Computers in Biology and Medicine 105 (2019) 151–156

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Automated detection of erythema migrans and other confounding skin lesions via deep learning

Philippe M. Burlina[a,b,*], Neil J. Joshi[a], Elise Ng[c], Seth D. Billings[a], Alison W. Rebman[d], John N. Aucott[d]

[a] Applied Physics Laboratory, Johns Hopkins University, United States
[b] Malone Center for Engineering in Healthcare, Johns Hopkins University, United States
[c] Department of Dermatology, Johns Hopkins University School of Medicine, United States
[d] Lyme Disease Research Center, Division of Rheumatology, Department of Medicine, Johns Hopkins University School of Medicine, United States

ARTICLE INFO

Keywords:
Erythema migrans
Lyme disease
Deep learning
Convolutional neural networks
Automated pre-screening

ABSTRACT

Lyme disease can lead to neurological, cardiac, and rheumatologic complications when untreated. Timely recognition of the erythema migrans rash of acute Lyme disease by patients and clinicians is crucial to early diagnosis and treatment. Our objective in this study was to develop deep learning approaches using deep convolutional neural networks for detecting acute Lyme disease from erythema migrans images of varying quality and acquisition conditions. This study used a cross-sectional dataset of images to train a model employing a deep convolutional neural network to perform classification of erythema migrans versus other skin conditions including tinea corporis and herpes zoster, and normal, non-pathogenic skin. Evaluation of the machine's ability to classify skin types was also performed on a validation set of images. Machine performance for detecting erythema migrans was further tested against a panel of non-medical humans. Online, publicly available images of both erythema migrans and non-Lyme confounding skin lesions were mined, and combined with erythema migrans images from an ongoing, longitudinal study of participants with acute Lyme disease enrolled in 2016 and 2017 who were recruited from primary and urgent care centers. The final dataset had 1834 images, including 1718 expert clinician-curated online images from unknown individuals with erythema migrans, tinea corporis, herpes zoster, and normal skin. It also included 116 images taken of 63 research participants from the Mid-Atlantic region. Two clinicians carefully annotated all lesion images. A convenience sample of 7 non-medically-trained humans were used as a panel to compare against machine performance. We calculated several performance metrics, including accuracy and Kappa (characterizing agreement with gold standard), as well as a receiver operating characteristic curve and associated area under the curve. For detecting erythema migrans, the machine had an accuracy (95% confidence interval error margin) of 86.53% (2.70), ROCAUC of 0.9510 (0.0171) and Kappa of 0.7143.

Our results suggested substantial agreement between machine and clinician criterion standard. Comparison of machine with non-medical expert human performance indicated that the machine almost always exceeded acceptable specificity, and could operate with higher sensitivity. This could have benefits for prescreening prior to physician referral, earlier treatment, and reductions in morbidity.

1. Introduction

Lyme disease is currently the most common tick-borne disease in the northern hemisphere, with over 300,000 new cases estimated annually in the United States alone [1–3]. The bacterial agent of Lyme disease, *Borrelia burgdorferi*, is transmitted into the skin through the bite of an infected tick. An average of 7–14 days later, a round or oval, red, centrifugally expanding skin lesion called erythema migrans (EM) typically appears [4]. EM marks the site of the initial infection of the skin in approximately 70–80% of cases [5]. Without antibiotic treatment, the EM lesion persists for a median of 4 weeks before resolving spontaneously [4].

Visual recognition of EM, along with a history of potential exposure to ticks, remain the primary criteria for diagnosis of early Lyme disease [6,7]. The two-tier serologic tests which are currently available to clinicians are not recommended for diagnosis during the early phase of

* Corresponding author. Applied Physics Laboratory, Johns Hopkins University, United States.
E-mail address: pburlin2@jhu.edu (P.M. Burlina).

https://doi.org/10.1016/j.compbiomed.2018.12.007
Received 16 October 2018; Received in revised form 26 November 2018; Accepted 8 December 2018

infection when EM is most likely to be present, due to their low sensitivity (less than or equal to 40%) [8]. Direct detection of *Borrelia burgdorferi* in blood or biopsy samples can be performed, but these tests are generally available only in research settings. In addition, they are not always practical for use by diagnosing clinicians given the extended processing time for results [9].

Despite its importance for the early diagnosis of Lyme disease, lesion identification remains a challenge, largely because EM often takes on a variety of appearances [10]. Notably, only 20% of patients with EM in the United States present with lesions that have classic central clearing ("ring-within-a-ring" or "bull's eye") manifestations [11]. Instead, the majority of lesions appear uniformly red or bluish-red in color [11,12]. Furthermore, 4–8% of EM have a small, central blister, which may lead to diagnoses of herpes zoster (HZ) [13]. Finally, approximately 20% of patients have multiple EM at the time of diagnosis, arising from hematogenous dissemination of the bacteria, which can be misdiagnosed as urticaria, erythema annulare centrifugum, or other annular skin disorders [14]. Among the general public, one internet-based survey found that respondents correctly identified a classic EM with central clearing 73% of the time, and vesicular, uniformly red, bluish-purple, or disseminated manifestations each less than 30% of the time [15]. Even general practitioners have difficulty recognizing EM. In one study, general practitioners correctly identified non-target EM lesions 64% of the time and classic target lesions 80% of the time [16].

In early, uncomplicated Lyme disease, treatment with oral antibiotics is highly effective at both rapidly resolving the EM lesion and preventing potentially serious long-term complications [12,17]. If not recognized or diagnosed, Lyme disease progresses through three stages, advancing from a skin-limited disease to possible later dissemination of the bacteria into the nervous, cardiac, and rheumatologic systems [17]. Consequently, accurate recognition of the EM rash by both patients and clinicians is crucial to early diagnosis and prompt initiation of appropriate treatment.

Across many disciplines within medicine, including dermatology, there has been increased interest in harnessing artificial intelligence and deep learning (DL) to assist doctors with individuals who may have to be routinely monitored (e.g., for skin cancer) and to possibly reduce errors in classification and diagnosis [18]. To our knowledge however, identification of Lyme disease through EM images has only been addressed in one study using classical machine learning (ML) approaches [19]. Machine-based screening of skin lesions for Lyme disease also has the potential to identify a high percentage of both typical and atypical EM.

Prior to 2012, image classification in medical image analysis was largely based on applying conventional classifiers combined with human-engineered image features. More recently, much progress has been demonstrated using DL techniques [20–25]. Deep convolutional neural networks (DCNNs, for example AlexNet [22] or ResNet [24]) have demonstrated significant improvement in image classification performance for computer vision tasks such as classification. Unlike the classical ML approach, the image features computed by the DL techniques that we employ in this work are learned directly from data in an autonomous fashion via an optimization procedure applied to a dataset of images that have been labeled with ground-truth classifications for training the neural network. These image features autonomously learned by the neural network provide a better representation than image features sub-optimally designed by engineers, and generally yield better performance with regard to the accuracy of classification. Recently, DL approaches have been successfully used for performing a number of medical image analysis and diagnostic tasks, including identifying skin cancer [26] or diagnosing retinal diseases [27,28], performing fine-grained disease severity classification [29,30] or prognosis of 5-year risk [30] and have largely supplanted earlier ML approaches [31,32].

This study aims to expand on the prior state of the art to allow computer-aided EM classification by leveraging DCNNs. This could have potential applications in the prescreening of skin lesions prior to clinical evaluation.

2. Methods

This study aims to perform a binary classification of EM versus non-EM from skin images taken under variable viewpoint, illumination, and acquisition conditions. Non-EM images considered here include normal skin as well as a mixture of confuser pathologies such as HZ and tinea corporis (TC). We chose HZ and TC as the confuser pathologies as they are both lesions that may be confused with Lyme disease but require different treatment modalities [13,16]. Additionally, TC was chosen because misdiagnosis of EM as TC may lead to inappropriate self-treatment with over-the-counter antifungal medicines, further delaying the diagnosis and treatment of Lyme disease.

While our end goal is the binary classification problem, often solving a more granular problem allows the machine to achieve higher performance. We therefore first addressed the 4-class classification problem of EM vs. normal vs. TC vs. HZ and then combined the non-EM classes for the desired 2-class classification (EM vs. non-EM).

2.1. Data sources

There are currently no annotated, publicly available clinical datasets of EM images available. Therefore, to generate a sufficient sample set of images, a dataset was created using publicly available images mined through online sources. The leveraging of online images was recently shown to be successful in generating DL classification models of referable skin cancers [26].

We performed Google searches of EM, HZ, TC, and normal skin images by using a combination of pre-determined search terms such as "Erythema migrans", "Lyme", "bullseye rash", and "leg", "face", or "African American". Next, we performed a machine-based removal of full or near duplicates. Finally, two clinicians (J. A. and E. N.) were tasked with the following for EM, HZ, and TC images: a) confirming and removing any remaining duplicate images; b) parsing out inappropriate or irrelevant images; c) excluding images where group classification was uncertain or of low probability; and d) carefully annotating the remaining images with moderate to high probability of accurate group classification based on visual appearance and the estimated size of the skin lesions. Image annotation for main pathologies was also followed with a more granular annotation of EM type (e.g., single vs. multiple, etc.). The annotation effort was divided between the clinicians with each image being annotated by one clinician. A research coordinator was also tasked with similar steps for the normal skin images with no lesion present.

A previously validated set of clinical EM images was also obtained from patients enrolled in a longitudinal cohort study of early Lyme disease. Research photographs were obtained at the time of initial Lyme disease diagnosis and study entry from participants after written informed consent was obtained. All participants in this study were recruited from a Lyme-endemic area during known seasonal tick activity months, and all EM cases were verified by a physician at the time of diagnosis. Use of the research clinical images from study participants was reviewed and approved by the Johns Hopkins Institutional Review Board and included written informed consent from the study participants. Participants did not receive a stipend for use of the photographs.

2.2. Deep convolutional neural networks

DL [23] has advanced due to a number of factors, including the development of large labeled datasets, the availability of markedly increased computational power via graphics processing units, and various algorithmic improvements. This study utilized DCNNs, which generate image feature representations at increased levels of abstraction via multiple layers of processing [23,25]. These features are learned

directly from data, as opposed to classical ML techniques that use human-engineered features. Here, a DCNN takes a skin image as input and processes the images through a cascade of operations to produce an output probability for each class. DCNNs perform linear operations (principally convolutions), followed by non-linear operations (e.g., rectified linear unit activations), to generate low-, mid-, and high-level feature representations of the input image. This processing has the effect of producing features that have meaning with increased levels of semantic abstraction at successive levels of the network. Features out of the last convolutional layer are then flattened into a single feature vector and further processed via fully connected layers whose effect is to implement additional feature refinement and classification logic, and finally output a probability value via SoftMax for each class label. The convolutional and fully connected layer weights are learned from the labeled training data via backpropagation using one of several optimization schemes (stochastic gradient descent, Adam, RSMProp, etc.). Therefore, this approach is thought of as being data-driven and end-to-end.

2.3. Experimental design

This study involved two experiments:

Comparison of Machine and Human Performance (E1). The goal of the first experiment was to: a) evaluate the machine's ability to operate as a pre-screener, and b) compare its performance against non-expert humans who would otherwise have performed this task and decided whether to seek care. Both machine and human performance were computed by comparing against the clinical criterion standard, and both were tested on the same test set of online images, with the exception that humans were trained under two differing protocols (A and B). Under protocol A, only EM images were used to train humans, in order to reenact a human suspecting EM performing an online search for guidance. Under protocol B, both EM and confuser (healthy skin, HZ, and TC) images were used to train humans, in this case to avoid confirmation bias that may occur under protocol A, and to allow for a balanced assessment of the lesion. Some humans were trained under protocol A only, some under B only, and some under both, in which case A was employed first, followed by testing on all images, then B was revealed and the human was asked to re-annotate all images from scratch. Humans could consult training material without any restriction prior to or during their classification task. In addition, humans were provided additional information on EM and Lyme disease from Wikipedia and CDC articles.

Assessing machine performance on joint online/clinical images (E2). E2 tested the performance of the machine against the same test set of online images as E1, but the data set was augmented with additional images from the clinical research study to test with images that show more subtle or variable presentations of EM. In E2 we report results both for the mixture test set (online and clinical), as well as the clinical-only test set. Some differences between the online and clinical images include the following: online images are 'in the wild' in the sense that view angle, illumination, and zoom factor are not as controlled. Additionally, they show a more diverse range of body parts and backgrounds when compared to the clinical images.

2.4. Data partitioning

Performance was computed against the clinician criterion standard. Partitioning of the online dataset involved subdividing images into training, validation, and testing subsets (70%, 10%, and 20%, respectively) for E1. For E2, the training/validation/testing datasets were identical to E1, but the additional clinical images were added to the test set in one case and used exclusively as the complete test set in a second case.

2.5. DCNN training

Our study uses the ResNet50 DCNN model [24]. ResNet was originally conceived as a means of producing deeper networks and including specific design patterns such as bottleneck and skip connections that make the output of a layer available to the next layer. Our implementation used the Keras and TensorFlow frameworks. We used transfer learning and fine-tuned the original ResNet50 weights (initially trained on the ImageNet dataset to classify 1000 different general object classes) for use on our skin classification problem. We used stochastic gradient descent with Nesterov momentum = 0.9 for training, with initial learning rate set to 1E-3. The training scheme used an early stopping approach, which terminated training after 10 epochs of no improvement of the validation set performance. We used a categorical cross entropy loss function. Dynamic learning rate scheduling was also used, in which we multiplied the learning rate by 0.5 when the training loss did not improve for 10 epochs. A batch size of 32 was used. Image preprocessing included rescaling and mean (ImageNet image) subtraction. Data augmentation for the images was also applied and consisted of horizontal flipping, blurring, sharpening, and changes to saturation, brightness, contrast, and color balance to further expand and generalize the dataset.

2.6. Metrics

Performance metrics used in this study included accuracy, F1, sensitivity, specificity, positive predictive value, negative predictive value, and Kappa score [33], which characterizes agreement with gold standard labels and discounts chance agreement. Since any classifier must balance a trade-off between sensitivity and specificity depending on the decision threshold, we used receiver operating characteristic (ROC) curves, which plot the detection probability (sensitivity) versus the false alarm rate (100% - specificity) for each algorithm/experiment. We then calculated the area under this curve (ROCAUC), which is considered a good metric for comparing classification methods.

3. Results

Image procurement. The initial online extraction produced a total of over 6000 images. After data curation, processing, and annotation, the final number of online images totaled 1718 images. In addition, 116 images from 63 unique, clinically validated research participants were used, resulting in a grand total of 1834 images. Table 1 shows the distribution of the final dataset across presentation types (see Fig. 1).

Experiment E1. Table 2 reports machine classification results. For the 4-class problem, accuracy was 84.00% and kappa was 0.7779 (the

Table 1
Data distribution for each skin type, and for experiments E1 and E2.

Data Set	Normal	Erythema Migrans (EM)	Herpes Zoster (HZ)	Tinea Corporis (TC)	Total
Training Set	151 (13.8%)	330 (30.2%)	410 (37.5%)	203 (18.6%)	1094
Validation Set	20 (16.1%)	43 (34.7%)	38 (30.6%)	23 (18.6%)	124
Test Set E1	108 (21.6%)	134 (26.8%)	190 (38.0%)	68 (13.6%)	500
Test Set E2	108 (17.5%)	250 (40.6%)	190 (30.8%)	68 (11.0%)	616

Fig. 1. Examples of erythema migrans with atypical (top) and classic bull's-eye (bottom) presentations (left: https://commons.wikimedia.org/wiki/Category:Erythema_migrans; right: JHU).

confusion matrix is shown in Fig. 2). For the 2-class problem, accuracy was 90.20%, Kappa was 0.7496, and specificity was 73.77% at 95% sensitivity. Overall results demonstrate promise as a pre-screener, with Kappa showing substantial agreement with criterion standard. ROC and human operating points in Fig. 3 show that machine performance almost always exceeded humans, with the exception of one participant.

Experiment E2. Table 3 reports results with accuracy of 86.53% and Kappa of 0.7143, which also demonstrate substantial agreement with the clinician-annotated criterion standard. We also report in the last row of Table 3 the results of using the clinical images exclusively as the test set. Since the clinical images contain only positive EM test examples, the sensitivity is the only metric that can be reported in this case (and is equal to the accuracy).

4. Discussion

This study of computer-assisted detection of EM via DL yielded promising results for early Lyme disease detection. We found that when utilizing a large dataset of images that included other non-EM lesions as comparators, this machine approach demonstrated substantial agreement with expert physician classification of lesions. Results in E1 and E2 yielded good performance across the 2- and 4-class problems. For the granular 4-class problem, machine misclassification of normal skin was highest with EM, and a modest misclassification of EM occurred with other confusers. As expected, performance for E2 was lower, since the machine was tested on fainter and less obvious clinical EM cases that it was never trained on. Also, it is noted that the sensitivity computed when using only clinical images for the test set in E2 is 71.55%, while the sensitivity computed when using only the online images for the test set is 81.34% in E1. This indicates that the online images are informative and representative of clinical images, but we anticipate better performance with the clinical data would be achieved given a sufficiently large dataset of clinical images to incorporate more subtle clinical cases in training. Still, the results showed good generalization, with kappa demonstrating substantial agreement with criterion standard across the board. Additionally, performance almost always improved over a group of non-physicians, suggesting a possible clinical application of prescreening skin lesions using photographs taken by individuals in non-medical settings.

Due to the lack of publicly available labeled datasets for EM machine-learning studies, the use of online images was necessary in order to obtain an adequate study set. As such, we followed the approach of a recent study investigating skin cancer detection using DCNNs from curated online images [26]. To our knowledge, only one prior study of computer-assisted detection of EM has been reported [19]. It used conventional machine learning methods including boosting, Support Vector Machines, naïve Bayes, and early neural nets (but not DCNNs) applied on human-engineered image features, and tested with a smaller dataset (143 images) [19]. Reported accuracies ranged from 69.23% to 80.42%, demonstrating the difficulties in addressing the detection of the varied presentations of the EM. By comparison, our study used a much larger dataset of complex images taken 'in the wild' and in the clinic and demonstrates enhancements in performance, for binary classification, ranging from 71.55% (E2, clinical only) to 86.53% (E2, mixture test set) and 90.20% (E1). For the more granular 4-class classification, performance ranges from 82.79% (E2) to 84.00% (E1). Additionally, to our knowledge, ours is the first study to test computer-assisted detection of other lesions, such as HZ and TC.

Table 2
Results for the machine classification performance for experiment E1 (95% confidence interval given in parenthesis). All values are % except ROCAUC and Kappa (which also characterizes the agreement with the gold standard).

Problem	Accuracy	Sensitivity	Specificity	Specificity (at 95% Sensitivity)	PPV	NPV	ROCAUC	Kappa
4-Class[a]	84.00 (3.21)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	0.7779
2-Class[b]	90.20 (2.61)	81.34 (6.60)	93.44 (2.54)	73.77 (4.51)	81.95 (6.54)	93.19 (2.58)	0.9539 (0.0184)	0.7496

E1 = Experiment 1, comparison of machine and human performance, PPV = Positive predictive value, NPV = Negative predictive value, ROCAUC = ROC area under the curve.

[a] Normal skin vs. erythema migrans vs. herpes zoster vs. tinea corporis.

[b] Erythema migrans vs. all other skin images (normal skin, herpes zoster, and tinea corporis) combined.

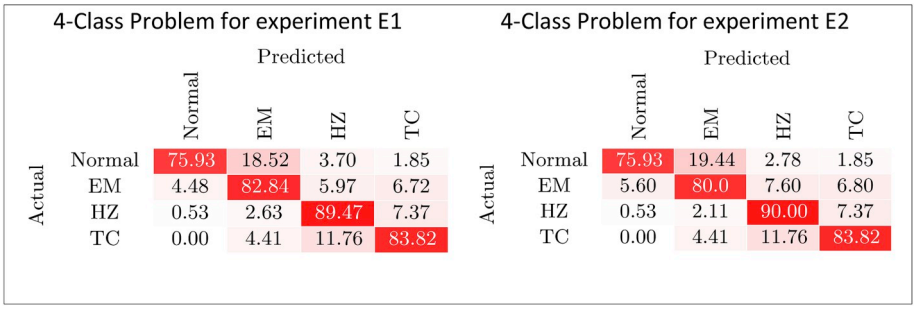

Fig. 2. Confusion matrices for machine classification performance results (%) for experiments E1 (left) and E2 (right) for the 4-class granular classification problem.

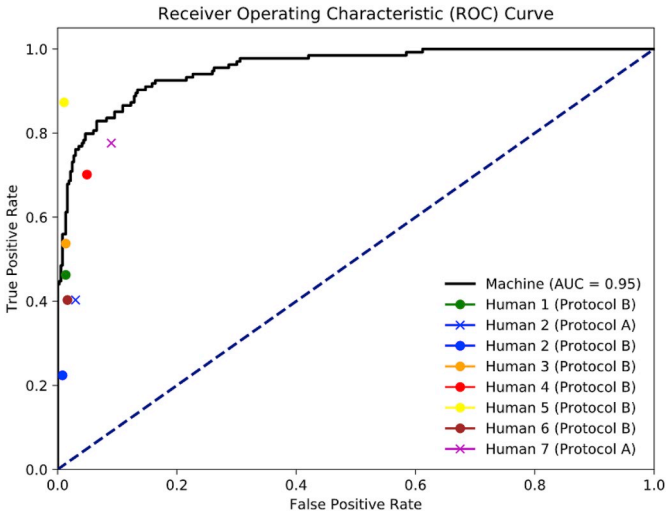

Fig. 3. ROC curve of 2-class problem (E1) with human comparison.

Studies have shown that neither physicians [16] nor the lay public [15] can identify EM with high accuracy, especially when this condition lacks the classic target appearance. However, suspicion of EM is the critical first step in prompting care-seeking behavior, and the lack thereof may lead to missed opportunities for early diagnosis at a time when Lyme disease is most easily treated and cured. Our study shows that computer-assisted diagnosis may have increased sensitivity over the average, non-medical observer. And while computer-assisted diagnosis may have a lower specificity, this is of less concern, as over-suspicion would only lead to physician evaluation.

We anticipate that computer-based approaches will lead to faster and more accurate lesion identification, and a decrease in incident misdiagnosis of early Lyme disease. Now that much of the general population has access to high-quality photography via cell phones, it has become increasingly common for patients to photograph their skin lesions. This presents the opportunity for future development of mobile apps to pre-screen these “in the wild” photos and alert users to suspect Lyme disease. This has the potential to lead patients to care they might not otherwise seek, which would be expected to decrease early missed diagnoses.

We now discuss limitations of this study. One is that, during curation, we parse out inappropriate or irrelevant images, excluding images where classification is uncertain or of low probability so as to decrease incorrect annotations that would affect both training and testing. However, doing so may also make the dataset easier to classify. Another notable limitation is the underrepresentation of darker-skinned individuals in our data set, despite inclusion of specific search terms geared towards incorporating a diversity of online images. Although EM is often found to be less frequently diagnosed among people of color due to a variety of factors [34], it is important that machine learning algorithms are trained on racially diverse images [35]. Additionally, variability in viewpoint, lighting, and resolution of the online images, as well as an inability to observe the lesions more closely, precisely estimate lesion size, and review corroborating clinical and laboratory patient data were all contributing factors which made clinical annotation more challenging than with a controlled set of images. However, this did have the benefit of rendering our final data set more representative of photos that may be taken by the lay public with their cell phones. Furthermore, the current diagnosis of EM and HZ rely primarily on visual inspection of the lesion and clinical suspicion. Given the unreliability of serologic testing in the early phase of infection with *B. burgdorferi* and the impractical nature of culture identification, this remains the gold standard. Finally, as previously noted, a significant minority of patients with early Lyme disease do not present with an EM rash. Although our approach may lead to earlier and more accurate diagnosis for those that do, additional approaches are also needed to address this subset of patients. Future work will entail expanding the binary and granular skin lesion classification to other types of skin pathologies that are clinically important to distinguish from EM such as cellulitis.

5. Conclusion

We applied DL for EM classification and demonstrated significant promise for deployment of automated methods for prescreening patients prior to physician referral. Our results suggest that the machine is likely more sensitive than patient self-assessment and has the potential

Table 3
Results for the machine classification performance for experiment E2 (95% confidence interval given in parenthesis). All values are % except ROCAUC and Kappa. For the last row which tests on the clinical positive EM examples only, the sensitivity is the only metric that can be reported (and is equal to the accuracy).

Problem	Accuracy	Sensitivity	Specificity	Specificity (at 95% Sensitivity)	PPV	NPV	ROCAUC	Kappa
4-Class[a]	82.79 (2.98)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	0.7561
2-Class[b]	86.53 (2.70)	76.40 (5.26)	93.44 (2.54)	75.96 (4.38)	88.84 (4.21)	85.29 (3.47)	0.9510 (0.0171)	0.7143
Clinical positive examples only	71.55 (8.21)	71.55 (8.21)	N/A	N/A	N/A	N/A	N/A	N/A

E2 = Experiment 2, assessing matching performance on joint online/clinical image set, PPV = Positive predictive value, NPV = Negative predictive value, ROCAUC = ROC area under the curve.

[a] Normal skin vs. erythema migrans vs. herpes zoster vs. tinea corporis.

[b] Erythema migrans vs. all other skin images (normal skin, herpes zoster, and tinea corporis) combined.

to be more accurate than diagnosis by a general non-specialist physician, who would ordinarily serve as the screening gatekeeper for acute onset rashes such as EM. Given the frequent underdiagnosis of EM, the use of automated detection would be beneficial by increasing the number of patients who seek further medical assessment for EM rashes and minimizing the number of cases that go unevaluated and undiagnosed. This could help prevent the otherwise serious long-term complications associated with late-stage Lyme disease.

Support

This work was supported by JHU APL internal research and development funds as well as JHU School of Medicine philanthropy funding.

Acknowledgements

We would like to thank Erica Mihm for assistance with annotation of normal skin images.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2018.12.007.

References

[1] B.M. Kuehn, CDC estimates 300,000 US cases of Lyme disease annually, J. Am. Med. Assoc. 310 (11) (2013) 1110.
[2] A.F. Hinckley, N.P. Connally, J.I. Meek, et al., Lyme disease testing by large commercial laboratories in the United States, Clin. Infect. Dis. 59 (5) (2014) 676–681.
[3] G. Stanek, G.P. Wormser, J. Gray, F. Strle, Lyme borreliosis, Lancet 379 (9814) (2012) 461–473.
[4] R.B. Nadelman, Erythema migrans, Infect. Dis. Clin. 29 (2) (2015) 211–239.
[5] A.C. Steere, V.K. Sikand, The presenting manifestations of Lyme disease and the outcomes of treatment, N. Engl. J. Med. 348 (24) (2003) 2472–2474.
[6] G.P. Wormser, R.J. Dattwyler, E.D. Shapiro, et al., The clinical assessment, treatment, and prevention of Lyme disease, human granulocytic anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious Diseases Society of America, Clin. Infect. Dis. 43 (9) (2006) 1089–1134.
[7] Centers for Disease Control and Prevention, Lyme Disease (Borrelia Burgdorferi) 2017 Case Definition, Published https://wwwn.cdc.gov/nndss/conditions/lyme-disease/case-definition/2017/, (2017) , Accessed date: 23 April 2018.
[8] M.E. Schriefer, Lyme disease diagnosis: serology, Clin. Lab. Med. 35 (4) (2015) 797–814.
[9] E.D. Shapiro, Clinical practice. Lyme disease, N. Engl. J. Med. 370 (18) (2014) 1724–1731.
[10] C. Bhate, R.A. Schwartz, Lyme disease: Part I. Advances and perspectives, J. Am. Acad. Dermatol. 64 (4) (2011) 619–636 quiz 637-618.
[11] C.D. Tibbles, J.A. Edlow, Does this patient have erythema migrans? J. Am. Med. Assoc. 297 (23) (2007) 2617–2627.
[12] R.P. Smith, R.T. Schoen, D.W. Rahn, et al., Clinical characteristics and treatment outcome of early Lyme disease in patients with microbiologically confirmed erythema migrans, Ann. Intern. Med. 136 (6) (2002) 421–428.
[13] D.R. Mazori, C.M. Orme, A. Mir, S.A. Meehan, A.L. Neimann, Vesicular erythema migrans: an atypical and easily misdiagnosed form of Lyme disease, Dermatol. Online J. 21 (8) (2015).
[14] R.R. Mullegger, M. Glatz, Skin manifestations of lyme borreliosis: diagnosis and management, Am. J. Clin. Dermatol. 9 (6) (2008) 355–368.
[15] J.N. Aucott, L.A. Crowder, V. Yedlin, K.B. Kortte, Bull's-Eye and nontarget skin lesions of lyme disease: an internet survey of identification of erythema migrans, Dermatol Res Pract (2012) 451727.
[16] D. Lipsker, A. Lieber-Mbomeyo, G. Hedelin, How accurate is a clinical diagnosis of erythema chronicum migrans? Prospective study comparing the diagnostic accuracy of general practitioners and dermatologists in an area where lyme borreliosis is endemic, Arch. Dermatol. 140 (5) (2004) 620–621.
[17] A.C. Steere, F. Strle, G.P. Wormser, et al., Lyme borreliosis, Nat Rev Dis Primers 2 (2016) 16090.
[18] Y. Fujisawa, Y. Otomo, Y. Ogata, et al., Deep learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumor diagnosis, Br. J. Dermatol. (2018).
[19] E. Čuk, M. Gams, M. Možek, F. Strle, V.M. Čarman, J.T. Tasič, Supervised visual system for recognition of Erythema Migrans, an early skin manifestation of Lyme Borreliosis, Strojniški vestnik - Journal of Mechanical Engineering 60 (2) (2014) 115–123.
[20] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, International Conference on Learning Representations, 2015 (San Diego).
[21] C. Szegedy, W. Liu, Y. Jia, et al., Going deeper with convolutions, IEEE Conference on Computer Vision and Pattern Recognition, 2015 (Boston, MA, USA).
[22] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 1, 2012 (Lake Tahoe, Nevada).
[23] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[24] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Paper Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27-30 June 2016, 2016.
[25] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning vol. 1, MIT Press, Cambridge, MA, 2016.
[26] A. Esteva, B. Kuprel, R.A. Novoa, et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118.
[27] P.M. Burlina, N. Joshi, M. Pekala, K.D. Pacheco, D.E. Freund, N.M. Bressler, Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks, JAMA Ophthalmol 135 (11) (2017) 1170–1176.
[28] P. Burlina, K.D. Pacheco, N. Joshi, D.E. Freund, N.M. Bressler, Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis, Comput. Biol. Med. 82 (2017) 80–86.
[29] P. Burlina, N. Joshi, K.D. Pacheco, D.E. Freund, J. Kong, N.M. Bressler, Utility of deep learning methods for referability classification of age-related macular degeneration, JAMA Ophthalmol 136 (11) (2018) 1305–1307.
[30] P.M. Burlina, N.J. Joshi, K.D. Pacheco, D.E. Freund, J. Kong, N.M. Bressler, Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration, JAMA Ophthalmol 136 (12) (2018) 1359–1366.
[31] S. Kankanahalli, P.M. Burlina, Y. Wolfson, D.E. Freund, N.M. Bressler, Automated classification of severity of age-related macular degeneration from fundus photographs, Invest. Ophthalmol. Vis. Sci. 54 (3) (2013) 1789–1796.
[32] P. Burlina, D.E. Freund, B. Dupas, N. Bressler, Automatic screening of age-related macular degeneration and retinal abnormalities, Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, 2011, pp. 3962–3966.
[33] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1) (1977) 159–174.
[34] A.D. Fix, C.A. Pena, G.T. Strickland, Racial differences in reported Lyme disease incidence, Am. J. Epidemiol. 152 (8) (2000) 756–759.
[35] A.S. Adamson, A. Smith, Machine learning and health care disparities in dermatology, JAMA Dermatology 154 (11) (2018) 1247–1248.