



E-SAN THAILAND CODING & AI ACADEMY

โครงการวิจัยโมเดลระบบนิเวศการเรียนรู้ที่บูรณาการ CODING & AI สำหรับเยาวชน
Model of Learning Ecosystem Platform integrate with Coding & AI for Youth

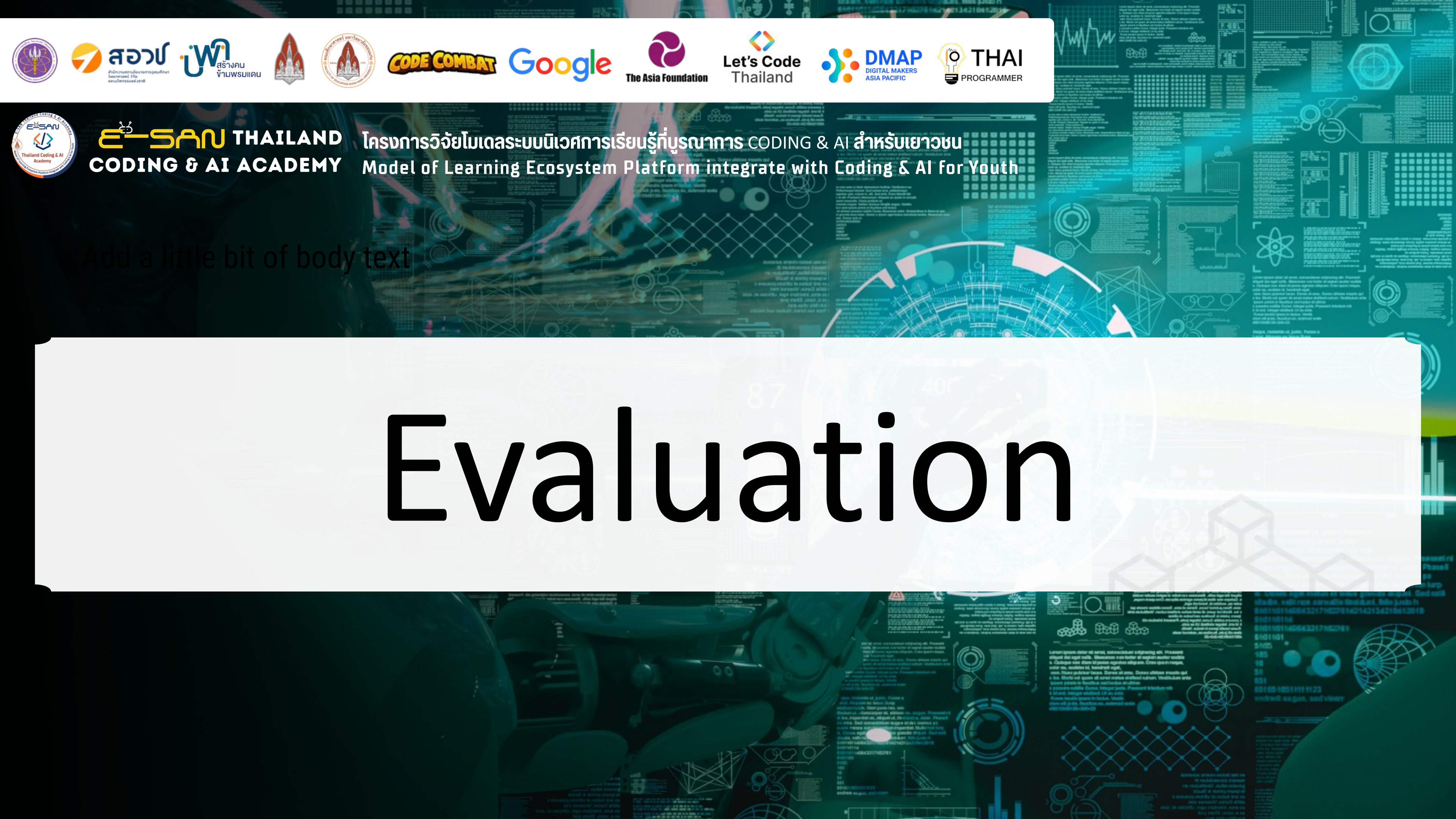


โครงการย่อยที่ 6

การพัฒนาเยาวชนเพื่อเข้าสู่วิชาชีพขั้นสูงด้าน Coding & AI
ร่วมกับ Coding Entrepreneur & Partnership: **Personal AI**

BiTNet: AI for Ultrasound Image Classification

ผศ.ดร.รณพงศ์ อินทะ
ผู้เชี่ยวชาญด้าน Computer Vision

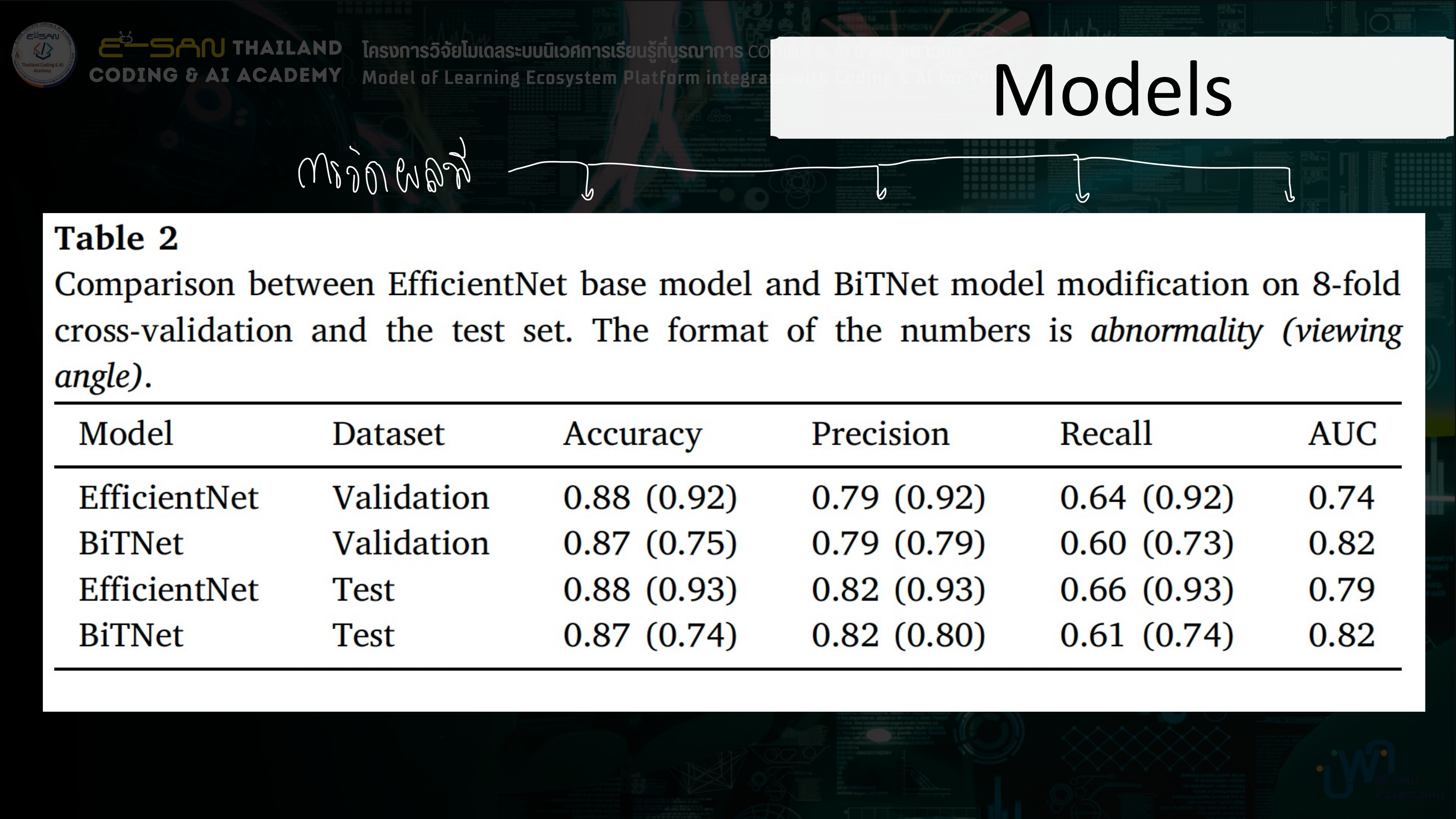


E-SAN THAILAND
CODING & AI ACADEMY

โครงการวิจัยโมเดลระบบนิเวศการเรียนรู้ที่บูรณาการ CODING & AI สำหรับเยาวชน
Model of Learning Ecosystem Platform integrate with Coding & AI for Youth

Add a little bit of body text

Evaluation



Models

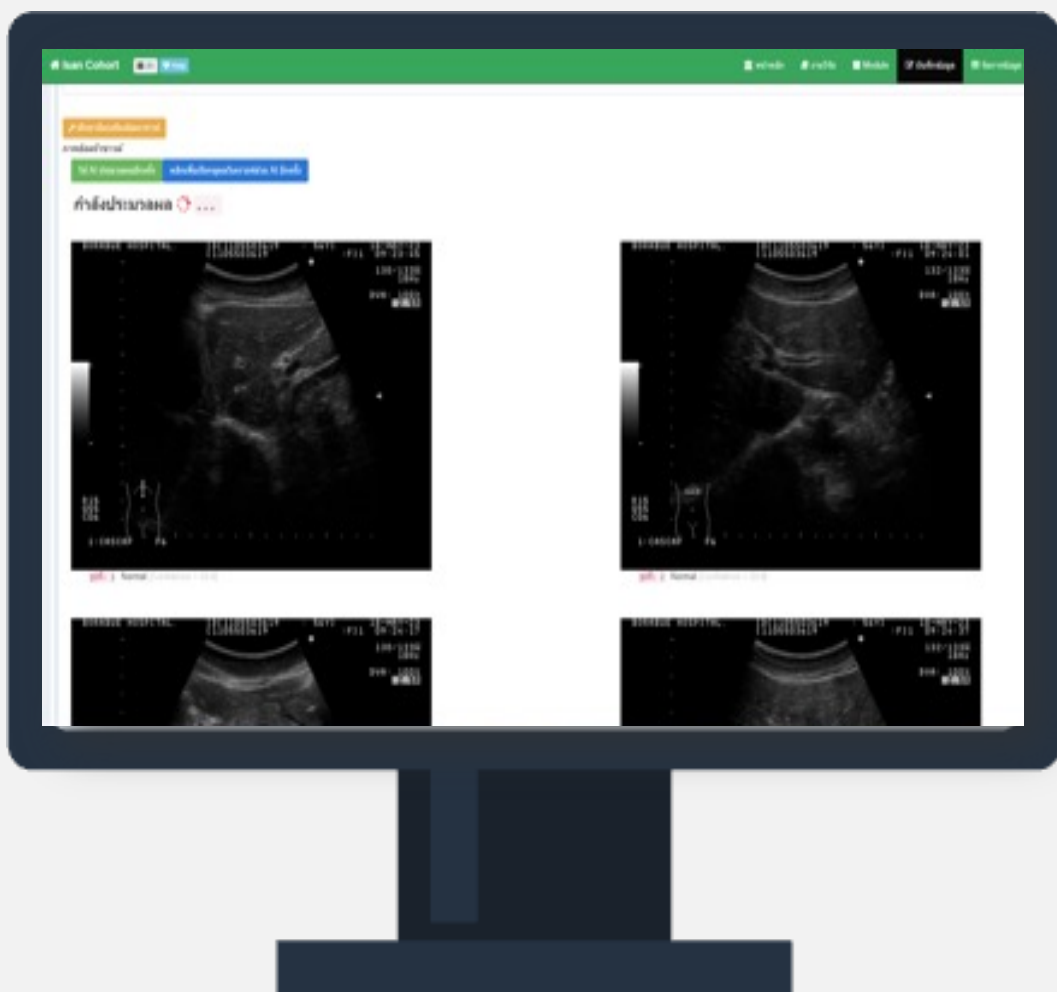
กรวัดผล



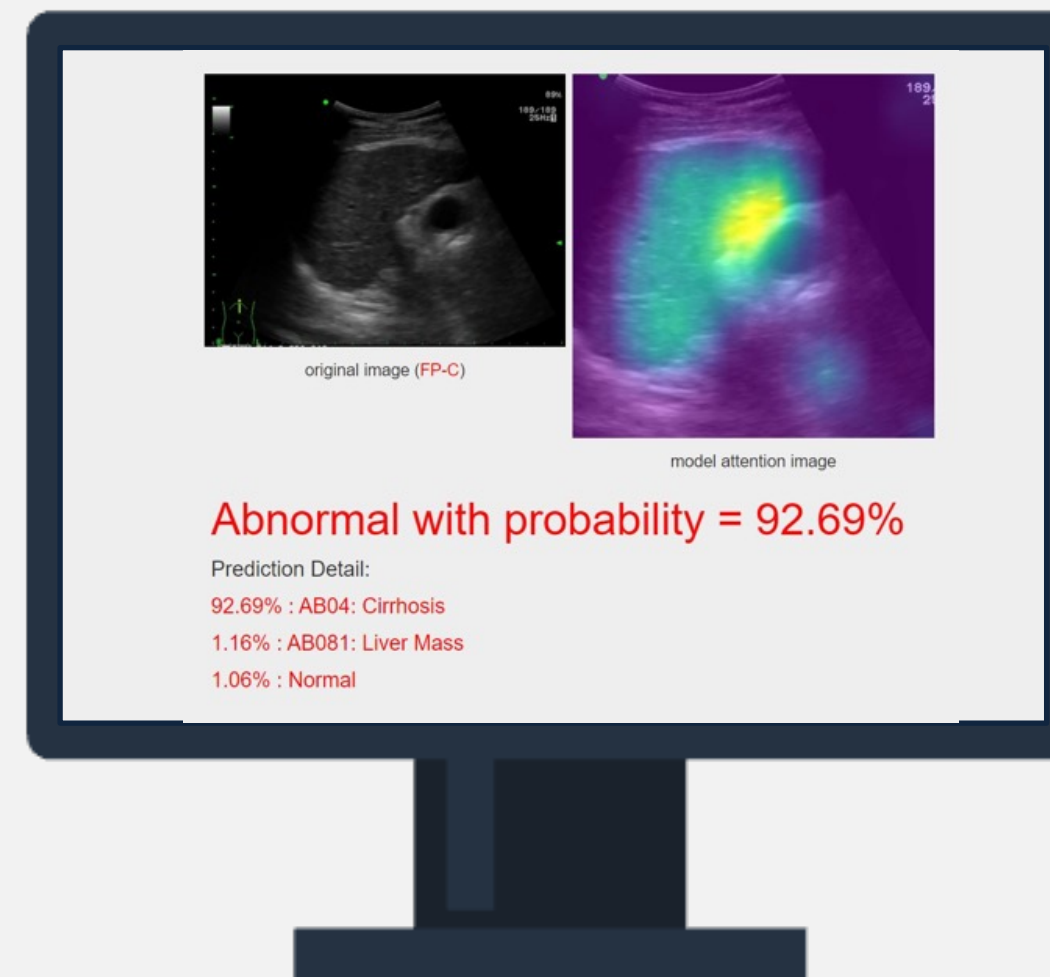
Table 2
Comparison between EfficientNet base model and BiTNet model modification on 8-fold cross-validation and the test set. The format of the numbers is *abnormality (viewing angle)*.

| Model | Dataset | Accuracy | Precision | Recall | AUC |
|--------------|------------|-------------|-------------|-------------|------|
| EfficientNet | Validation | 0.88 (0.92) | 0.79 (0.92) | 0.64 (0.92) | 0.74 |
| BiTNet | Validation | 0.87 (0.75) | 0.79 (0.79) | 0.60 (0.73) | 0.82 |
| EfficientNet | Test | 0.88 (0.93) | 0.82 (0.93) | 0.66 (0.93) | 0.79 |
| BiTNet | Test | 0.87 (0.74) | 0.82 (0.80) | 0.61 (0.74) | 0.82 |

2 Applications

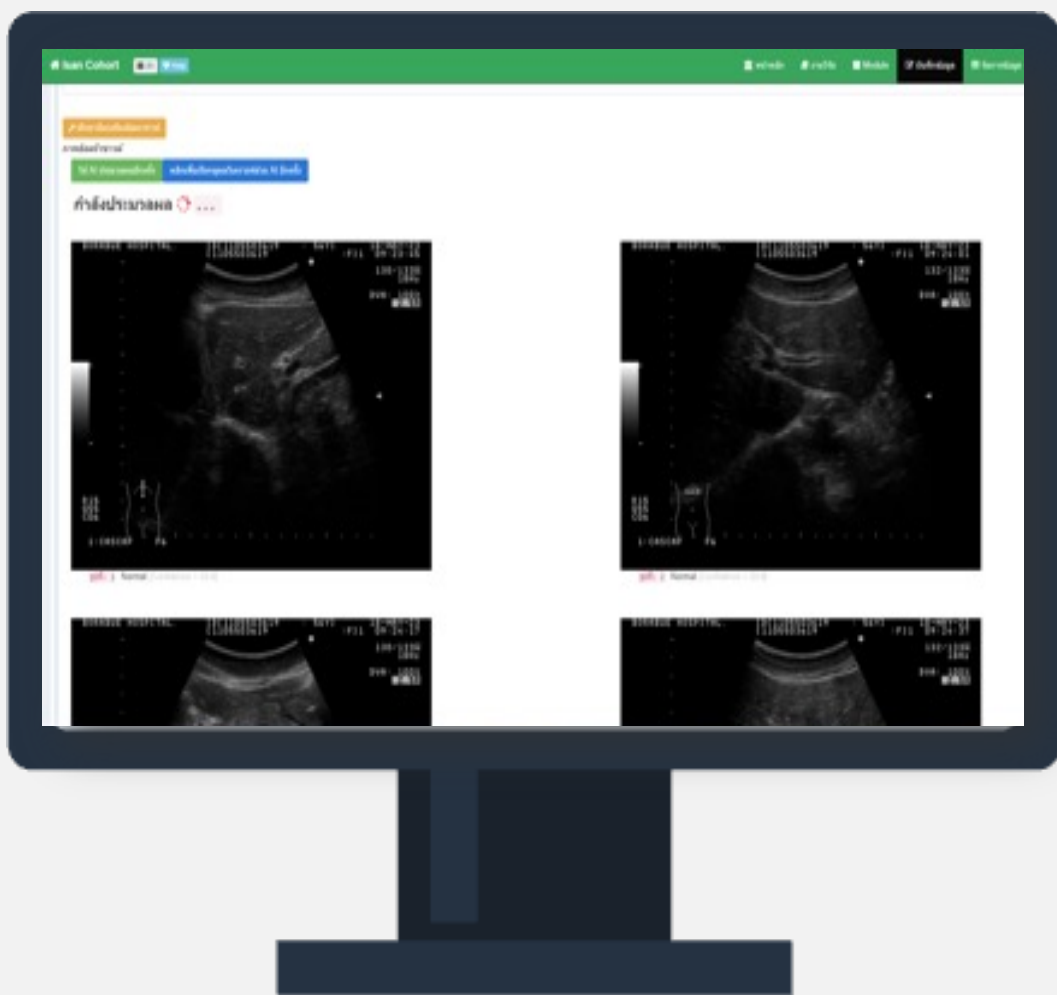


Auto Pre-screening



Assisting tool

1st Application



Auto Pre-screening

100% confidence normal

Abnormal with probability = 92.69%

Prediction Detail:

92.69% : AB04: Cirrhosis

1.16% : AB081: Liver Mass

1.06% : Normal

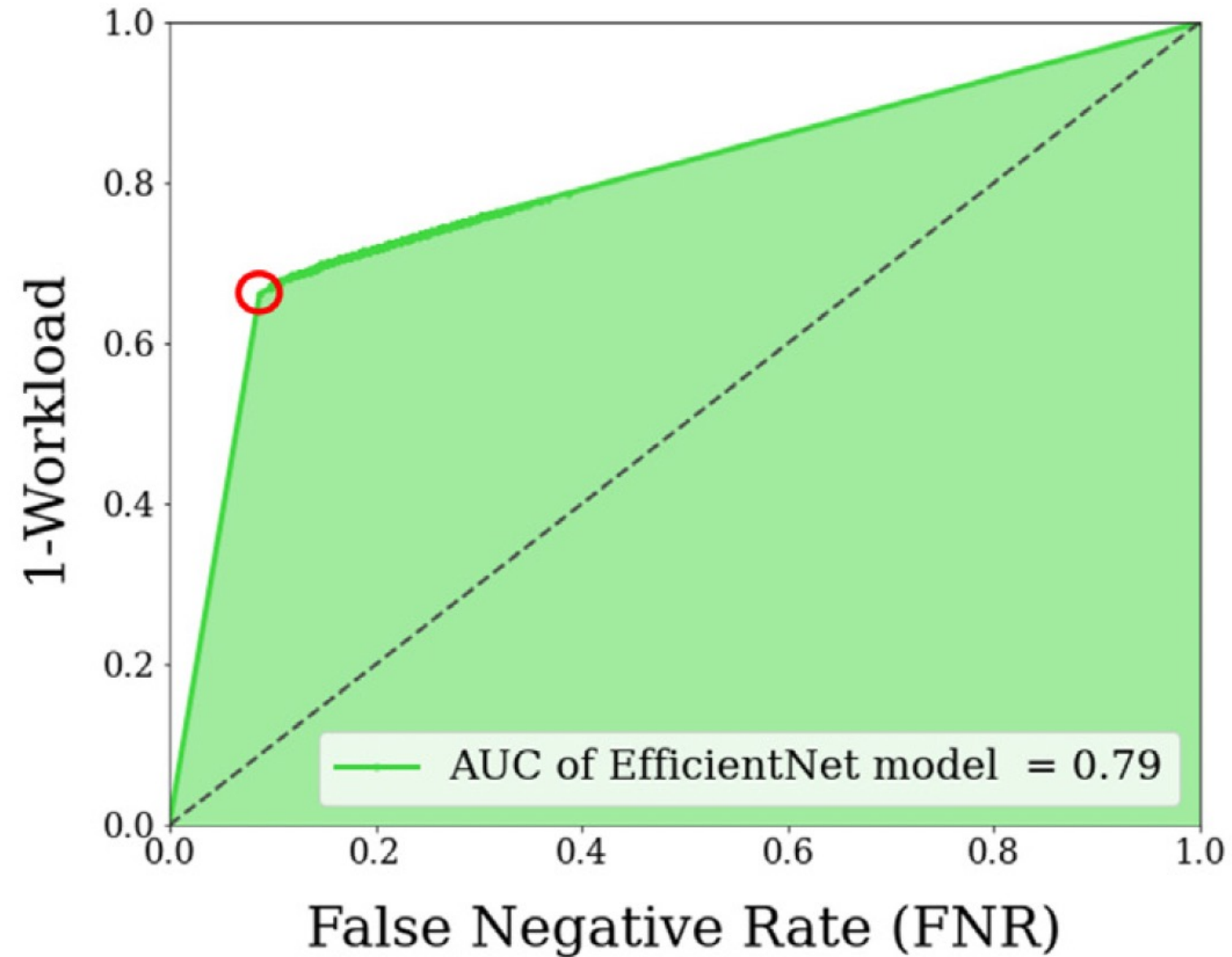
or

Otherwise

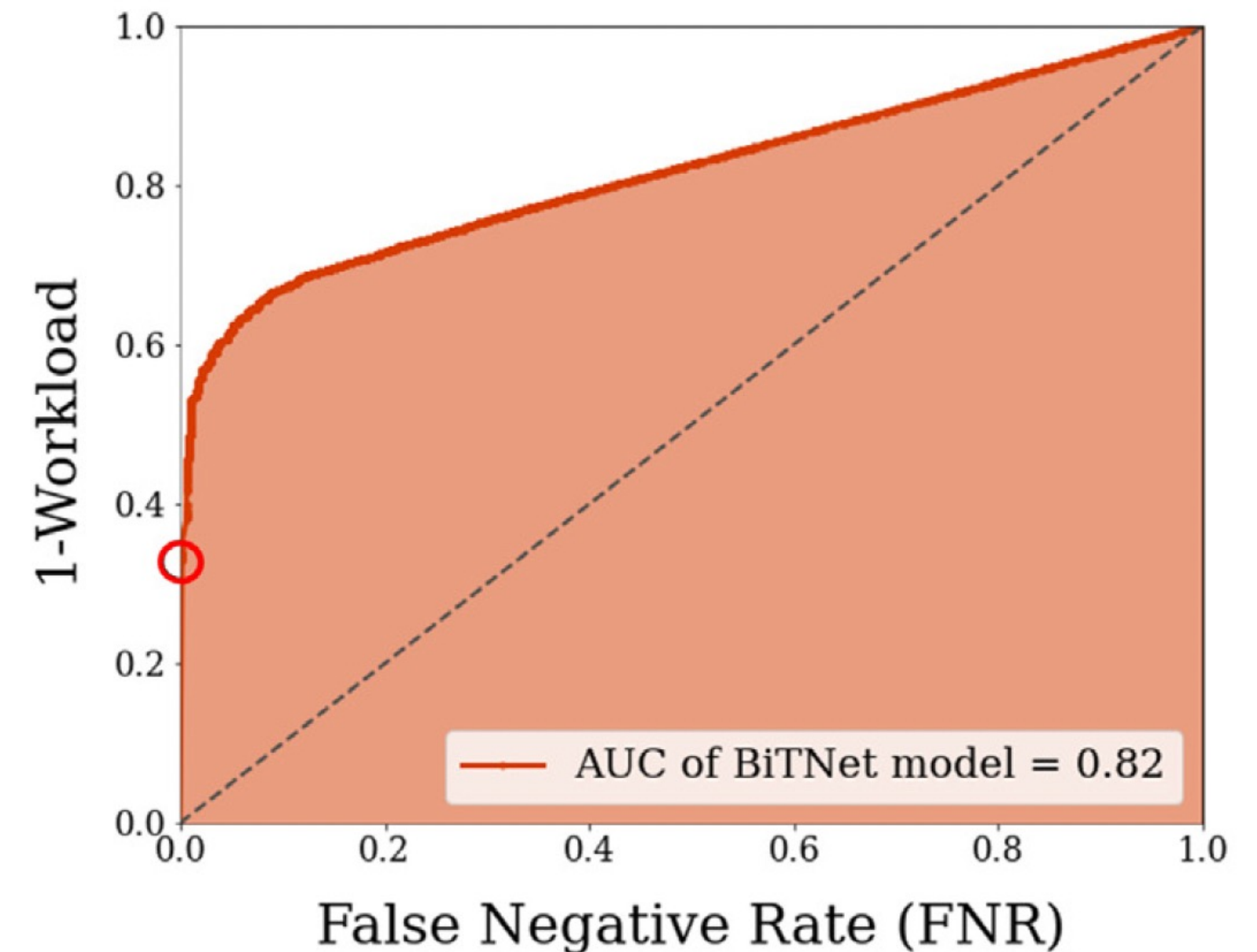
Assisting tool

Auto Pre-screening

Comparison between workload reduction-rate and false negative rate when varies-thresholds of the model.

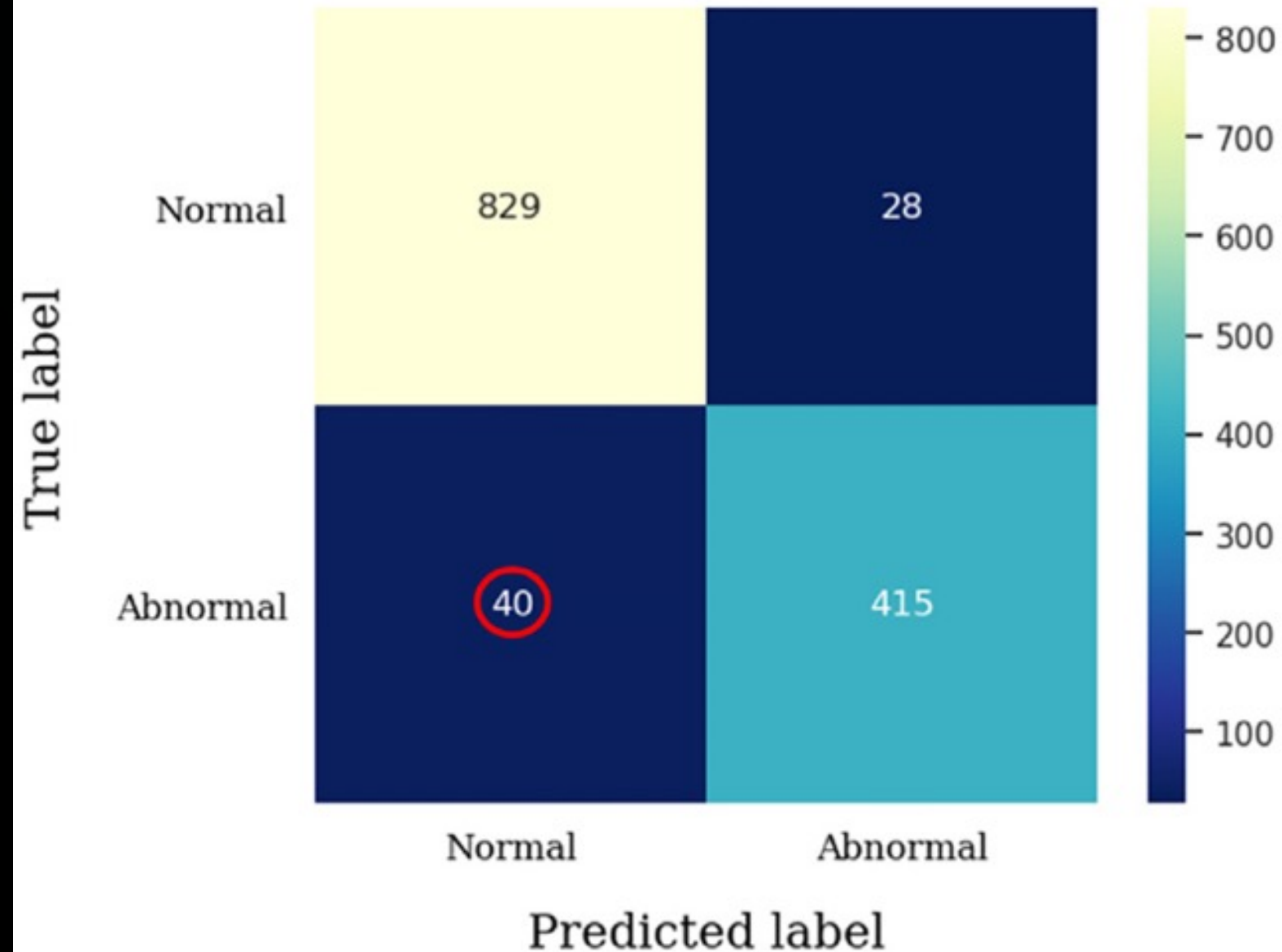


Comparison between workload reduction-rate and false negative rate when varies-thresholds of the model.

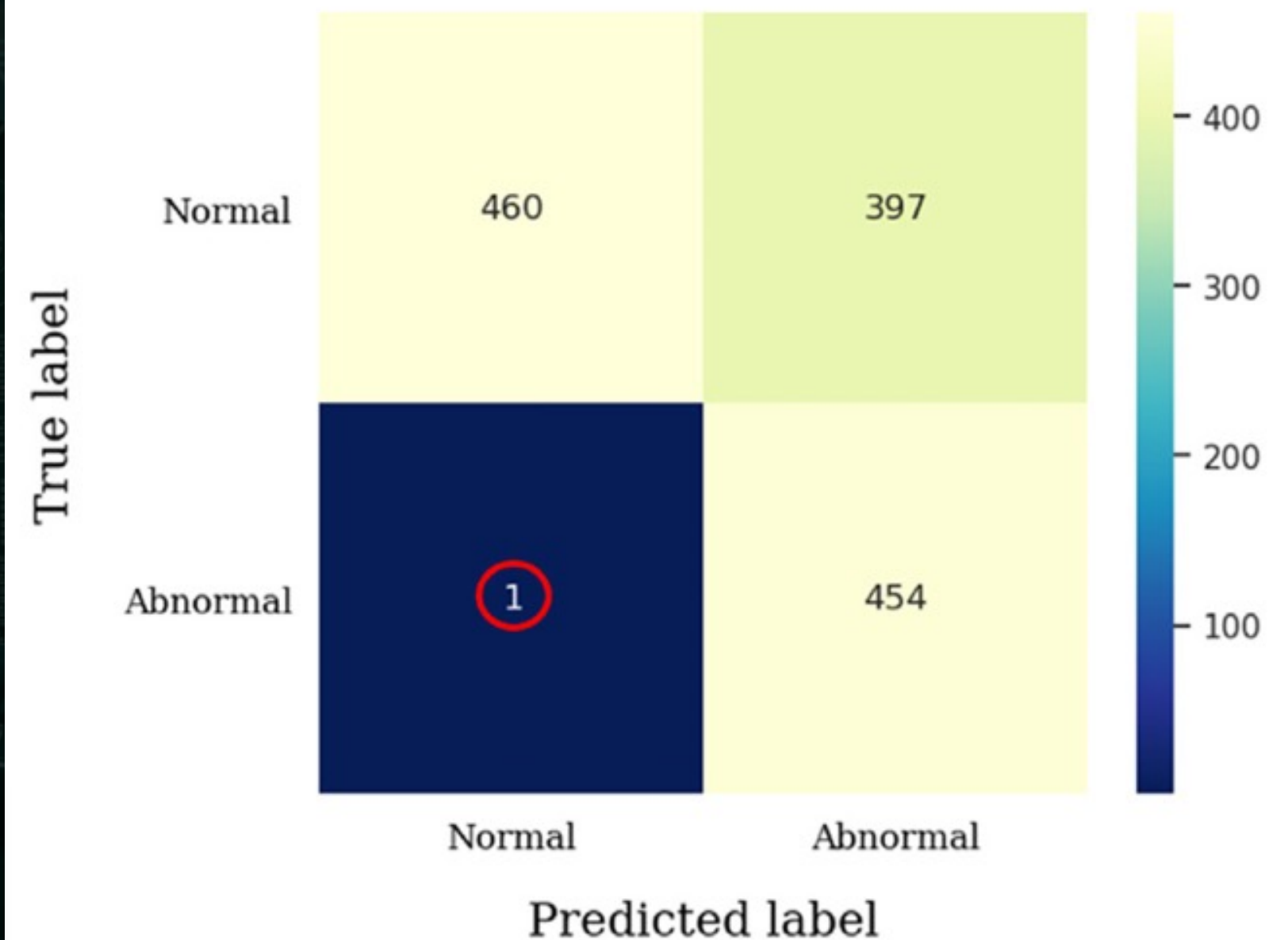


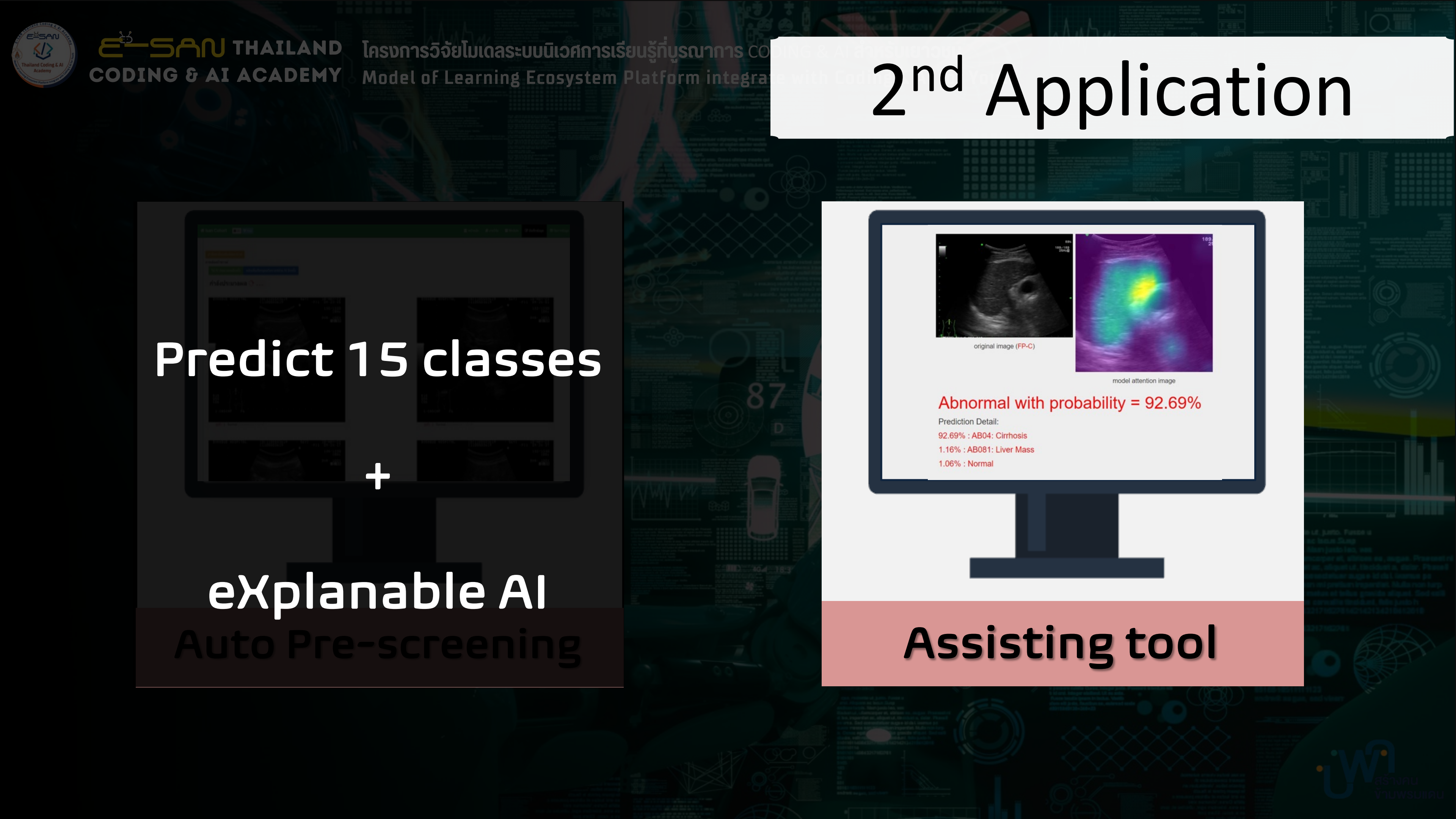
Auto Pre-screening

Confusion matrix (Normal/Abnormal class)
by EfficientNet model



Confusion matrix (Normal/Abnormal class)
by BiTNet model



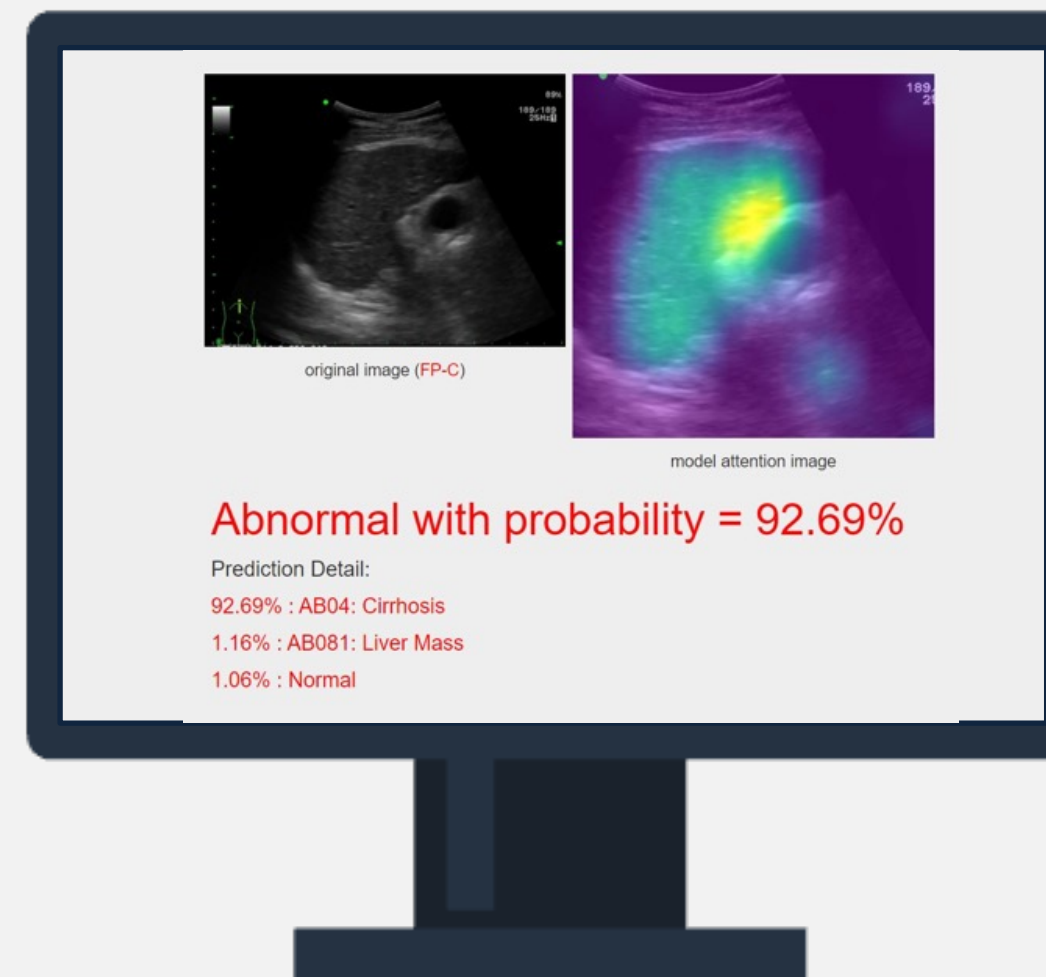


2nd Application

Predict 15 classes

+

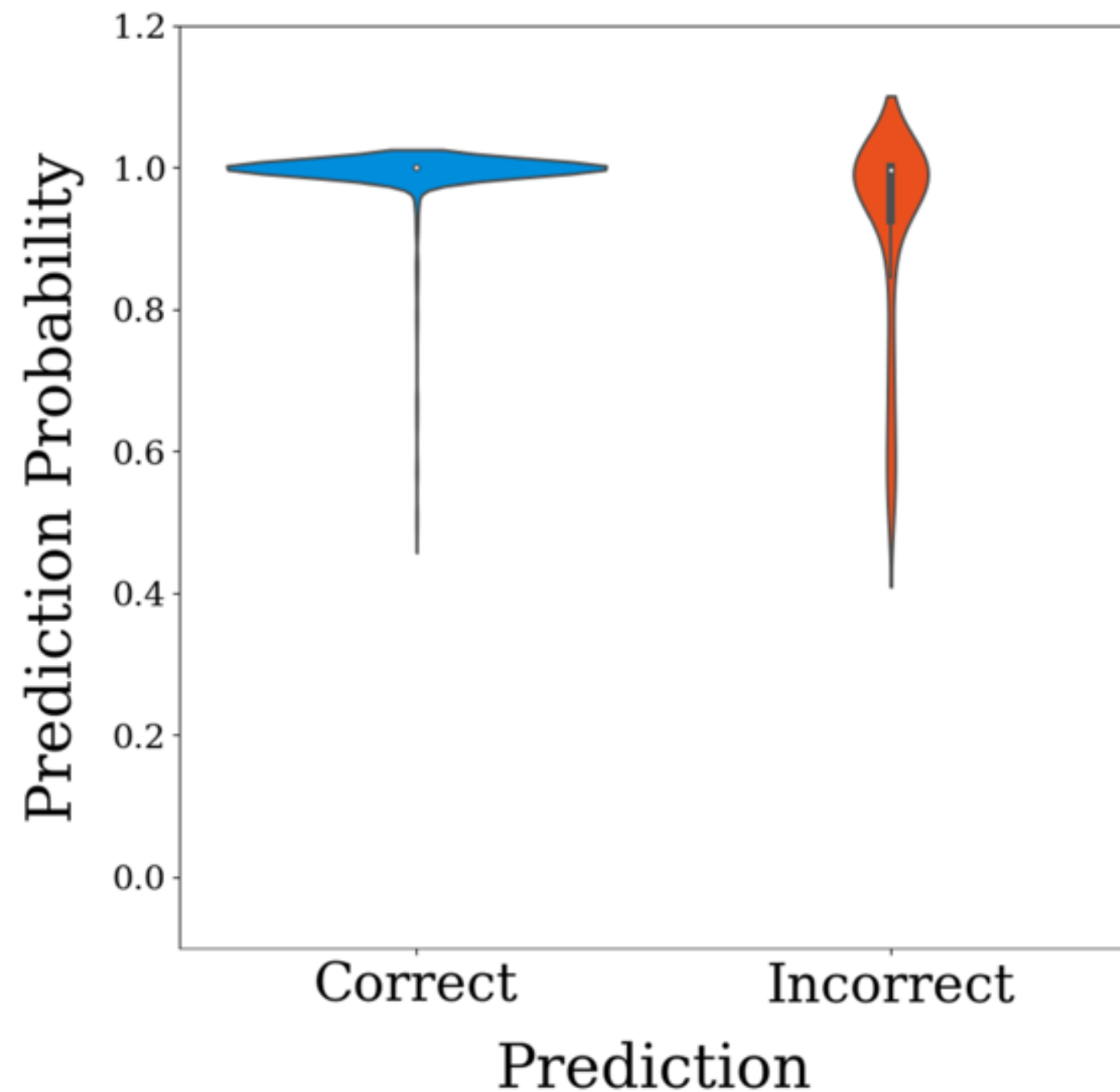
eXplanable AI
Auto Pre-screening



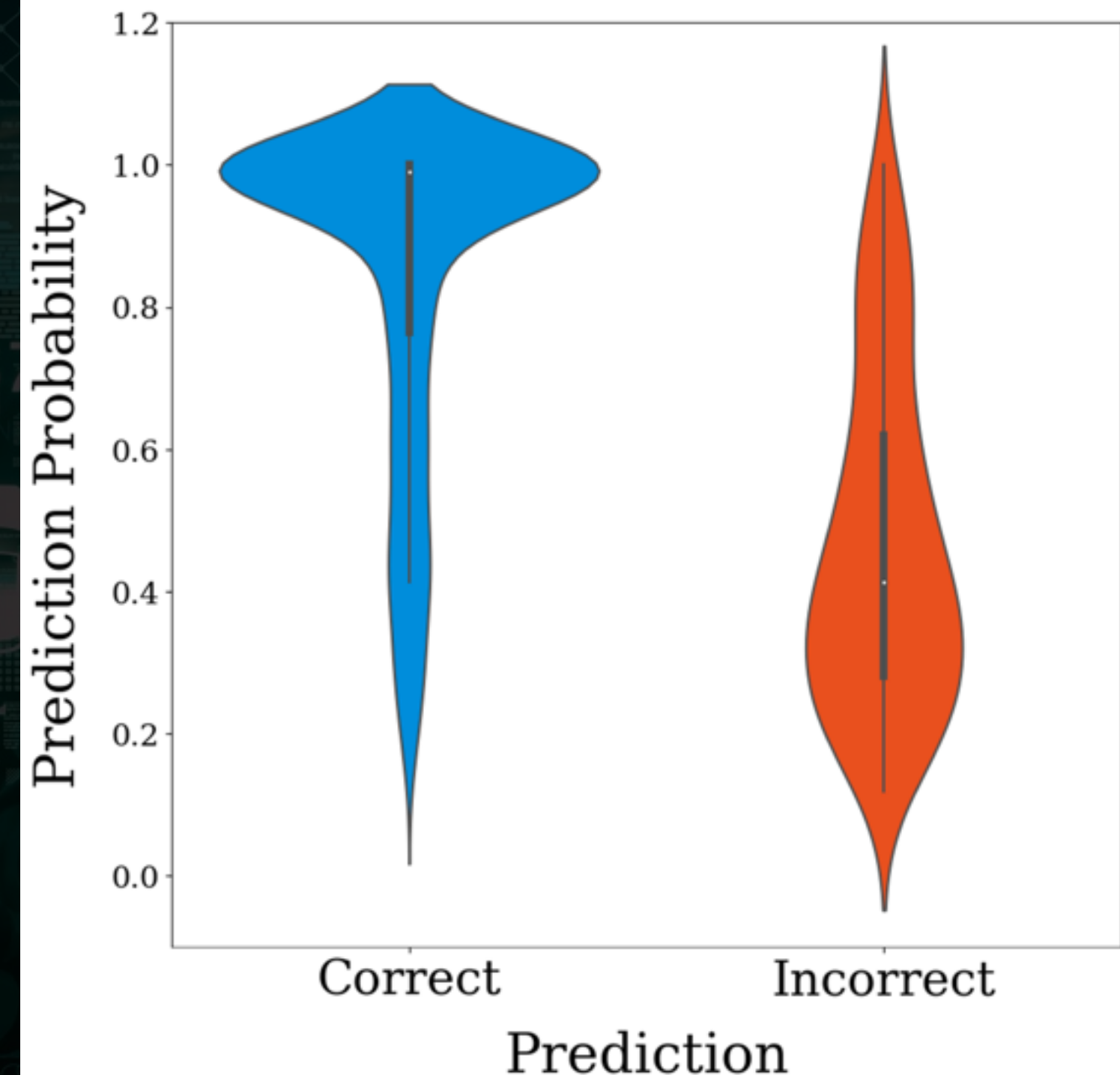
Assisting tool

Assisting tool




EfficientNet model



BiTNet model



Assisting tool



Data distribution (150 test images)

| | FP-A | FP-B | FP-C | FP-D | FP-E |
|----------|------|------|------|------|------|
| AB01 | 1 | 1 | 1 | | |
| AB02 | 1 | 1 | 1 | | |
| AB03 | 1 | 1 | 1 | | |
| AB04 | 1 | 1 | 1 | 1 | |
| AB05 | 1 | 1 | 1 | | |
| AB06 | 1 | 1 | 1 | | |
| AB07 | 1 | 1 | 1 | | |
| AB081 | 1 | 1 | 1 | | |
| AB082 | 1 | 1 | 1 | | |
| AB083 | 1 | 1 | 1 | | |
| AB09 | | 2 | 1 | | |
| AB10 | | | 3 | | |
| AB11 | | | 1 | 2 | |
| AB12 | | | | 3 | |
| Abnormal | 11 | 12 | 14 | 6 | 0 |
| Normal | 22 | 24 | 28 | 12 | 21 |

Total : 150 images
Abnormal : 43 images
Normal : 107 images

Group1 Group2

assisted unassisted




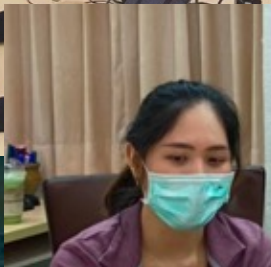

unassisted assisted

Training session

Session 1:
Diagnose 150 test images

Washout Period of 4 Weeks

Session 2:
Diagnose 150 test images



5 general practitioners (GP's), 2 residence radiologists, 2 non-hepatobiliary radiologists and 2 hepatobiliary radiologists.

Assisting tool

1. The independent samples T-Test

- Compare the means of **mean difference** in prediction confidence of the **correct and incorrect** groups between the BiTNet model and the EfficientNet model.
 - **Hypothesis**: The means of mean differences of the BiTNet model were significantly higher than those of EfficientNet.

Paired Samples T-Test

Compare of mean accuracy, precision, and recall of the diagnostic performance of the participants with and without assistance.

Hypothesis: The mean accuracy, precision, and recall scores of the diagnostic performance of the participants with assistance were significantly higher than those without assistance.

Compare of mean accuracy between the first round of the experiment and the second round of the experiment with the participants.

Hypothesis: The mean accuracy scores no significant difference between the first round and the second round of the experiment.

Compare of mean similarity scores between AI suggestion (prediction) and the final decision of the participants when assisted/unassisted.

Hypothesis: The mean similarity score of the assisted participants was significantly higher than that of the unassisted participants.

Assisting tool

1. The independent samples T-Test

1.1. Compare the means of mean difference in prediction confidence of the correct and incorrect groups between the BiTNet model and the EfficientNet model.

1.1.1. Hypothesis: The means of mean differences of the BiTNet model were significantly higher than those of EfficientNet.

2. The Paired Samples T-Test

➤ Compare of mean **accuracy**, **precision**, and **recall** of the diagnostic performance of the participants **with and without** assistance.

○ **Hypothesis**: The mean **accuracy**, **precision**, and **recall** scores of the diagnostic performance of the participants with assistance were significantly higher than those without assistance.

2.1. Compare of mean accuracy between the first round of the experiment and the second round of the experiment of the participants.

$$H_c: \mu_1 - \mu_2 = 0$$

2.1.1. Hypothesis: The mean accuracy scores no significant difference between the first round and the second round of the experiment.

$$H_a: \mu_1 - \mu_2 > 0$$

2.2. Compare of mean similarity scores between AI suggestion (prediction) and the final decision of the participants when assisted/unassisted.

2.2.1. Hypothesis: The mean similarity score of the assisted participants was significantly higher than that of the unassisted participants.

Assisting tool

1. The independent samples T-Test

➤ Compare the means of mean difference in prediction confidence of the correct and incorrect groups between the BiTNet model and the EfficientNet model.

○ **Hypothesis**: The means of mean differences of the BiTNet model were significantly higher than those of EfficientNet.

2. The Paired Samples T-Test

➤ Compare of mean accuracy, precision, and recall of the diagnostic performance of the participants with and without assistance.

○ **Hypothesis**: The mean accuracy, precision, and recall scores of the diagnostic performance of the participants with assistance were significantly higher than those without assistance.

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 - \mu_2 \neq 0$$

➤ Compare of mean **accuracy** between the **first round** of the experiment and the **second round** of the experiment with the participants.

○ **Hypothesis**: The mean accuracy scores no significant difference between the first round and the second round of the experiment.

➤ Compare of mean similarity scores between AI suggestion (prediction) and the final decision of the participants when assisted/unassisted.

○ **Hypothesis**: The mean similarity score of the assisted participants was significantly higher than that of the unassisted participants.

Assisting tool

1. The independent samples T-Test

➤ Compare the means of mean difference in prediction confidence of the correct and incorrect groups between the BiTNet model and the EfficientNet model.

➤ Hypothesis: The means of mean differences of the BiTNet model were significantly higher than those of EfficientNet.

2. The Paired Samples T-Test

➤ Compare of mean accuracy, precision, and recall of the diagnostic performance of the participants with and without assistance.

➤ Hypothesis: The mean accuracy, precision, and recall scores of the diagnostic performance of the participants with assistance were significantly higher than those without assistance.

$$H_0: \mu_1 - \mu_2 = 0$$

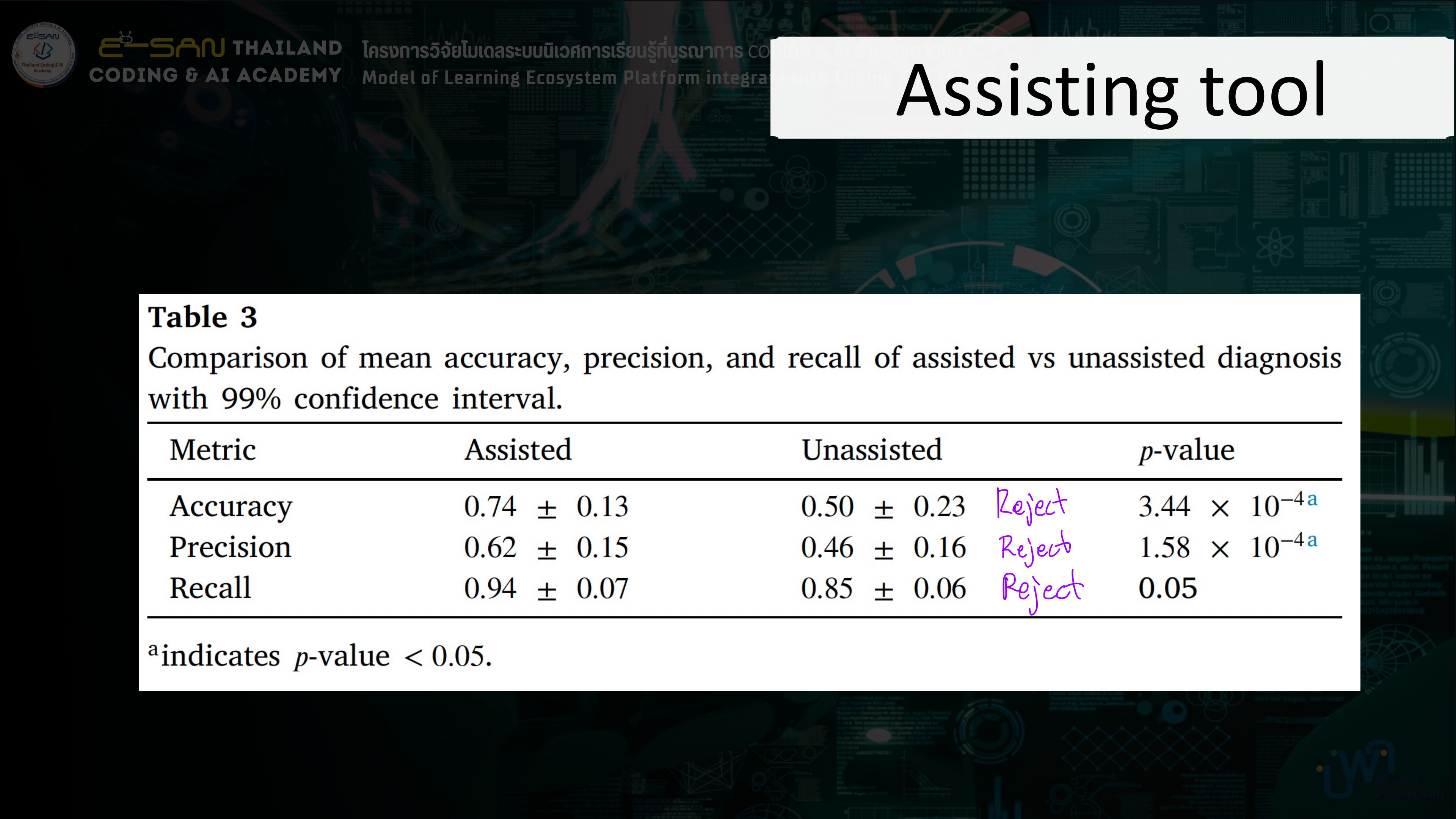
➤ Compare of mean accuracy between the first round of the experiment and the second round of the experiment with the participants.

$$H_a: \mu_1 - \mu_2 > 0$$

➤ Hypothesis: The mean accuracy scores no significant difference between the first round and the second round of the experiment.

➤ Compare of mean **similarity scores** between **AI suggestion** (prediction) and the final decision of the participants when **assisted/unassisted**.

○ **Hypothesis:** The mean similarity score of the assisted participants was significantly greater than that of the unassisted participants.



Assisting tool

Table 3

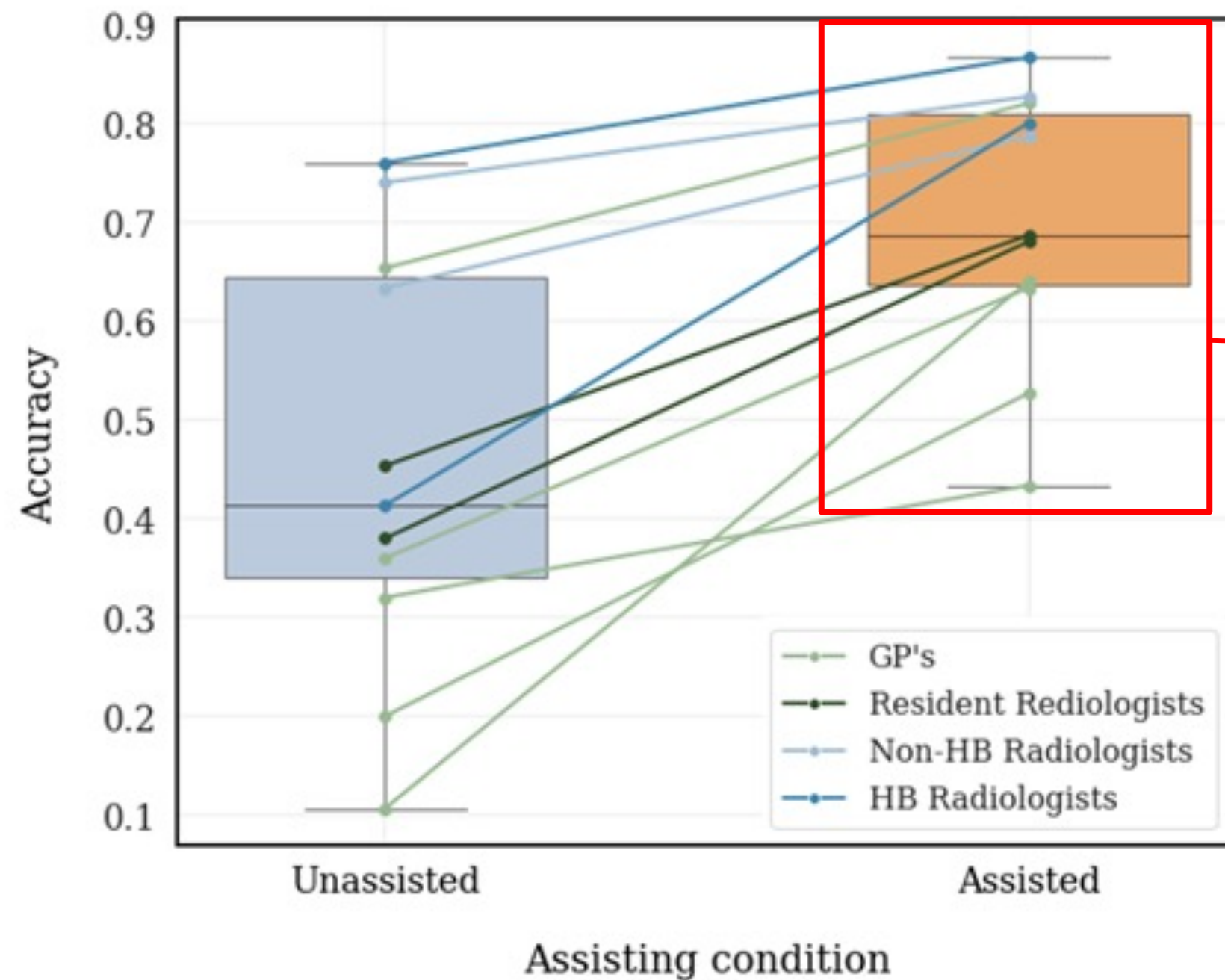
Comparison of mean accuracy, precision, and recall of assisted vs unassisted diagnosis with 99% confidence interval.

| Metric | Assisted | Unassisted | | <i>p</i> -value |
|-----------|-------------|-------------|--------|---------------------------|
| Accuracy | 0.74 ± 0.13 | 0.50 ± 0.23 | Reject | $3.44 \times 10^{-4}{}^a$ |
| Precision | 0.62 ± 0.15 | 0.46 ± 0.16 | Reject | $1.58 \times 10^{-4}{}^a$ |
| Recall | 0.94 ± 0.07 | 0.85 ± 0.06 | Reject | 0.05 |

^aindicates *p*-value < 0.05.

Assisting tool

Comparing accuracies between unassisted vs assisted



increase **overall's** accuracy
by **18%**

increase **GP's** accuracy
by **26%**

