# Towards Classifying Histopathological Microscope Images As Time Series Data

*Sungrae Hong[1], Hyeongmin Park[2], Youngsin Ko[3], Sol Lee[1], Bryan Wong[1], Mun Y Yi[1]**

[1]Graudate School of Data Science, KAIST, Daejeon, South Korea
[2]Department of Industrial Engineering, KAIST , Daejeon, South Korea
[3]Seegene Medical Foundation, Seoul, South Korea

ISBI 2025
2025 IEEE International Symposium on Biomedical Imaging
April 14-17, 2025 | Houston, TX, USA

Poster No. Session 1 - 68 (1571081670)

## INTRODUCTION

✓ Background

- Cancer remains a leading cause of death worldwide, prompting the deep-learning based computer vision community to develop various models to address this critical issue [1].
- However, we note that the deep learning community has neglected microscopy images, which have significant practical advantages over scanned images. Specifically, nearly every existing research has been conducted on scanner-based images [2]. Microscopes are significantly more affordable than scanners thereby offering significantly different accessibility to patients in third-world nations [3].

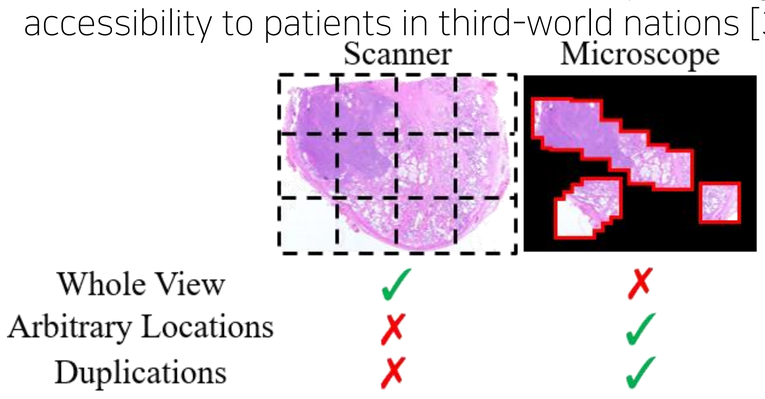


**Fig. 1**. denotes an image patch acquired by the scanner and a region captured by a microscope. Microscopy images are composed of overlapping sequences manually captured by domain experts

1. *ArbitraryLocations* : Microscopic images are periodically produced, making it impossible to accurately capture the coordinates of where the produced images were acquired.
2. *InconsistentLength* : The number of images captured per slide varies due to the automatic image acquisition process.
3. *Duplications* : The image acquisition process often results in a large number of redundant images.
4. *WeakLabel* : The absence of a whole-view, hundreds of images produced, make it very difficult for experts to annotate each images separately.

## METHOD

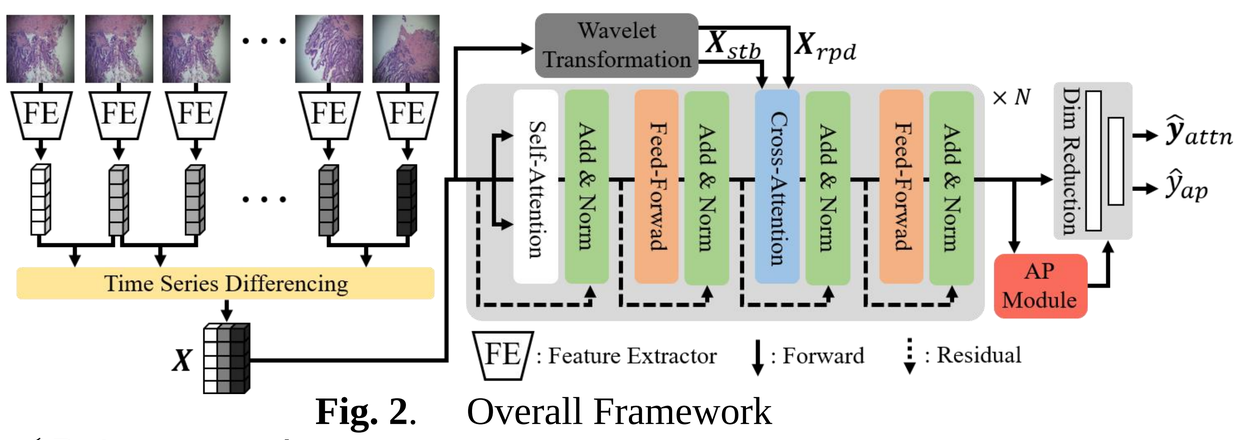


**Fig. 2**. Overall Framework

✓ Data processing

$$\Delta(\boldsymbol{x}_i, \boldsymbol{x}_{i+1}) = ||\boldsymbol{x}_{i+1} - \boldsymbol{x}_i||_2^2 \qquad (1)$$

- **Time series differencing** We extract a feature sequence $X = \{x_i \epsilon \mathbb{R}^d | i = 1,2, \dots, n\}$ using pretrained extractor. We applied time series differencing to remove duplications (**Equation 1**).
- **Wavelet transformation** Experts often capture narrow regions, causing similar symptoms to be filmed repeatedly. On the other hand, they occasionally move the slide glass to acquire images from entirely different regions. To mimic this pattern, we employ Wavelet transformation

✓ Attention module

$$\text{Attention}(Q, K, V) = \text{sigmoid}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (2)$$

$$\text{CrossAttention} = \lambda_{stb} \cdot \text{Attention}(Q, K_{stb}, V_{stb}) + \lambda_{rpd} \cdot \text{Attention}(Q, K_{rpd}, V_{rpd}) \qquad (3)$$

- We used self-attention module for input sequence and corss-attention for $X_{stb}, X_{rpd}$.

✓ Attention Pooling (AP) module

$$\boldsymbol{P} = \sum_{i=1}^{n} a_i \boldsymbol{x}_i \in \mathbb{R}^d \qquad (4)$$

$$a_i = \frac{\exp\{\boldsymbol{w}^T(\tanh(\boldsymbol{A}_t\boldsymbol{x}_i) \odot \text{sigmoid}(\boldsymbol{A}_s\boldsymbol{x}_i))\}}{\sum_{j=1}^{n}\exp\{\boldsymbol{w}^T(\tanh(\boldsymbol{A}_t\boldsymbol{x}_j) \odot \text{sigmoid}(\boldsymbol{A}_s\boldsymbol{x}_j))\}} \qquad (5)$$

- This module aggregates sequence into a point prediction.

✓ Objective functions

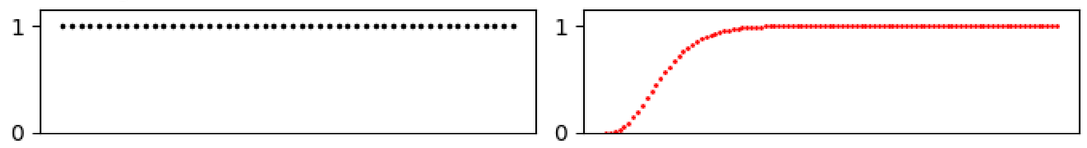


**Fig. 4**. Static target and proposed implicit target.

$$\mathcal{L}_{dtw} = ||D(\hat{\boldsymbol{y}}_{attn}, \boldsymbol{y}_l) - D(\boldsymbol{y}_{ideal}, \boldsymbol{y}_l)||_1 \qquad (6)$$

$$\mathcal{L}_{align} = || \underset{C_i \in \{C_1, \cdots, C_n\}}{\text{argmax}} (\hat{\boldsymbol{y}}_{attn}) - \underset{C}{\text{argmax}}(\hat{y}_{ap})||_1 \qquad (7)$$

$$\mathcal{L} = \lambda_{dtw}\mathcal{L}_{dtw} + \lambda_{ap}\mathcal{L}_{ap} + \lambda_{align}\mathcal{L}_{align} \qquad (8)$$

- The algorithm maps dynamic-length sequences closer to pre-defined target sequence. Because this is non-differentiable dynamic programming, we leverage the novel solution, Soft-DTW [4] $D(\cdot,\cdot)$.
- we propose using the cumulative distribution function of a Beta(3,20) as an implicit target to address the weakly supervised nature of our image sequences (**Fig. 4.**). It is grounded in the empirical observation that experts find it challenging to capture regions exhibiting symptoms in a first frame due to the manual, high-zoom nature of the microscope.

## EXPERIMENT

✓ Settings

| | Magnitudes | *n*(Images) | | *n*(Cases) | | Average Images/Case | |
|---|---|---|---|---|---|---|---|
| | | N | M | N | M | N | M |
| BreakHis | 40× | 652 | 1,370 | 24 | 58 | 27 | 24 |
| | 100× | 644 | 1,437 | | | 27 | 25 |
| | 200× | 623 | 1,390 | | | 26 | 24 |
| | 400× | 588 | 1,232 | | | 15 | 21 |
| *Aanonymous-Center* | 40×, 100×, 200× | 52,076 | 84,100 | 492 | 423 | 106 | 199 |

**Table. 1**. Dataset description. N and M indicates normal and malignant respectively.

✓ Quantitative results

**Table. 2**. Comparison results

| | | *AnonymousCenter (n=186)* | | BreakHis *(n=28)* | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 40× | | 100× | | 200× | | 400× | |
| | | F1 | Accuracy | F1 | Accuracy | F1 | Accuracy | F1 | Accuracy | F1 | Accuracy |
| LSTM [12] | | 0.975 | 0.976 | 0.940 | 0.914 | 0.923 | 0.893 | 0.938 | 0.914 | 0.925 | 0.893 |
| GRU [13] | | 0.978 | 0.980 | 0.916 | 0.879 | 0.945 | 0.921 | 0.933 | 0.907 | 0.927 | 0.900 |
| Transformer [9] | | 0.981 | 0.983 | 0.928 | 0.900 | 0.943 | 0.921 | 0.932 | 0.907 | 0.949 | 0.929 |
| ABMIL [14] | | 0.982 | 0.984 | 0.891 | 0.850 | 0.913 | 0.871 | 0.911 | 0.871 | 0.904 | 0.857 |
| TransMIL [15] | | 0.977 | 0.978 | 0.941 | 0.914 | 0.924 | 0.893 | 0.918 | 0.886 | 0.924 | 0.893 |
| Ours | AP | **0.990** | **0.991** | 0.954 | 0.936 | 0.947 | 0.929 | 0.945 | 0.921 | 0.954 | 0.936 |
| | DTW Distance | 0.988 | 0.989 | 0.962 | **0.950** | 0.951 | **0.936** | 0.947 | 0.929 | 0.969 | **0.957** |
| | KNN | **0.990** | **0.991** | 0.954 | 0.936 | **0.952** | **0.936** | 0.953 | 0.936 | **0.970** | **0.957** |
| | Voting | **0.990** | **0.991** | **0.964** | **0.950** | **0.952** | **0.936** | **0.958** | **0.943** | 0.964 | 0.950 |

- **Various inference strategies** Our proposed method offers four inference strategies: $\hat{y}_{ap}$ prediction, DTW distance between $\hat{y}_{attn}$ and $y_l$, KNN using $\hat{y}_{attn}^{train}$ and $\hat{y}_{attn}$, and a majority voting of them.
- **Quantitative results** Our proposed method can predict the entire sequence without truncation and aggregate the entire sequence for a single prediction, combining the advantages of both approaches.

✓ Ablation study

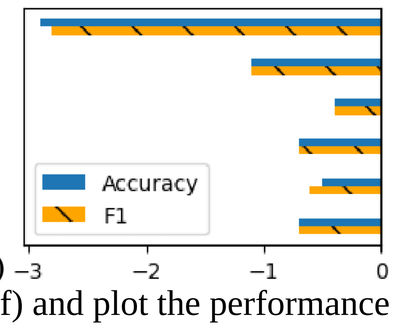


**Fig. 4**. We ablate (i) Wavelet transformation (ii) Implicit target (iii) Reference sequence yideal (iv) $\mathcal{L}_{align}$ (v) Attention module (vi) AP module. We present combinations of ablations from (a) to (f) and plot the performance decreases compared to the fully equipped model.

## CONCLUSION

- We propose a novel framework that treats microscopic image sequences as time-series data. By employing various inference strategies and a voting mechanism, we achieved superior results.

**REFERENCES: [1]** Amelie Echle, Niklas Timon Rindtorff, Titus Josef Brinker, Tom Luedde, Alexander Thomas Pearson, and Jakob Nikolas Kather, "Deep learning in cancer pathology: a new generation of clinical biomarkers," British journal of cancer, vol. 124, no. 4, pp. 686–696, 2021. **[2]** Ruiwei Feng, Xuechen Liu, Jintai Chen, Danny Z Chen, Honghao Gao, and Jian Wu, "A deep learning approach for colonoscopy pathology wsi analysis: accurate segmentation and classification," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 10, pp. 3700–3708, 2020. **[3]** Yifan Zhang, Hanbo Chen, Ying Wei, Peilin Zhao, Jiezhang Cao, Xinjuan Fan, Xiaoying Lou, Hailing Liu, Jinlong Hou, Xiao Han, et al., "From whole slide imaging to microscopy: Deep microscopy adaptation network for histopathology cancer image classification," in International conference on medical image computing and computer-assisted intervention. Springer, 2019, pp. 360–368. **[4]** Marco Cuturi and Mathieu Blondel, "Soft-dtw: a differentiable loss function for time-series," in International conference on machine learning. PMLR, 2017, pp. 894-903.

**COMPLIANCE WITH ETHICAL STANDARDS** : This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Review Board (SMF-IRB-2020-007) and (KAIST-IRB-22 335). Also, this research study was conducted retrospectively using human subject data made available in open access by [18]. Ethical approval was not required as confirmed by the license attached with the open access data.

**Acknowledgement:** This research was supported by the Seegene Medical Founda tion, South Korea, under the project "Research on Develop ing a Next Generation Medical Diagnosis System Using Deep Learning" (Grant Number: G01180115).

KIRC KNOWLEDGE INNOVATION RESEARCH CENTER
KAIST ISE Department of Industrial and Systems Engineering
KAIST GSDS Graduate School of Data Science
Seegene

○ : Presenter, * : Corresponding Autor