Problem Statement for Rheumatoid Arthritis Mobile App

Chase Denecke

CS 461 Fall 2018

I. Abstract

Rheumatoid Arthritis (RA), is a common disease that has serious implications for the quality of life for those who suffer from it. There is no cure for RA, but there are various treatments for the disease. Effective administration of these treatments depend to a large degree on being able to measure disease activity. There are various tools for doing this, but they are all expensive an often require a trip to the doctors office.

Our team plans to make an application that can track disease activity using a smartphone camera. If we are able to get this application working, it could be either a supplementary test that could be used in conjunction with existing measurements of disease activity, or it could replace those tests with a much cheaper, more convenient alternative.

To create this application, we will need a large set of labelled images (at least 20,000), a knowledge of how to create, train, and evaluate a convolutional neural network, and a means of measuring the accuracy of the network. This last point will be accomplished by means of cross validation.

II. Definition and Description of Problem

Rheumatoid Arthritis (RA), is a disease that affects somewhere between 200,000 and 1.3 million people in the United States. It is an autoimmune disease characterized by swelling and pain in joints. The immune systems of people with RA attack their joints, causing thickening of the synovium (the tissue on the inside of the joints). Over a longer period of time, RA causes severe loss of joint mobility by destroying connective tissue and causing fusion of bones at the joint. It is very painful. There is no cure for Rheumatoid Arthritis.

One of the biggest barriers to effective treatment of RA is the lack of a cheap test capable of classifying disease activity. The immune system of people with the condition goes into a highly active state called a "flare", during which joints often swell to several times their normal size. The pain and loss of mobility experienced during these times is debilitating. It is still not well understood what causes flares. There are long lists of "risk factors" that have shown positive correlation with flare activity, but the lack of a cheap and accurate means of measuring disease activity makes it difficult and expensive to conduct clinical studies. In the words of one RA researcher, "there is a need for criteria and tools that can be used to reliably identify and quantify RA flares that represent clinically important worsening."

The state-of-the-art means of measuring disease activity is an extended clinical study involving patient and physician assessments, tender joint count, xray interpretation, ultrasounds, and blood work. Some of the factors involved in calculating the disease activity index are useful by themselves, such as swollen joint count, or a blood test. The existing tests are all either expensive (blood tests), or not very accurate (patient questionnaires). This causes suboptimal outcomes for patients. For example, when a patient goes on a new treatment plan, there is not a very good way of measuring whether or not it is working. Also, the expense of high-quality testing methods makes it very hard to pin down the risk factors for flares.

The problem of classifying RA activity can essentially be thought of as an issue of picking a subset of all possible datastreams and mapping those streams to a clinically useful output via some sort of decision making process.

Traditionally, classification methods such as DAS28 have condensed this output to a single number. Outputs of similar values are grouped together to form segments, such as high, moderate, or low disease activity. This grouping is done to make it clearer to the doctor reading the report which action should be taken in response to a test result.

But the output in its native, non-condensed form is actually a vector. The projection of this vector onto the real number line (which is what we are doing when we express DAS28 results as a number from 0 to 10) is an inherently lossy process. This loss of information results in lower quality, less individualized treatment. An individual who has a high erythrocyte sedimentation rate and a low level of joint swelling should not get the same treatment as one who has a high level of swelling and a low rate of sedimentation. Yet the number output by the DAS28 treats those two patients the same. So a future better solution would ideally go straight from the the vector output of the test to a proposed treatment.

III. Proposed Solution

The goal of our software is simple: we want to improve outcomes for individuals with Rheumatoid Arthritis. ThereâĂŹs several ways in which we believe we can add value to the existing RA treatment ecosystem. The biggest way we believe we can add value is by creating a tool to classify RA activity that costs significantly less than existing tools, yet maintains comparable accuracy.

Since RA flares entail significant visual changes to the hands and feet of those affected by it, a piece of software capable of classifying images is probably the best approach. To that end, we plan to use a convolutional neural network to output a vector with at least a disease activity score and possibly other information as well (such as current joint degradation). The neural network will run in a smartphone app. A user of this app who has rheumatoid arthritis will take a picture of their hand with the app, and the application will show (and log) information about their current RA activity level.

Neural networks need a lot of labeled training data to work properly, so we will be collecting images of hands that are accompanied by a survey taken concurrently. Karate Health (our client), will be collecting this data set for us. The data set will need to contain at least 20,000 images. The image labels will be provided by patients filling out surveys.

Neural networks are limited in their accuracy by the accuracy of the labels given to the data they are trained with. So although our network will come close to matching patient self-diagnosis, it is unlikely it will be able to match the diagnostic accuracy of a blood test. If we are able to show that is indeed possible to classify RA activity with a phone application, it would provide justification for a future application with much more commercial viability for which labels are provided by blood tests or other tests of greater diagnostic accuracy.

IV. Performance Metrics

Since we are building a neural network, we will rely on the standard means of classification of neural network performance: cross validation. In cross validation, we partition our training data into two subsets: one large one (perhaps 15000-18000 images) to train the network, and one smaller one to check the accuracy of our network. Since our output will be real-valued rather than categorical, the goal of the neural network should be to minimize loss. Average loss and standard deviation of

loss will be our main performance metrics. If we see the average loss of our neural network go down, we will know we are making progress.