Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Doctoral Consortium Track

Early Diagnosis of Lyme Disease by Recognizing Erythema Migrans Skin Lesion from Images Utilizing Deep Learning Techniques

Sk Imran Hossain
Université Clermont Auvergne, CNRS, ENSMSE, LIMOS, F-63000 Clermont-Ferrand, France
sk_imran.hossain@uca.fr

Abstract

Lyme disease is one of the most common infectious vector-borne diseases in the world. We extensively studied the effectiveness of convolutional neural networks for identifying Lyme disease from images. Our research contribution includes dealing with lack of data, multimodal learning incorporating expert opinion elicitation, and automation of skin hair mask generation.

1 Research Problem

Lyme disease is an infectious tick-transmitted disease that manifests itself in most cases with erythema migrans (EM) skin lesions in the early stage. Better diagnosis of these early forms would allow improving the prognosis by preventing the transition to a severe late form thanks to appropriate antibiotic therapy. The unavailability of well-labeled public EM datasets because of privacy concerns of medical data may be the reason for the lack of extensive studies in this field. Existing works on machine learning based Lyme disease prediction only utilize images of EM lesions whereas doctors believe that corresponding patient metadata should be incorporated in the analysis [Burlina *et al.*, 2020; Čuk *et al.*, 2014]. Even though images of EM can be collected from the internet and hospitals, it will be time-consuming to collect corresponding patient metadata associated with each image to jointly train a multimodal deep learning model. Skin hair over lesions can be a challenge for real-life application as we do not have sufficient training images with skin hair to cover many real-life skin hair variations. Standard image processing based hair removal is not beneficial for real-time application and removing hair does not give new features to the network. Augmenting images with skin hair can be of interest. So, we are dealing with three research questions: Q1: How to deal with the lack of a well-labeled publicly available dataset? Q2: How to utilize patient metadata in the absence of enough training data? Q3: How to efficiently deal with skin lesion hair?

2 Research Contribution and Plan

Our research contribution and plan for the identified research questions are described in the following subsections.

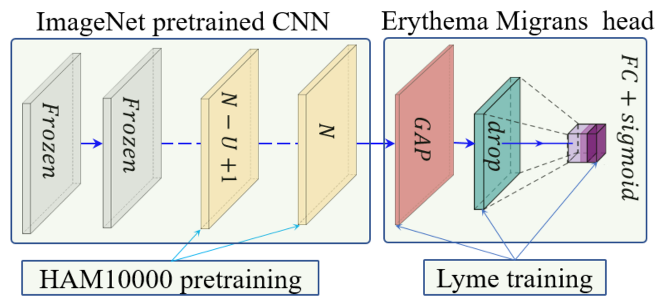

Figure 1: Transfer learning workflow. GAP, drop, and FC stand for Global Average Polling, dropout, and Fully Connected layers respectively. N is the number of ImageNet pre-trained layers and U represents the number of layers used for fine-tuning.

2.1 Dealing with the Lack of Publicly Available Dataset

As there is no publicly available EM image dataset for Lyme disease predictions, we utilized an EM dataset consisting of images collected from Clermont-Ferrand University Hospital Center (CF-CHU) of France and the internet. CF-CHU collected the images from several hospitals in France. This dataset was labeled by expert dermatologists and infectiologists from CF-CHU. First, we benchmarked this dataset for twenty-three well-known CNN architectures in terms of predictive performance metrics, computational complexity metrics, and statistical significance tests. Second, to improve the performance of the CNNs, we used transfer learning from ImageNet pre-trained models as well as pre-trained the CNNs with the skin lesion dataset "Human Against Machine with 10000 training images (HAM1000)" [Tschandl *et al.*, 2018] as shown in Figure 1. Our study confirmed the effectiveness and potential of even some lightweight CNNs to be used for Lyme disease prescanner mobile applications. We also made all the trained models and part of the dataset publicly available at https://dappem.limos.fr/download.html. The study has been published at [Hossain *et al.*, 2022]. Now, considering the performance and activation map visualization of the different architectures we are planning to create a custom lightweight architecture for our problem utilizing a residual building block consisting of dilated, depthwise separable convolutions, and attention mechanisms.

2.2 Utilizing Patient Metadata in the Absence of Training Data

As we do not have corresponding patient metadata for each of the images of our dataset so initially, we are working on incorporating expert opinion elicitation. The questionnaires and weight attribution were prepared by collecting data from 16 expert dermatologists. As it is difficult for the doctors to provide estimates for the parameters of a probability distribution, they assigned weight values to different symptoms in the range of -1 to +3 (a higher value represents a higher contribution of the symptom towards the possibility of the disease). We summarized each case as the min-max normalized average weight sum. Based on all possible cases we proposed three approaches to the experts for converting the weight sum to a probability score. The approaches are 1. cumulative prob-ability of a normalized weight sum based on the kernel density estimate, 2. cumulative probability of a normalized weight sum based on the density estimate from a Gaussian mixture with two components, and 3. posterior probability of a normalized weight sum belonging to the second component of Gaussian mixture (assuming the second component represents the ill subpopulation) as shown in Figure 2. The experts opted for the second approach. We are still working on finalizing the elicitation process.

2.3 Efficiently Dealing with Skin Lesion Hair

Skin hair augmentation techniques require a hair mask to generate hair in given locations. These masks are created either manually, with random curves or lines and segmentation. Generative Adversarial Network (GAN) can be utilized to automate the creation of hair masks. We are planning to create a GAN to automate the skin hair mask generation process. For that purpose, we are creating a hand-curated skin lesion hair segmentation dataset with a custom segmentation model which can be utilized for preparing the skin lesion hair mask dataset for the GAN training as shown in Figure 3.

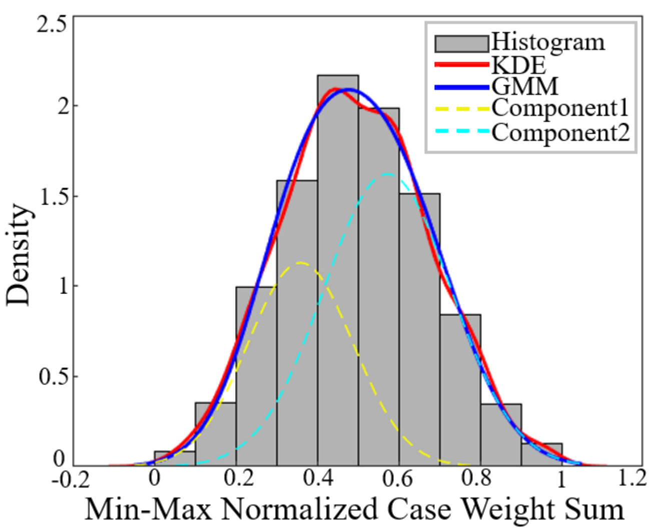

Figure 2: Expert opinion elicitation process. KDE and GMM stand for Kernel Density Estimation and Gaussian Mixture Models respectively.

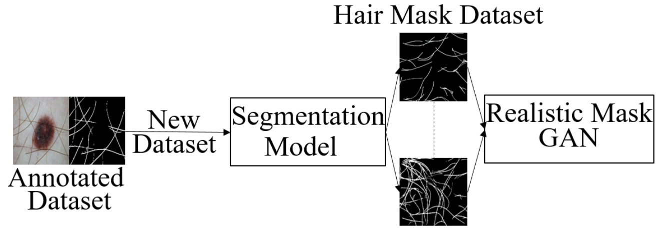

Figure 3: Realistic skin hair mask generation. GAN stands for Generative Adversarial Netowrk.

Acknowledgements

This research is funded by the European Regional Development Fund, project DAPPEM –AV0021029. This work is su-pervised and co-supervised by Prof. Engelbert MEPHU NGUIFO (https://perso.isima.fr/ enmephun/) and Jocelyn de Goër de Herve (https://jgoer.epia.clermont.inrae.fr/) respectively.

References

[Burlina *et al.*, 2020] Philippe Burlina, Neil Joshi, Philip Mathew, William Paul, Alison Rebman, and John Aucott. AI-based detection of erythema migrans and disambiguation against other skin lesions. *Computers in Biology and Medicine*, 125:103977, oct 2020.

[Čuk *et al.*, 2014] Erik Čuk, Matjaž Gams, Matej Možek, Franc Strle, Vera Maraspin Čarman, and Jurij Tasič. Supervised visual system for recognition of erythema migrans, an early skin manifestation of lyme borreliosis. *Strojniski Vestnik - Journal of Mechanical Engineering*, 60(2):115–123, feb 2014.

[Hossain *et al.*, 2022] Sk Imran Hossain, Jocelyn de Goër de Herve, Md Shahriar Hassan, Delphine Martineau, Evelina Petrosyan, Violaine Corbin, Jean Beytout, Isabelle Lebert, Jonas Durand, Irene Carravieri, Annick Brun-Jacob, Pascale Frey-Klett, Elisabeth Baux, Céline Cazorla, Carole Eldin, Yves Hansmann, Solene Patrat-Delon, Thierry Prazuck, Alice Raffetin, Pierre Tattevin, Gwenaël Vourc'h, Olivier Lesens, and Engelbert Mephu Nguifo. Exploring convolutional neural networks with transfer learning for diagnosing Lyme disease from skin lesion images. *Computer Methods and Programs in Biomedicine*, 215:106624, mar 2022.

[Tschandl *et al.*, 2018] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5(1):1–9, aug 2018.