Explainable Feature Embedding using Convolutional Neural Networks for Pathological Image Analysis

Kazuki Uehara, Masahiro Murakawa, Hirokazu Nosato, Hidenori Sakanashi
National Institute of Advanced Industrial Science and Technology (AIST), Japan

Background

Pathology
- ➢ To determine treatment of cancer
- ➢ Requiring considerable time and effort
  - ✓ Evaluating vast number of cells in a tissue on a glass slide with a microscope

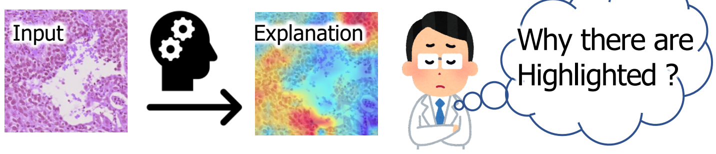
Computer Aided Diagnosis (CAD)
- ➢ Relieve pathologists' burden
- ➢ The use of Convolutional Neural Networks (CNN)
  - ✓ High accuracy for pathological CAD systems
  - ✓ Basis of its decision is hardly interpretable

Accuracy and **Explainability** are required to ensure reliability

Motivation

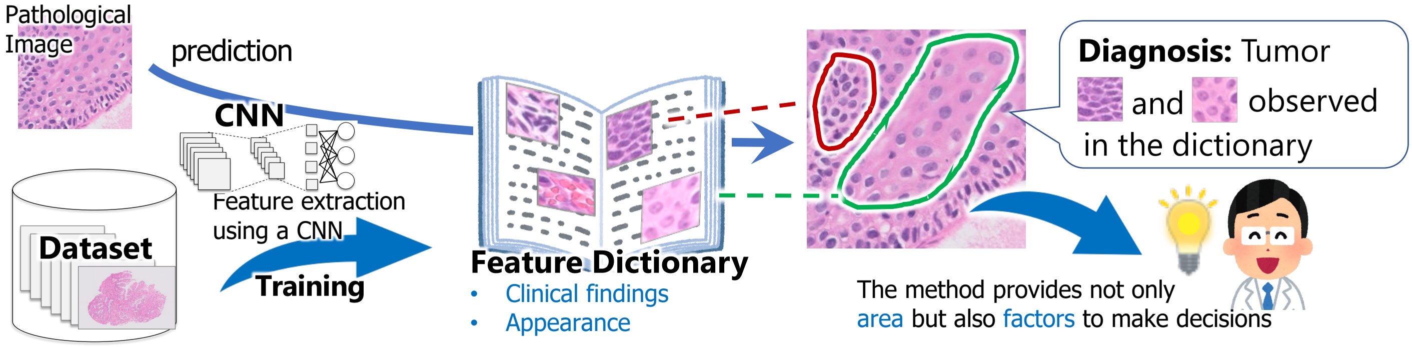
Explainability:
Basis of diagnoses can be interpreted by humans

Related studies for Explainability
Activation based explanations are popular but they cannot tell the reasons for their decisions

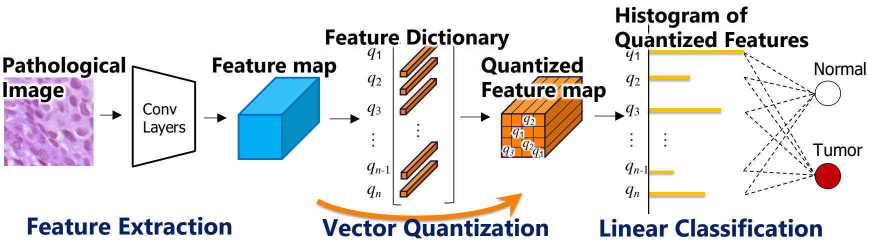


Proposed Method

Main idea: Making decisions by referring a dictionary, which is a collection of interpretable features learned by CNN



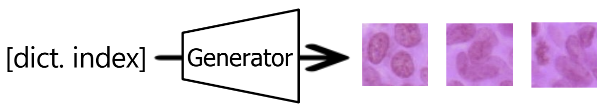
Diagnosis Network
Learning feature extraction and constructing its dictionary in end-to-end fashion
→ Enabling accurate classification



Visualization Network
This network generates images via dictionary's index.
→ Making CNN features interpretable in visual

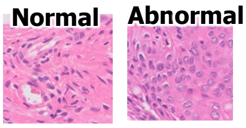
Autoregressive models can generate clear images by directly inferencing image distribution

Conditional Pixel CNN



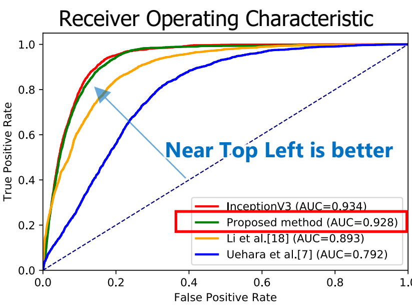
Result and Discussion

Tissue classification
- ➢ Uterine cervix
  - ✓ Normal : Train 84,194 Test 27,601
  - ✓ Abnormal : Train 12,088 Test 2,286
  - ✓ Image size : 256x256 pixel

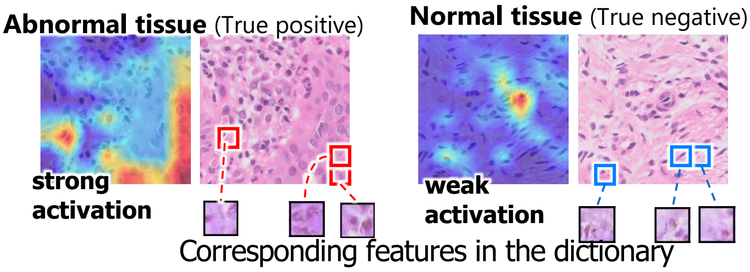


Comparison methods
- ➢ Inception V3 [Liu +, 2017] : **State of the art**
- ➢ ProtoNet [Li +, 2018] : Explainable neural network
- ➢ Dictionary based CNN [Uehara+, 2019] : Explainable neural network

Classification result
- ➢ Our method yielded **0.93 of AUROC**
  - ✓ Comparable to SOTA
  - ✓ Better than other explainable CNNs
- ➢ Learning dictionary in end-to-end fashion was effective



Explanations
Redder area is important for classifying atypical tissue



Conclusion
- Our method brings great advantages compared with the conventional methods in terms of accuracy and explainability
- We plan to confirm the features in the dictionary from a pathology viewpoint