Application of deep learning in thyroid cancer classification

based on cytological images

Phạm Ngọc Hải

Hanoi University of Science, VNU

April 18, 2024

Contents

Introduction

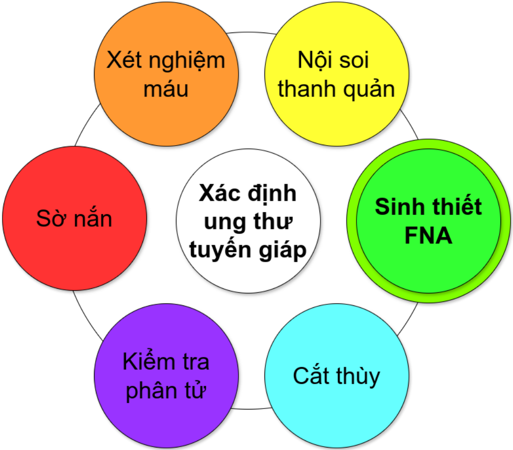
Problem Statement

Early and accurate prediction of thyroid cancer helps achieve highly effective treatment

- Thyroid cancer is a common type of cancer, especially in women and individuals over the age of 45
- According to research from the US National Cancer Institute:
  - Detecting thyroid cancer at an early stage increases the patient's chance of survival up to 98% compared to only 66% when detected at a later stage
  - Accurate diagnosis of the type and stage of cancer increases the chances of selecting suitable treatment methods
- As a result, survival rates and quality of life after treatment have significantly improved

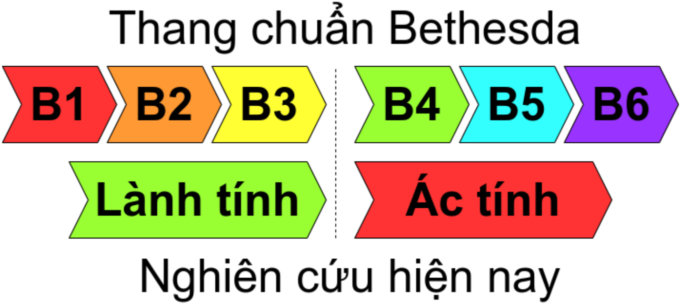
Problem Statement

FNA aspiration (FNA) method has many advantages in predicting disease status according to Bethesda standards



Problem Statement

Limitations of current studies in automating thyroid cancer classification based on cytology images



Problem Statement

Limitations of current studies in automating thyroid cancer classification based on cytology images

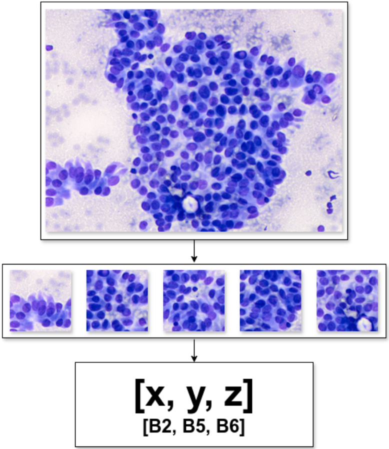
TABLE 2 A summary of recent research applying AI to thyroid cytology specimens.

Study	Year	Aim	Technique	Level	Sample Size	Reported Metrics	Results
Savala et al. (60)	2018	FTC vs FA	Neural network	Slide	57	Accuracy	100%
Margari et al. (61)	2018	Predict TBS diagnosis	Classification and regression trees	Slide	521	Accuracy	91%
		Benign vs malignant				Accuracy Sensitivity Specificity	93.0% 92.4% 93.6%
Sanyal et al. (62)	2018	PTC vs non-PTC	CNN	Image	370	Accuracy Sensitivity Specificity	85.1% 90.5% 83.3%
Guan et al. (63)	2019	PTC vs benign	CNN	Slide	279	Accuracy	95.0%
				Image	887	Accuracy Sensitivity Specificity	97.7% 100% 94.9%
Maleki et al. (64)	2019	PTC vs NIFTPs and noninvasive EFV-PTC	Support vector machine	Slide	59	Accuracy Sensitivity Specificity	76.1% 72.6% 81.6%

Research Objectives

- Applying some pre-trained models in image processing to classify thyroid cancer cytology images
- The classification standard is based on the Bethesda standard with 3 labels (B2, B5, B6) of the data set provided by 108 Central Military Hospital
- The remaining labels are missing because the doctor has not cleaned the data yet

Research Objectives



Literature Review

Literature Review

The methods are classification and regression trees (CARTs), support vector machines(SVM), convolutional neural network (CNN)

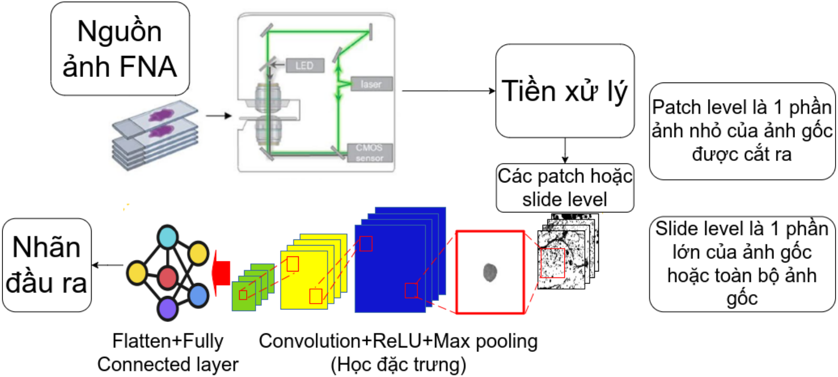
TABLE 2 A summary of recent research applying AI to thyroid cytology specimens.

Study	Year	Aim	Technique	Level	Sample Size	Reported Metrics	Results
Savala et al. (60)	2018	FTC vs FA	Neural network	Slide	57	Accuracy	100%
Margari et al. (61)	2018	Predict TBS diagnosis	Classification and regression trees	Slide	521	Accuracy	91%
		Benign vs malignant				Accuracy Sensitivity Specificity	93.0% 92.4% 93.6%
Sanyal et al. (62)	2018	PTC vs non-PTC	CNN	Image	370	Accuracy Sensitivity Specificity	85.1% 90.5% 83.3%
Guan et al. (63)	2019	PTC vs benign	CNN	Slide	279	Accuracy	95.0%
				Image	887	Accuracy Sensitivity Specificity	97.7% 100% 94.9%
Maleki et al. (64)	2019	PTC vs NIFTPs and noninvasive EFV-PTC	Support vector machine	Slide	59	Accuracy Sensitivity Specificity	76.1% 72.6% 81.6%

Research Methods

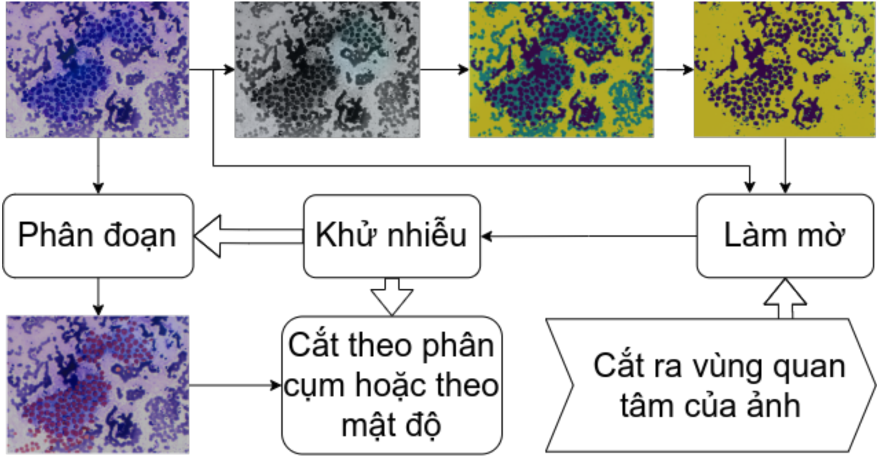
Research Background

General idea of activity flow



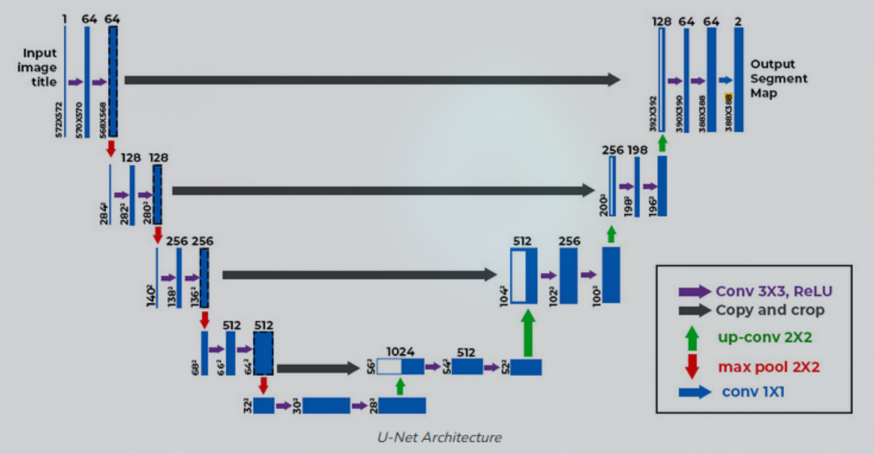
Methods

Problem 1. Identify important regions of the image



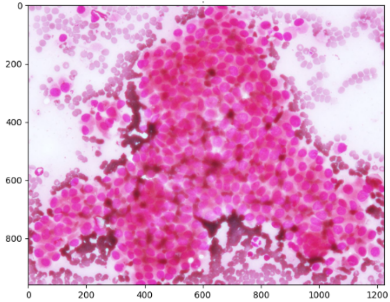
Methods

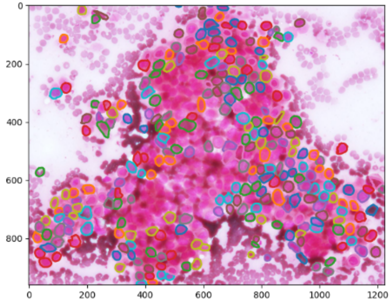
Problem 1. Identify important regions of the image U-Net architecture



Methods

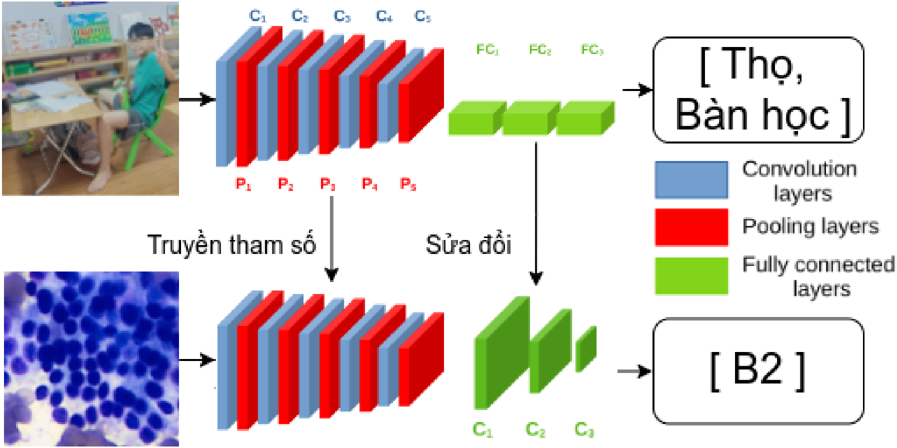
Problem 1. Identify important regions of the image U-Net result





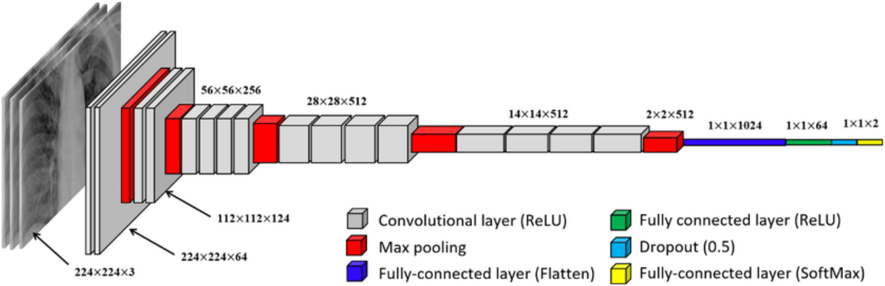
Methods

Problem 2. Using transfer learning to classify images



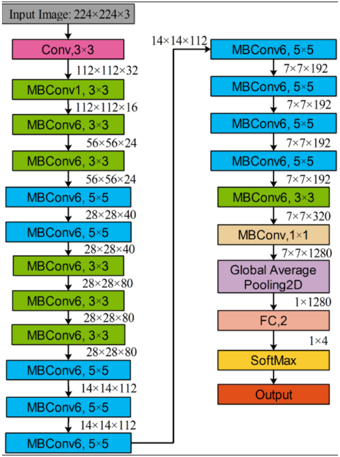
Methods

Problem 2. Using transfer learning to classify images VGG19 architecture



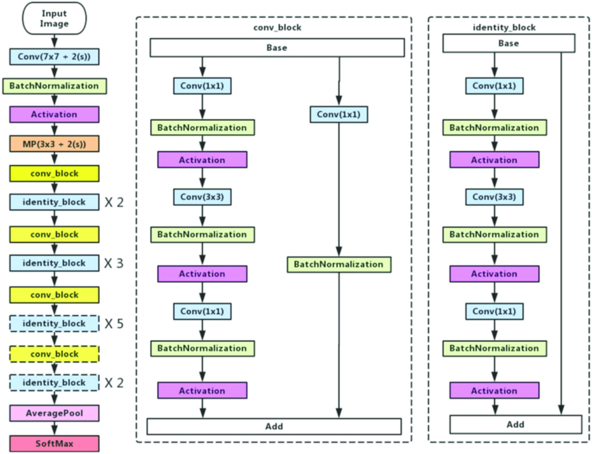
Methods

Problem 2. Using transfer learning to classify images EfficientNet architecture



Methods

Problem 2. Using transfer learning to classify images ResNet architecture



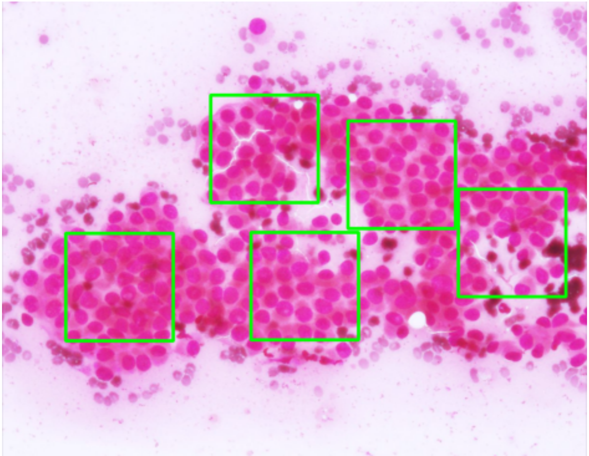
Material, Results and Discussion

Material

- Image data from 108 Central Military Hospital about 3 labels B2, B5, B6
- Labels B1, B3, B4 are missing because the hospital does not have clean data about these 3 labels
- The number of labels are: [103, 541, 777]

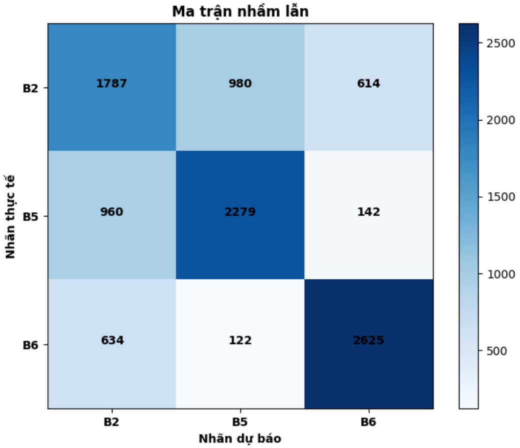
Results

Problem 1. Crop, balance data, and split into training, validation, and test sets



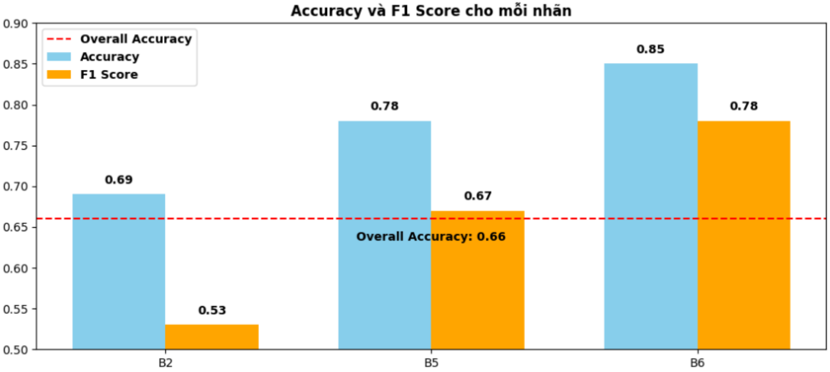
Results

Problem 2. Evaluate the trained model (patch level)



Results

Problem 2. Evaluate the trained model (patch level)



Results

Problem 2. Evaluate the trained model (slide level)

```
[ ] # Test cho 20 epoch tiếp theo (lần 6) (giữ nguyên 100 tầng đầu)
    # Đánh giá hiệu suất của mô hình trên tập test
    test_loss, test_accuracy = model.evaluate(test_generator)

    # In ra kết quả
    print("Test loss:", test_loss)
    print("Test accuracy:", test_accuracy)

    3/3 [==============================] - 0s 103ms/step - loss: 0.6822 - accuracy: 0.7349
    Test loss: 0.6821742057800293
    Test accuracy: 0.7349397540092468
```

Discussion

At patch level

- Just crop the image - don't blur other areas: ResNet > EfficientNet > VGG
- Blur, denoise areas of interest, then crop the image: Data has been prepared but not yet trained
- After cropping the image, blurring and denoising unimportant areas in the cropped image: Data has not been prepared yet

At slide level

- Achieve higher accuracy than the patch level on the test set
- However, it is difficult to create a model that avoids learning redundant features in the entire image

Conclusion and Future work

Conclusion

- With previous research (even though it was only binary classification on FNA images)
- Combined with current research results

$\rightarrow$ It shows the potential to improve deep learning to classify multi-label FNA images according to the Bethesda standard

Future work

- Test two image processing directions before training the model at patch level
- Test the direction of blurring and denoising before training the model at slide level
- Research data characteristics for training using traditional machine learning methods
- Future research needs to expand the problem to include data from multiple centers as well as develop the ability to explain the causes of classification to build trust with doctors

Thank you!

Source code

HARITO

SCAN ME

Reference