MSE Thesis

Task for:

Alexander Peter (Master student)

Data Science (Specialization)

from   Prof. Dr. Marc Pouly (Advisor)
Fabian Gröger (2nd supervisor)
Tobias Mérinat (Expert)

1. Title of Work

Investigating the Impact of Plant Disease Images on Dermatology Diagnostics

2. Externally-financed Research/Development Project

Strategic Collaboration with University Hospital Basel / Medtech Joint Venture

3. Industry/Business Partner

University Hospital Basel

4. Specialization Focus

- *Computer Vision*
- *Deep Learning*
- *Self-Supervised Learning*
- *Medical Imaging*

5. Contents

- Supervised learning for medical image analysis requires a large volume of annotated data. This concerns tasks related to the classification, localization, and segmentation of diseases as well as the scoring of severity degrees. However, supervised learning typically requires vast amounts of annotated data to achieve the high performance and robustness required in the medical field. In this domain, annotated samples can be hard to obtain. For example, the acquisition process must consider strict regulations and the fact that collecting standardized images generally requires special equipment. Additionally, there is a strong geographic influence on the data collection. For example, patients in Swiss hospitals predominantly have light skin pigmentation and insect bites are rarely seen, while they rank among the most prevalent skin diseases in African countries.

- Many approaches exist that aim to use transfer learning in order to improve predictive power of these models and counteract data scarcity in the medical domain. Popular approaches include transferring features learned from the general image domain, such as

ImageNet, or features learned from the target domain with techniques such as Self-Supervised Learning (SSL). Both approaches have drawbacks. For example, ImageNet pretraining does not always lead to improved performance, and SSL pretraining requires a large corpus of domain-specific images.

- Transfer learning has been shown effective when application domains are related to each other. Thus one possible improvement of the scarcity situation in dermatology consists in finding a similar domain with access to large collections of public images.

- This project aims to investigate if the domain of plant diseases can be used as a means to target data scarcity in digital dermatology. The community around plant diseases is very active and semantically similar to dermatology, as plants can have leaves with different pigmentation, and diseases are visible from the surface. Furthermore, large collections of datasets are available along with already trained foundation models, making it an optimal candidate for further investigation. This leads to the following research questions:

  1. How does transfer learning from plant diseases influence the predictive performance of classification and segmentation models in digital dermatology?

  2. What is the most effective transfer strategy?

  3. Does this transfer strategy lead to a reduction of annotated samples?

  4. Are there simple strategies to compare samples between domains, i.e., given a skin image, what are the most informative or influentiial plant disease images from the training set?

Deliverables

- Project documentation with no more than 100 pages (all inclusive)
- Dataset and source code repository

6. Execution of Work

Start of Work:	04.03.2024
Submission of project report:	until 18.10.2024 at 5pm directly to the advisor and expert
Closing Presentation:	tbd, max. 2 weeks after hand in.
Grade Notification	immediately after presentation

→ Addtional dates according to the separate schedule

7. Documentation

2 copies of the project report are to be prepared (for the advisor and expert). The report also contains the mandatory inclusions:

- The following self-declaration:

  „I hereby declare that I have prepared the present work independently and have used nothing other than the specified aids. All text sectons, citations or contents of other authors used have been explicitly marked as such. Horw, date, personal signature"

– The submission of the complete electronic data (reports, presentation, measurement data, programs, evaluations, etc.).

MSE work carried out within the framework of a research project should be included in the PPDB. For the MSE thesis, the category University Writing (Hochschulschrift) is to be used; specialization project reports are categorized as reports / working papers. The publication should be linked to the corresponding research project. The "CC BY-NC-ND" is recommended as the licence (see ZHB > Research > E-Media > Open Access > OA Publications Lucerne > Master theses). The PDF is to be added via the PPDB and included in the LORY open-access repository (via button in the PPDB; will be released by the ZHB after a few days). The MSE thesis must be archived. If this is not possible via the PPDB, a print version must be deposited in the library.

8. Additional Comments

Rotkreuz, 26. Februar 2024

Advisor	Expert	Student

The original task description for the specialization project with all signatures has to be submitted to the MSE-secretariat **no later than the end of the first week of the semester via mse.informatik@hslu.ch**. Possible amendments can be passed on electronically.