

Automated scoring of bone erosion for rheumatoid arthritis with deep neural networks

Zurich University of Applied Sciences

Author: Janick Rohrbach

Supervisor: Dr. Oliver Dürr
Prof. Dr. Beate Sick

Industrial Partner: Seantis GmbH

External Supervisor: Fabian Reinhard
Dr. Tobias Reinhard

Date: December 22, 2017

Declaration of originality

Project Thesis at the School of Engineering

By submitting this project thesis, the undersigned student confirms that this thesis is his own work and was written without the help of a third party.

The student declares that all sources in the text (including Internet pages) and appendices have been correctly disclosed. This means that there has been no plagiarism, i.e. no sections of the project thesis have been partially or wholly taken from other texts and represented as the student's own work or included without being correctly referenced.

Any misconduct will be dealt with according to paragraphs 39 and 40 of the General Academic Regulations for Bachelor's and Master's Degree courses at the Zurich University of Applied Sciences (Rahmenprüfungsordnung ZHAW (RPO)) and subject to the provisions for disciplinary action stipulated in the University regulations.

Zurich, December 22, 2017

Janick Rohrbach

Abstract

Abstract goes here

Acknowledgements

I would like to express my sincere thanks to my supervisors Beate Sick and Oliver Dürr who provided me with guidance and support during the writing of this thesis.

I would also like to thank Fabian Reinhard and Tobias Reinhard (Seantis GmbH) for their valuable inputs.

Further, I want to acknowledge the SCQM foundation which made this thesis possible by providing the comprehensive dataset. A list of rheumatology offices and hospitals that are contributing to the SCQM registries can be found on www.scqm.ch/institutions. The SCQM is financially supported by pharmaceutical industries and donors. A list of financial supporters can be found on www.scqm.ch/sponsors.

Contents

1	Introduction	6
1.1	Background	6
1.2	Related literature	6
1.3	Aim and scope of this thesis	7
1.4	Outline	8
2	Theory	9
2.1	Finger joints	9
2.2	Ratingen-score	9
2.3	Artificial neural networks	10
2.3.1	Feed forward neural networks	11
2.3.2	Convolutional neural networks	11
3	Methods	13
3.1	Data preparation	13
3.2	Modeling	13
3.3	Classification	14
3.4	Regression	14
4	Results	15
5	Discussion	16
6	Conclusion	17

1 Introduction

This thesis shows a method for the automated scoring of x-ray images of patients with rheumatoid arthritis.

1.1 Background

Rheumatoid arthritis is an autoimmune disease. Which means that the disease is caused by a malfunctioning immune system. The immune system attacks healthy tissue instead of bacteria and viruses. This causes inflammation in the joints. Irreversible damage to the bone in the joint can occur, if the inflammation lasts for a long time. [1] Rheumatoid arthritis is incurable, merely the symptoms can be treated.

Today, the severity of the bone erosion is assessed by a trained rheumatologist by using x-ray images of hand and feet. This process takes several minutes per patient. This thesis shows, how recent advances in computer vision make it possible to automate this task. This leads to time savings which in return help the rheumatologist to spend more time with the patient.

The Swiss Clinical Quality Management in Rheumatic Diseases (SCQM) Foundation runs a national registry of inflammatory rheumatic diseases. [2] They have collected anonymized patient data for over 10 years and provide the x-ray images used for this analysis.

Seantis GmbH, the industrial partner for this thesis, is a Swiss company that develops data driven web applications for medical research, public administration and aviation. [3] For their customer SCQM they want to automate the bone erosion assessment. They already have a working algorithm, which detects the body part shown in the x-ray image. A second algorithm detects the joints in the image and extracts them as single images. These images are then used together with the bone erosion scores to train our model.

1.2 Related literature

There are several applications where convolutional neural networks are used in medical research.

A recent paper from Tajbakhsh et al. [4] investigated whether fine-tuning a pre-trained CNN is better than training a CNN from scratch when applied to medical images. They find that pre-trained networks with fine-tuning always outperformed or at least performed as well as CNNs trained from

scratch. They further recommend a layer-wise fine tuning which seems to outperform shallow and deep tuning.

A study by Paul et al. [5] tried to classify osteoporosis by considering x-ray images of the bone. This task proved to be very difficult as the x-ray images from healthy patients look very similar to the ones of patients with the disease. By using a transfer learning approach they achieved a validation accuracy of 44.82 %.

Zhou et al. [6] used a two-level ensemble of neural networks to identify lung cancer cells on x-ray images of the chest. The first-level ensemble classifies whether a cell is a cancer cell or not by using full voting. The second-level ensemble is used only on cells classified by the first-level as cancer cells. It differentiates between different cancer classes as well as a non-cancer class. This ensemble works with plurality voting. The authors state that this method achieves a high accuracy and a low rate of false negatives.

A report from Chen [7] showed the application of convolutional neural networks on x-ray images of hands to predict the developmental bone age. He achieves a top one and two accuracy of 46 % and 70 % respectively. This result is close to previously used methods which use manual segmentation and handcrafted features.

In a degree project Hensman and Masko [8] looked at the impact of imbalanced training data for CNNs. They find, that heavy imbalances have a strong impact on the performance and suggest oversampling of minority classes to improve the performance of the network.

1.3 Aim and scope of this thesis

The aim of this thesis is to predict bone erosion scores from x-ray images. We further examine how the bone erosion and the disease activity are correlated and use a time series of images to predict the course of the disease.

The work is based on images of the left hand only. There also exist images of right hands as well as images of left and right feet. But at this point in time, only the joints of left hands have been extracted from the images. It is assumed that the model will perform similar on the joints of the right hand. By fine-tuning the model on the images of joints from feet it should also perform well for those images.

1.4 Outline

Section 2 describes how the joints in the hand are scored and introduces artificial neural networks (ANNs) and the subcategory convolutional neural networks (CNNs). In section 3 the steps from data preparation to predicting the scores are shown. section 4 presents the results of the predictions which are then discussed in section 5. Finally section 6 concludes this thesis

2 Theory

2.1 Finger joints

Figure 1 shows an x-ray image of two hands similar to the images received from the SCQM foundation. The five proximal interphalangeal (PIP) joints and the 5 carpometacarpal (MCP) joints per hand are shown with blue bounding boxes. These are the joints, that are assessed with a score. Usually, the wrist joint is also scored but we limited our analysis to the five PIP joints and the 5 MCP joints.

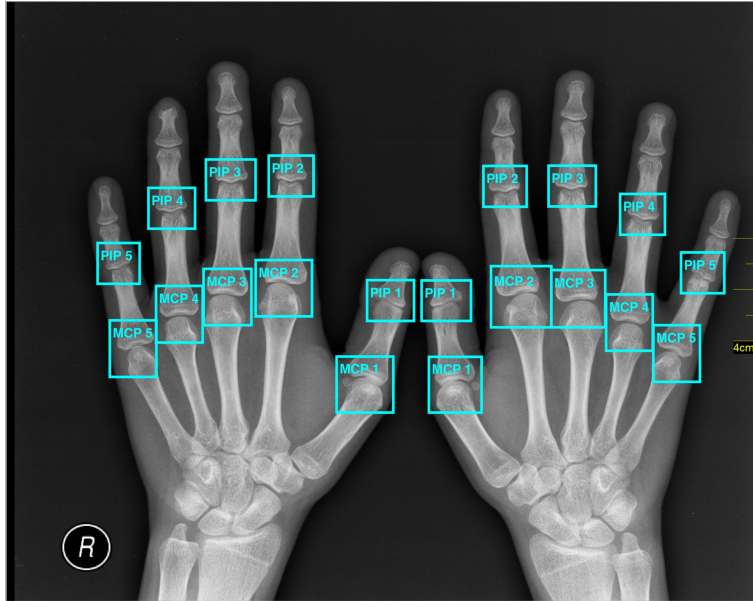


Figure 1: Proximal interphalangeal (PIP) joints and carpometacarpal (MCP) joints.

Original image by Nevit Dilmen (CC BY-SA) https://commons.wikimedia.org/wiki/File:Medical_X\discretionary{-}{-}{-}Ray_imaging_OPC06_nevit.jpg

2.2 Ratingen-score

The most important criteria for the effectiveness of a treatment is the influence on the radiological progression. To quantify the irreversible bone erosion in the joint, several scoring methods were developed. The score used in this

thesis is called Ratingen-score, it estimates the percentage of eroded joint surface. [9]. The labels of our data lie within 0 and 100 and correspond to the percentage of joint surface erosion. These values can easily be converted to Ratingen-Scores according to Table 1.

Stage	Description
0	Normal joint
1	One or more erosions, less than 20 % of the joint surface is eroded
2	21 % - 40 % of the joint surface is eroded
3	41 % - 60 % of the joint surface is eroded
4	61 % - 80 % of the joint surface is eroded
5	More than 80 % of the joint surface is eroded

Table 1: Disease stages of Ratingen-score [9]

2.3 Artificial neural networks

This section offers a very brief introduction to artificial neural networks. A more in-depth explanation can be found in Andrey Karpathy’s course notes for the Stanford class CS231n. [10]

The structure of an Artificial neural network (ANN) is inspired by the human brain. It consists of approximately 100 billion neurons which form an interconnected network. The neurons can communicate with each other by transmitting electrical potential. If the potential in a neuron reaches a certain threshold, it fires and transmits the potential to connected neurons. [11]

An ANN is a very simplified model of this biological process. A single neuron can be described by following equation.

$$f\left(\sum_i w_i * x_i + b\right)$$

Where x_i are the inputs, w_i are the weights and b is a bias term. The activation function f models the firing of the neuron. The activation function used in this thesis is the rectified linear unit (ReLU) which is calculated as follows: $f(x) = \max(0, x)$

These single neurons can then be combined to networks. The most simple network is the feed forward neural network described in the following section.

2.3.1 Feed forward neural networks

A feed forward neural network (FFNN) has an input layer, arbitrarily many hidden layers and one output layer. The neurons of each layer are connected to every neuron of the next layer. It is called feed forward because data can only flow in one direction and there are no feedback loops. Figure 2 shows a possible structure for an FFNN.

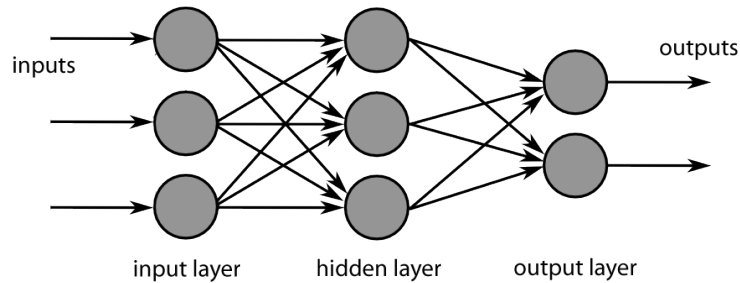


Figure 2: Feed forward neural network with one hidden layer.

Image by Chrislb (CC BY-SA) https://commons.wikimedia.org/wiki/File:MultiLayerNeuralNetworkBigger_english.png

For supervised learning the weights and biases of this network can be trained by using back-propagation. The input is fed through the network with randomly initialized parameters. The output is then compared to the true values by using a loss-function. The loss is then back-propagated through the network to adjust the parameters.

A special type of the FFNN, used for image recognition, is the convolutional neural network, which is described in the next section.

2.3.2 Convolutional neural networks

Convolutional neural networks (CNNs) take an image as an input. The image can be seen as a 3-dimensional matrix, where the third dimension includes the different color channels. Instead of fully connected layers, convolutional layers are used. Convolutions work as filters that detect different features in the image. These filters usually have a small size (e.g. 3x3) and are moved over the image. The parameters for this filter are always the same, it does not matter where in the image the filter is applied. Figure 3 shows a

possible architecture of a CNN with multiple convolutional layers and a fully connected layer at the end.

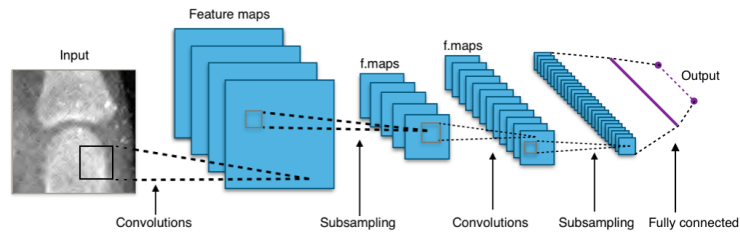


Figure 3: Structure of a convolutional neural network.

Original image by Aphex34 (CC BY-SA) https://commons.wikimedia.org/wiki/File:Typical_cnn.png

3 Methods

Keras

3.1 Data preparation

The pre-processing step formats the images into a suitable format that can be used as an input for the CNN.

The images of the joints have varying exposure. Some images are very dark while others are very bright. It was therefore considered to apply a histogram equalization, which is a linear transformation that maps the lightest pixel to 255 and the darkest pixel to 1. However, this transformation did not improve the accuracy of the model and was not used for the final model.

The data was randomly split into a training set (70 % of the data), a test set (20 % of the data) and a validation set (70 % of the data). It was split such that all images of the same person are in the same set.

The Rau scores of the labeled joints are highly imbalanced. As seen in Figure 4 most of the joints are healthy and received a score of 0. There are quite a few observations with scores between 0 and 25 but there are very little observations with scores higher than 25. Because the CNN minimizes the overall loss-function, it would perform bad for the underrepresented part of the data. In this case, the model would be bad in predicting high scores. In order to make the model a good predictor for all cases, we introduced ...???????????

3.2 Modeling

Convolution

BatchNorm

Relu

Convolution

BatchNorm

Relu

MaxPooling

6x

2 Dense layers

Output layer Softmax

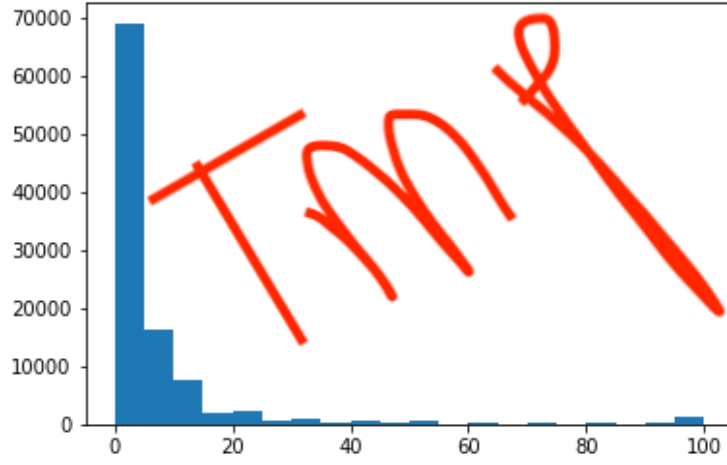


Figure 4: Distribution of the Rau scores

3.3 Classification

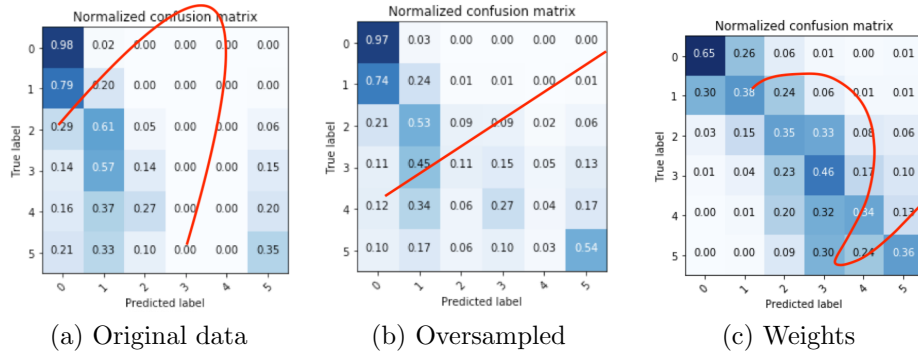


Figure 5: Confusion matrixes for the predictions of the classification models.

3.4 Regression

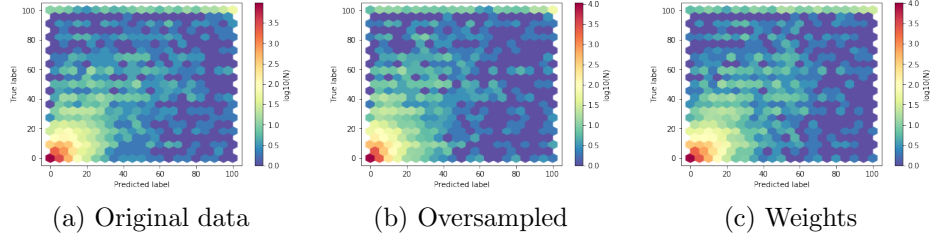


Figure 6: Predictions of the regression models.

4 Results

5 Discussion

6 Conclusion

References

- [1] American College of Rheumatology. *Rheumatoid Arthritis*. [Online; accessed 26-September-2017]. URL: <https://www.rheumatology.org/I-Am-A/Patient-Caregiver/Diseases-Conditions/Rheumatoid-Arthritis>.
- [2] SCQM. *About Us*. [Online; accessed 11-October-2017]. URL: <https://www.scqm.ch/en/ueber-uns/>.
- [3] Seantis GmbH. *Data Driven Web Applications*. [Online; accessed 12-October-2017]. URL: <https://www.seantis.ch>.
- [4] Nima Tajbakhsh et al. “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?” In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1299–1312. URL: <https://doi.org/10.1109/TMI.2016.2535302>.
- [5] Rahul Paul et al. *Make Your Bone Great Again : A study on Osteoporosis Classification*. University of South Florida, 2017. URL: <https://arxiv.org/abs/1707.05385>.
- [6] Zhi-Hua Zhou et al. “Lung cancer cell identification based on artificial neural network ensembles”. In: *Artificial Intelligence in Medicine* 24.1 (2002), pp. 25–36. URL: [https://doi.org/10.1016/S0933-3657\(01\)00094-X](https://doi.org/10.1016/S0933-3657(01)00094-X).
- [7] Matthew Chen. *Automated Bone Age Classification with Deep Neural Networks*. Stanford University, 2016. URL: http://cs231n.stanford.edu/reports/2016/pdfs/310_Report.pdf.
- [8] Paulina Hensman and David Masko. *The Impact of Imbalanced Training Data for Convolutional Neural Networks*. KTH Royal Institute of Technology, 2015. URL: https://www.kth.se/social/files/588617ebf2765401cfcc478c/PHensmanDMasko_dkand15.pdf.
- [9] R. Rau and S. Wassenberg. “Scoringmethoden bei der rheumatoiden Arthritis”. In: *Bildgebende Verfahren in der Rheumatologie*. Ed. by Deutsche Gesellschaft für Rheumatologie. Steinkopff, 2007. Chap. 2, pp. 27–46. URL: https://doi.org/10.1007/978-3-7985-1721-9_2.
- [10] Stanford University Andrej Karpathy. *CS231n Convolutional Neural Networks for Visual Recognition*. [Online; accessed 26-September-2017]. URL: <https://cs231n.github.io/>.

- [11] R. Kruse et al. *Computational Intelligence*. Springer, London, 2016.
URL: https://doi.org/10.1007/978-1-4471-7296-3_2.

List of Figures

1	Proximal interphalangeal (PIP) joints and carpometacarpal (MCP) joints.	9
2	Feed forward neural network with one hidden layer.	11
3	Structure of a convolutional neural network.	12
4	Distribution of the Rau scores	14
5	Confusion matrixes for the predictions of the classification models.	14
6	Predictions of the regression models.	15