VALIDATION TEST OF ARTIFICIAL INTELLIGENCE ASSISTED IN PIGMENTED SKIN LESIONS DIAGNOSIS

1THORFUN TREEWATANAKUL, 2CHARTURONG TANTIBUNDHIT, 3PANLOP CHAKKAVITTUMRONG, 4SAROJ SUVANASUTHI

1Division of Dermatology, Chulabhorn International College of Medicine Thammasat University, Pathum Thani, Thailand 2Center of Excellence in Intelligence Informatics, Speech and Language Technology, and Service Innovation (CILS), Faculty of Engineering, Thammasart University, Thailand 3Department of Dermatology, Faculty of Medicine, Thammasart University, Thailand 4Dermatology Clinic, Samitivej Sukhumvit Hospital, Thailand

E-mail: 1 Omamp_aha@hotmail.com, 2chturt@engr.tu.ac.th, 3panlop078@yahoo.com, 4doctorhe@gmail.com

Abstract - Artificial intelligence (AI) has been reported in high potential in classification of pigmented skin lesions (PSLs) under experimental conditions recently. Thai researchers team has trained the first AI which enable to diagnose PSLs in Thailand. AI showed high sensitivity, specificity, and accuracy in diagnosis up to 80-90%. Although the computer diagnosis of PSLs is accurate under experimental conditions, the practical value of automated diagnostic tool compared to standard is currently unknown. This study aimed to validate the diagnosis performances of AI assisted in PSLs diagnosis compared to Board-certified dermatologist using dermoscopic images.

Index Terms - Artificial intelligence, dermoscopy, pigmented skin lesions

I. INTRODUCTION

Pigmented skin lesions (PSLs) are included both benign and malignancy. They are very similar in morphologies, colors, and textures. To distinguish between malignant melanomas and benign moles is a challenging task for dermatologists. Dermoscopy, a non-invasive tool for microscopic examination of pigmented skin lesions, significantly improved the diagnostic accuracy in comparison with inspection by unaided eyes but only in specialized well-trained physicians. [1]-[3]. Recently, computer aided diagnosis system has been introduced. [4]-[5] Deep learning, a neural network paradigm of artificial, has been used in several fields particularly in visual task such as face recognition, object classification, playing strategic board game like Go, and medical screening which has been shown to exceed human performances. [6]-[8].

Thai researchers team has trained AI in classification of melanoma, nevus, and seborrheic keratosis by using dermoscopic images from International Skin Imaging Collaboration (ISIC) Challenge 2017 dataset. The result showed high performance in sensitivity, specificity, and accuracy up to 80-90%. Although the computer diagnosis of PSLs is accurate under experimental conditions, the practical value of automated diagnostic tool under clinical conditions is currently unknown.

In this study, we aimed to validate the diagnostic performances of the first artificial intelligence assisted pigmented skin lesions diagnosis in Thailand and compare the diagnostic ability to Board-certified dermatologist using dermoscopic images.

II. MATERIALS AND METHODS

This study was approved by Institutional Review Board of Bangkok Hospital Medical Center (BMC-IRB). Our researchers trained AI using DenseNet-121 algorithm with dermoscopic images of pigmented skin lesions (PSLs) from International Skin Imaging Collaboration (ISIC) Challenge 2017 dataset (see Fig. 1) We conducted cross-sectional study using 150 randomly selected dermoscopic images of PSLs as test dataset; 78 malignant melanomas, 42 seborrheic keratosis, and 30 nevi from ISIC Challenge 2017 source of data. All images were saved in JPG format. Inadequate image qualities such as poor focus or too much artifacts and images which are included multiple lesions were excluded. No overlap exists between test dataset and trained dataset for AI. Each dermoscopic image was provided one correct answer out of three choices; melanoma, seborrheic keratosis, and nevus. The test dataset was read by our AI system vs Board-certified dermatologist. No additional clinical information was provided to the dermatologist.

III. STATISTICAL ANALYSIS

Sensitivity, specificity, and accuracy were calculated according to standard formulae. Computers submitted predictions of each dermoscopic image with one out of three choices including 0, 1, 2. Each score was checked with correct (1.0) and incorrect (0.0) answer. Statistical analyses used SPSS (version 10.0).
IV. RESULTS

A. Diagnostic performances of Board-certified dermatologists

Table 1 Diagnostic performances of Board-certified dermatologist in diagnosis of PSLs

<table>
<thead>
<tr>
<th></th>
<th>MM</th>
<th>SK</th>
<th>NV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>90.0</td>
<td>69.0</td>
<td>48.7</td>
</tr>
<tr>
<td>Specificity</td>
<td>58.8</td>
<td>92.9</td>
<td>98.2</td>
</tr>
<tr>
<td>Accuracy</td>
<td>65.2</td>
<td>83.9</td>
<td>69.6</td>
</tr>
</tbody>
</table>

From Table 1, sensitivities of Board-certified dermatologist in melanoma, SK, and NV diagnosis were 90%, 69%, and 48.7%. Specificities were 58.8%, 92.9%, and 83.9% respectively. Accuracies were 65.2%, 83.9%, and 69.6%.

B. Diagnostic performances of artificial intelligence (AI)

Table 2 Diagnostic performances of artificial intelligence in diagnosis of PSLs

<table>
<thead>
<tr>
<th></th>
<th>MM</th>
<th>SK</th>
<th>NV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>83.3</td>
<td>85.7</td>
<td>73.1</td>
</tr>
<tr>
<td>Specificity</td>
<td>80.9</td>
<td>96.5</td>
<td>89.7</td>
</tr>
<tr>
<td>Accuracy</td>
<td>81.4</td>
<td>92.9</td>
<td>80.8</td>
</tr>
</tbody>
</table>

Our AI’s diagnostic performances showed sensitivity of 83.3% in melanoma diagnosis, 85.7% in SK diagnosis, and 73.1% in NV diagnosis. Specificities in diagnosis of melanoma, SK, and NV were 80.9%, 96.5%, and 89.7% respectively. Accuracies were 81.4%, 92.9%, and 80.8%. (see Table 2)

C. Comparison of diagnostic performances of Board-certified dermatologist vs artificial intelligence system

Sensitivity
In melanoma diagnosis, AI system showed lower sensitivity compared to Board-certified dermatologist (83.3% vs 90.0%). In contrast, AI has higher sensitivities in diagnosis of SK and NV. (see Fig. 1)

DISCUSSION
This is the first research to validate AI assisted in PSLs diagnosis compared to Board-certified dermatologist in Thailand. Our validation test shows that AI system has higher diagnostic performances compared to Board-certified dermatologist in almost all parameters, except for sensitivity in melanoma diagnosis and specificity in nevus diagnosis. The more images you’ve trained AI, the more accuracy you will get in the future.
CONCLUSION

Our artificial intelligence system achieves performance in diagnosis of pigmented skin lesions with a level competence comparable to Board-certified dermatologists. I suggest that other types of PSLs should be considered to be included in the validation test. Also, larger sample size and clinical dermoscopic images may be necessary for the further studies.

REFERENCES


★★★