

Original Article

Deep Neural Network for Early Image Diagnosis of Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis

Atsushi Fujimoto, MD, PhD^{a,b}, Yuki Iwai, MD^a, Takashi Ishikawa, MD, PhD^c, Satoru Shinkuma, MD, PhD^a, Kosuke Shido, MD^d, Kenshi Yamasaki, MD, PhD^d, Yasuhiro Fujisawa, MD, PhD^e, Manabu Fujimoto, MD, PhD^f, Shogo Muramatsu, PhD^g, and Riichiro Abe, MD, PhD^a Niigata, Nagaoka, Sendai, Tsukuba, and Suita, Japan

What is already known about this topic? Primary care physicians in most medical facilities rely on clinical information when Stevens-Johnson syndrome (SJS/TEN) is suspected. No artificial intelligence tools have been developed to assist in the diagnosis of SJS/TEN.

What does this article add to our knowledge? We have developed the first deep convolutional neural network (DCNN) for the image diagnosis of cutaneous adverse drug reactions. In addition to clinical information, as a diagnostic test, the DCNN may provide useful supportive data for a comprehensive diagnosis.

How does this study impact current management guidelines? The DCNN demonstrated superior performance in screening for SJS/TEN compared with dermatologists. Using the DCNN routinely in the diagnosis of cutaneous adverse drug reactions can reduce the burden on patients and physicians.

BACKGROUND: Stevens-Johnson syndrome (SJS)/toxic epidermal necrolysis (TEN) is a life-threatening cutaneous adverse drug reaction (cADR). Distinguishing SJS/TEN from nonsevere cADRs is difficult, especially in the early stages of the disease.

OBJECTIVE: To overcome this limitation, we developed a computer-aided diagnosis system for the early diagnosis of SJS/TEN, powered by a deep convolutional neural network (DCNN).

METHODS: We trained a DCNN using a dataset of 26,661 individual lesion images obtained from 123 patients with a diagnosis of SJS/TEN or nonsevere cADRs. The DCNN's accuracy of classification was compared with that of 10 board-certified dermatologists and 24 trainee dermatologists.

RESULTS: The DCNN achieved 84.6% sensitivity (95% confidence interval [CI], 80.6-88.6), whereas the sensitivities

of the board-certified dermatologists and trainee dermatologists were 31.3 % (95% CI, 20.9-41.8; $P < .0001$) and 27.8% (95% CI, 22.6-32.5; $P < .0001$), respectively. The negative predictive value was 94.6% (95% CI, 93.2-96.0) for the DCNN, 68.1% (95% CI, 66.1-70.0; $P < .0001$) for the board-certified dermatologists, and 67.4% (95% CI, 66.1-68.7; $P < .0001$) for the trainee dermatologists. The area under the receiver operating characteristic curve of the DCNN for a SJS/TEN diagnosis was 0.873, which was significantly higher than that for all board-certified dermatologists and trainee dermatologists.

CONCLUSIONS: We developed a DCNN to classify SJS/TEN and nonsevere cADRs based on individual lesion images of erythema. The DCNN performed significantly better than did dermatologists in classifying SJS/TEN from skin

^aDivision of Dermatology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan

^bMedical Bit Valley Aile Home Clinic, Nagaoka, Japan

^cDepartment of Medical Informatics, Niigata University Medical and Dental Hospital, Niigata, Japan

^dDepartment of Dermatology, Tohoku University Graduate School of Medicine, Sendai, Japan

^eDepartment of Dermatology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan

^fDepartment of Dermatology, Osaka University Graduate School of Medicine, Suita, Japan

^gDepartment of Electrical and Information Engineering, Niigata University Graduate School of Science and Technology, Niigata, Japan

This work was supported by the ICT infrastructure establishment and implementation of artificial intelligence for clinical and medical research from the Japan Agency for Medical Research and Development (JP201k1010036).

Conflicts of interest: The authors declare that they have no relevant conflicts of interest.

Received for publication October 31, 2020; revised September 5, 2021; accepted for publication September 8, 2021.

Available online ■■

Corresponding authors: Riichiro Abe, MD, PhD, Division of Dermatology, Niigata University Graduate School of Medical and Dental Sciences, 1-757 Asahimachidori, Chuo-ku, Niigata 951-8510, Japan. E-mail: aberi@med.niigata-u.ac.jp. Or: Atsushi Fujimoto, MD, PhD, Division of Dermatology, Niigata University Graduate School of Medical and Dental Sciences, 1-757 Asahimachidori, Chuo-ku, Niigata 951-8510, Japan; Department of Dermatology, Aile Home Clinic, Medical Bit Valley Aile Home Clinic, 1-10-13 Shimoyanagi, Nagaoka, Niigata 940-2013, Japan E-mail: afujimoto@med.niigata-u.ac.jp. Or: Shogo Muramatsu, PhD, Department of Electrical and Information Engineering, Niigata University Graduate School of Science and Technology, Niigata, Japan. E-mail: shogo@eng.niigata-u.ac.jp.

2213-2198

© 2021 American Academy of Allergy, Asthma & Immunology

<https://doi.org/10.1016/j.jaip.2021.09.014>

Abbreviations used

AI- Artificial intelligence
 AUC- Area under the curve
 cADR- Cutaneous adverse drug reaction
 DCNN- Deep convolutional neural network
 EM- Erythema multiforme
 MPE- Maculopapular exanthema
 NPV- Negative predictive value
 PPV- Positive predictive value
 SJS- Stevens-Johnson syndrome
 TEN- Toxic epidermal necrolysis
 t-SNE- t-Distributed stochastic neighborhood embedding

images. © 2021 American Academy of Allergy, Asthma & Immunology (J Allergy Clin Immunol Pract 2021;■:■-■)

Key words: Stevens-Johnson syndrome/toxic epidermal necrolysis; Cutaneous adverse drug reaction; Deep convolutional neural network; Artificial intelligence; Image diagnosis; Early diagnosis

INTRODUCTION

Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are rare, life-threatening mucocutaneous reactions most commonly triggered by medications, showing severe and extensive skin detachment.¹ The overall mortality rate among patients with SJS/TEN is approximately 30%, ranging from 10% for SJS up to 50% for TEN.² The diagnosis of SJS/TEN is difficult in the early phase and the start of appropriate treatment for SJS/TEN is often delayed, resulting in high mortality rates. Therefore, management of patients with SJS/TEN requires a rapid diagnosis. Skin lesions typically begin with ill-defined, coalescing, erythematous macules with purpuric centers. However, in the early stage, skin lesions of SJS/TEN may appear to be similar to those of nonsevere cutaneous adverse drug reactions (cADRs) such as maculopapular exanthema (MPE) or erythema multiforme (EM). In addition, because incidences of SJS and TEN are rare (for TEN, 0.4-1.2 cases/million),² most dermatologists lack sufficient experience in treating SJS/TEN, which complicates the clinical diagnoses of SJS/TEN, especially in the early stages. Although a few methods and biomarkers to distinguish between early-stage SJS/TEN and nonsevere cADRs have been established, including soluble Fas ligand³ and granulysin,^{4,5} their availability is limited to research-based analysis in particular laboratories and specialized centers. Therefore, primary care physicians in most medical facilities currently rely on clinical information to suspect and diagnose SJS/TEN. In other words, physicians decide whether to conduct further examinations or transfer patients to higher-order medical facilities based on clinical findings alone.

To overcome this limitation, we aimed to develop a computer-aided diagnostic system for the clinical diagnosis of early SJS/TEN powered by a deep convolutional neural network (DCNN). Recently, deep learning has attracted attention as an artificial intelligence (AI) strategy.^{6,7} A DCNN is commonly applied in image classification tasks. Furthermore, it has been reported that DCNN models showed high performance in pattern recognition not only for general images but also for medical images.⁸ Considering these observations, it was assumed that a DCNN might contribute to

predicting SJS/TEN using images from individual skin lesions. In the future, AI technologies may enable the development of efficient methods to diagnose rare hereditary diseases more efficiently.⁹ Furthermore, imaging diagnosis of rare diseases such as SJS/TEN using a DCNN is an attractive application of AI.

This study assessed the feasibility and accuracy of a computer-aided diagnostic system powered by a DCNN for the early diagnosis of SJS/TEN. We primarily aimed to determine whether a DCNN could achieve a level of SJS/TEN detection comparable to that of dermatologists.

METHODS**Dataset and reference standards**

This study was approved by the Institutional Review Board of Niigata University (No. 2018-0284). To develop the DCNN, we constructed a dataset consisting of clinical photos of patients with a final diagnosis of cADRs. We retrospectively reviewed all patients with a confirmed diagnosis of cADRs who visited a specific facility (Niigata University, Tohoku University, University of Tsukuba, and Osaka University) from January 2006 to September 2018. This study included only patients whose causative drug was identified based on a chronologic relationship between drug exposure and the onset of the rash, and drug withdrawal and resolution of the rash, or testing for causative drug identification, such as the lymphocyte transformation test. Using previously defined criteria,^{10,11} the dermatologists classified 123 included patients into three groups: MPE, EM, and SJS/TEN. Patients with other types of cADRs, such as those with drug reactions with eosinophilia and systemic symptoms, acute generalized exanthematous pustulosis, and unclassified types of cADRs, were excluded. We labeled MPE and EM as nonsevere cADRs.

Subsequently, we annotated individual lesions on clinical pictures of each patient (Figure 1, A). Two dermatologists performed annotation visually using a rectangular design that included individual lesions in the largest area. The original clinical images contained multiple individual lesions of various phases such as erythema, papule, blister, and erosion. Because dermal erosions and blisters are not early rashes, their presence makes it easy for dermatologists to diagnose SJS/TEN. In this study, we excluded erosions and blisters and used only erythema and papules as early rashes because we aimed to develop an AI capable of diagnosing SJS/TEN from clinical images with no other information before typical TEN features appeared. Thus, we used annotated lesion images rather than complete digital images to develop the DCNN model. As a result, we obtained approximately 100 to 200 individual lesion images from each patient, and 26,661 individual lesions in total (Table I). The individual lesion images were saved in folders for each patient, and training data with correct answers were created.

Deep convolutional neural network architecture and training

We performed transfer learning with fine-tuning using GoogLeNet (GoogLeNet was the winner in image classification tasks at the ImageNet Large Scale Visual Recognition Challenge 2014 competition), which is a DCNN pretrained with 12,000,000 general images.⁸ We removed the final classification layer and replaced it to fit our dataset (Figure 1, B). Before training the model, to fit the image size to GoogLeNet, we trimmed all individual lesion images to 224 × 224 pixels so that individual lesions were centered. If the DCNN makes judgments based on extra-lesional information such as the skin texture of the normal area or the quality of photographs, the accuracy rate and performance of the DCNN may be overestimated. To prevent this

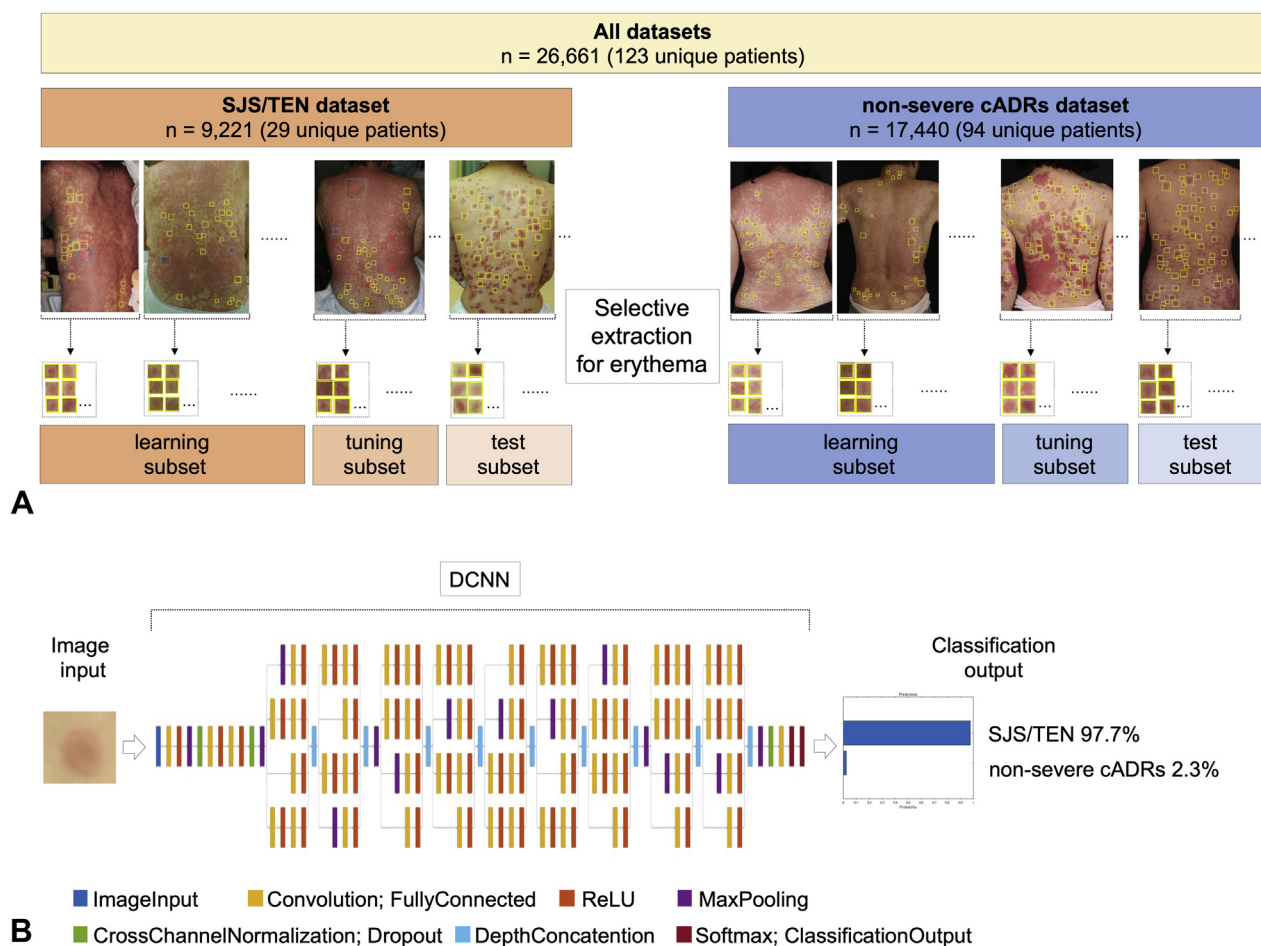


FIGURE 1. Development, validation, and network architecture of the deep convolutional neural network (DCNN). **(A)** Workflow diagram showing the datasets used to develop and validate the DCNN. Only erythema (yellow square) was annotated. Erosions (blue square) and blisters (red square) were excluded from annotation. **(B)** Network architecture of the DCNN. *cADRs*, cutaneous adverse drug reactions; *ReLU*, rectified linear activation function; *SJS/TEN*, Stevens-Johnson syndrome/toxic epidermal necrolysis.

TABLE I. Characteristics of patients

Characteristic	Stevens-Johnson syndrome/toxic epidermal necrolysis (n = 29)	Nonsevere cutaneous adverse drug reactions (n = 94)	Total (n = 123)
Sex			
Male	13 (45%)	37 (39%)	50 (40.7%)
Female	16 (55%)	57 (61%)	73 (59.3%)
Age, y			
Median (mean ± SD)	55.0 (53.7 ± 19.8)	62.0 (59.7 ± 18.0)	61.0 (58.3 ± 18.5)
Total images, n	9221	17,440	26,661

effect, the data were divided not based on an image, but rather on patients, into three sets at a ratio of approximately 80:10:10 for training, validation, and testing, respectively. For data augmentation, each image was randomly rotated at 0° to 359°. Experimental settings are summarized as follows.

The solver used stochastic gradient descent with momentum, a minibatch size of 128, a maximum of 100 epochs, an initial learning rate of 1e-4, and a validation frequency of 100.

The model was saved and used for subsequent testing. To visualize the internal features learned by the DCNN, we projected the

last hidden layer of the DCNN consisting of 1024-dimensional vectors onto two dimensions using t-distributed stochastic neighborhood embedding (t-SNE).

Statistical analysis and performance test

We evaluated the accuracy of the trained DCNN using a testing dataset for binary classification. The DCNN classified each input image into two disease classes (SJS/TEN or nonsevere cADRs) and the accuracy of a diagnoses of SJS/TEN was calculated. The classification accuracy for SJS/TEN diagnosis was calculated using the

TABLE II. Performance of deep convolutional neural network

Deep convolutional neural network output	Diagnosis	
	SJS/TEN	Nonsevere cADRs
SJS/TEN	642	664
Nonsevere cADRs	50	1,673

cADRs, cutaneous adverse drug reactions; SJS/TEN, Stevens-Johnson syndrome/toxic epidermal necrolysis.

formulas: sensitivity (%) = (number correctly classified as SJS/TEN)/(total number of SJS/TEN) $\times$ 100; specificity (%) = (number correctly classified as nonsevere cADRs)/(total number of nonsevere cADRs) $\times$ 100; accuracy (%) = ([number correctly classified as SJS/TEN] + [number correctly classified as nonsevere cADRs])/([total number of SJS/TEN] + [total number of nonsevere cADRs]) $\times$ 100; positive predictive value (PPV) (%) = (number correctly classified as SJS/TEN)/(total number classified as SJS/TEN) $\times$ 100; and negative predictive value (NPV) (%) = (number correctly classified as nonsevere cADRs)/(total number classified as nonsevere cADRs) $\times$ 100.

Subsequently, we compared the classification accuracy of DCNN with that of 10 board-certified dermatologists and 24 trainee dermatologists (doctors who had not yet obtained board certification). A total of 10 DCNNs were trained using randomly divided data for each training set; accuracy was verified using testing images for each DCNN, and the average value and 95% confidence interval for each parameter described earlier were calculated. The dermatologists classified 100 randomly presented images from the same image dataset that the DCNNs used for the test. Only one image of an individual lesion was provided to the dermatologist; other attribute information such as age, sex, or clinical course was not provided. The dermatologists' classification accuracy was compared with results of the DCNN using the Welch (two-tailed) *t* test; *P* values less than .05 were considered significant. The receiver operating characteristic curve was generated by changing the operating threshold of the DCNN classifier, and the area under the curve (AUC) was calculated. To compare the classification accuracy of the DCNN and that of the dermatologists, we also plotted the classification accuracy of the dermatologists. For transfer learning and statistical analysis, we used MATLAB R2018b Statistics and Machine Learning Toolbox and Deep Learning Toolbox (MathWorks Inc, Natick, Mass) and the Python library scipy.stats (Python Software Foundation, Wilmington, DE), respectively.

RESULTS

Performance of the DCNN

The performance of the DCNN was evaluated using 3029 testing images. Table II shows the DCNN's performance in differentiating SJS/TEN from nonsevere cADRs. The sensitivity of the DCNN (the proportion of individual lesions correctly predicted as SJS/TEN) was 92.8% (642 of 692 images), and the NPV was 97.1% (1673 of 1723 images). The specificity was 71.6% (1673 of 2337 images), the PPV was 49.2% (642 of 1306 images), and the overall accuracy was 76.4% (2315 of 3029 images). The DCNN showed high sensitivity and NPV for SJS/TEN, indicating high screening performance for SJS/TEN.

Comparison between DCNN and dermatologists

Ten board-certified dermatologists and 24 trainee dermatologists participated in this study. We compared their accuracy on

the image classification task with that of the DCNN (Table III). Regarding the sensitivity in the binary classification (SJS/TEN or nonsevere cADRs), we found no statistically significant differences between trainee dermatologists and board-certified dermatologists (27.8% vs 31.3%, Welch's *t* test [32 degrees of freedom] = -0.61 ; $P = 1.00$) (Figure 2, A). The DCNN achieved a sensitivity of 84.6% (t [32 degrees of freedom] = -18.1 ; $P < .0001$ for the DCNN model vs trainee dermatologists and t [18 degrees of freedom] = -9.29 ; $P < .0001$ for DCNN vs board-certified dermatologists). The NPVs of the trainee dermatologists, board-certified dermatologists, and DCNN were 67.4%, 68.1%, and 94.6%, respectively. The NPVs were not significantly different between the trainee and board-certified dermatologists (t [32 degrees of freedom] = -0.54 ; $P = 1.00$); however, the DCNN showed significantly higher values than that of the dermatologists (t [32 degrees of freedom] = -27.6 ; $P < .0001$ for DCNN vs trainee dermatologists and t [18 degrees of freedom] = -21.4 ; $P < .0001$ for DCNN vs board-certified dermatologists) (Figure 2, B). In contrast, the specificities of the trainee dermatologists, board-certified dermatologists, and DCNN were 80.2%, 77.7%, and 74.0%, respectively. Only the trainee dermatologists showed significantly higher results than the DCNN in terms of specificity (t [32 degrees of freedom] = 3.49 ; $P < .01$ for DCNN vs trainee dermatologists; t [32 degrees of freedom] = 1.01 ; $P = .10$ for DCNN vs board-certified dermatologists) (Figure 3, A). No significant difference was found for the PPV among the trainee dermatologists, board-certified dermatologists, and DCNN (Figure 3, B). The accuracy rates of the trainee dermatologists, board-certified dermatologists, and DCNN were 61.8%, 61.4%, and 76.6%, respectively, and the DCNN showed significantly higher results than the dermatologists (t [32 degrees of freedom] = -16.1 ; $P < .0001$ for DCNN vs trainee dermatologists and t [18 degrees of freedom] = -14.8 ; $P < .0001$ for DCNN vs board-certified dermatologists) (Figure 3, C). The AUC for the DCNN was 0.873, significantly outperforming both the trainee dermatologists and board-certified dermatologists (Figure 2, C).

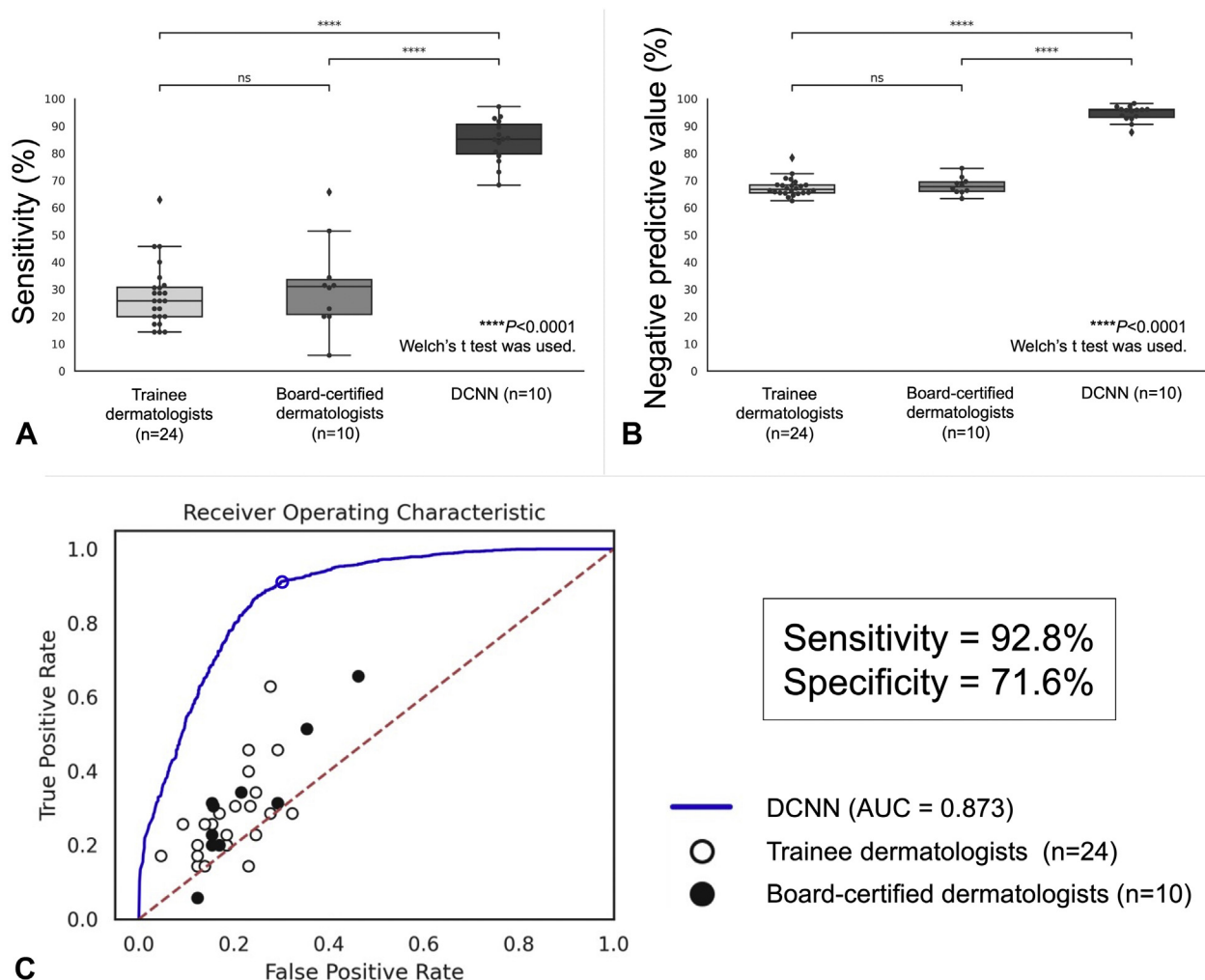
As indicators of screening performance for SJS/TEN, the DCNN's sensitivity and NPV were superior to those of the dermatologists. In contrast, the dermatologists slightly outperformed the DCNN or showed no significant difference in terms of specificity and PPV, respectively; the latter are parameters that indicate the ability to make a definitive diagnosis. To improve specificity, the DCNN should be trained for a large number of iterations using several images from various skin lesions, including viral exanthem and other non-drug induced rashes, as well as immune-related adverse events.

Visualizing the DCNN

Internal features learned by the DCNN were visualized using the dimensionality reduction method with *t*-SNE (Figure 4). Each point represents an individual lesion image projected in two dimensions from the output of the last hidden layer of the DCNN. Points belonging to the same class tended to cluster together, and SJS/TEN gathered on the opposite side to the nonsevere cADRs. However, some individual lesions of SJS/TEN and nonsevere cADRs were mixed, especially in the center of the *t*-SNE plot, which implies that some nonsevere cADR lesions were difficult to distinguish from SJS/TEN, even for the DCNN.

TABLE III. Diagnostic accuracy of deep convolutional neural network for Stevens-Johnson syndrome/toxic epidermal necrolysis lesions

Group	% Value (95% confidence interval)				
	Sensitivity	Specificity	Accuracy	Positive predictive value	Negative predictive value
Trainee dermatologists	27.8 (23.2-32.4)	80.2 (77.4-83.1)	61.8 (60.2-63.4)	43.5 (39.8-47.2)	67.4 (66.1-68.7)
Board-certified dermatologists	31.3 (20.9-41.8)	77.7 (70.8-84.5)	61.4 (59.6-63.2)	41.9 (36.2-47.6)	68.1 (66.1-70.0)
Deep convolutional neural network	84.6 (80.6-88.6)	74.0 (71.9-76.0)	76.6 (75.8-77.5)	48.2 (44.5-52.0)	94.6 (93.2-96.0)

**FIGURE 2.** Comparison of diagnostic accuracy between the deep convolutional neural network (DCNN) and dermatologists. Bee swarm boxplots of (A) sensitivity and (B) negative predictive values of image classification by trainee dermatologists, board-certified dermatologists, and the DCNN. Single measurements of the diagnostic accuracy of the trainee dermatologists, board-certified dermatologists, and DCNN are presented as filled circles. Outliers are plotted as diamonds. (C) Receiver operating characteristic analysis of the system. AUC, area under the curve; ns, not significant.

DISCUSSION

To the best of our knowledge, this is the first study to validate the use of a deep learning system to detect SJS/TEN from images of individual lesions. Our results demonstrate that the DCNN achieved generalizable detection of SJS/TEN from single images of individual lesions without having to specify explicit rules or features. The DCNN detected SJS/TEN with significantly higher sensitivity and NPV than the dermatologists based on single images of individual lesions. In contrast, the specificity and

PPV were insufficient to make a definitive SJS/TEN diagnosis. Thus, the DCNN could overdiagnose SJS/TEN. Nevertheless, considering the high mortality rate of SJS/TEN, high sensitivity and NPV may be more important than the specificity and PPV for a screening tool, and the DCNN meets this requirement in a real-world primary care setting. In practice, clinical diagnoses are based not only on images of a skin rash but also on other clinical information (ie, duration, symptoms, and medical history). Along with the clinical information for each patient, our findings

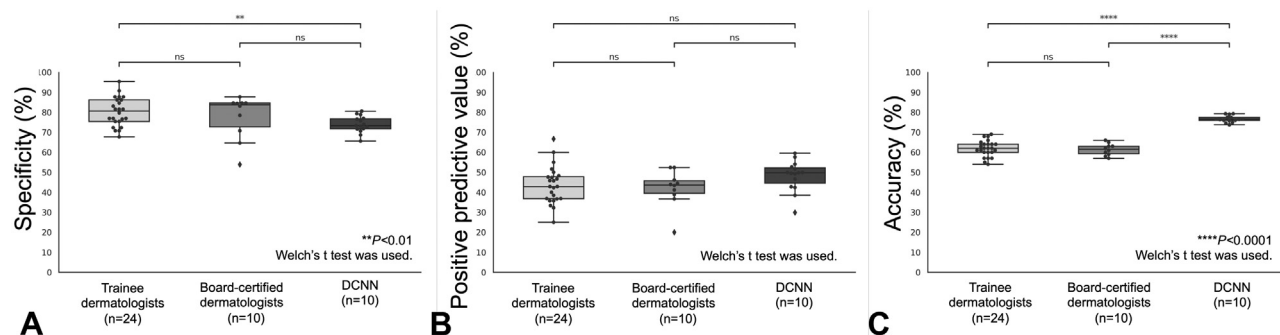


FIGURE 3. Comparison of diagnostic accuracy between the deep convolutional neural network (DCNN) and dermatologists. **(A)** Specificity of image classification by trainee dermatologists, board-certified dermatologists, and the DCNN. **(B)** Positive predictive values of image classification by trainee dermatologists, board-certified dermatologists, and the DCNN. **(C)** Accuracy of image classification by trainee dermatologists, board-certified dermatologists, and the DCNN. *ns*, not significant.

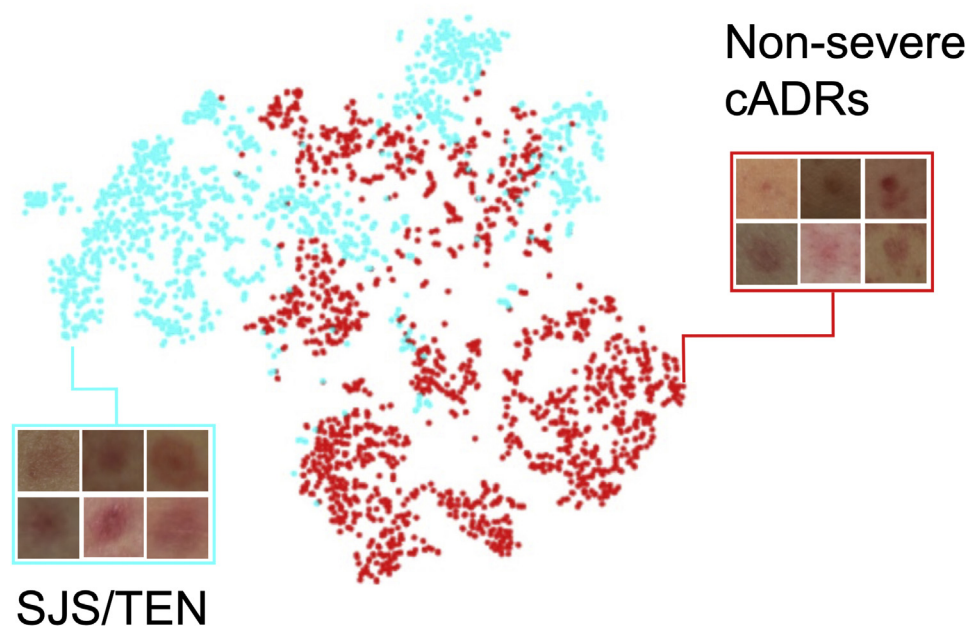


FIGURE 4. t-Distributed stochastic neighborhood embedding visualization of the last hidden layer representations in the deep convolutional neural network for Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) and nonsevere cutaneous adverse drug reactions cADRs).

suggest that as a diagnostic test, the DCNN may provide useful supportive data for a comprehensive diagnosis.

We have developed the first DCNN for the image diagnosis of cADRs. Esteva et al¹² first reported a DCNN with a higher AUC of 0.91 and an accuracy of $72.1\% \pm 0.9\%$ (mean $\pm$ SD) for detecting skin tumors using clinical or dermoscopic images of individual lesions. Their DCNN classified keratinocyte cancer and melanoma at the dermatologist level. Fujisawa et al¹³ also demonstrated that their DCNN classified images of skin tumors more accurately than did board-certified dermatologists. However, to date, no practical DCNN has been developed to classify inflammatory skin disorders, including cADRs. Here, we developed a DCNN to classify SJS/TEN and nonsevere cADRs from individual lesion images of erythema. Currently, we have demonstrated the possibility only of binary discrimination

between SJS/TEN and nonsevere cADRs. The DCNN should be used as a screening tool for SJS/TEN only under conditions in which a physician has clinically considered either SJS/TEN or nonsevere cADRs. Nevertheless, this study suggests the viability of AI for the diagnosis of inflammatory skin diseases.

Using the one-by-one method, we successfully developed AI system focused on a rare skin disease such as SJS/TEN. To apply deep learning to SJS/TEN diagnoses, sufficient training data are required. However, owing to the rarity of the disease, early SJS/TEN images are limited. To overcome this problem, we annotated multiple lesions individually from a single clinical image. Subsequently, we excluded non-early cutaneous lesions such as blisters and erosions from the target of annotation such that only erythematous lesions were annotated. Using this method, it was possible to obtain sufficient training data of

SJS/TEN lesions from relatively few patients. Furthermore, this method may allow the development of diagnostic imaging tools for other rare skin diseases.

We will try to look for evidence of SJS/TEN; specifically, dermatologists diagnose diseases by finding something more than erythema. This means that it is difficult to diagnose SJS/TEN when dermatologists find no evidence of TEN. Therefore, in this study, we aimed to develop an AI capable of diagnosing SJS/TEN from clinical images reliably and accurately with no additional information, before typical TEN features appeared. Thus, we did not use the general images.

Regarding versatility, the diagnostic assistance test of our DCNN has some advantages over previously reported tests. It can be used at any facility, provides fast results, and is noninvasive. Patients with SJS/TEN often visit primary care physicians during early onset. Even when cADRs may be suspected from drug history and symptoms, it is difficult to distinguish early SJS/TEN from nonsevere cADRs. To date, some early diagnostic markers of SJS/TEN have been reported, such as soluble Fas ligand and granulysin.^{3,5} For instance, a rapid immunochromatographic test for serum granulysin showed a sensitivity of 80% and a specificity of 95.8% for SJS/TEN versus nonsevere cADRs. However, these biomarkers are not commercially available and can be measured only in research facilities. Although histologic epithelial cell necrosis on skin biopsy may be diagnostically useful,¹⁴ it takes considerable time to obtain results requires and imposes a burden on patients. Our DCNN showed a high sensitivity of 93.6% and specificity of 70.7% for an SJS/TEN diagnosis without being restricted by facilities and invasiveness to patients. If our DCNN is put to practical use in the field of primary care, it may improve the early diagnosis of SJS/TEN.

In this study, we developed a tool to distinguish nonsevere cADR with drug-induced SJS/TEN. We focused on drug-induced exanthem because it was difficult to diagnose viral exanthem. Although there was evidence of viral infection, the possibility of drug exanthema was not excluded in the medication received by the patient. Therefore, we tested only patients whose causative drug was identified based on a chronologic relationship between drug exposure and the onset of the rash, and drug withdrawal and resolution of the rash, or testing for causative drug identification, such as the lymphocyte transformation test.

The proposed DCNN model has limitations. Lesion images used for learning were not true early lesions. The original SJS/TEN images are not necessarily taken early in the first 1 to 2 days of onset of the disease. In this study, erythema was extracted based on stored photographs taken at the time of the first visit. Thus, it might be that the developed approach cannot diagnose SJS/TEN based on true early lesions. It is typically predicted that most erythema in many cases of SJS/TEN will progress to become blisters or erosions without therapeutic intervention. Therefore, we assumed that erythema is an early rash of SJS/TEN. However, there is a possibility that erythema may not change to blisters or erosions. Furthermore, the DCNN is likely to predict the progression to SJS/TEN by assessing subtle changes in skin color owing to the presence of subclinical blisters beneath the erythema and papules. It is

necessary to verify the performance of the DCNN in future works. This study did not focus on the differential diagnosis between viral exanthem and SJS/TEN. The DCNN system cannot be used to distinguish between viral exanthem and severe drug eruptions; further development is necessary to distinguish for this. It is interesting to analyze whether the DCNN system distinguishes raised typical target lesions (a mild form of EM) from flat atypical target lesions (a severe form of EM). However, in this system, it is difficult to recognize raised skin lesion (a three-dimensional shape) from a clinical photograph (a two-dimensional shape) accurately. A future study should add three-dimensional information to the layer analysis method.

We developed a DCNN to classify SJS/TEN and nonsevere cADRs from individual lesion images of erythema. The DCNN had superior performance in screening for SJS/TEN compared with dermatologists. Use of the DCNN routinely in the diagnosis of cADRs can reduce the burden on patients and physicians. Further prospective studies are warranted to confirm the usefulness of the DCNN for the early diagnosis of life-threatening severe cADRs.

REFERENCES

1. Stern RS, Divito SJ. Stevens-Johnson syndrome and toxic epidermal necrolysis: associations, outcomes, and pathobiology—thirty years of progress but still much to be done. *J Invest Dermatol* 2017;137:1004-8.
2. White KD, Abe R, Ardern-Jones M, Beachkofsky T, Bouchard C, Carleton B, et al. SJS/TEN 2017: building multidisciplinary networks to drive science and translation. *J Allergy Clin Immunol Pract* 2018;6:38-69.
3. Murata J, Abe R, Shimizu H. Increased soluble Fas ligand levels in patients with Stevens-Johnson syndrome and toxic epidermal necrolysis preceding skin detachment. *J Allergy Clin Immunol* 2008;122:992-1000.
4. Abe R, Yoshioka N, Murata J, Fujita Y, Shimizu H. Granulysin as a marker for early diagnosis of the Stevens-Johnson syndrome. *Ann Intern Med* 2009;151:514-5.
5. Fujita Y, Yoshioka N, Abe R, Murata J, Hoshina D, Mae H, et al. Rapid immunochromatographic test for serum granulysin is useful for the prediction of Stevens-Johnson syndrome and toxic epidermal necrolysis. *J Am Acad Dermatol* 2011;65:65-8.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
7. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med* 2020;26:900-8.
8. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017;284:574-82.
9. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med* 2019;25:60-4.
10. Auquier-Dunant A, Mockenhaupt M, Naldi L, Correia O, Schroder W, Roujeau JC, et al. Correlations between clinical patterns and causes of erythema multiforme majus, Stevens-Johnson syndrome, and toxic epidermal necrolysis: results of an international prospective study. *Arch Dermatol* 2002;138:1019-24.
11. Roujeau JC. Clinical heterogeneity of drug hypersensitivity. *Toxicology* 2005;209:123-9.
12. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-8.
13. Fujisawa Y, Otomo Y, Ogata Y, Nakamura Y, Fujita R, Ishitsuka Y, et al. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br J Dermatol* 2019;180:373-81.
14. Pereira FA, Mudgil AV, Rosmarin DM. Toxic epidermal necrolysis. *J Am Acad Dermatol* 2007;56:181-200.