

Using Image Recognition based on Machine Learning to help identify Skin Disorders

By Tudor Neacsu

Abstract

The aim of this project is to develop a tool which will be able to use images from large galleries contained in dermatological databases in order to provide the user with an estimated diagnosis of a specific skin condition. The main use case for this program would be that the user submits a set of pictures of their skin abnormalities, together with their location on their body. To this, the program replies with an ordered list of the most possible conditions that the user is experiencing/suffering from.

The program will use image recognition and machine learning algorithms to identify and learn the characteristics specific to each disease used to train the tool. Any extra information that the user provided will be used to narrow down the search process. Furthermore, I intend to write the tool in such a manner that allows an easy modification of any algorithms used, together with the easy option to expand upon the initial parameters of the software (i.e. the administrator of the application should be able to modify the algorithms if necessary and expand upon the types of the data used, for example during the introduction of a new skin condition found on a part of the body unspecified so far in the software).

To carry this out, I intend to use several image recognition algorithms and techniques such as edge detection and boundary finding and pattern matching and shape detection. The machine learning algorithm/algorithms used would be of a supervised classification nature. Furthermore, I intend to limit my data set to a select few skin conditions.

Finally, this tool will be polished with a GUI to permit easy usage. I am undecided, but it may be a simple web interface.

Experiments

Useful experiments which I believe would be beneficial to perform would include different approaches to extracting information from the images such that they're identifiable and analysable by the program. One specific early experiment will involve the demonstration of my capacity to write and understand machine basic machine learning algorithms. I intend to begin with a supervised binary classification algorithm on two dimensional graphs.

Timeline

This is the timeline I will attempt to abide by. I will re-evaluate it if necessary during the course of this project.

1. Understanding the data (03/10/2016)
 - Selecting a representative set of data
 - Identify a set of rules that relate to the existing data
 - Report upon a resilient database model to use which allow expansion (Consider Non – SQL databases) and a concrete data set.
 - Bonus: Start experimenting with image recognition
2. Algorithms and development – Experimentation (17/10/2016 – 18/11/2016)
 - Identify suitable solutions to extract data from images
 - Experiment with ways of deriving information from the images and decide upon the most useful relevant one(s)
 - Experiment with machine learning algorithms
 - Understand the key values derived from the data
 - Define the first model for the data to be structured in
 - Report with a clear definition of the database model used, together with the most successful algorithms

|| Extra time in case I fall behind ||

3. Beginning to write software (5/12/2016)
 - Define clear set of Software Engineering techniques to aid the creation of a reusable application
 - Attempt to implement found results from experimentation into the actual program
 - Report upon the first version of the tool's release.
4. Design and optimisation (06/02/2017)
 - Decide upon an interface and conclude with a GUI (Consider Django)
 - Expand the data set
 - Further Optimise algorithms

Risk Assessment

Here I will list the possible troubles I will face during this ongoing project:

- Poor time management leading me to greatly fall behind on the project
 - This can lead to an unfinished project.
 - Solution: define a time in each day where I spend time on the project; Reassess progress weekly.
- Loss of focus on stage of the timeline.
 - This can lead me to spiral off into working on sections of the program which aren't relevant or important at that moment in time.
 - Solution: constantly reflect on current task and required task by consulting with the timetable
- Difficulty to fully understand the mathematical concepts used.
 - This can have serious consequences if the algorithms do not work as expected or need a more in-depth modification
 - Solution: define a period of time during a day per week to progress mathematical understanding
- Failure to write the code for the mathematics itself.
 - Solution: Have back-up APIs to use instead
- Loss of documentation
 - This can either come in the form of poor repository commits or a poor diary being kept.
 - Solution: Dedicate at least one hour at the start and middle of the week to complete diary and document all my actions; Ensuring to use good Software Engineering practices.
- This project could also prove to be more than I can realistically handle/am capable of.
 - Solution: Trimming down the project such that the most relevant goals are achieved. This has to be determined at most by the middle of November.

Bibliography

Software Engineering for Image Processing Systems By Philip A. Laplante

Computer Vision Algorithms in Image Algebra By Gerhard X. Ritter and Joseph N. Wilson

Problog: A probabilistic prolog and its application in link discovery By L. De Raedt, A. Kimmig, and H. Toivonen

Bayesian networks in biomedicine and health-care. Artificial Intelligence in medicine By P.J.F. Lucas, L.C. van der Gaag, and A. Abu-Hanna

URLs:

<http://opencv.org/>

<https://www.mongodb.com/>

<https://www.djangoproject.com/>

<https://dtai.cs.kuleuven.be/problog/>

Image data will be first taken from: http://www.dermweb.com/photo_atlas/default.htm

Existing similar project already in development: <http://www.lubax.com/>

My git repository: <https://github.com/VLaDOS/ImageDer.git>