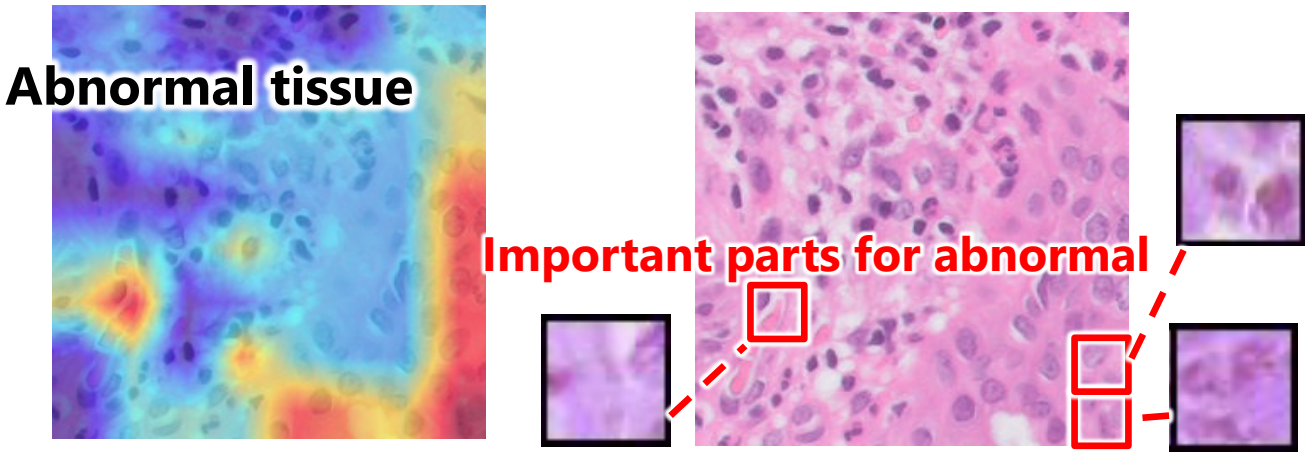
Explainable Feature Embedding using Convolutional Neural Networks for Pathological Image Analysis

National Institute of Advanced Industrial Science and Technology (AIST), Japan

○Kazuki Uehara, Masahiro Murakawa, Hirokazu Nosato, Hidenori Sakanashi

Contributions

- We propose an **explainable** feature learning method for pathological image analysis using a CNN:
  - ➢ Constructing a feature dictionary with vector quantization
  - ➢ Visualizing dictionary items as images using a generator
- Experimental results showed **0.93 of AUROC** on detecting atypical tissues in pathological images



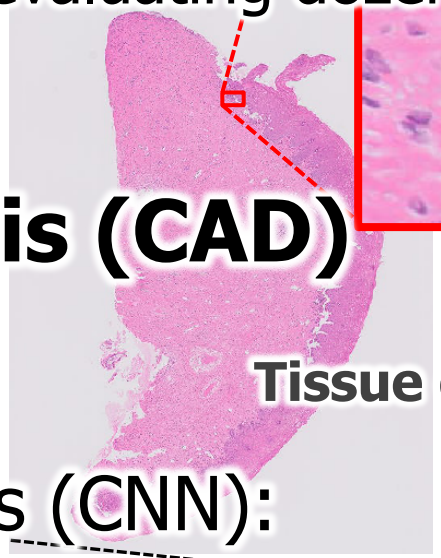
Explanations
Providing learned features by the CNN in the dictionary

Introduction

Pathology

- To determine treatment of cancer
- Requiring considerable time and effort

  Exploring tiny atypical cells by evaluating dozens of cells on a slide image



Tissue on a glass slide

Computer Aided Diagnosis (CAD)

Relieve pathologist's burden

Convolutional Neural Networks (CNN):

High accuracy for pathological image analysis

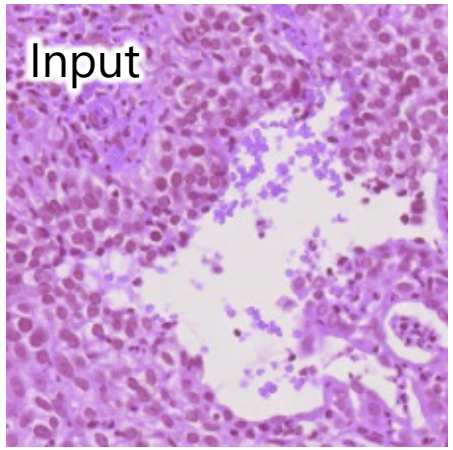
Explainability of CNNs

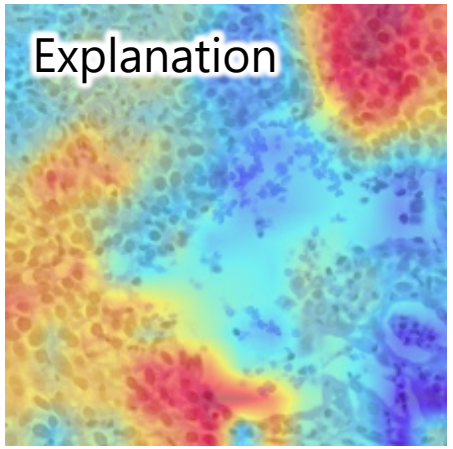
CAD systems should be accurate and explainable to ensure their reliability

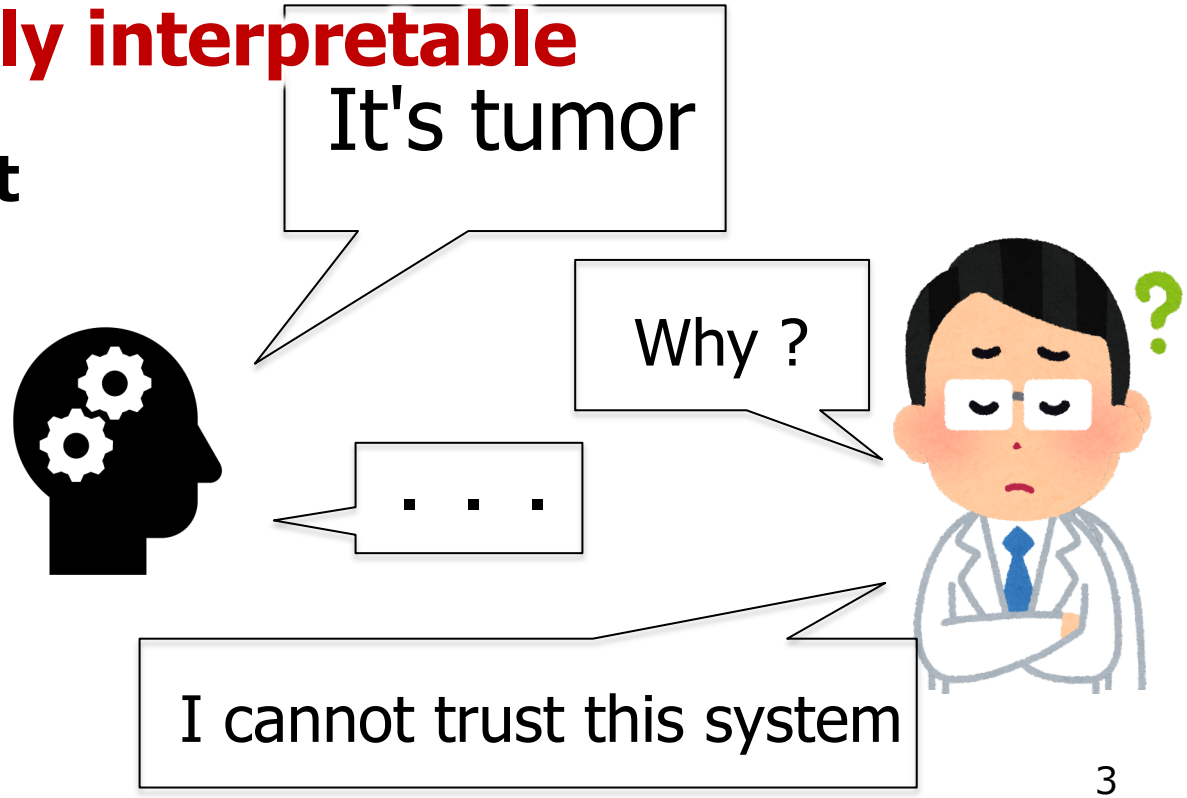
Explainability: Basis of diagnoses can be interpreted by humans

Decisions made by CNNs are hardly interpretable

Activation based explanations cannot tell the reasons for their decisions

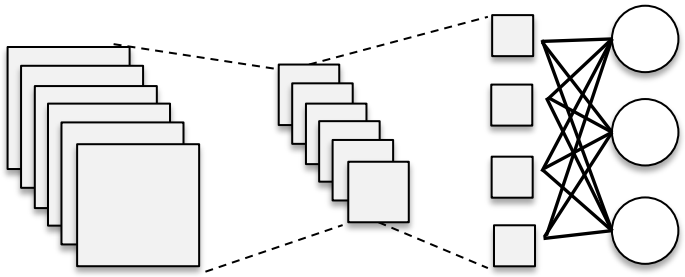




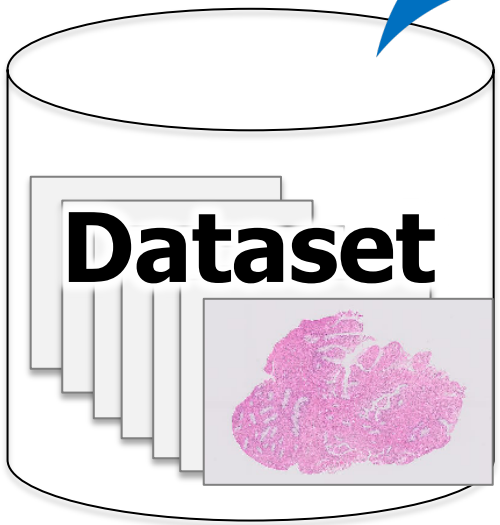


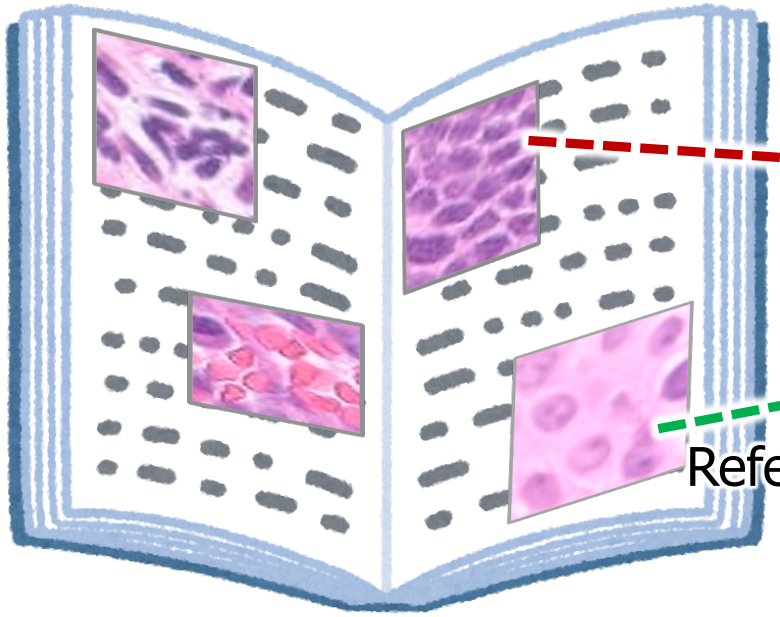
Objective

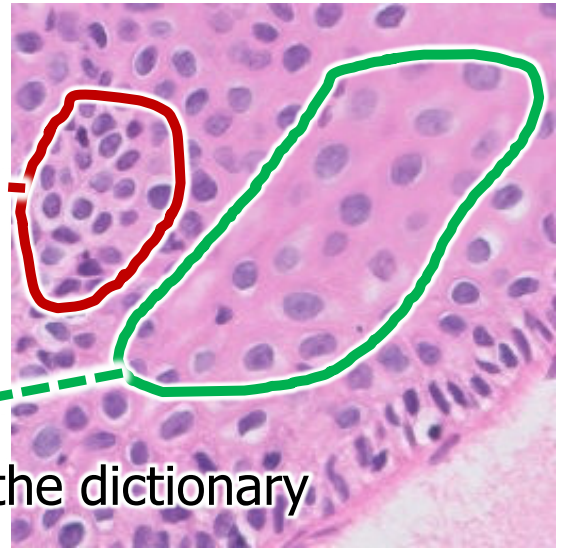
Accurate and **Explainable** diagnosis based on a dictionary using a CNN



Important features for diagnosis



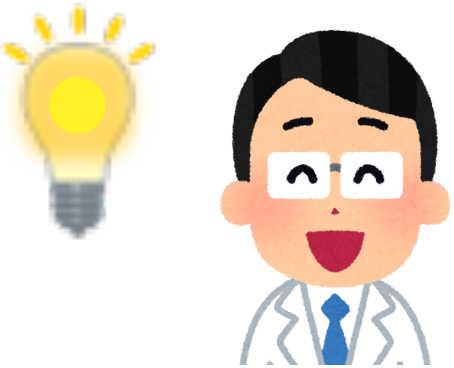




Referring the dictionary

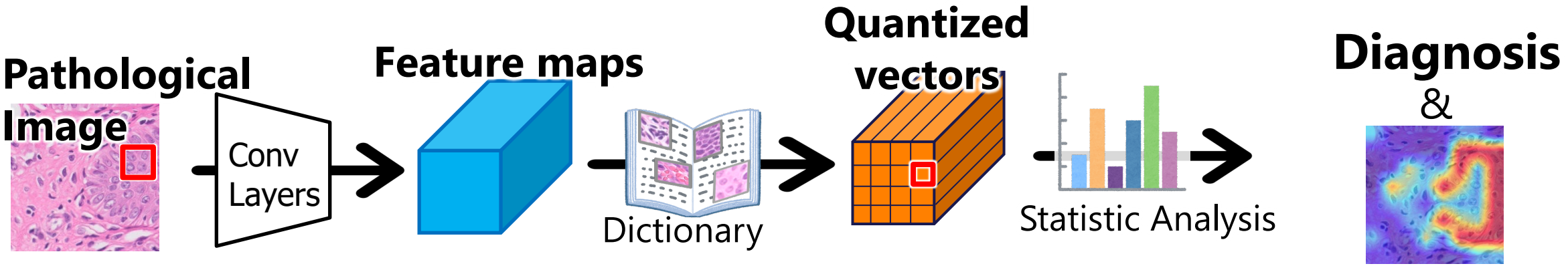
Feature Dictionary

- Clinical Findings
- Appearance



Explainable Feature Learning

Diagnosis Network for accurate diagnosis



Visualization Network for explanation

Visualize dictionary items (learned features)

Experimental Setup

Comparison methods

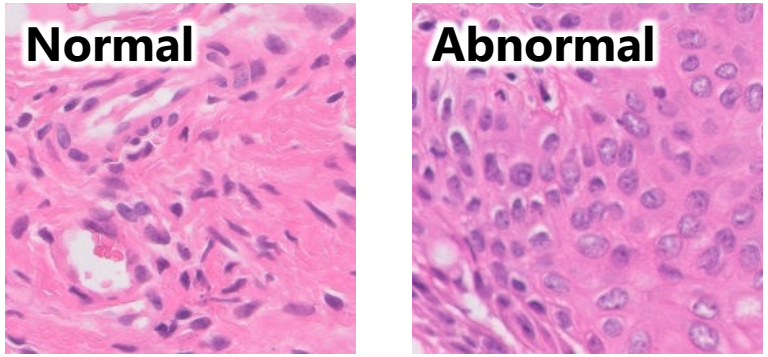
- ✓ Inception V3 [Liu +, 2017] : **State-of-the-art** classification, no explanation
- ✓ ProtoNet [Li +, 2018] : **Explainable** CNN
- ✓ Dictionary based CNN [Uehara+, 2019] : **Explainable** CNN

Dataset

Pathological image patches of a uterine cervix
Each image patch has 256x256 pixel

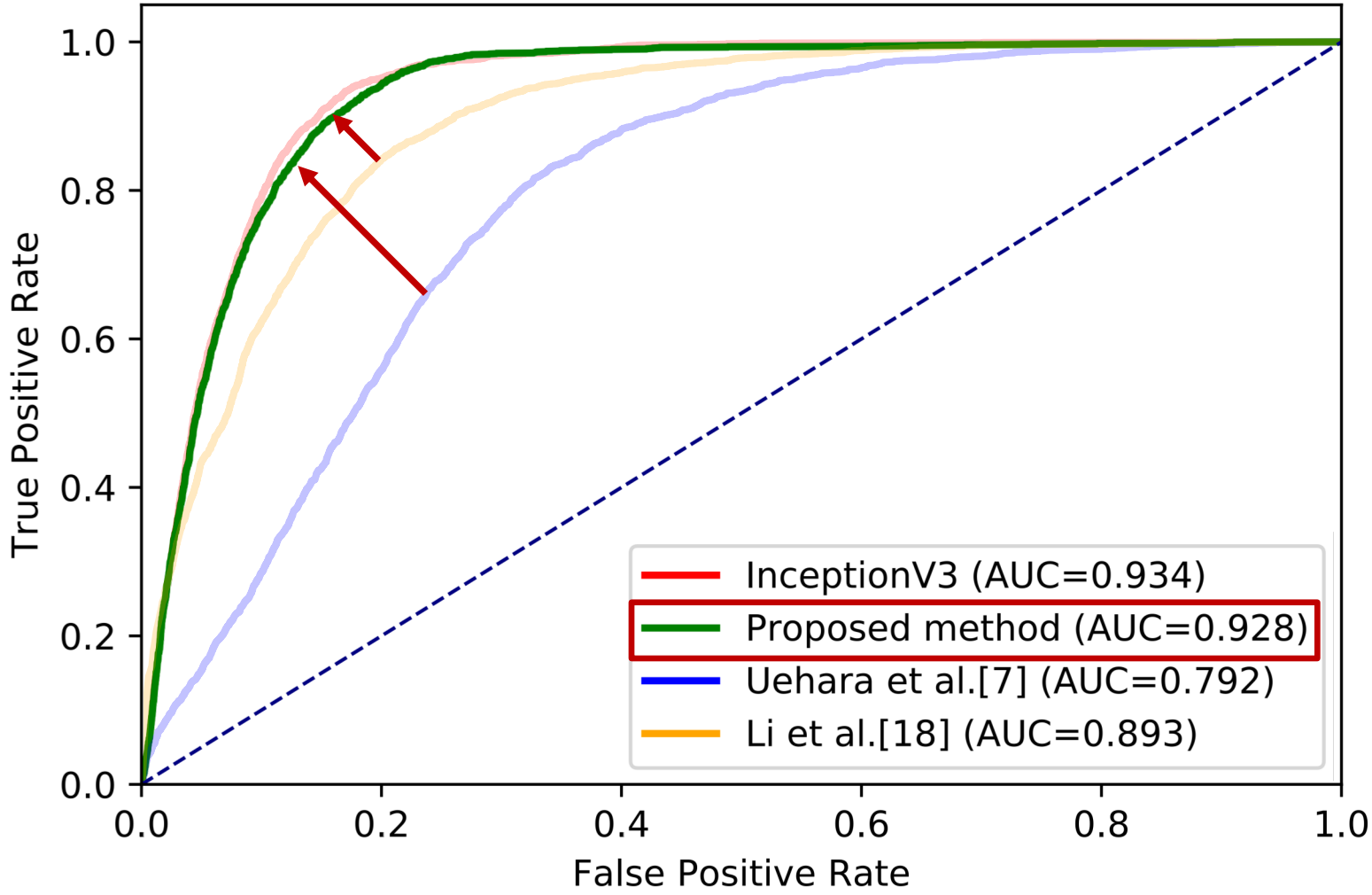
Normal	: Train (84,194)	Test (27,601)
Abnormal:	Train (12,088)	Test (2,286)

Classification



Classification Result

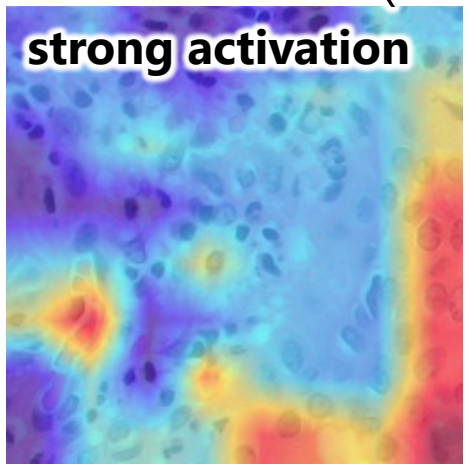
Our method yielded high classification accuracy

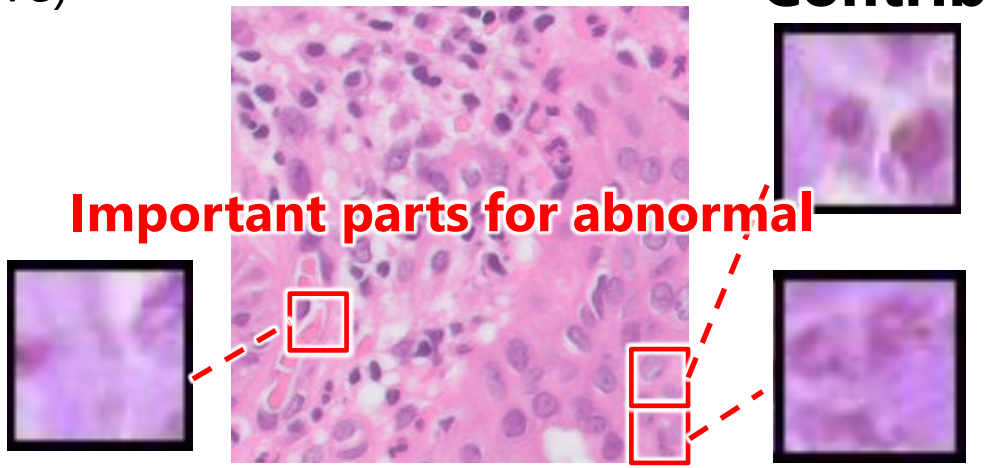


Results of Visual Explanations

Redder area is important for classifying atypical tissue

Abnormal tissue (True Positive)

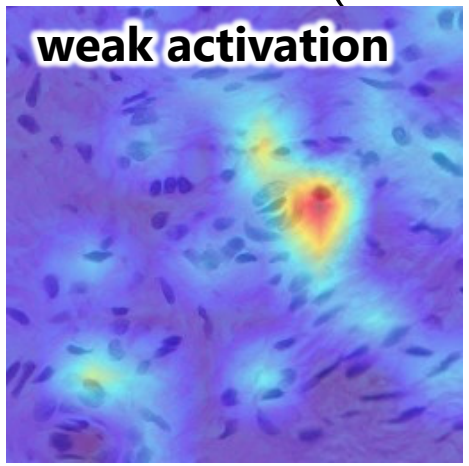


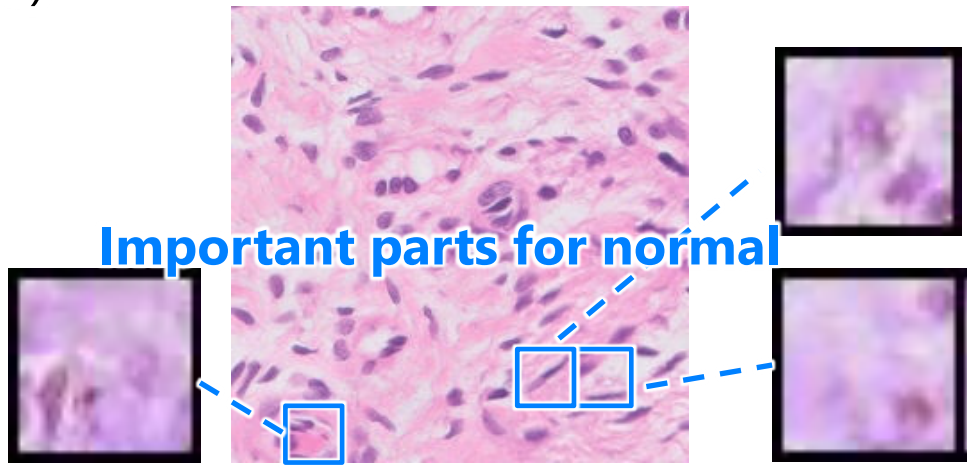


Contributed features

large nucleus

Normal tissue (True Negative)





small nucleus

Conclusion

We have proposed **explainable feature** learning method to ensure reliable diagnosis for **pathological image analysis**

- ✓ **Accurate diagnosis**
  - ➢ End-to-end dictionary learning
- ✓ **Easy to interpret its basis of diagnosis**
  - ➢ Linear combination of cooccurrence of items in the dictionary
  - ➢ Visualize the items as images

Experimental result demonstrated that our method has advantages of **explainability** compared with the conventional **black-box** models