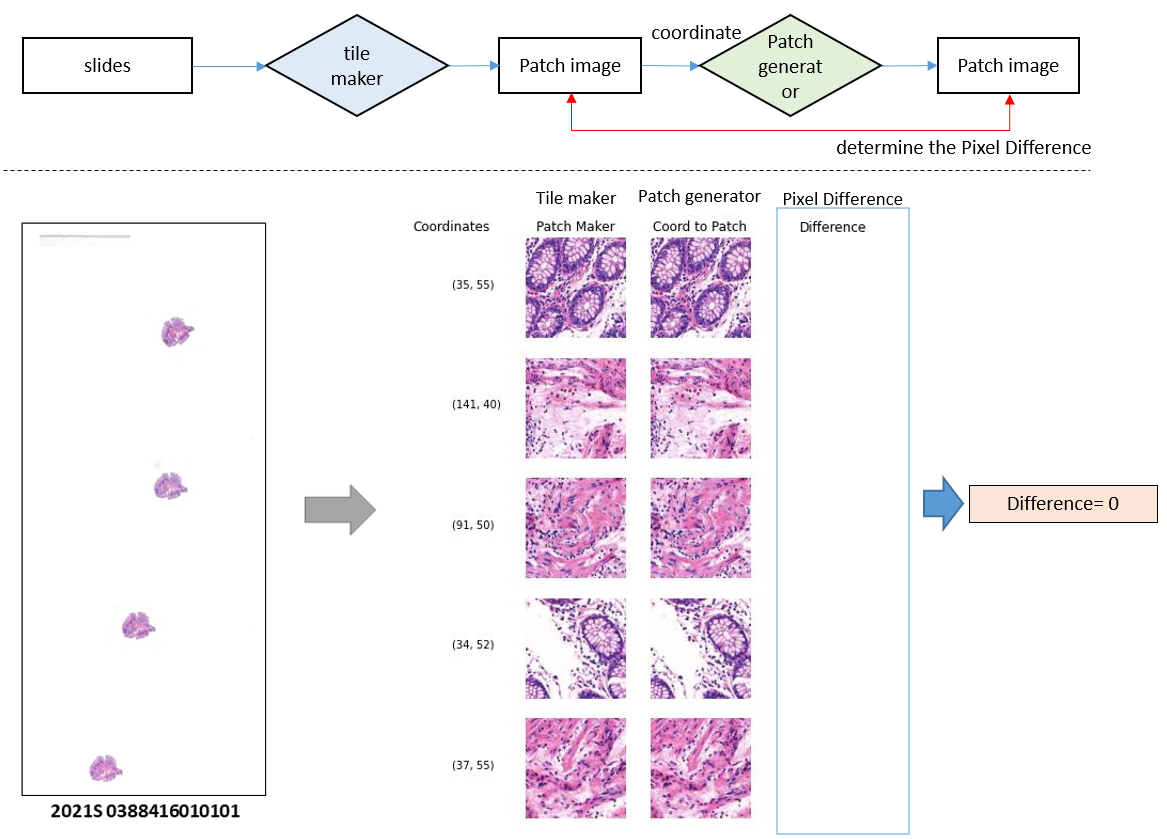
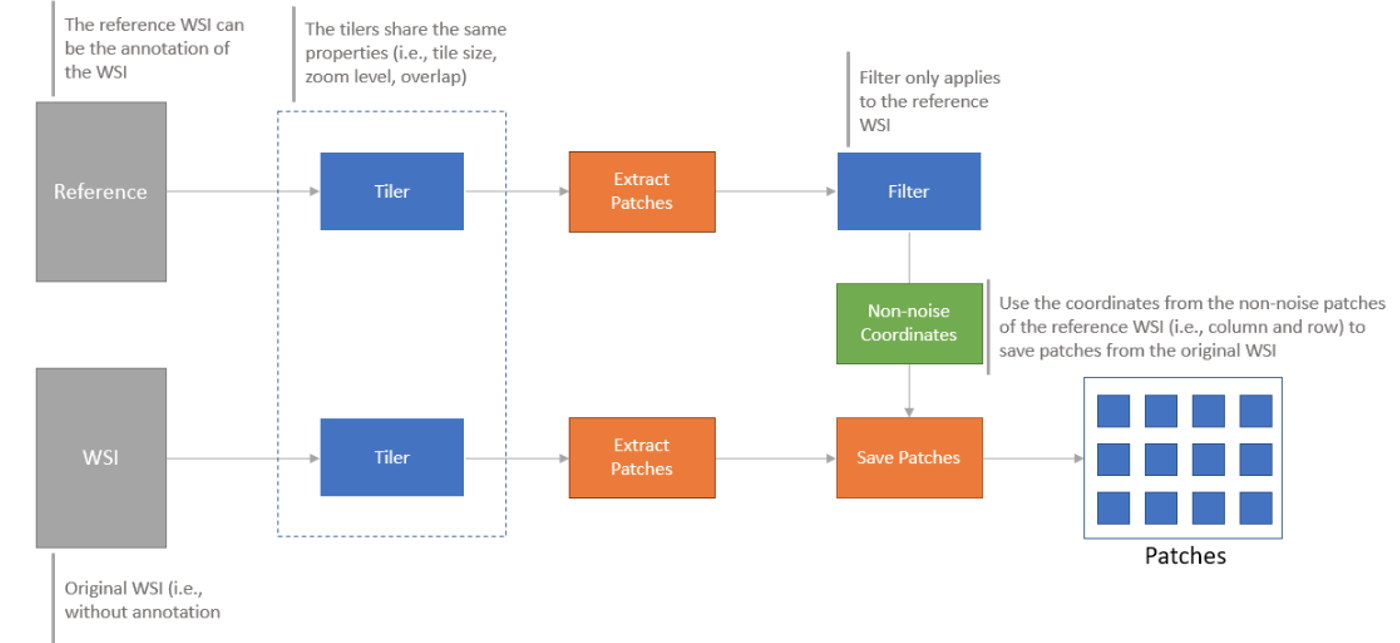
* **Sys\_oracle selection =1 인 것 위주로 선택. 이유도 같이 설명해주기**
* **여기에 앞 딴에서 DB 변경 스토리 라인을 설명해줘여 하나?** 
  + **Lossdiff : noisy-label 및 noisy-data 처리의 목적**
  + **AL system의 특수성** 
    - **AL 추가될 데이터는 noise 발생 가능성이 현저히 낮음 : 사실상 전문의에 의한 전수 검사**
    - **D/M case : 전수 검사에 가까운 데이터가 추가될 예정**
    - **N case : 일부 데이터의 자동 선택이 발생함.**
    - **=> Lossdiff를 따로 거치지 않아도 된다.**
  + **기능** 
    - **Train.py** 
      * **Start\_date : 학습할 범위의 시작 날짜**
      * **End\_date : 학습할 범위의 끝 날짜**
    - **Config.py** 
      * **Ori\_data\_path : 오리지널 데이터 위치**
      * **Selec\_data\_path : 선택 데이터 폴더들이 저장된 상위 폴더**
    - **Output** 
      * **학습 모델의 주요 정보를 담은 csv 생성 (ex- Densenet201\_20220801\_150654.csv)**
      * **학습된 모델의 pkl 파일 (ex- 20220801\_220854\_model.pkl)**
* **WSI classifier train module**



* + **기능 : 모듈이 학습 대상 기간에 해당하는 각 folder를 읽어와서 학습할 예정** 
    - **Database : 현황 등**
    - **User Input :**
  + **학습된 Densenet201 저장**

+ 좌표 기반 모듈 개발(Tiler)

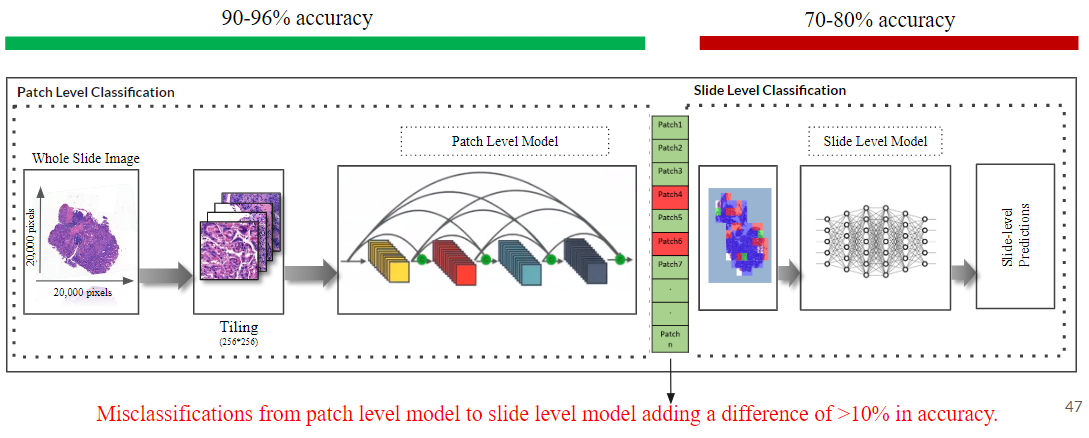
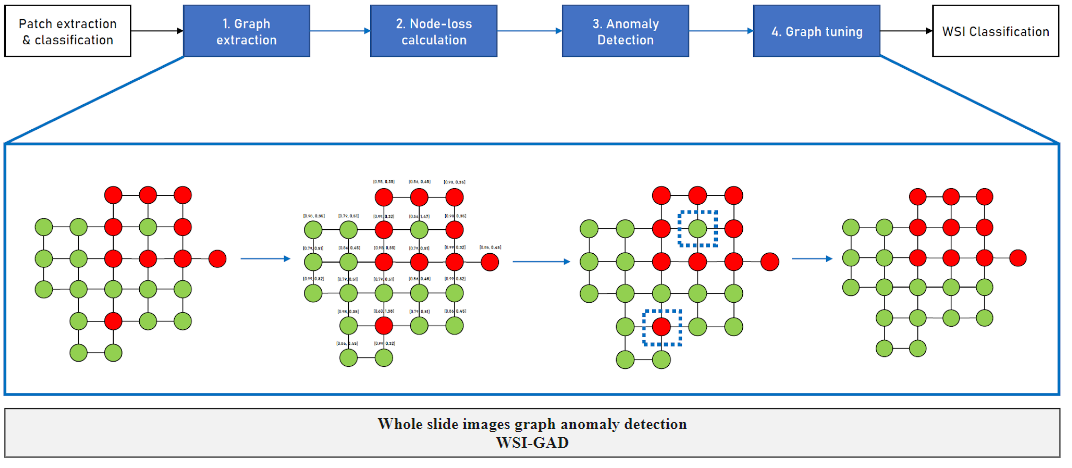
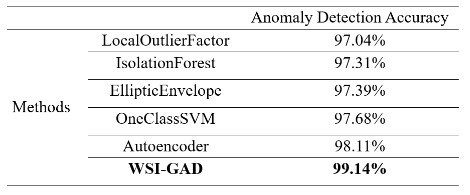
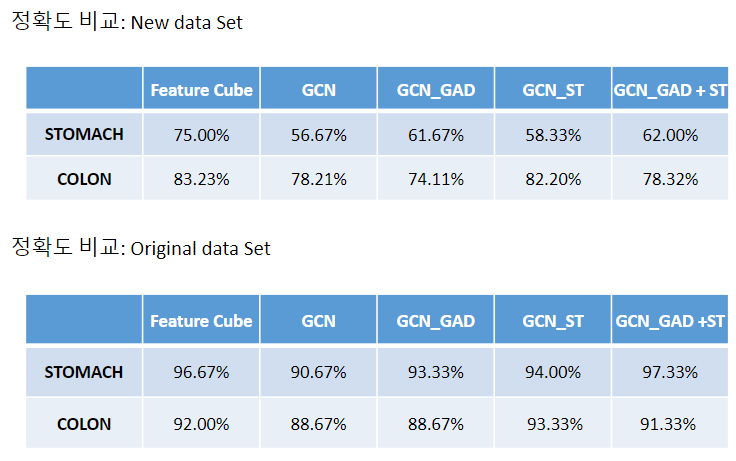
문제점 : 1)

* Annotation only 데이터의 경계선 학습 이슈의 우려 발생
* 
  + **좌표 기반 패치 선별 방법 도입** 
    - **[배경] Real-world 이미지와 차이가 존재 / Edge 파트에 대한 Bias 유발** 
      * **실제 Prediction 시에는 Annotation 정보가 없으나, 학습에 사용되는 이이미제는 포함이 되어있었음**
      * 
      * **조치 : 좌표 기반 패치 선별 방법 고안**
      * **기대 사항 :** 
        + **Edge bias 해결 기대**
        + **[데이터 차이 완화]Active learning 시 생성되는 병변 패치와 동일한 형태**
        + **Active learning의 패치 추천을 위한 새로운 calibration 방법 적용 가능**
    - **모듈 점검 간 생성 간 위치 차이가 없음을 확인함** 
      * 
    - **[조치] Tiler 모듈 도입 및 통합 완료 (추가 확인 필요)** 
      * 
* AL 파트 테스트 데이터 변환
  + 슬라이드 기준으로 변환
  + Region-based의 적용 => 슬라이드 기준으로 변환?
  + 실제 현업 업데이트 시나리오와 유사한 환경
    - 기존) 1) 무작위 WSI의 무작위 Patch 데이터 Pool / 2) 인공적 절단면 존재
    - 현재) 1) WSI 단위 데이터 Pool 생성, 2) 인공적 절단면 제거

**모듈 추가 개발**

* **Region based 모듈**
* **Ex)- Overlap 기능 추가**

**3.1.2 Graph CNN 모델 업데이트**

* **[문제점 식별 / 현황 분석]** 일부 슬라이드에서 다수의 그래프 생성이 확인되어 대응 방법 연구중
* **[목표] 기존의 분류하지 못하던 암세포의 작은 영역을 포함하는 Colon 슬라이드를 예측함으로써 기존 WSI Classifcation 방법인 Feature cube을 지원한다.**
* **[조치1] Majority voting 성능 실험 (결론이 나온게 있나?)**
  + Hard voting : 각 클래스 레이블에 대한 예측을 합하고, 가장 많은 투표로 클래스 레이블을 예측
  + Soft voting : 각 클래스 레이블에 대해 예측된 확률을 합하고 가장 큰 확률을 가진 클래스 레이블을 예측
  + Hard voting + WSI Classifier : Hard voting 이후 Tie 발생 시 WSI 사용
  + Hard voting + Soft voting : Hard voting 이후 Tie 발생 시 Soft voting 사용
* **[조치 2] 신규 Framework Update : 5 종류의 WSI Classification framework 성능 비교 실험** 
  + 신규 모델 : WSI-GAD 제안
    - 
  + [WSI- GAD 목표] Patch 기반에서 Slide-level 모델로 넘어갈 때 Error propagation을 줄이자. Patch 딴에선 정확도가 90~96% 달성해도, Slide 딴으로 넘어가면 성능이 악화된다.
    - 
  + [방안] Patch Classification의 결과가 잘못된 것(Noise)를 Filtering 하자!
    - 
  + [결과] Patch Anomaly detection 방법 중 정확도가 가장 높았다. / 단, WSI 딴에서는 성능이 Feature cube가 더 높았다.
    - 
    - Colon and Stomach 정확도 결과 모두 Feature cube > WSI\_GAD(무타자 코드) > WSI 순으로 확인. WSI\_GAD(무타자 코드) + Tissue-level classification 결합형 모델 더 높은 정확도를 제공. Colon 에서 정확도가 더 높은 것을 확인할 수 있음
* [문제점 식별] WSI Classification에서 신규 데이터 셋에서 모든 모델의 성능 악화 식별됨.
  + 
  + **Active learning 데이터 수집 이후로 연기** 
    - **AL로 수집된 데이터를 기반으로 사전 셋업된 모델에 학습하여 정량적 모델 평가 진행**
* **[조치] Tissue commonality 연구 진행** (지금 진행도는 어떻게 되어가고 있나?)
  + **목표 : 이미지 간의 공통 패턴 탐색을 통해 input 정보 향상**

**3.1.3 프레임워크 고도화 :** 1) 4 class 모델에 대한 연구, VIT 기반의 시스템, 대상 장기의 범위를 확장

1) **[Class 확대] (진행되고 있나?)**

2) **[VIT] (확인 필요 )**

3) **[적용 범위 확장]** (진행되고 있나?)

+ Small dataset experiment?

* **[조치?] Lossdiff [이거 2022년도에 개발된 방법이 맞나? 즉, 보고서에 담는게 맞나?]**
  + [배경] 슬라이드 조직 마킹 과정에서 패키 기반의 Label Noise 발생 가능성이 높음
  + [방안] Training loss 값을 기준으로 학습에 도움이 되는 유효한 패치들을 분류하고 저장
    - Memorize Effect : Deep Network가 초기 Epoch에서는 Clean하고 쉬운 패턴을 학습하고 이후에는 세부 패턴을 학습하는 현상. 이로 인해 초기에는 Clean data를 학습하고, epoch가 증가할수록 noisy data에 overfit 한다.
  + **=> 사후가 아닌 ‘사전’에 Harmful 데이터는 피하는 동시에 모델이 잘 모르는 데이터를 선택하는 방법을 찾자** 
    - **LossDiff 는 Harmful 데이터의 정보(기준)를 생성**
    - **Leraning loss는 Training loss 값을 예측하여, 데이터 selection에 활용**
  + **LossDiff : Noisy-label 및 noise-data 처리의 목적** 
    - **AL 추가될 데이터는 noise 발생 가능성이 현저히 낮음 : 사실상 전문의에 의한 전수 검사 => Low confidence**

3.3. 연구 추가 성과

3.3.1 추가 달성 모델

* **[추가 달성 유무]**
* **[설명]**
* **[기존 한계점]**
* **[문제 분석]**
* **[새로운 대안]**
* **[결과]**

**4. 2차년도 연구 개발 계획**

**4.1. 2차년도 기술 개발 범위**

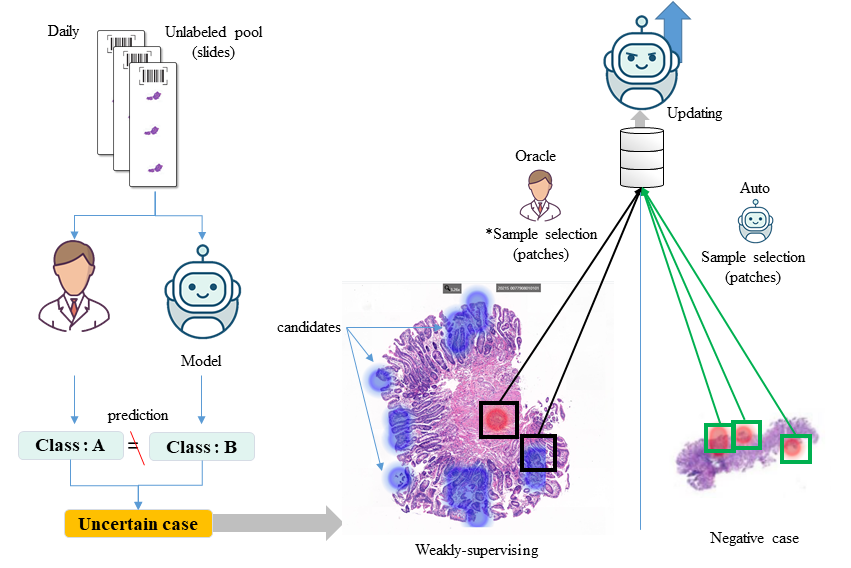
* **[큰 단락] 2차년도의 기술 개발 범위는 크게 1) 스캐너 이미지 기반의 조직 병리 자동 판독 모델 고도화 부분과 2) 현미경 이미지 기반의 조직 병리 자동 판독 모델 연구부분으로 나눌 수 있음**
* **[영역 구체화 – 시스템 발전] 또한, 각 연구 범위에 대한 주요 연구 내용은 스캐너 기반 판독 시스템 고도화와 현미경 기반의 판독 시스템 연구를 포함함**
* **[다음년도 주요 골자] 스캐너 이미지 기반의 조직 병리 자동 판독 모델 연구**
* 스캐너의 지속적인 업데이트를 위한 **incremental learning 시스템**을 구축하고자 함
* **Graph CNN 모델 업데이트**를 통해서 작은 병변에 대한 대응력 강화하고자 함
* 추가적으로, 프레임워크 고도화를 위해, **4 class 모델에 대한 연구, VIT 기반의 시스템 등을 고려하고 있으며, 대상 장기의 범위를 확장하는 연구를 시도**해보고자 함 (예시, 유방암, 전립선암 등)
* **[표 26~27]는 2차년도 기술 개발 계획에 대한 간트 차트이며, 각 연구 개발의 내용과 각 내용에 대한 추진 일정을 나타냄**
* 해당 표는 추진 일정에 대한 대략적인 기간을 포함하며, 일부 시스템의 경우 설치 및 사용법 전달을 위해 해당 과정에 대한 기간이 확장될 수 있음

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2차년도 – 스캐너 기반 조직 병리 분류 시스템 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 일련  번호 | 연구 개발 내용 | 추진 일정 | | | | | | | | | | | | | | | | | | | | | | | | 기간  (주) |
| 3 | | 4 | | 5 | | 6 | | 7 | | 8 | | 9 | | 10 | | 11 | | 12 | | 1 | | 2 | |
| 1 | 요구 패치 생성 모듈 구현 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 16 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | ‘프레임 워크 학습 시스템 구현 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 16 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Increamental learning system 테스트 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 12 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | WSI 그래프 유사도 모듈 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 12 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | 조직 chunk 단위 voting 모듈 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 12 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Voting 방식 학습 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 14 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Spatial 모델과 앙상블 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 10 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | 대상 장기 확장을 위한 전략 연구 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 16 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 | 추가 대상 모델 가능성 테스트 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 14 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

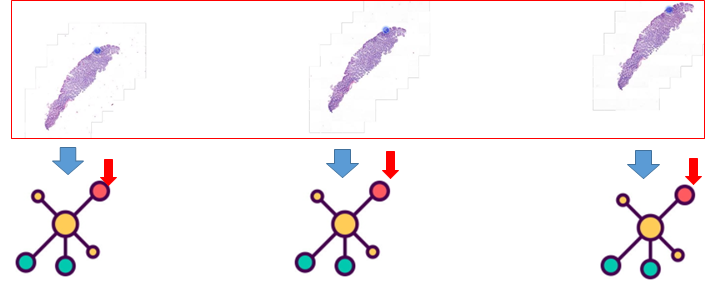
**[프로젝트 일정 기한 제시 필요]**

**4.2. 스캐너 기반 판독 시스템 고도화**

* **스캐너 이미지 기반의 조직 병리 자동 판독 모델 연구**
* **1) 스캐너의 지속적인 업데이트를 위한 incremental learning 시스템을 구축하고자 함 [세부 설명]**
* **[목표]** 해당 방식은 active learning의 방법을 기반으로 전문의와 모델의 예측 결과가 다른 경우에 대해서 모델이 학습하기 위한 환경을 생성하고 이를 통해 지속가능한 시스템이 될 수 있게 개선하기 위한 방법임
* **[방식]** <그림 33>과 같이 negative case에 대해서는 모델이 스스로 학습하도록 하며, 병변이 있는 case에 대해서는 전문의의 도움을 받아 특정한 패치 이미지들을 학습용으로 생성하고 모델이 학습하게 할 계획임



* **2) Graph CNN 모델 업데이트를 통해서 작은 병변에 대한 대응력 강화하고자 함**
* **[목표]** 현재 하나의 슬라이드에서 생성되는 여러개의 조직 chunk들에 대해서 종합적으로 정보를 고려하고 있지 못하기 때문에, 해당 chunk들에 대한 정보를 종합적으로 고려하는 시스템을 생성하고자 함
* **[방법]** <그림 34>과 같이 하나의 슬라이드에서 각 조직 chunk에 대해서 별개의 graph를 생성하고 graph간의 유사도를 측정할 계획임. 측정된 유사도를 기반으로 높은 유사도를 보이는 사례에 대해서는 병변 조직의 spatial 정보 이외에 작은 병변이 발생해도 해당 슬라이드를 병변이 있음을 인식하도록 모델을 구성하고 실험할 계획임
* **[기대효과]** 해당 방법을 통해서 모델의 안정성을 향상시키고 작은 병변에 대한 대응력을 향상 시킬 수 있을 것으로 기대함



* [추가 목표] 추가적으로, 프레임워크 고도화를 위한 방법을 연구하고 업데이트를 계획하고 있음. 해당 고도화를 위해서 전문의와 실제 AI 프레임 워크를 사용하는 사용자들과 인터뷰를 진행하고, 이슈들을 파악하여 대응할 계획임.