

MIL과 CLAM 방법론을 이용한 병리 슬라이드 분류

김가은, 나다경, 윤우성, 이주현

Abstract

Along with the development of various hardware, the development of artificial intelligence made it possible to introduce deep learning technology in the medical field. Among them, the Whole Pathological Slide Image (WSI) plays an important role in diagnosing the disease, including the cause of stroke. However, since WSI is very large in size, whereas the size of abnormal areas is microscopically small, the process of selecting and analyzing WSI is quite difficult. Recently, the objective of diagnosis using deep learning models has become detecting the abnormal area without patch-level or pixel-level label, but only with slide-level label, which is called weakly-supervised learning. Accordingly, multi-instance learning may be a great advantage in diagnosing the cause of the stroke. In this paper, we address this problem by introducing a multi-instance deep learning method (MIL) and Clustering-constrained Attention Multiple instance learning (CLAM), which introduced the Attention Network using the MAYO STRIP blood clot dataset. We first select discriminant instances and then use these instances to diagnose diseases based on the proposed MIL approach. In addition, CLAM, which applied the Attention Pooling method to the basic MIL methodology, was confirmed to utilize the information at patch-level (instance-level) more abundantly by utilizing a technique called clustering.

Keywords: MIL, WSI, CLAM, Stroke, Weakly-Supervised Learning, Attention Network

1. 서론

뇌졸중은 전 세계적으로 사망 원인 2위이며, 매년 700,000명 이상의 사람들이 뇌로 가는 동맥을 막는 혈전으로 인한 허혈성 뇌졸중을 경험한다. 재발한 뇌졸중(A second stroke)은 환자의 생존 가능성을 낮추지만, 의사가 뇌졸중 발병 후 치료 관리에 영향을 미치는 뇌졸중 원인을 알 수 있는 경우 후속 뇌졸중 발병을 낮출 수 있다. 기계적 혈전 절제술(mechanical thrombectomy)은 큰 혈관 폐색으로 인한 급성 허혈성 뇌졸중의 치료 기준으로 자리매김했다. (문지현, 2021) 그 결과, 회수된 응고(retrieved clots)는 분석에 적합하게 되었고 의료 전문가들은 허혈성 뇌졸중 병인과 혈전 기원을 예측하기 위해 딥러닝 기반 방법을 적용하고 있다. (유재학 등, 2020)

우리는 Whole Slide Image(=Whole Slide digital pathology Image)를 사용하여 두 가지 주요 급성 허혈성 뇌졸중(AIS) 원인 유형인 심장 및 대동맥 아테롬성 동맥 경화증(cardiac and large artery atherosclerosis)을 구별하는 모델을 구축하는 것이 최종 목표이다. 병리학 슬라이드를 활용해 분류를 진행해야 하는 특수성을 지녔기에 일반적인 분류 모델(CNN)을 사용한다면 WSI의 특성을 고려하지 못해 효과성이 매우 떨어질 수 있다. 따라서 우리는 이런 문제를 해결하기 위해 작은 패치로 나눈 후 Shujun Wang 등 (2020)이 제안한 MIL 방법론을 이용해서 분류를 허혈성 뇌졸중(ischemic stroke)의 혈전 기원을 분류하고자 한다. 기본적인 MIL 방법을 3가지 모델(EfficientNetb0, ResNet18, RegNet16)로 실험을 진행했다. 추가로, MIL 방법론 후에 Attention Pooling을 진행한다면 더욱 효과적인 성능을 낸다는 (Ming Y.Lu 등, 2020) 논문에 근거해 CLAM이라는 방법론으로 실험을 진행했다.

본 연구에서는 병리학 이미지 분류 문제에서의 기본적인 CNN 모델과 MIL과 CLAM이라는 방법론의 성능을 비교하고자 한다. WSI에 최적화된 효과적인 분류 모델을 개발함으로써 치명적인 뇌졸중의 혈전 기원을 더 잘 식별할 수 있게 하여 의사들이 뇌졸중 후 최상의 치료 관리를 처방하게 하고, 후속 뇌졸중의 가능성을 줄일 수 있게 할 수 있는 의의가 있다.

본 보고서는 총 8장으로 구성되어 있다. 2장에서는 선행 연구로 기존의 모델들에 관하여 설명하고, 3장에서는 제안하는 모델의 세부적인 부분 및 구성을 자세하게 설명한다. 4장은 실험에 사용한 데이터에 관한 설명과 실험에 적용한 파라미터 처리 및 결과를 제시한다. 5장은 이 논문의 결론을 서술한다.

2. 선행 연구

2.1 병리 슬라이드 딥러닝 연구

딥러닝 기술이 개발되기 이전에는 WSI는 통상적으로 Hematoxylin and Eosin으로 염색된 후에 병리학자들이 현미경을 통해 진단을 내리는 방식으로 진행됐다. 그러나 방대한 조직의 표본 수와 실험 중간에 발생하는 오염이나 패치의 소실로 인해 문제점이 많았다. 하지만 최근에는 스캐너 기계의 발전으로 저렴한 비용과 더불어 빠른 속도로 데이터를 확보할 수 있었고, 이를 바탕으로 WSI에 대한 대규모 데이터 세트를 구축할 수 있게 되었다.

이는 자동화된 진단 시스템을 위한 기계학습 구축에 이용되고 있지만, 크게 세 가지 문제에 직면해 있다. 첫째, WSI의 데이터가 매우 커서 자동화된 슬라이드 분석 파이프라인에 구현하는 것은 시간이 길고 비용이 많이 든다. 둘째, 광범위한 이미지 사전 처리가 필요하다. 비일관성과 인체 조직 고유의 이질성을 해결하기 위해 색상 정규화(Color Normalization), 공간 필터링(Spatial Filtering), 노이즈 제거(Denoising), 등의 작업을 거쳐야 한다. (Khan AM, 2014) 이후 관심 영역탐색(Detection of region of interest : ROI) 과정을 진행해야 한다. 자동화된 세포 이미지 탐색 및 분할은 세포 계측학에 있어서 매우 중요하고 기본적인 단계이며 각종 암 진단 및 예후 평가에 있어서도 마찬가지이다. (Chen JM, 2017) 셋째, WSI 워크플로의 재현성이 보장되지 않는다. 환자데이터를 기반으로 한 새로운 통찰력의 경우 결과를 재현할 수 있는 것이 매우 중요하지만 현재 솔루션들은 대부분 임시변통인 경우가 많다. 또한 기계학습 모델학습에 사용되는 여러 버전의 데이터와 실험을 추적하는 데에도 효율적이지 않다는 문제점이 있다. (송상용, 2017). 따라서 이를 해결하기 위해 수많은 기계학습 전문가들이 연구를 진행하고 있으며 MIL 도입 및 다양한 분류기법들이 개발되고 있다. (Hau Chen, 2019)

2.2 MIL(Multiple Instance Learning)

MIL은 개별적으로 레이블이 지정된 Instance Set를 받는 대신 단일 클래스 레이블이 할당된 Instance Bag을 처리한다. MIL의 주요 목표는 의료 진단과 같은 Bag Label을 예측하는 모델을 학습하는 것이다. MIL과 기존 Image classification의 다른 점은 개별 데이터가 하나의 Instance가 아닌 여러 개의 Bag로 이루어졌다는 점이다. 따라서 Bag의 instance가 모두 Positive일 필요는 없고, 하나라도 Positive Instance가 있다면 Bag의 Label은 Positive가 된다. Attention-Based로 Multiple Instance Learning을 진행할 때 평균 및 최대 연산자보다 더 높은 수준을 해석할 수 있게 한다(Maximilian Ilse 등, 2018). 다시 말해, MIL-Pooling 연산자는 예측된 Bag Label에 대한 각 Instance의 기여에 대한 통찰력을 제공한다.

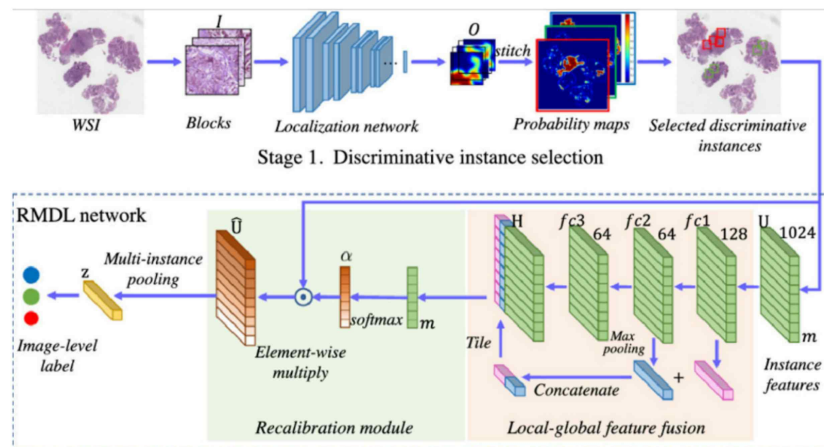


Figure Structure of RMDL

재보정 된 다중 인스턴스 딥러닝 기법(RMDL)을 이용해 규모가 매우 크고 비정상 영역의 크기가 다양한 정보영역을 분석하는 문제를 해결했다(Shujun Wang 등, 2020). 먼저 판별 인스턴스를 선택한 다음 이러한 인스턴스를 사용하여 제안된 RMDL 접근방식을 기반으로 질병을 진단하는 방식을 사용한다. 제안된 RMDL 접근방식은 판별적 인스턴스 스택과 이미지 수준 예측 두 가지 연속적인 단계로 진행된다. 판별적 인스턴스는 먼저 지역화 네트워크에서 생성된 예측 확률 맵을 기반으로 판별 패치를 추출한다. 특히, 탐지 프로세스의 속도를 높이기 위해 패치 기반의 Full-Convolutional Localization Network를 설계한다. 그런 다음, 판별 패치의 예측을 직접 집계하는 단점을 극복하기 위해 이미지 수준 분류를 위한 새로운 재보정 된 다중 인스턴스 딥러닝 네트워크(RMDL)를 개발한다. 이미지 수준 예측은 개발된 새로운 RMDL 네트워크는 인스턴스 기능을 자동으로 재보정하여 최종 이미지 수준 레이블을 예측하기 위해 패치의 다양한 기여도를 고려한다.

2.3 CLAM

CLAM, Clustering-constrained Attention Multiple Instance Learning, 은 MIL의 구조를 바탕으로 좀 더 발전된 모델이다. CLAM의 중요한 특징 2가지는 어텐션 네트워크를 사용해서 자동으로 WSI를 분류하는 데에 진단적 가치가 높은 영역을 알아낼 수 있다는 점이다. 또한 클러스터링 기법을 활용해서 패치-레벨(인스턴스-레벨)에서의 정보를 더 풍부하게 활용할 수 있게 되었다는 점이다.

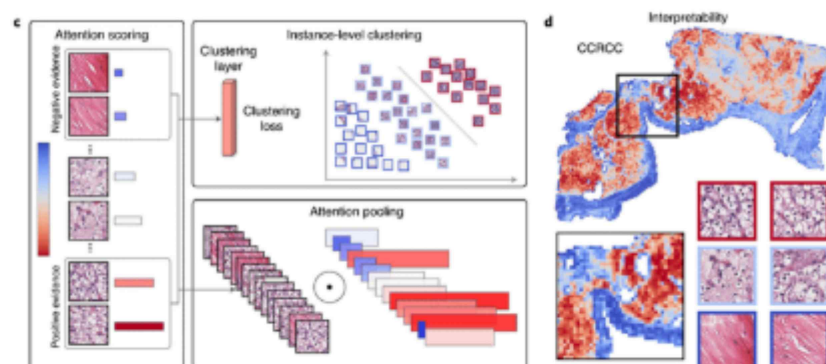


Figure Binary Clustering and Attention Pooling

이진 클러스터링은 인스턴스-레벨 클래스의 특징을 더 명확하게 알기 위해서 추가한 기법이다. 완전 연결 층 W1 이후에 n개의 클래스마다 512개의 은닉층을 가진 클러스터링 층을 놓는다. 어텐션 네트워크의 결과를 이용해서 어텐션 점수가 높은 패치 몇 개만을 활용한다. 실제 레이블과

대응되는 어텐션 브랜치를 “in-the-class”, 그 레이블을 제외한 나머지 $n-1$ 개의 어텐션 브랜치를 “out-of-the-class” 브랜치라고 명명했다. 클러스터링 태스크를 통해서 Patch-level 특징 공간을 양의 증거와 음의 증거로 선형적으로 분리할 수 있다. 암의 하위 유형을 진단할 때는 상호 배제의 원리(mutual exclusive)가 적용된다. “in-the-class” 브랜치에서 양의 증거와 음의 증거로 나눈 그 패치들을 클러스터링한 것이, 나머지 $n-1$ 개의 “out-of-the-class” 브랜치에 추가적인 제약을 가할 수 있다. 상호 배제의 원리를 적용할 수 없는 문제들에 있어서는 “out-of-the-class branch”를 사용하지 않는다. 인스턴스-레벨 클러스터링의 비용함수로는 Smooth Top1 SVM loss를 사용한다. 과적합을 줄이고, 데이터 레이블에 노이즈가 많거나, 데이터 개수가 많지 않을 때 유용하다. 슬라이드-레벨 분류에서는 잘 알려진 Cross-Entropy Loss를 사용한다.

3. 모델 제안

3.1 MIL(Multiple Instance Learning)

본 장에서는 2장에서 기술한 MIL(Multiple Instance Learning) 모델을 WSI(Whole slide Image)에 적용하는 새로운 모델을 기술한다. 앞선 선행연구에서 MIL 모델은 permutation-invariant를 가정했다. MIL의 scoring function은 인스턴스 X 의 집합 $S(X)$ 는 대칭함수로 정의했다. score function을 대칭적으로 만들기 위해서는 MIL pooling이 필요하며, permutation-invariant를 가정한다.

아래는 WSI에 적용한 MIL 모델의 구조이다. 입력 이미지는 tile generation 과정을 통해 얻은 tile을 center_id 별로 나누어 입력하는 방식으로 구성되어 있다. center_id가 불균형해 metric 값이 안정적이지 못했기 때문에 클러스터링을 이용해 총 6개의 그룹으로 묶어 주었다. 이후 사전 훈련된 CNN Encoder 구조를 통과시켜 각각 center_id 별로 CE와 LAA의 feature extraction을 진행한다. 이때 얻어진 feature 들은 ResNet을 통해 추출된 최종 단계의 특징들이다. 이후 feature를 Concat시켜 각각의 features 들을 합한다. 마지막으로 Classifier를 통과시켜 최종적으로 Weighted multi-class logarithmic loss를 기준으로 삼아 확률로 CE에 기인한 뇌졸중인지, LAA에 기인한 뇌졸중인지 판단한다.

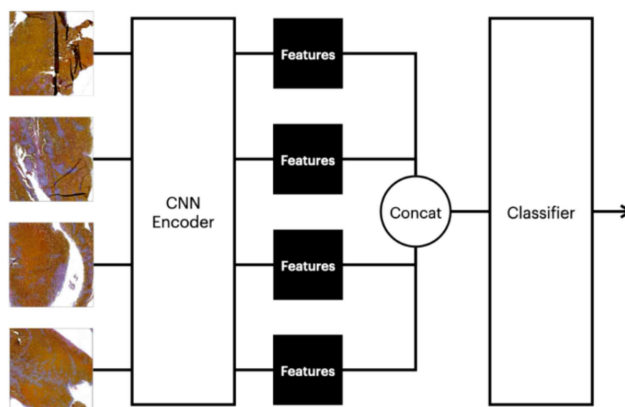


Figure Architecture of MIL

3.2 CLAM

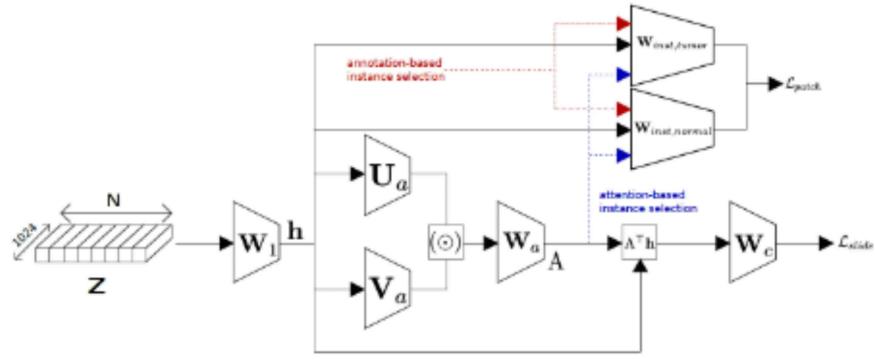


Figure 4 Architecture of CLAM

앞선 MIL과 같이 패치를 생성한 후 feature extraction 과정에서 생성한 input을 모델 아키텍처에 넣는다. 첫 번째 완전 연결 층인 W1을 거치면 512 X 64 차원이 된다. W1을 거친 후 2개의 어텐션 네트워크에 들어간다. 위의 그림에서 U와 V가 어텐션 네트워크로, feature vector h와 V를 곱해서 하이퍼볼릭 함수를 취해준 것과 feature vector h와 다른 어텐션 층인 U를 곱해서 시그모이드 함수를 취해준 값에 행렬 곱을 한다. 그리고 이것을 스케일링시켜주면 $a_{k,m}$ (k는 tile index, m은 mth 클래스를 의미)이 나온다.

$$a_{k,m} = \frac{\mathbf{W}_{a,m}(\tanh(\mathbf{V}_a \mathbf{h}_k^\top) \odot \text{sigm}(\mathbf{U}_a \mathbf{h}_k^\top))}{\sum_{i=1}^N \exp\{\mathbf{W}_{a,m}(\tanh(\mathbf{V}_a \mathbf{h}_i^\top) \odot \text{sigm}(\mathbf{U}_a \mathbf{h}_i^\top))\}}$$

Figure 5 Attention Computation

CLAM의 인스턴스-레벨 클러스터링은 클러스터링 가중치와 특징 벡터 h를 연산해서 클러스터링 스코어를 구한다. 가짜 레이블을 만들 때는 어텐션 네트워크 결과값을 이용하는데 이것의 어텐션 스코어가 높으면 가짜 레이블을 1로 주고, 어텐션 스코어가 낮으면 가짜 레이블을 0으로 준다. 정리하자면, CLAM은 어텐션 네트워크를 적용해서 슬라이드-레벨 분류를 하고, 이진 클러스터링에서 어텐션 스코어를 다시 한번 활용해서 가짜 레이블 0과 1을 생성해서 패치를 분류하여 마치 패치-레벨 레이블이 있는 것과 같은 효과를 줄 수 있다.

4. 실험

우리는 1,000개 이상의 WSI에 대하여 실험을 진행하였다. WSI 분류에 적합한 MIL 모델, CLAM 모델을 이용하였을 때 좋은 성능을 보이는지 확인하기 위하여 기본 CNN 모델과 MIL 모델, CLAM 모델을 사용하여 실험을 진행하였다. 추가로 하이퍼파라미터 튜닝을 해서 성능을 더 높이고자 하였다.

4.1 데이터 설명

1,000개 이상의 고해상도 Whole slide 디지털 병리학 이미지로 구성되며, 각 슬라이드는 급성 허혈성 뇌졸중을 겪은 환자의 혈전을 묘사한다. Training, test sets를 구성하는 슬라이드는 CE(심장 색전증) 또는 LAA(대동맥 아테롬성 동맥 경화증)로 알려진 원인(기원)이 있는 혈전을 나타낸다. Task는 각 환자에 대해 테스트 세트에 있는 슬라이드의 원인(CE 또는 LAA)을 분류하는 것이다.

Train 데이터는 총 754개로, Test 데이터는 4개로 각각 5개의 열로 구성되어 있다. image_id는 {patient_id}_{image_num}, center_id는 의료 센터 식별, patient_id는 환자 식별, image_num는 동일한 환자로부터 얻은 혈전의 이미지를 열거, label은 혈전의 원인(CE 또는 LAA)이다. Submission은 patient_id 별로 label을 추측해서 확률값으로 나타내어 결과를 제출하는 방식이다.

아래는 우리가 분류해야 하는 이미지의 예시이다.

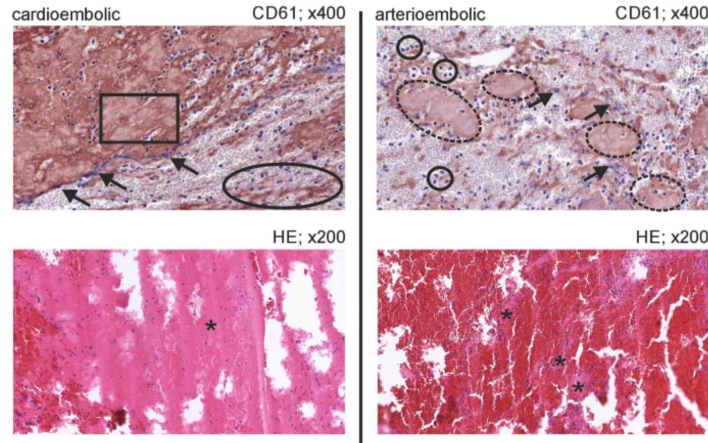


Figure 6 Properties of CE(Left) Properties of LAA(Right)

TROMBEX 혈전 분류에서 사용되는 다양한 혈소판 분포 방식에 따르면 CE와 LAA를 구분하는 특징은 다음과 같다. 왼쪽 사진들을 보면 CE는 심혈관계 혈전은 혈소판이 많다는 특징을 가지고 있다. 그들의 분포 패턴은 충상(네모)과 원환(타원) 구름 형성을 지니고 있으며 섬유망의 밀도가 높은 특징이 있다. 오른쪽 사진들을 보면 LAA에 관한 내용이다. 동맥 색전 혈전은 배열된 소량의 혈소판을 지니고 있다. 일부 분해된 호중구뿐만이 아닌 손상되지 않은 호중구(원)가 발견되며 매끈하지 않고 거친 면을 가지기도 한다는 특징이 있다.

Training 데이터에서 label의 분포는 CE가 72.5%, LAA가 27.5%로 약 2.6배가량 차이가 있으며 불균형함을 알 수 있다. 또한, 데이터를 얻은 출처인 center_id 또한 불균형하며 11번 center의 데이터가 가장 많았고 이후 4번째와 7번째 센터가 그 뒤를 잇고 있다.

4.2 평가지표

$$\text{Log Loss} = - \left(\frac{\sum_{i=1}^M w_i \cdot \sum_{j=1}^{N_i} \frac{y_{ij}}{N_i} \cdot \ln p_{ij}}{\sum_{i=1}^M w_i} \right)$$

Figure 7 Log Loss

Weighted multi-class logarithmic loss를 사용하여 평가를 진행했다. 각 이미지에는 CE 또는 LAA의 병인학 클래스가 라벨로 표시되어 있으며, 각 이미지에 대한 각 클래스에 대한 확률을 구해야 한다.

4.3 실험 준비

먼저 test 데이터가 4개 밖에 존재하지 않기에 Trainset을 Train set(50%), Validation set(20%), Test set(30%)으로 나누었다. 또한, CE의 데이터가 LAA의 데이터보다 2.6배 많아

데이터 불균형 문제가 발생하여 결과에 영향을 줄 수 있다. WSI는 색깔이 매우 중요하게 작용하는 특성을 이용하여 LAA라벨의 Train 데이터만을 채도 0.2, 0.4, 0.6로 차이를 주어 3배로의 증강시켰다. 이후 WSI의 특성에 맞게 tile generation을 진행했다. 이때 block size는 28, block 당 crop은 8개, crop size는 block size*block 당 crop으로 정의했다. crop과 block은 모두 사각형 이미지이며 block은 이미지 위에 grid를 만들고 crop은 이 grid 이후에 사용된다. 이미지를 바로 resize 하지 않고 grid를 생성한 후 cropping을 한 이유는 연산 속도를 높이고, specific한 케이스에서 좀 더 robust하게 만들기 위함이다. 그래서 복잡성과 연산 속도와의 trade off 관계가 있었지만, 연산 속도와 용량의 이점을 위해서 2-level grid를 사용했다. 이후 8,16,32 등 다양하게 이미지를 resize했지만 24가 가장 결맞음이 좋아 최종 결정했다. OpenCV는 이미지를 RGB형태가 아닌 BGR 형태로 로딩하기 때문에 색감이 원본 이미지와 다르게 나타날 수 있다. 따라서 WSI의 색상 특성을 고려해 타일의 색상을 RGB에서 BGR로 바꾸어주고 이후 tile을 불러올 때 다시 RGB로 바꾸어주었다.

입력받은 tile들을 resnet18 구조를 통과시켜 각각 center_id별로 CE와 LAA의 feature extraction을 진행한다. 하지만 일부 center_id의 tile수가 너무 적기에 값의 metric 값이 안정적이지 못하였다. 아래는 center_id별 feature의 수를 Euclidean distance를 이용하여 K-means Clustering 한 결과이다. Silhouette 기법을 사용한 결과 k가 6개일 때 가장 좋은 클러스터링임을 알 수 있었고, 이를 기준으로 6개의 그룹으로 클러스터링을 진행하였다. 각각의 그룹은 [(11), (4), (7), (1, 5), (10, 3), (6, 2, 8, 9)] 이다.

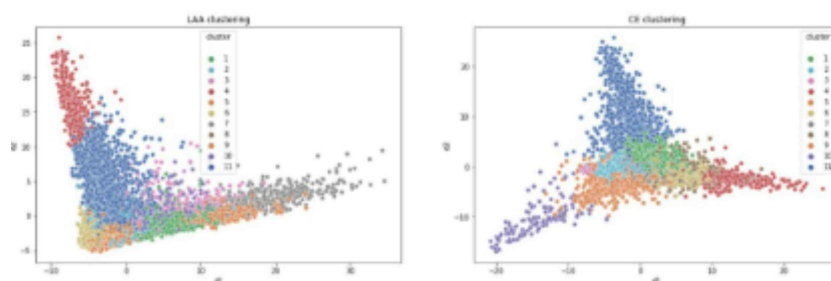


Figure Feature Clustering

CNN Encoder를 통과시킬때 feature extraction의 성능 향상을 위하여 레이어 동결을 이용한 실험을 진행하였다. 또한, Classifier를 통과시킬 때 좋은 성능을 보이는지 확인하기 위하여 Resnet18, EfficientNetb0, RegNet으로 실험을 진행하였다. 추가적으로, 불균형문제 해소를 위해 Stratified K-Fold 기법을 적용해해 3,5,7,10으로 실험하였다.

4.4 결과

고해상도 WSI를 분류하기 위해 우리는 base CNN model, MIL model, CLAM model의 비교실험을 진행하였다. Data augmentation은 validation set에는 적용하지 않았고, 기본 train set에서 5:3:2로 Train:Valid:Test로 떼서 사용해서 실험을 진행하였다. 실험 결과는 Train-Loss, Train-Acc, Val-Loss, Val-Acc, Test-loss를 이용하여 평가한다. Tables는 실험의 결과이다.

Model	Train-Loss	Train-Acc	Val-Loss	Val-Acc
Base CNN model	1.0272	0.4768	1.0355	0.4722
MIL + Resnet	0.6570	0.6134	0.7131	0.5728
MIL + Regnet	0.6550	0.6162	0.6952	0.5704
MIL + EfficientNet	0.6343	0.6541	0.7493	0.5828

CLAM	0.2457	0.7218	0.2457	0.6815
------	--------	--------	--------	--------

Table Result of Experiment

Model	Learning_rate	Weight_decay	Batch_size	Epoch
Base CNN model	0.05	0.0001	64	50
MIL + Resnet	0.01	0.0005	128	10
MIL + Regnet	0.01	0.0001	256	5
MIL + EfficientNet	0.05	0.0001	128	10
CLAM	0.005	0.0001	1	20

Table Result of Hyperparameter tuning

Table 1은 우리 모델을 다른 기본 이미지 모델과 비교해서 성능을 비교한 결과 표이다. Table 1을 보면 비교 모델인 base CNN model에 비해 MIL 방법론을 섞은 이미지 모델이나, CLAM 모델의 성능이 더 좋은 것을 알 수 있으며, 그중에서도 CLAM이 전체 성능이 가장 좋은 것을 확인할 수 있었다. 이를 통해 기본적인 이미지 분류 모델보다 데이터 특성에 맞는 MIL 방법론이나 CLAM이 병리학적 feature를 뽑아내고 분류하는 데 효과적임을 확인할 수 있다.

5. 결론

본 논문에서는 WSI의 병원을 분류하는 방법론으로 MIL과 CLAM을 제안하였다. 기본 이미지 분류 모델에 비해서 WSI라는 특성에 맞춰서 병원을 더 잘 분류해내는 것을 보여줄 수 있다.

MIL 방법론을 활용하는 것은 이미지를 타일 단위로 나눔으로써 용량을 줄일 수 있고, WSI에서 병원의 특성을 더 잘 파악하는 것을 기대하게 한다. 기본 MIL 방법론에 어텐션 네트워크를 사용해서 기본 MIL-Pooling보다 더 유연하게 적용됨으로써 성능이 좋을 것으로 기대한다. 또한 클러스터링 기법으로 Patch-level에서의 정보를 더 풍부하게 활용할 수 있었다.

Kaggle에서 제공된 Mayo Clinic - STRIP AI 데이터 셋에 대하여 두 가지 주요 급성 허혈성 뇌졸중 원인 유형인 심장 및 대동맥 아테롬성 동맥 경화증을 구별하는 실험을 진행하였다. 비교 모델로는 base CNN model(ResNet)을 사용하여 제안된 MIL이나 CLAM을 사용한 모델이 비교모델보다 좋은 성능을 보이는 것을 확인하였다. 또한 어텐션 네트워크를 사용한 CLAM 모델이 모든 모델 중에서 가장 좋은 성능을 보이는 것을 확인하였다.

우리는 WSI에서 MIL, CLAM 방법론을 활용함으로써 성능을 높일 수 있었고, 어텐션 네트워크를 적용하는 것이 성능향상에 도움이 되는 것을 확인하였다. 특징 추출을 할 때 Resnet50 모델을 사용하는데 이때 ImageNet에서 훈련된 가중치를 사용한다. 그러나 이는 병리학 도메인에서의 가장 큰 약점이 되어왔기 때문에 후속 연구로 자기-지도적 학습 알고리즘을 이용한 실험을 진행하고자 한다. (Dehaene 등, 2020). 또한 ViT 모델을 활용함과 동시에 WSI의 패치들을 층위별로 특징을 추출해서 훈련을 하는 방식도 생각해볼 수 있다. (Chen 등, 2022).

참고문헌

Bhattacharyya S. A brief survey of color image preprocessing and segmentation techniques. *J Pattern Recognit Res* 2011;6:120–9.

Khan AM, Rajpoot N, Treanor D, Magee D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans Biomed Eng* 2014;61:1729–38.

. Chen JM, Li Y, Xu J, Gong L, Wang LW, Liu WL, Liu J. Computer-aided prognosis on breast cancer with hematoxylin and eosin histopathology images: A review. *Tumour Biol* 2017;39:1-12.

송삼용, 인공지능 병리슬라이드 서론 *Hanyang Med Rev* 2017;37:77-85

. Farahani N, Parwani AV, Pantanowitz L. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathol Lab Med Int* 2015;7:23–33.

Rojo MG, García GB, Mateos CP, García JG, Vicente MC. Critical comparison of 31 commercially available digital slide systems in pathology. *Int J Surg Pathol* 2006;14:285–305.

ShujunWangaYaxiZhubLequanYuaHaoChencHuangjingLinacXiangboWandXinjuanFanbPheng-AnnHenga RMDL: Recalibrated multi-instance deep learning for whole slide gastric image classification. *medical image analysis volume 58*, 2019; 101549

M. Yang, Z. Xie, Z. Wang, Y. Yuan and J. Zhang, "Su-MICL: Severity-Guided Multiple Instance Curriculum Learning for Histopathology Image Interpretable Classification," in *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3533-3543, Dec. 2022, doi: 10.1109/TMI.2022.3188326.

Y. Zhou and Y. Lu, "Multiple Instance Learning with Task-Specific Multi-Level Features for Weakly Annotated Histopathological Image Classification," *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 1366-1370, doi: 10.1109/ICASSP43922.2022.9747121.

Z. Wang, C. Saoud, S. Wangsiricharoen, A. W. James, A. S. Popel and J. Sulam, "Label Cleaning Multiple Instance Learning: Refining Coarse Annotations on Single Whole-Slide Images," in *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3952-3968, Dec. 2022, doi: 10.1109/TMI.2022.3202759.

T. -C. Do, H. -J. Yang, S. -R. Kang, S. -H. Kim, G. -S. Lee and J. -J. Min, "Positional Multi-Cross-Attention for Bone Age Estimation Using Deep Multiple Instance Learning," *2022 26th International Conference on Pattern Recognition (ICPR)*, 2022, pp. 4285-4291, doi: 10.1109/ICPR56361.2022.9956223.

A. Tarkhan, T. K. Nguyen, N. Simon, T. Bengtsson, P. Ocampo and J. Dai, "Attention-Based Deep Multiple Instance Learning with Adaptive Instance Sampling," *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, 2022, pp. 1-5, doi: 10.1109/ISBI52829.2022.9761661.

S. Iyer, A. Blair, C. White, L. Dawes, D. Moses and A. Sowmya, "Vertebral Compression Fracture detection using Multiple Instance Learning and Majority Voting," *2022 26th International Conference on Pattern Recognition (ICPR)*, 2022, pp. 4630-4636, doi: 10.1109/ICPR56361.2022.9956309.

B. Li, Y. Li and K. W. Eliceiri, "Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-supervised Contrastive Learning," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 14313-14323, doi: 10.1109/CVPR46437.2021.01409.

J. S. Leiby, J. Hao, G. H. Kang, J. W. Park and D. Kim, "Attention-based multiple instance learning with self-supervision to predict microsatellite instability in colorectal cancer from histology whole-slide images," *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 3068-3071, doi: 10.1109/EMBC48229.2022.9871553.

J. Moranguinho, T. Pereira, B. Ramos, J. Morgado, J. L. Costa and H. P. Oliveira, "Attention Based Deep Multiple Instance Learning Approach for Lung Cancer Prediction using Histopathological Images," *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021, pp. 2852-2855, doi: 10.1109/EMBC46164.2021.9631000.

A. Schmidt, J. Silva-Rodríguez, R. Molina and V. Naranjo, "Efficient Cancer Classification by Coupling Semi Supervised and Multiple Instance Learning," in *IEEE Access*, vol. 10, pp. 9763-9773, 2022, doi: 10.1109/ACCESS.2022.3143345.

H. Zhang *et al.*, "DTFD-MIL: Double-Tier Feature Distillation Multiple Instance Learning for Histopathology Whole Slide Image Classification," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 18780-18790, doi: 10.1109/CVPR52688.2022.01824.

M. E. Tschuchnig *et al.*, "Evaluation of Multi-Scale Multiple Instance Learning to Improve Thyroid Cancer Classification," *2022 Eleventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2022, pp. 1-6, doi: 10.1109/IPTA54936.2022.9784124.

Maximilian Ilse, Jakub M. Tomczak, Max Welling. "Attention based Deep Multiple Instance Learning". 2018. 6.28,