Lyme detection with Deep Learning

Soham Roy

April 3, 2024

Introduction

- **Task:** Detecting early-stages of Lyme disease based on images using supervised-learning model with convolutional neural networks.
- **Introduction:**

  1. Disease cause by a bacteria called *Borrelia*
  2. Transmitted to humans through the bite of infected ticks.
  3. Difficult to diagnose rapidly.
  4. Spreads to joints, heart and nervous system.
  5. Can take up to several months or even years to have its strongest symptoms.

- **Motivation:**
  1. Early detection helps preventing people getting health complications.
  2. Limited public image-based research on this task.
  3. Traditional clinician accuracy was found to be approximately 80%.
  4. Current approaches make use of data collected from private sources.

Figure: Erythema Migrans rash. Image: CDC EM, 2022.

Related work and Challenges

Challenges:

1. Not enough publicly available datasets for Lyme disease.
2. Difficult to achieve a good performance. Complicate to avoid overfitting.

Related work:

- *Čuk et al., 2014*
- *Burlina et al., 2018*
- *Koduru and Zhang, 2021*
- *Hossain et al., 2022*

Proposed Method (I)

Data

- Perform transfer learning with *public* datasets
- Fine-tuning phase using *public* data

1. **Fine-tuning:**
   - Lyme Disease Rashes (CDC EM, 2022)
2. **Transfer Learning:**
   - Dermnet (Dermnet, 2020)
   - HAM 10000 (Skin Cancer MNIST: HAM10000, 2018)
   - **Problems encountered:** The dataset IEEDataPort, 2020. Publicly available, *at a cost!*

Proposed Method (II)

Models:

- **ResNet50**: Best performance in previous works.
- **MobileNetV2**: Low computational, fair results for number of parameters.

Augmentation:

- **Augmentor:** (Bloice, Roth, and Holzinger, 2019) Relevant augmentations to biomedical imaging.
- **Torchvision:** PyTorch native. 30 types image transformations. **Normalization. PIL image to Tensor.**

Proposed Method (III)

- **Siamese networks**

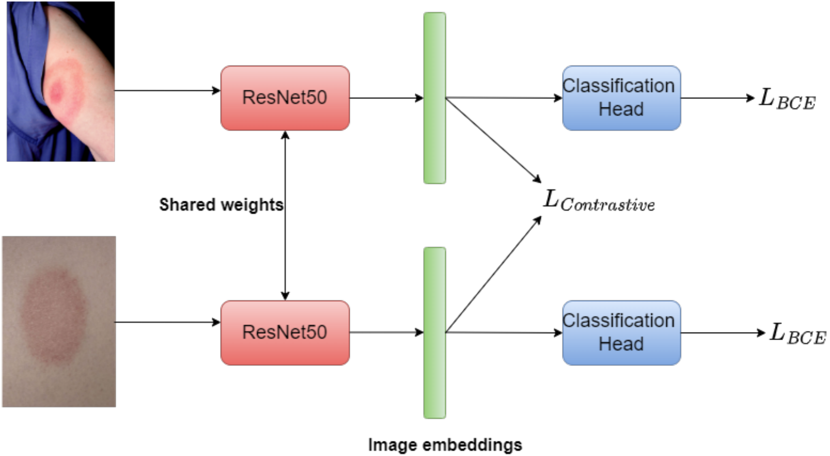

Figure: Proposed Siamese Network architecture.

Results

- **Accuracy metric**

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Reported results:

Models	Accuracy
Čuk et al., 2014	69.23 to 80.42
Burlina et al., 2018	76.05±0.74
Hossain et al., 2022	84.42 ±1.36
Koduru and Zhang, 2021	91.95 (validation)

Table: Reported results of previous work. **Note:** some models are private and their performance cannot be tested.

Results

Models	Pre-trained Dataset	Val. Accuracy	Test Accuracy
ResNet50	ImageNet	91.66	66.66
ResNet50	HAM1000	**93.05**	62.07
ResNet50	Dermnet	91.6	**67.35**
MobileNetV2	ImageNet	87.5	81.6
MobileNetV2	HAM1000	**90.27**	75.86
MobileNetV2	Dermnet	86.1	**85.0**
ResNet50_SN	ImageNet	100	86.20
ResNet50_SN	HAM1000	**100**	88.5
ResNet50_SN	Dermnet	99.3	**89.6**
MobileNetV2_SN	ImageNet	99.65	85.05
MobileNetV2_SN	HAM1000	**100**	**89.65**
MobileNetV2_SN	Dermnet	98.9	87.3

Table: Obtained results on Lyme dataset CDC EM, 2022. **Note:** SN models correspond Siamese Network implementations

Analyses I

- **Trouble:** very different dataset with very little data
- Transfer learning on skin data has a significant impact on performance.
- Simple transfer learning is not enough. High training, but low test. Overfitting.
- ResNet50 and MobileNetV2 have *similar* results.
- In our setting, ResNet50(non SN) is slightly worse, due to *overfitting*. More parameters than MobileNetV2(non SN) for really small dataset.

Analyses II

- **Solution:** Siamese Networks. Perera and Patel, 2018
- Blend of Contrastive loss and BCE loss.
- **Constractive loss:** Minimize/maximize variance between features of the same/different class.
- **BCE loss:** Minimize missclassification loss.
- It prevents overfitting. Higher validation and test than training accuracy.
- Our proposed network achieves best performance regardless of the type of the network used for pretraining.

Analyses III

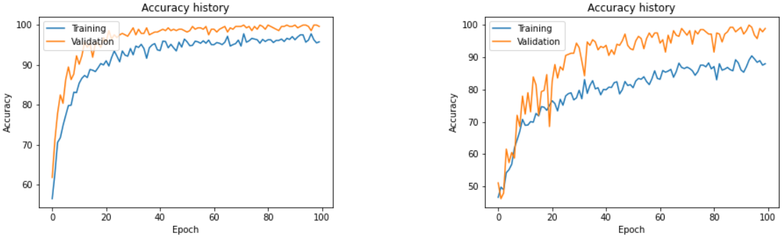

Figure: Accuracy MobileNetV2 vs ResNet50SN.

- MobileNetV2 has a faster convergence than RestNet50_SN due to small amount of parameters.
- Thus, MobileNetV2 is more efficient and has better generalization capability. **Best** choice for our problem.

Conclusions

- Lyme dataset is small, not enough for plain optimization.
- Adequate data augmentation for biomedical images improves performance.
- Choice of transfer learning dataset has a crucial impact on performance.
- **Siamese network** performance outperforms previous work.
- With our approaches, good accuracy is achieved using only public datasets.
- Best model: MobileNetV2 pre-trained on Skin Cancer MNIST: HAM10000, 2018.

Bloice, Marcus D, Peter M Roth, and Andreas Holzinger (Apr. 2019). "Biomedical image augmentation using Augmentor". In: *Bioinformatics* 35.21, pp. 4522–4524. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btz259. eprint: https://academic.oup.com/bioinformatics/article-pdf/35/21/4522/30330763/btz259.pdf. URL: https://doi.org/10.1093/bioinformatics/btz259.

Burlina, P. et al. (2018). "Skin Image Analysis for Erythema Migrans Detection and Automated Lyme Disease Referral". In: *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Ed. by Danail Stoyanov et al. Cham: Springer International Publishing, pp. 244–251. ISBN: 978-3-030-01201-4.

CDC EM (2022). https://www.kaggle.com/datasets/sshikamaru/lyme-disease-rashes.

Dermnet, Kaggle (2020). https://www.kaggle.com/datasets/shubhamgoel27/dermnet.

Hossain, Sk Imran et al. (2022). "Exploring convolutional neural networks with transfer learning for diagnosing Lyme disease from skin lesion images". In: *Computer Methods and Programs in Biomedicine* 215, p. 106624. ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2022.106624.

IEEDataPort (2020). https://ieee-dataport.org/documents/image-dataset-various-skin-conditions-and-rashes#files.

Koduru, Tejaswi and Edward Zhang (2021). "Using Deep Learning in Lyme Disease Diagnosis". In: *Journal of Student Research* 10.4. DOI: 10.47611/jsrhs.v10i4.2389. URL: https://jsr.org/hs/index.php/path/article/view/2389.

Perera, Pramuditha and Vishal M. Patel (2018). "Learning Deep Features for One-Class Classification". In: *CoRR* abs/1801.05365. arXiv: 1801.05365. URL: http://arxiv.org/abs/1801.05365.

Skin Cancer MNIST: HAM10000, Kaggle (2018). https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000.

Čuk, Erik et al. (2014). "Supervised Visual System for Recognition of Erythema Migrans, an Early Skin Manifestation of Lyme Borreliosis". In: *Strojniški vestnik - Journal of Mechanical Engineering* 60.2, pp. 115–123. ISSN: 0039-2480. DOI: 10.5545/sv-jme.2013.1046. URL: https://www.sv-jme.eu/article/supervised-visual-system-for-recognition-of-erythema-migrans-an-early-skin-manifestation-of-lyme-borreliosis/.