
Title

Written by
Group 17gr7402

1. Semester

School of Medicine and Health

Biomedical Engineering and Informatics

Fredrik Bajers Vej 7

9220 Aalborg

Title:

title title

Theme:

Biomedical Signals and Information

Project Period:

Fall semester 2017

Project Group:

17gr7402

Abstract

Participants:

Birgithe Kleemann Rasmussen

Ignas Kupcikevičius

Linette Helena Poulsen

Mads Kristensen



Supervisor(s)

Shellie Boudreau

Lasse Riis Østergaard

Page Numbers:

Appendix:

Date of Completion: 20/12/2017

The content of this report is freely available, but publication may only be pursued due to agreement with the author.

Preface

Bla bæa

Contents

Chapter 1 Introduction	1
1.1 Primary aim	2
1.2 Secondary aim	2
Chapter 2 Background	3
2.1 Anatomy of the Knee	3
2.2 Pain	5
2.3 Patellofemoral pain syndrome	6
2.4 Identify and interpret pain	7
2.4.1 Identify pain	7
2.4.2 Pain interpretation	8
2.5 Knee regions	9
2.6 Machine learning	11
2.6.1 Deep Learning	11
2.6.2 Back-propagation	12
2.6.3 Dropout	15
2.6.4 Core Neural Network layers	16
Chapter 3 Methodology	18
3.1 Data	18
3.1.1 Software application: Navigate Pain	19
3.1.2 Data representations	19
3.2 Pre-analysis	21
3.2.1 Classification of data	21
3.2.2 Threshold selection	22
3.2.3 Simple regression models	24
3.3 Pre-processing	26
3.3.1 Morphology-representation	26
3.3.2 Regions-representation	27
3.3.3 Combined-representation	28
3.4 Programs	29
3.5 Deep learning (but I would like to think another name for this section)	29
3.5.1 Sofware and hardware	30
Chapter 4 Description of models	31
4.1 Data handling and design choices of the models	31
4.1.1 Data handling in python	31
4.1.2 Optimization techniques	31
4.1.3 Training of the networks	33
4.2 Vector image model	33

4.3 Raw binary representation model	33
4.4 Combined representation model	35
Chapter 5 Results	36
Bibliography	37
Appendix A Knee injury and Osteoarthritis Outcome Score (KOOS) [?]	41
Appendix B Threshold analysis	45

Chapter 1

Introduction

Patellofemoral pain syndrome (PFPS) is a painful musculoskeletal condition that is presented as pain behind or around the patella [1, 2]. PFPS affects 6-7 % of adolescents, of whom two thirds are highly physically active [3]. Additionally the prevalence is more than twice as high for females than males [3, 4]. PFPS may be present over a longer period of time where a high number of individuals experience a recurrent or chronic pain [5] and may also lead to osteoarthritis [4, 6].

Patellofemoral pain (PFP) is often described as diffuse knee pain, that can be hard for individuals to explain and localize [5]. Despite the fact that individuals feel pain in the knee, there is no structural changes in the knee such as significant chondral damage or increased Q-angle. Because of this there is no definitive clinical test to diagnose PFPS and it is thereby often diagnosed based on exclusion criterias [4] to which PFPS is also described as an orthopaedic enigma, and is one of the most challenging pathologies to manage [7]. To assist diagnosis of PFPS, pain maps may be used as a helpful tool for the individuals to communicate their pain by drawing pain areas on a body outline [8].

Another method to measure pain is visual analog scale (VAS), that scores pain intensity between no-pain to the worst pain imaginable [10]. A study by Boudreau et al. shows that through the use of pain maps it is possible to find a correlation between the symptom duration and the size as well as morphology of pain area in individuals with a symptom duration longer than five years. Furthermore a correlation between pain intensity and the size of the pain was found.[9] It is a known problem that chronic pain is considered as multidimensional, because the perceived pain of an individual is influenced by biomedical, psychosocial and behavioral factors [?].

Because there is a lack of knowledge in the area of PFP, and it has been shown that there is a correlation between pain maps and symptom duration as well as pain intensity, it is interesting to investigate if pain maps can be used to classify and predict PFP related information.

A method that has not been found used in this context before is deep learning. The deep learning method is chosen for this study because it is a state-of-the-art method, that has shown greater performance in specific computation fields, compared to other machine learning methods [11]. Furthermore, the method is chosen because of its ability to find non-linear connections between input and output data [11], which is found relevant for this study mainly based on the fact that pain is subjective and may be affected by the multidimensionality of chronic pain.

The goals of this project is to explore how accurate a deep learning model can classify pain maps according to symptom duration and pain intensity by using a limited dataset. Because of the imbalance in prevalence between females and males, the gender is included as a feature in the deep learning model. Furthermore, morphology of the pain is considered to be

relevant, based on the indication that morphology and size of pain area increase with prolonged symptom duration. To accommodate the difficulty of localizing pain related to PFP, the pain is divided into different knee regions. This might indicate whether regions has a relation to either symptom duration or pain intensity. To investigate the influence of morphology and location of the pain, three types of pain map representations are created: a binary representation which reflect the morphology, a simplified representation of morphology based on knee regions that give information about the pain location, and a combined representation that contains the morphology divided into knee regions.

1.1 Primary aim

The aim of this study is to explore classification performance of a deep learning model, using PFP pain maps and gender as input to classify either symptom duration or pain intensity.

It is hypothesized that a deep learning model that uses pain maps and gender as input parameter has a higher performance when classifying according to symptom duration than pain intensity.

1.2 Secondary aim

The secondary aim of this project is to investigate if multiple pain map representations, which reflect the morphology and location of the pain, affect the classification performance of the deep learning models.

It is hypothesized that different data representations of pain maps, reflecting morphology and location of pain, affect the performance accuracy of a deep learning model when classifying according to symptom duration or pain intensity.

Chapter 2

Background

This chapter presents the background knowledge that optimizes the understanding of essential topics in this project, such as patellofemoral pain and deep learning. Regarding patellofemoral pain it is relevant to get knowledge about the anatomy of the knee as well as pain and pain measurements if a deeper understanding of the syndrome is considered necessary. Furthermore, the chapter is essential for getting a basic understanding of some properties and optimizers in the neural network models used in this project.

2.1 Anatomy of the Knee

The knee is the largest synovial joint in the body and consists of a hinge and a gliding joint. The hinge joint is placed between the lateral and medial femoral condyles and the lateral and medial tibial condyles. Between the patella and femur is the gliding joint formed. The structure of the knee is illustrated in figure 2.1.[12]

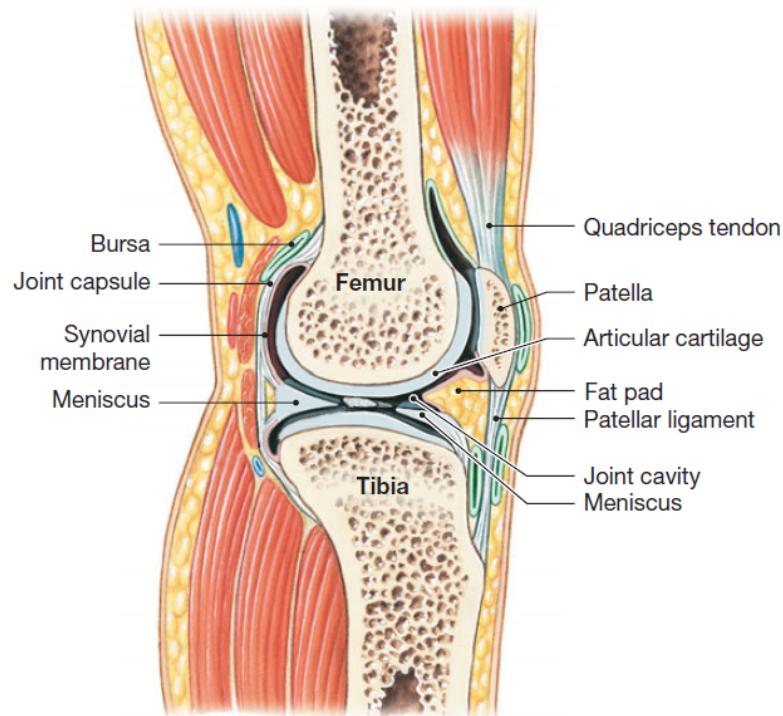


Figure 2.1: The anatomy of the knee. Edited [12].

It is shown in figure 2.1 that the patella is a sesamoid bone. At birth the patella consists of cartilaginous and ossifies when the child's extremities gets stronger, which typically proceeds

between age two or three and the beginning of puberty.

The patella is surrounded by the tendon of the quadriceps femoris. Quadriceps femoris is the muscles which controls the extending of the knee. The quadriceps tendon is combined to the surface anterior and superior of patella. Tibia is combined to the anterior and inferior surface of the patella by the patellar ligament. The bones, tibia and femur, are covered by articular cartilage with the purpose of protecting the bones from friction. The articular cartilage on the two bones are separated from one another by synovial membranes that contains synovial fluid, that further reduce the friction. The primary functions of the synovial fluid is to lubricate, distribution of nutrient and absorption of shock.[12]

The fat pads and menisci are placed between the articular cartilages. The fat pads' function is to protect the cartilage and fill out space as result of the joint cavity changes. The menisci stabilize the knee and acts like pads, that conform shape when femur moves. In addition to fat pads and menisci, the bursa acts as friction minimization between patella and tissues.[12] There are three separate articulations in the knee joint. The first is between the patella and the patellar surface of the femur and the rest are between the femoral and tibial condyles. Additionally, the knee consist of seven major ligaments that stabilize the knee joint, which are shown in figure 2.2.[12]

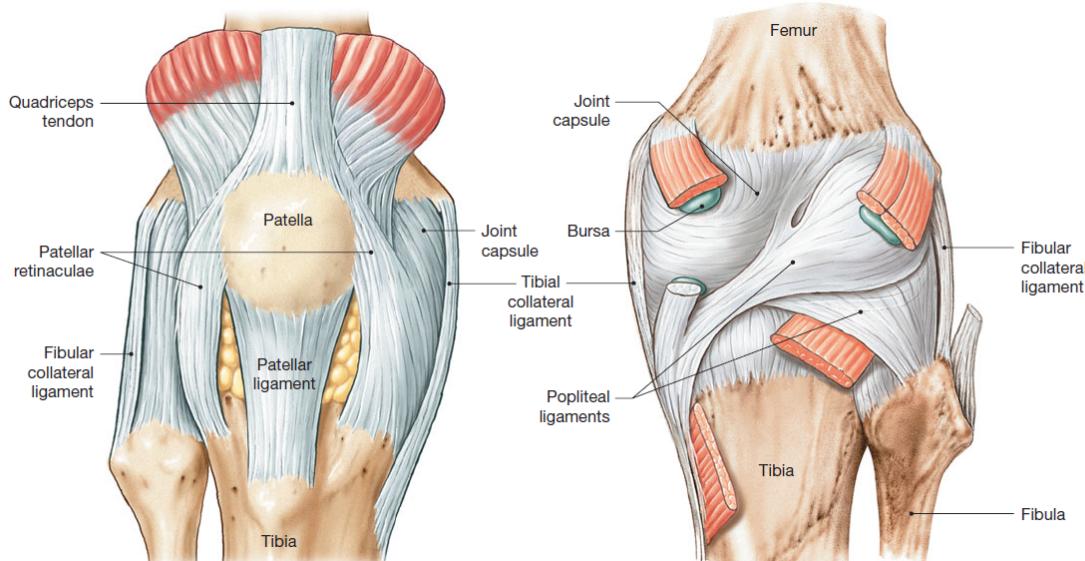


Figure 2.2: The anatomy of the knee with focus on the ligaments. Edited [12].

The ligaments patellar retinaculae and patellar ligament support the anterior surface of the knee. When the knee is fully extended, the tibial and fibular collateral ligament are responsible for stabilizing the joint. Between femur and the two lower bones in the leg, tibia and fibula, is the location of the two popliteal ligaments, which stabilize the posterior surface of the joint. In addition to the visible ligaments in figure 2.2 there are the anterior cruciate ligament (ACI) and posterior cruciate ligament (PCL) in the joint capsule. The two ligaments cross each other and are connected to the tibial and femoral condyles, which reduce the movement, anterior and posterior.[12]

2.2 Pain

The International Association for the Study of Pain (IASP) has defined pain as being “an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage” [13, 14].

Humans are aware of the surroundings and threats to their bodies because of the pain. The pain indicates that there might be a risk for permanent damage on the body, which refrain humans from danger and therefore increases the chances of survival.

Pain can be either nociceptive or neuropathic. Nociceptive pain is associated with tissue damage. This type of pain is related to the nociceptors, which are receptors with a high threshold that when stimulated may give the perception of pain in tissues [15]. Neuropathic pain occurs central from the nervous system. This pain can be caused by illness or physical damage.[15]

Furthermore, pain can be divided into three categories: acute pain (less than three months), persistent or chronic pain and cancer pain.[16] Additionally, the pain can be divided into qualities, which is shown in figure 2.3.

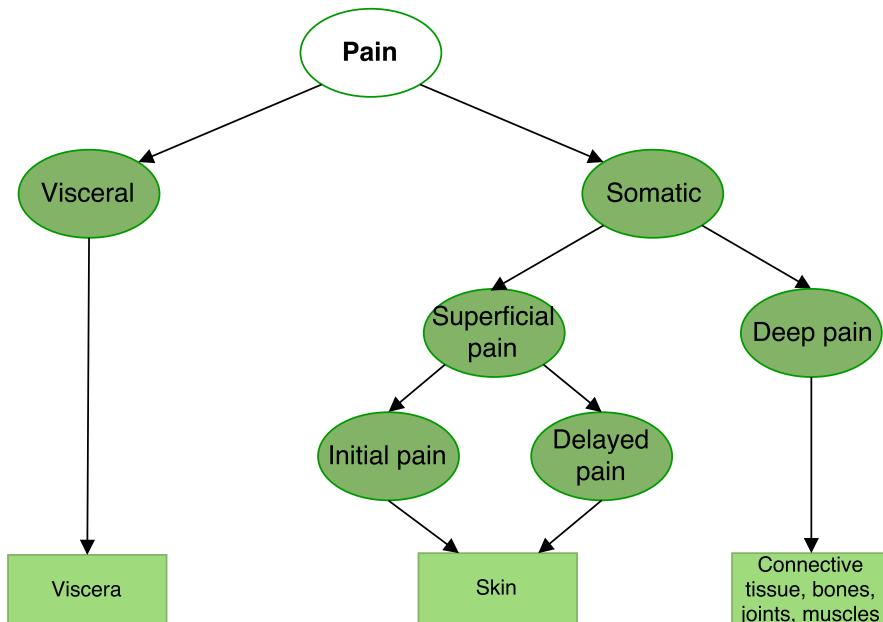


Figure 2.3: Model of pain qualities. Ovals with green background represent qualities of pain. The rectangles show where the pain originates. Edited [15].

Pain can be divided into two qualities; visceral and somatic pain. Examples of visceral pain include pain associated with gallstone and appendicitis. This pain can be characterised as a dull or diffuse feeling. Somatic pain is subdivided into superficial pain and deep pain. If the pain derives from the skin it is superficial pain, which furthermore is divided into initial pain and delayed pain. The initial pain is the first pain that is received, and characterised as sharp and localizable. The delayed pain, also known as the second pain, is a dull or burning pain that occur after a half to one second. This pain is more difficult to localise than the initial pain and lasts longer.[14, 15] The other type of somatic pain is deep pain, which is associated with pain from the muscles, bones, joints and connective tissue. This pain is described as a dull pain and it radiates into the surrounding tissue, which makes the exact pain area hard to point out.[14, 15]

2.3 Patellofemoral pain syndrome

Patellofemoral pain syndrome (PFPS) is a painful musculoskeletal condition [1, 2], which presents pain behind or around the patella. The patellofemoral pain (PFP) is often known as anterior knee pain and runner's knee. The pain is often described as diffuse knee pain, which is provoked by patellofemoral loaded activities like climbing stairs, running on hard or slanted surfaces, hiking, squatting or just prolonged sitting in the same position.[2, 6, 9, 17] However, this pain is not caused by previous trauma [6]. Knee pain is not the only symptom of PFPS, the individuals often complain about knee stiffness, patellofemoral crepitus, swelling knee and having trouble with common daily activities [6, 12]. The individuals may limit or stop physical activity because of the pain, and that can lead to an increase in weight [4, 6]. Physiologically PFPS is associated with incorrect movement of the patellar, that occurs when the patella moves outside of its ordinary track, which for instance can be movement in lateral direction instead of movement in superior-inferior direction.[12]

PFP is mostly prevalent in adolescents and younger adults who are physically active, but it can affect people of all ages and activity levels [1, 6, 17]. Additionally, females are affected about more than twice as often as males [4]. Furthermore, the PFPS may persist for up to 20 years and lead to osteoarthritis [4, 6].

Despite the fact that individuals feel pain in the knee, there is not any structural changes in the knee such as significant chondral damage or increased Q-angle [4]. The Q-angle, also known as the quadriceps angle, is the angle between a line that follows the longitudinal axis of femur, and a line that follows a line from the tibial tubercle through the center of patella [18].

There is no definitive clinical test to diagnose PFP, but there is a test to elicit the knee pain by doing a squatting manoeuvre. The PFP is evident in 80% of people who are tested positive in this test.[6, 17] Therefore the diagnosis PFPS is often based on exclusion [4]. After the diagnosis, evidence based treatments may reduce pain and improve function that allows individuals to maintain physical activity [17]. The aetiology of PFPS still remains unclear [2].

HUSK MARTINI! *** Since the aetiology of PFPS still remains unclear [2], it is hard to place this type of pain in addition to nociceptive and neuropathic pain. But PFPS can be classified as deep pain and acute or chronic pain. Since the symptom duration often is longer than six month it is described as a chronic deep pain. ****

2.4 Identify and interpret pain

There are many ways to identify and interpret pain. To identify pain and find some physical damage that causes the knee pain, objective methods may be used. Subjective methods are used to interpret pain for collecting knowledge of the individual's pain intensity, behavior and how it is experienced.[19]

2.4.1 Identify pain

An objective pain measurement is often used when an individual experiences knee pain where a clinical examination of the knee can occur. This examination involves i.a. provocative tests, such as anterior and posterior drawer test, Lachman's test and pivot test that examines the integrity of the ACL and PCL. Furthermore is McMurray test which tests for meniscal tear.[20] Illustrations of the tests are shown in figure 2.4.

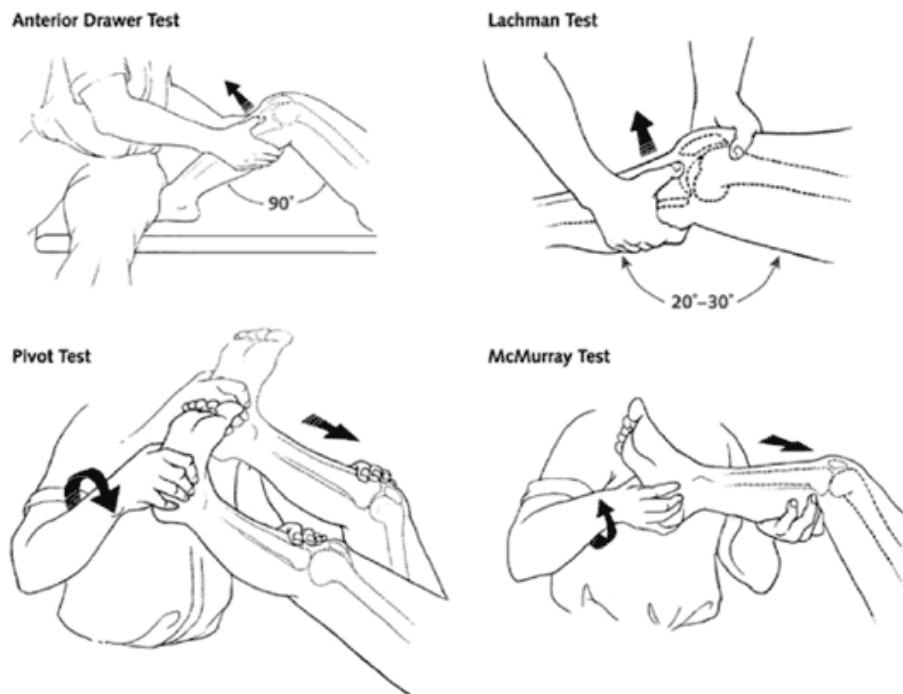


Figure 2.4: Clinical examination with provocative tests: Anterior Drawer Test, Lachman Test, Pivot Test and McMurray Test.[20]

In addition to clinical tests there is some paraclinical tests such as X-ray and MRI, but as mentioned PFPS does not show any structural changes in the knee [4], which makes it difficult for healthcare personnel to treat the individuals.

2.4.2 Pain interpretation

Pain is experienced and perceived subjectively [13, 19] and is dependent on personality and character [15], which is why it is important to measure the pain from the individual's perspective.

One of the most commonly method used to measure pain intensity is Visual Analogue Scale (VAS) [21]. VAS is often used in clinical and research settings, where the individuals mark their pain on a scale from no-pain to the worst pain they can imagine.[10] An illustration of a VAS is shown in figure 2.5.

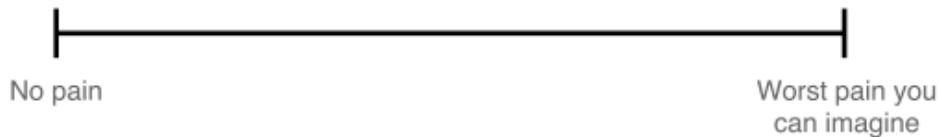


Figure 2.5: Visual Analogue Scale (VAS). Edited [10].

Additionally to mark pain on a scale is questionnaires used to define individual's pain. An example on a questionnaire is Knee injury and Osteoarthritis Outcome Score (KOOS), which contains questions about symptoms, stiffness, pain, daily living, function, sports and recreational activities and quality of life. When the individuals fill the scheme a score between zero and one hundred is achieved. A score at zero represents extreme knee problems, whereas a score at one hundred represents no knee problems.[22] The KOOS questionnaire can be seen in Appendix A.

Pain mapping

Individuals indicate their pain by 'placing both hands over their knees', because of the PFP characteristics, which makes it hard to precisely communicate their pain. Thereby pain mapping is a method for individuals to better indicate and communicate their pain. Pain mapping is a technique, that Harold Palmer introduced in 1949 [23], which is used to transfer an individual's perceived pain into an objective graph or map by drawing the pain area. Pain drawings can be made by the individuals who draw their pain areas on a body outline. Pain drawings can also be made by observers who observe the individuals and then draw from the signs the individuals are showing. An example of a body outline is shown in figure 2.6. Sometimes a subjective questionnaire is added to the pain drawings to get a more detailed overview of the pain to determine parameters associated with the pain. These parameters can also be useful in determining the source of the pain.[24]



Figure 2.6: An anterior body outline of pain drawing taken from the software application Navigate Pain.

Pain mapping are commonly used in clinical practice [24], and can be useful for individuals when they try to describe their pain. Pain maps may also be helpful in diagnosing individuals and follow-ups during or after treatment to get an indicator of the individual's response to the treatment.[8] According to Schott, there are some issues with the graphical representations of pain, some of which are problems with drawing a three-dimensional feeling of pain on a two-dimensional surface, and distinguishing between internal and external perceived pain on a map [24].

2.5 Knee regions

When looking at pain drawing samples from multiple individuals it is evident that there is a high variability in the distribution of pain patterns across different areas of the knee, which might be related to the PFP characteristics.

To distinguish between different pain areas, the knee can be divided into various regions as seen in figure 2.7, where the division of the left and right anterior knees are illustrated.

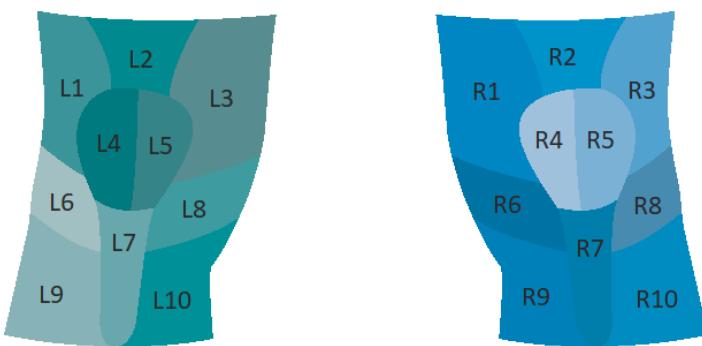


Figure 2.7: The regions of the left and right knees, where each knee is split into ten regions. Edited [25].

The divisions are inspired by Photographic Knee Pain Map (PKPM) which are designed to categorise the location of knee pain, for diagnostic and research purposes. PKPM represents both knees that makes it possible to identify unilateral and bilateral pain.[25]

The regions are based on the anatomical structures according to the areas where individuals often indicate pain. There are ten regions, where region 1 and 3 represent the superior lateral and superior medial areas for patella. Region 2 refers to quadriceps tendon. The patella is divided into lateral and medial regions, which are region 4 and 5. Region 6 and 8 are lateral and medial joint line areas. Patella tendon is region 7 and the two last regions, 9 and 10, are tibia lateral and medial.[25]

2.6 Machine learning

Machine learning describes the use of algorithms to make a system able to identify different data types, like images or text, for transcription of speech into text, matching news items, posts or selection of relevant results of search [11]. Machine learning identifies rules in a dataset from given input and output. If a computer learns this feature, it can be used to make intelligent decisions and predict specific outcomes.[26] Machine learning is a field that has seen a lot of progress over the past decades, partially because developers recognize the ease in training a system only using examples of the desired in- and output behavior. By using this method it is easier than trying to manually write a piece of code that anticipates different scenarios from different input types.[27]

2.6.1 Deep Learning

Deep learning is a branch of machine learning. The main difference between the use of machine learning and deep learning, is that deep learning is more suitable for handling raw data forms. Instead, a machine learning system often needs a feature extractor, that will generate a feature vector from the data that can be used as an input. Deep learning is based on different techniques that makes it able to handle raw data, mainly because of its structure.[11, 28] Because of this the system will automatically detect the necessary representations needed for classification and detection.

Neural network is a structure of deep learning which consists of different layers, that can be divided into input- and output-layers, with one or more hidden layers in between [28]. The key aspect of these layers is that the features are not defined by programmers, but they are found and learned from raw data using a general-purpose learning procedure.[11] An example of a neural network structure can be seen in figure 2.8.

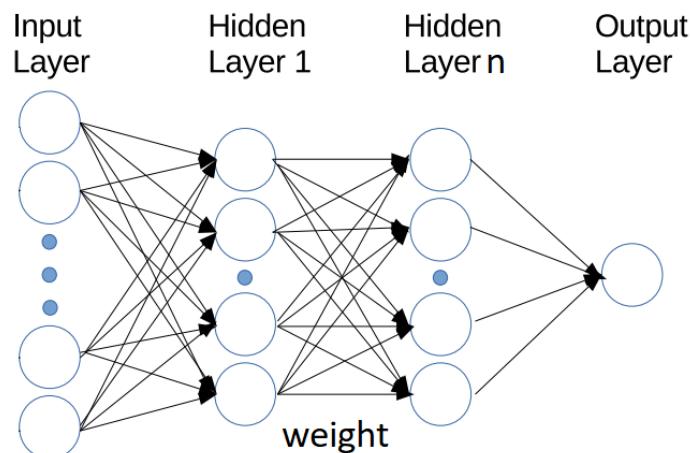


Figure 2.8: Example of the structure of a neural network. Edited [29].

The different layers consist of a series of nodes, where each node is connected by weights ****EXPLAIN!**** to one or several other nodes in different layers. In the input-layer the nodes receive data. The second layer will then receive the output from the previous layer, and this process continues through the layers until the output-layer is reached.[28] An example of how the hidden layers may affect an image in one type of neural network can be explained in the following. Firstly, the system detects minor changes like edges. Secondly, the edges are

compared and put together to make up different kind of shapes. In the third hidden layer, it will be further combined to make up an object that can be identified.[11]

Learning scenarios

****WRITE A INTRODUCTION! ITS NEED TO BE TRAIN, OPTIMIZED ECT.*****
There are different approaches for training a neural network, where the two main learning scenarios are supervised and unsupervised learning.

Supervised learning is the most common way of training in machine learning. When applying this learning method the neural network is trained with input data that has a corresponding label. The network calculates an output through the forward pass, where the data is simply passed through the network. This output may then be compared to the label, and used to evaluate the performance of the system. As a result of the evaluation, the network may learn from the data by doing a backward pass through the network, also known as back-propagation.[11] Overall supervised learning may be described as teaching the network how to associate a given input to a specific output [31], and is mostly associated with classification, regression, and ranking problems [30].

Unsupervised learning is when training is performed with data that has no output label. Instead of learning associations between input and output, the network organizes the data by searching for common characteristics [30]. An example of an unsupervised learning algorithm is k-mean clustering, where the unlabeled dataset goes through a classification, and splits data into clusters that are near each other [31].

2.6.2 Back-propagation

***** CORRECT FORWARD AND BACKWARD + EQUATIONS***** Back-propagation is a popular learning algorithm in neural networks, that is based on gradient descent, and used because of its simplicity and computationally efficiency.[32, 33] It is the learning process where the weights of a neural network are adjusted in order to reduce the error calculated between the calculated output of the network and the expected output. This makes back-propagation closely related to supervised learning, to which back-propagation is the most general method used.[33] When a neural network is initialized the weight may be set with a random value, meaning that the neural network may perform very poorly through the first iterations of the training. Based on a loss function a loss is calculated for every input that passes through the network, that may be used as a part of back-propagation to make the adjustments on the weights to reduce this loss. As training progresses, the loss should decrease as a result of the weight adjustments, and improve the performance of the neural network.[11, 31, 33] This learning process continues until optimal weights with minimum error is reached.[34] The basic concept is that gradients can be computed efficiently by propagation from the output to the input in order to minimize the overall output error as much as possible during the learning stage. This algorithm process is divided in two main stages: forward and backward. In the first process (forward), the back-propagation architecture is described as the inputs and weights multiplication of separate node summed with additional coefficient called bias.[34, 35]

MAYBE TALK ABOUT ACTIVATION FUNCTIONS IN RELATION TO BACKPROPAGATION.

IF WE WILL HAVE TIME WE NEED TO ADD ACTIVATION GRAPHS WITH EXPLANATION

A PROBLEM OFTEN RELATED TO OPTIMIZATION IS THE LARGER THE NETWORK THE HARDER IT IS TO OPTIMIZE, BUT IN ANOTHER WAY IF THE NETWORK IS SMALL AND SIMPLE THE EASIER IT IS TO OPTIMIZE.

Gradient Descent

Gradient descent is one of the most common techniques for optimizing neural networks. It is a way to minimize the loss function by updating the parameters, like weights, in the opposite direction of the gradient of the objective function.[36] The principle of the gradient descent could be explained as a "ball climbing down a hill" until a local minimum is reached as shown in figure 2.9. At each step, the opposite direction of the gradient is taken and the step size is determined by the value of the learning rate together with the slope of the gradient until the convergence is reached. Convergence means that oscillations of the value are small enough to call it the minimum value.[37]

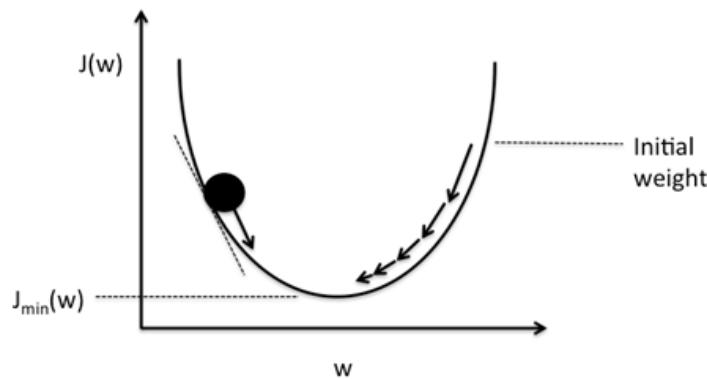


Figure 2.9: Illustration of gradient descent working principle, where $J(w)$ is the loss function, $J_{min}(w)$ is final approximation to the local minimum of $J(w)$, w is value of the parameter. The arrows indicate the step direction, i.e. the negative gradient.[37]

There are three variants of gradient descent: batch gradient descent, stochastic gradient descent, and mini-batch gradient descent. They differ on the amount of data used to compute the gradient of the loss function. Depending on which of the gradient descent variants is used, the trade-off between the accuracy and the runtime could be seen.

Batch gradient descent computes the gradient of the cost function with regards to the parameters for the entire training dataset.[36] Batch gradient descent has the significant deficiency, it takes a single step for one pass over the training set, meaning the larger dataset, the slower algorithm updates the weights and the longer it will take to reach global minimum.[quota] In cases like these, stochastic gradient descent is being used more commonly.
**** SOURCE ****

Stochastic gradient descent (SGD) performs a parameter update for each training example and label. It is therefore much faster and it also performs frequent updates with a high variance causing loss function to fluctuate. These fluctuations enable it to jump to new potentially better local minima, but it may complicate the convergence to reach the exact minimum because of overshooting.[38]

Mini-batch gradient descent performs the parameters update for every mini-batch of training examples, specified by a batch size. By that, the variance of the parameter updates are reduced leading to more stable convergence and fast performance.[36]

Additionally, there could be few challenges while using gradient descent as an optimizer. It is difficult to pick a proper learning rate so few gradient descent optimization algorithms were invented. The most popular optimizers are adagrad, adadelta and adam. **** SOURCE **** Adagrad is an algorithm for gradient descent optimization which adapts the learning rate to the parameters. It performs larger updates for frequent and smaller for infrequent features. It has one weakness if the learning rate shrinks too much, the algorithm is no longer able to adapt.[36]

Adadelta is an updated version of Adagrad, but the learning rate is monotonically decreasing. Using this optimizer, there is no need to tune the parameters of optimizer meaning that it can be applied in a variety of situations.[36]

Adam stands for Adaptive Moment Estimation. It is the most used method for computing adaptive learning rate and updating the parameters. This optimizer calculates the learning rate for each parameter and stores momentum changes separately. This helps to reach the convergence faster with a decent learning speed.[40]

A study Patacchiola and Cangelosi showed the evaluation of the performance on different optimizers on a dataset containing 21977 female and male head pose images. The result is shown in figure 2.10, where the Adam optimizer had the fastest convergence rate and it reached the lowest loss values.

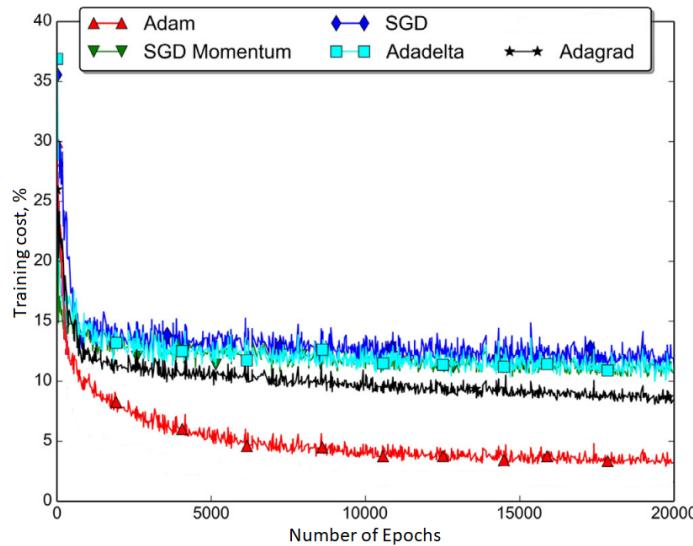


Figure 2.10: Comparison of the convergence speed between different optimizers used to train architecture on AFLW dataset. The loss values are the mean of the five attempts.[41]

However, the results are not always similar. All of the optimizers perform differently depending on the problem and parametrization, which in the majority of the cases is the most challenging part. This leads to the conclusion that there are not a winning optimizer and it has to be chosen based on every problem.[42]

Learning curves

**** EXPLAIN WHY IT'S GOOD TO SEPARATE DATA AND WHY? MOVE IT UP BEFORE GRADIENT **** During the beginning of training, the training error of network will typically be relatively high, but during training the error decreases monotonically, as the weights are adjusted in the network [33]. An illustration of how the error values are affected during training can be seen in figure 2.11.

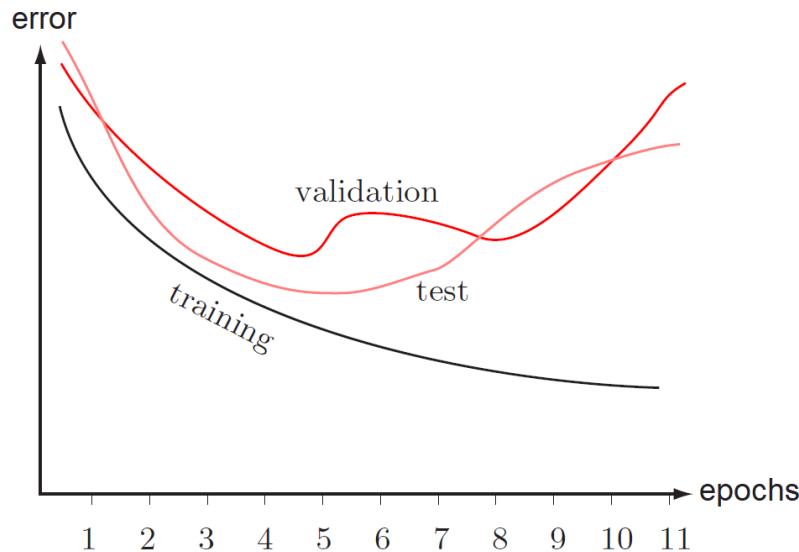


Figure 2.11: Illustration of how training (black), validation (red), and test (orange) error is affected by the increase in epochs. Edited [33].

From the figure it can be seen how the error value of the validation, can be used to evaluate the network. Near the fifth epoch the validation and the test error starts to rise, indicating that the network is overfitting to the training data, thereby decreasing the generalization abilities. Validation error can therefore be used as stop criterion for when the training is optimal, and prevent overfitting. Typically the validation and test error will always be higher than the training error, which is also seen in figure 2.11.[33]

2.6.3 Dropout

Dropout is a regulation technique used to reduce overfitting of the neural network. In a network the dropout is applied to the individual layers, and works by randomly dropping different nodes temporarily in the given layer during training. An illustration of the principle of dropout can be seen in figure 2.12. This parameter is specified with a percent value, which defines the fraction of the nodes that drop [43].

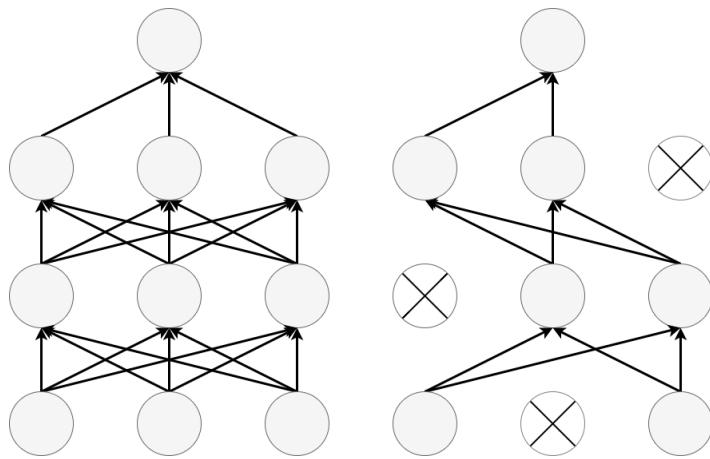


Figure 2.12: Dropout effect: The figure on the left illustrates a fully connected network, without dropout, and the right shows a network where dropout is enabled on the first three layers. Edited [44].

This reduces co-adaptation, where nodes compute the same features, to where this may increase the generalization capabilities for a neural network. A study Srivastava et al. has tested the use of dropout in different neural networks, and indicates that the most optimal range of dropout is 20% of the nodes in the visible layers, and 50% in the hidden layers.[44]

2.6.4 Core Neural Network layers

**** INTRODUCTION! ****

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) is a type of special neural network for processing data with a grid-like topology [31]. CNNs perform highly in several tasks, including digit recognition, image classification and face recognition. The key aspect of CNNs is to automatically learn a complex pattern by extracting visual features from the pixel-level content.[29, 35]

The architecture of a typical CNNs is always combined of two types of layers: convolutional and pooling [31, 11].

The purpose of the convolutional layer is to recognize the features in the input by taking e.g. an image and scan it, then split it up into the feature maps. The terminology regarding the output of a convolutional layer can be referred to a feature map [31, 35]. A complete convolutional layer consists of several feature maps, so multiple features can be extracted at each location in the image [35]. The other layer called pooling, typically follows after convolutional layer. It consists of the same number of feature maps as the previous convolutional layer had. Each feature map is used as a new input in a pooling layer. Depending on the network's depth, the convolutional and pooling layers alternate until the last pooling layer is reached.[35]

The combination of convolutional and pooling layers defines the part of the network which performs feature extraction while the classification part is made by fully connected layers. A typical CNNs architecture for character recognition of the images, called LeNet-5, can be seen in figure 2.13.

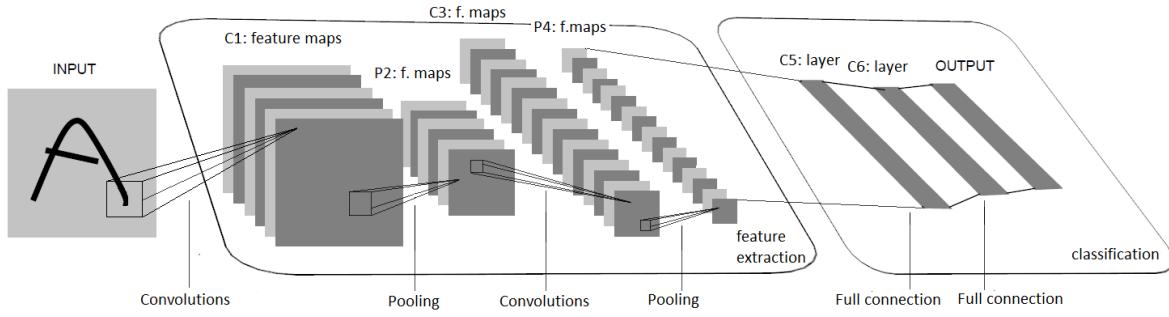


Figure 2.13: Architecture of LeNet-5, a CNN for character recognition. Each plane is a feature map and the size of the feature map differs throughout the layers. The model consists of two convolutional, two pooling and two fully connected layers [35].

Units in the first hidden layer of LeNet-5 are arranged in planes, which are separate feature maps respectively. Once a feature has been detected, its location becomes unimportant and only its approximate position relative to other features is relevant. This is because the positions are likely to vary for different images of the same object and it is irrelevant for identifying the pattern. To reduce the importance of the position, the spacial resolution of the feature map is decreased by the use of pooling layer.[35]

Pooling layer

MOVING THE EXAMPLE TO THE END, SO CNN IS FIRST, THEN POOLING, FULLY CONNECTED LAYERS AND THE EXAMPLE As previously mentioned convolution are typically followed by pooling [11, 31]. Pooling can be used to reduce the size of the dataset, which may increase computation speed, because the amount of data passed to the next layer is smaller. By pooling the input, a smaller representation is given, that still contains the relevant features.[31, 35] The pooling process can be defined as a window that passes over e.g. a feature map from convolution, where a value within the window is extracted. One type of pooling layer is max pooling that takes the maximum value within the window [31?]. A pooling layer may be defined simply by its window (Kernel) size, padding size and a stride length, where stride length is the number of values the window jumps as shown in figure 2.14.[?]

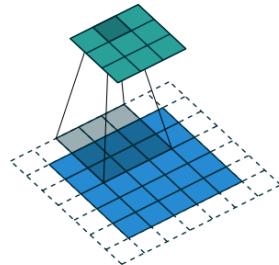


Figure 2.14: Illustration of pooling with 3×3 kernel over a 5×5 input using padding and strides (i.e., input = 5, kernel = 3, stride = 1, padding size = 1) [?].

Chapter 3

Methodology

This chapter creates an understanding of the available data, how it is analysed and prepared before including it in deep learning models. Additionally, the program Navigate Pain and programs used to development of the deep learning models are described. Furthermore, the implementation of the deep learning model is presented consisting of hardware, software and architecture of used neural network.

3.1 Data description

Data used in this project were collected beforehand from an on-going FOXH trial which is conducted in collaboration with Danish and Australian universities. The data consists of pain maps which were drawn by individuals with PFPS through the use of an application, Navigate Pain, in a clinical setting. Navigate pain is further described in section 3.1.1. The pain maps are both from individuals with uni- and bilateral PFP. An example of a pain drawing with bilateral pain is shown in figure 3.1.



Figure 3.1: An example of a pain drawing from individual with PFPS. The red markings indicate the area of pain perceived by the individuals. In this case the PFP is bilateral (on both knees).

In addition to the pain maps a corresponding dataset was available. This contained information regarding the individuals in terms of i.a. age, gender, symptom duration, pain intensity and the most prominent knee for pain. However, not all of the information was

present for all individuals. Before using the data in the deep learning models, a manual data handling was necessary. This incorporates matching the given pain maps and associated ID regarding the individuals, which resulted in 217 pain maps. Furthermore, specific information like gender, symptom duration and pain intensity are collected. The number of pain maps and associated gender and symptom duration, is 205. Additionally, there were 197 pain maps with associated gender and pain intensity.

**** MAKE A TABLE WITH THE NUMBER OF PAIN MAPS ****

3.1.1 Software application: Navigate Pain

Navigate Pain is a software application that is used to visualise the location, shape and spatial distribution of pain from patient to healthcare personnel. The application permits individuals to draw their pain into a body outline with different colors and line thickness. Navigate Pain android was developed at Aalborg University and a commercial web application is available at Aglance Solutions (Denmark).[45] Figure 3.2 illustrates the process using the application.

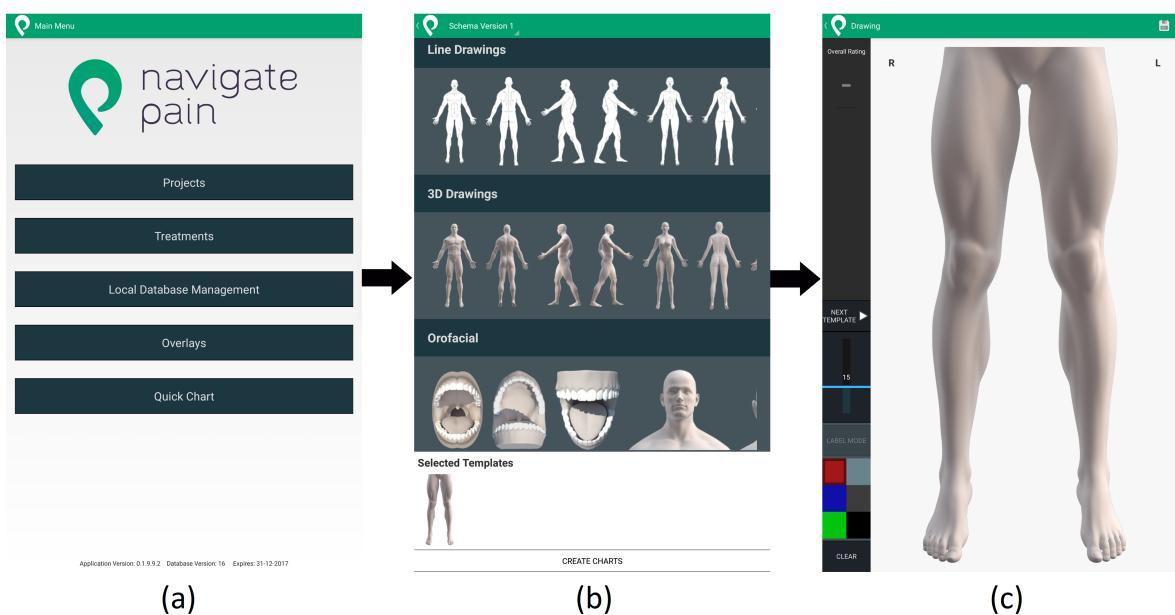


Figure 3.2: The process for making a pain map with Navigate Pain. (a) shows the main screen, (b) categories of body outlines and (c) body outline for lower extremities.

The left screen in figure 3.2(a) is the main screen. By clicking on "Project", a folder with individuals is created. From each individual information like name, age, height is saved. Before the individual can draw their pain areas, the body outline has to be chosen, which is illustrated in figure (b). The body outlines are divided into five categories: Line Drawings, 3D Drawings, Orofacial, Special Zooms and Knee Pain. In the bottom the selected templates are shown. When clicking on "CREATE CHARTS" the screen in figure (c) is shown. Here it is possible to draw the pain areas with different colors and line thickness, which can be seen in the left side of the screen. Afterwards the pain map can be saved.

3.1.2 Data representations

It is presumed that different representations of the pain maps affect the performance accuracy of a deep learning models, which is why different data representations are created. A study by Boudreau et al. found a correlation between a prolonged symptom duration and the size of the pain area. It was shown that the pain area increased for individuals that have a symptom duration for longer than five years compared to those with a symptom duration below. Likewise, pain intensity had a correlation with the size of pain area for individuals with a symptom duration for more than five years. Furthermore, the shape of the pain developed from a U-shape to an O-shape for individuals with a symptom duration above five years.[9] Based on this study the morphology of PFP is considered to be relevant to investigate, which is why morphology is the first data representation.

The PFP is often described as diffuse pain and it is therefore difficult to describe and localise [5]. To accommodate this is it chosen to divide the pain into different knee regions, which may indicate whether a specific region of the knee influence the PFP. This is converted to a simplified data representation that indicate active pain regions. A combination of the two data representations is created to achieve a third data representation which both include the morphology of the pain and the different active knee regions. Furthermore is gender an interesting parameter to use as an input, because the prevalence is more than twice as high for females than males. Thereto perceived pain is subjective and depends on the individual's character and personality. The distribution of gender is investigated by creating a histogram, which is shown in figure 3.3.

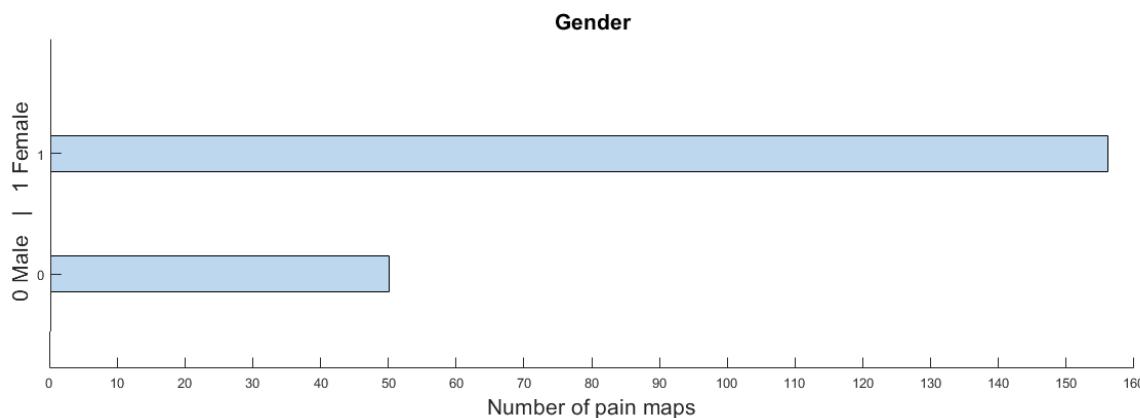


Figure 3.3: Histograms of the distribution of gender.

According to the given data the prevalence is higher for females than males. The females constitute 156 of the 206 individuals.

Since the symptom duration of PFP seems to affect the size and shape of the pain area, is it chosen to classify the three data representations in proportion to symptom duration. Likewise is it chosen to classify pain intensity because of the influence from the size of pain area. Thereto is it interesting to investigate whether there is a connection between knee regions and either symptom duration or pain intensity. The three data representations are referred to as morphology-, regions- and combined-representation.

3.2 Pre-analysis

The pain maps and associated symptom duration as well as pain intensity are analysed to get an overview of the data. The data is analysed in MatLab, where the distribution of the outputs, symptom duration and pain intensity, are investigated whereafter classification to the deep learning models are decided. Furthermore are different threshold values analysed according to five pain maps to select the threshold which should define when a region is active. Lastly, a reference to the deep learning models is created to investigate whether the size of pain areas can predict either symptom duration or pain intensity by using a simple linear regression.

3.2.1 Classification of data

The deep learning models should classify the input, pain maps and gender, in different categories in relation to symptom duration or pain intensity intervals. To find these intervals are histograms of the both symptom duration and pain intensity created.

A histogram of the symptom duration associated with the pain maps is illustrated in figure 3.4.

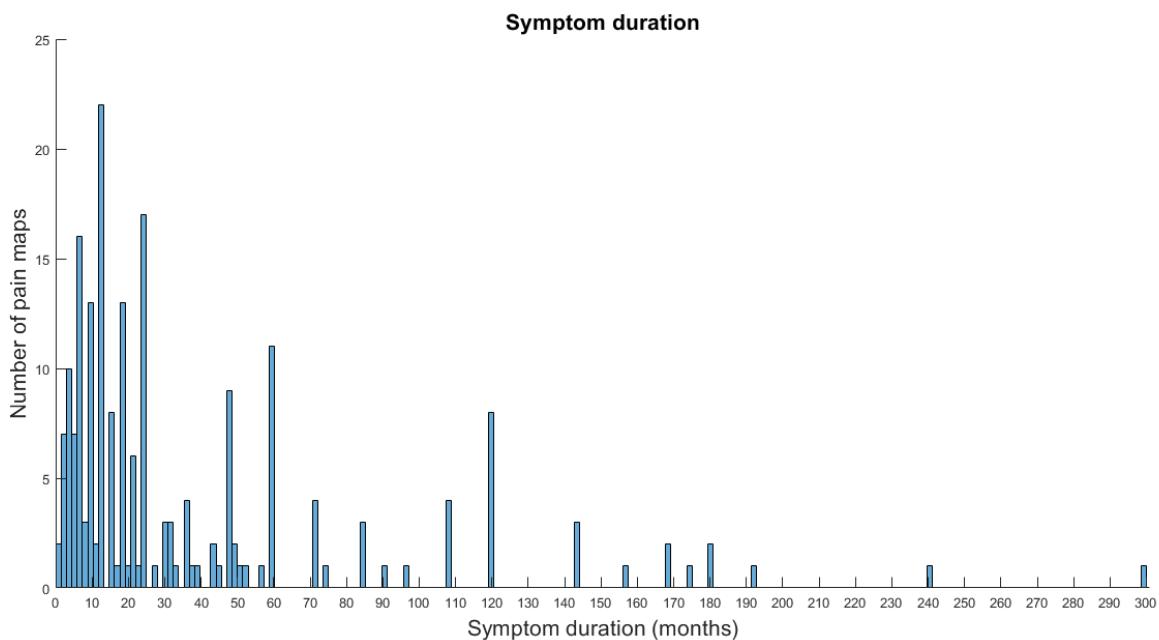


Figure 3.4: A histogram of the symptom duration.

The symptom duration is divided into some classes which the models should classify in addition to. To test the models are the data firstly divided into two extremes, since it is assumed that if the models predict badly with the extremes, the models will not predict better with multiple classifications of symptom duration. The histogram is evaluated and thereby are the extremes chosen to be 0 to 12 and 36 to 300 months. Afterwards is the interval between used as a third classification. In the appurtenant data to the pain map the individuals have stated their pain intensity as the worst pain in the last 24 hours and the last seven days. It is not assumably that the individuals have performed any PFP provoked activity in the last 24 hours before drawing their pain, therefore it is chosen to use the worst pain intensity in the last seven days to get a more average value for the worst pain intensity.

To explore the difference between the individuals' stated pain intensity in the last seven days is a histogram created which can be seen in figure 3.5.

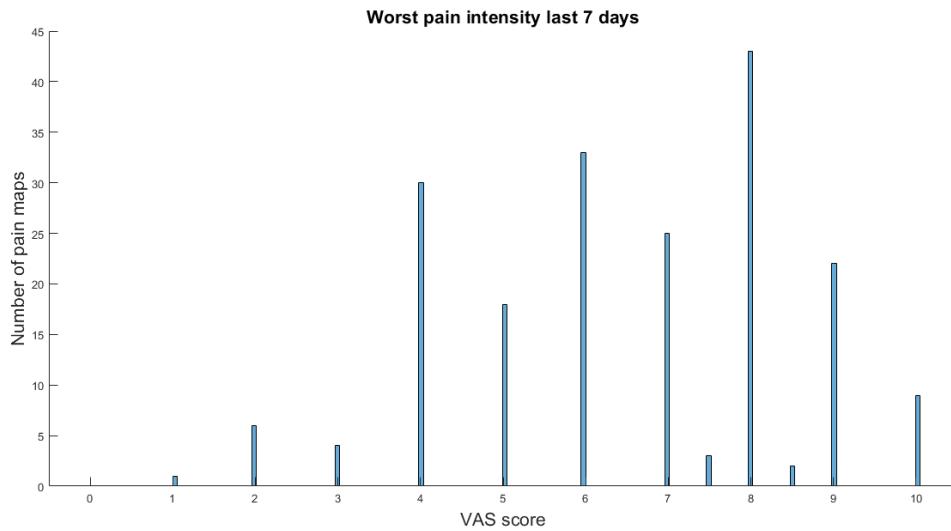


Figure 3.5: Histograms of the pain intensity the last seven days.

Likewise to the symptom duration classification, is the worst pain intensity divided into some into the extremes, which are chosen to be intervals 1 to 4 and 8 to 10. The third classification constitute the interval between the extremes.

3.2.2 Threshold selection

In relation to the data representation that contains information about the active pain regions, it is necessary to find a threshold that decides when a knee region contains enough pain pixels to be considered active. A threshold is required to increase the confidence of an active pain region by avoiding minimal contributions e.g. small pain areas in the associated regions. Simultaneously the threshold may not be too large so that potential pain regions will not be incorporated. The threshold to indicate active pain regions is decided based on an analysis, where threshold values of 0, 5, 10 and 15 percent are tested. The analysis of the threshold is tested on five random pain maps to get a general impression of the data. To better illustrate the division of the pain regions are the regions in figure 2.7 colored in different colors that are easier to distinguish, which is shown in figure 3.6.

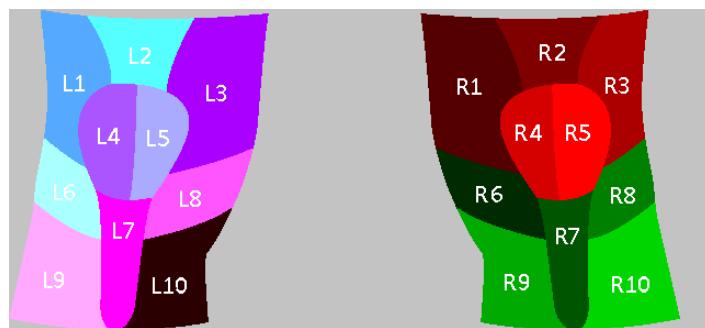


Figure 3.6: Knee regions colored in colors that are easier to distinguish.

An example of pain maps and appurtenant bar chart are illustrated in figure 3.7. The pain maps, figure 3.7(a-d), are likewise colored in the same colors as figure 3.6 to indicate which regions that are affected according to 0, 5, 10 and 15 percent threshold. The last figure (e) is the bar chart that indicates how many and which active regions there are according to the threshold values.

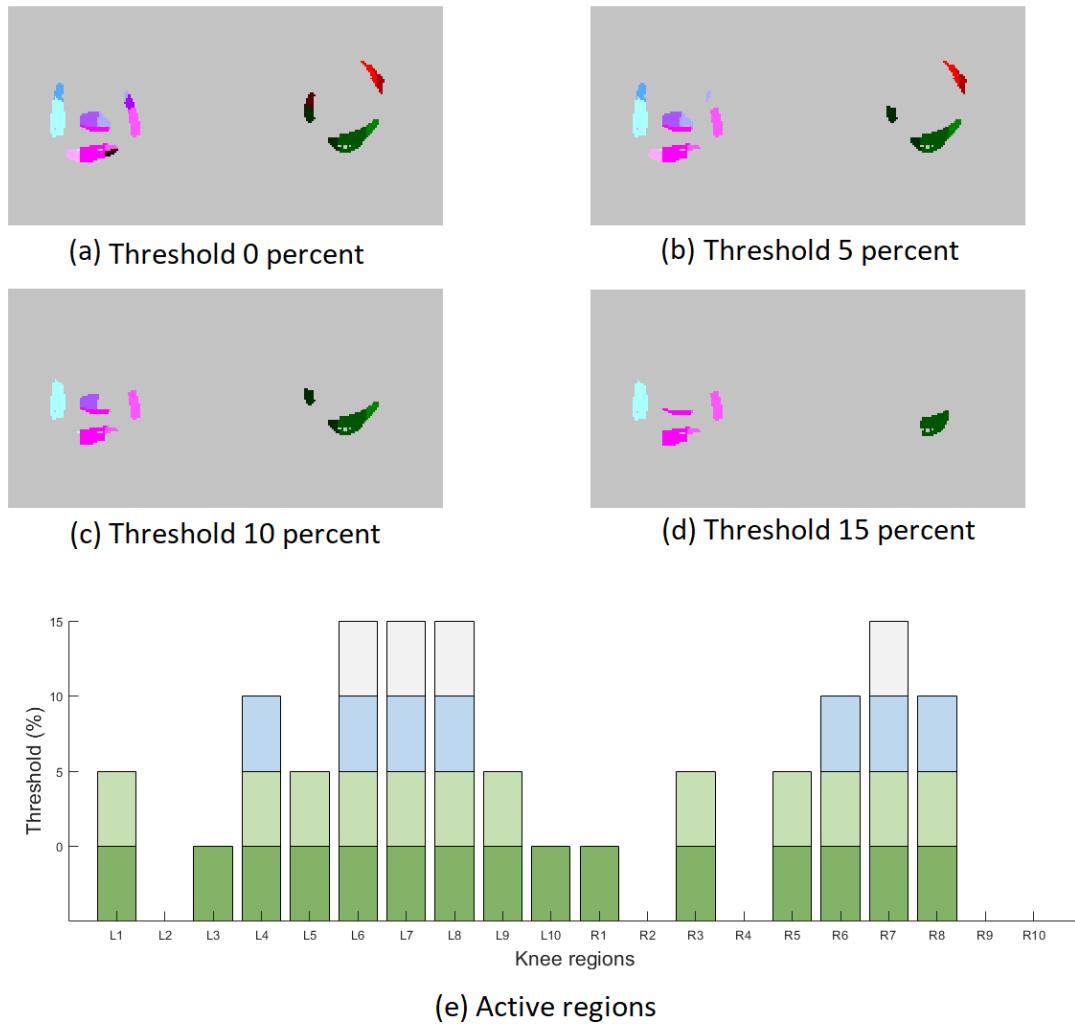


Figure 3.7: The active knee regions when the threshold is (a) 0 percent, (b) 5 percent, (c) 10 percent and (d) 15 percent. (e) is the bar chart that indicates how many and which knee regions that are considered active.

According to figure 3.7 (a) and (e) it is shown that the knee to the left has nine active regions and the knee to the right has six active knee regions, when the threshold is zero. In proportion to the active regions are region L3, L10 and R1 very small and are thereby the first regions to be discarded when the threshold is increased by 5 percent, which is shown in figure (b). By comparing figure (a) and (b) can minor changes according to the missing regions be seen, compared to figure (c) and (d) where greater areas disappears after increasing the threshold to 10 and 15 percent. Based on analysis of the five pain maps and bar charts, figure 3.7 and appendix B, is a threshold on 5 percent chosen to avoid including minor pain areas, like region L10, as active knee regions, and to avoid discarding too many and large areas, like regions R3 and R5.

3.2.3 Simple regression models

To explore whether there is a simple correlation between a single feature and either symptom duration or pain intensity, are plots created. In relation to the data representations, morphology and regions, is it chosen to investigate two single features in relation to the representations. The features are number of pain pixel and active pain regions. The two features are visualised according to each output, symptom duration and pain intensity. If these simple features have a correlation to the symptom duration or pain intensity, it may not be significant to investigate morphology and division of regions as features in the deep learning models. To test if there is a linear correlation between the number of pain pixel and symptom duration is a linear regression model made, which is shown in figure 3.8.

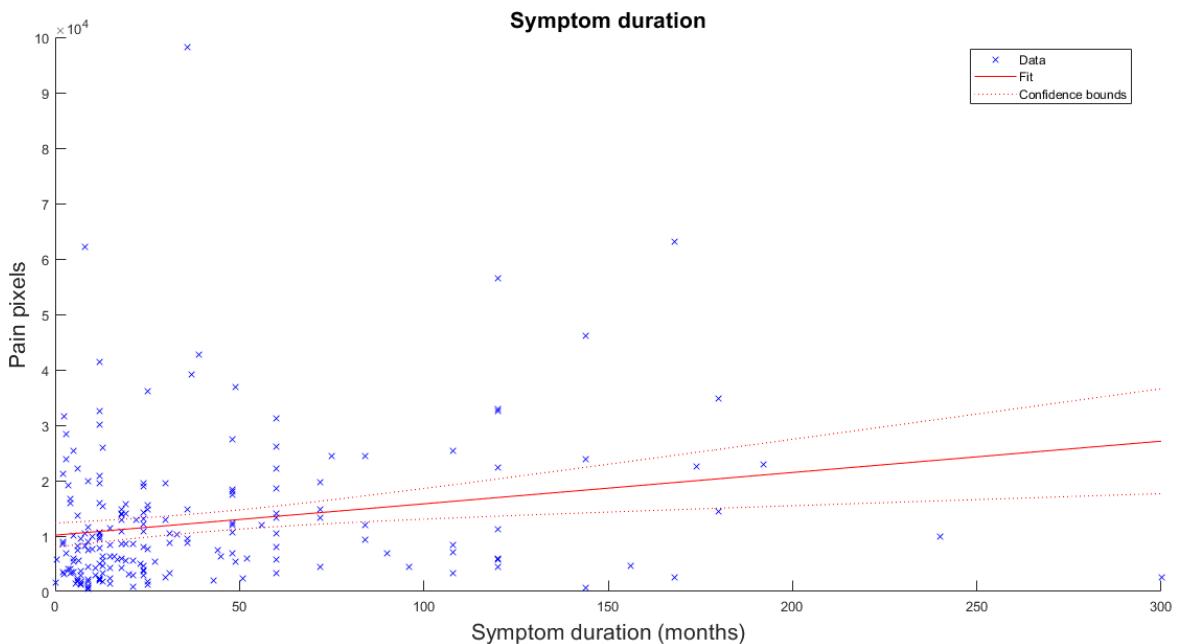


Figure 3.8: A linear regression of the symptom duration and the number of pain pixels.

According to the figure is it assumed that there is not a correlation between the symptom duration and number of pain pixels, since there is wide spread between in the data. As a result of the linear regression model is an R-squared value on 0.046, which indicate that there is not linear correlation, because the value is close to zero. A linear regression model of pain intensity and the size of pain area is also made, which is illustrated in figure 3.9.

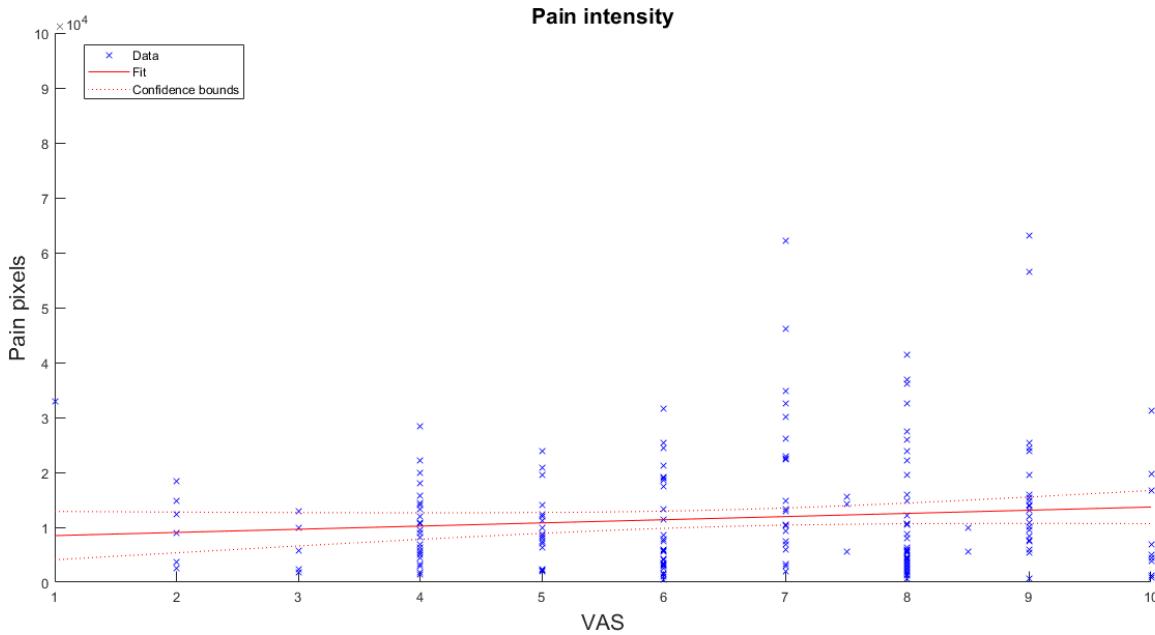


Figure 3.9: A linear regression fit of pain intensity stated in VAS and the size of pain area.

Figure 3.9 also indicates a wide spread in the data. The R-squared value of this model is 0.0117, which is very low, so there is no linear correlation between pain intensity and the number of pain pixels. These linear regression models are not very suitable when trying to predict symptom duration or pain intensity from the size of pain area. However, they can be compared to the performance of the deep learning models. Beside the linear regressions is the correlation between the number of active regions and either symptom duration or pain intensity investigated, which is shown in figure 3.10 and figure 3.11.

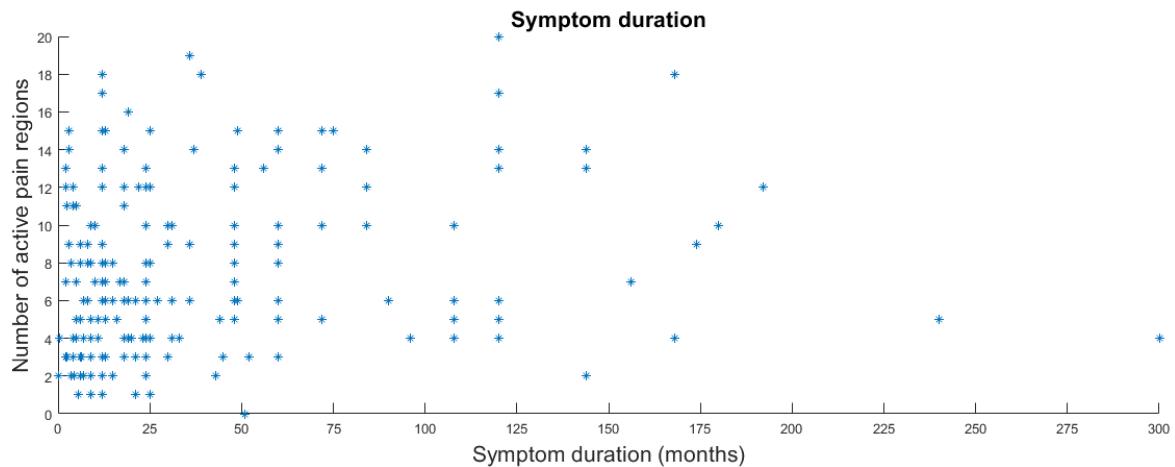


Figure 3.10: Plot of the correlation between the number of active pain regions and symptom duration.

In figure 3.10 is a dispersion of the data, which presumably means that there is not a correlation to be found between the number of regions and for how long the individuals have had PFP. Thereto is a figure of the correlation between the number of active pain regions and pain intensity shown in figure 3.11.

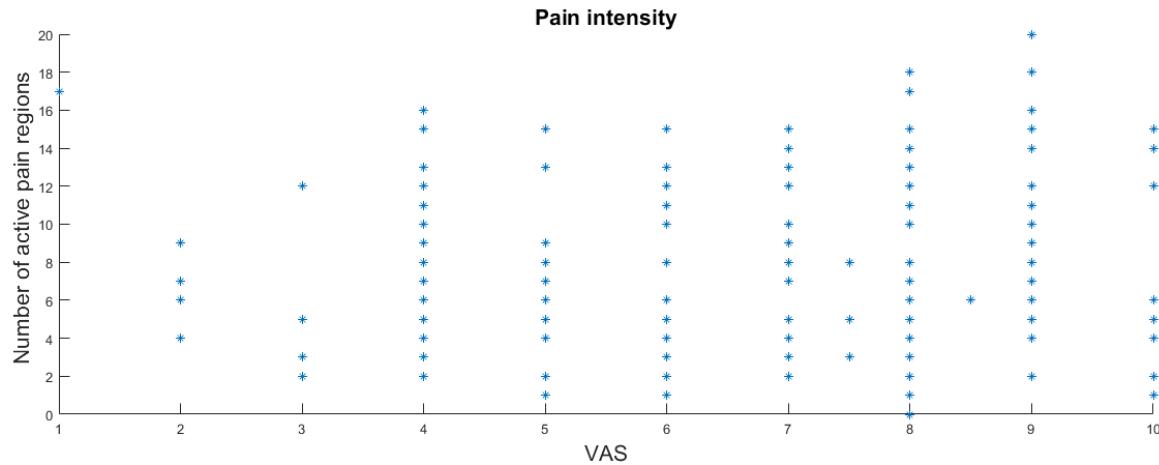


Figure 3.11: Plot of the correlation between the number active pain regions and pain intensity.

The figure shows once again a wide dispersion of the data. There is therefore no simple correlation between the number of active regions and pain intensity. Based on the four figures is it assumed that a single feature, number of pain pixels or regions, may not have a simple correlation with the outputs, symptom duration or pain intensity. Thereto is it interesting to investigate the morphology of pain and knee regions as features in a more complicated model.

3.3 Pre-processing

The data is pre-processed in MatLab to prepare it to the three different deep learning models. Each model has an appurtenant data representation which are prepared in three different ways. The three data representations are morphology, regions and combined, which are described in section 3.1.2. Common for the data representation is that the pain maps are imported as image-matrices whereafter the matrices are resized, since the given data was collected at different resolutions (screen sizes). Furthermore, the matrices are cropped to sort out unnecessary data like the areas inferior and superior to the knee. Before the data is used as an input in the deep learning models each matrix, which represent an image, is converted into a vector whereafter they are assembled in one matrix for each data representation. To get additional information associated with the pain maps, is gender added by including a column vector to the three matrices. In addition to the input, the deep learning models need an output to train the models. The output, which is either symptom duration or pain intensity, is likewise added as a column vector. The following sections describe the pre-processing of the individual data representations.

3.3.1 Morphology-representation

The first representation of data is a binary matrix of the original pain maps. Firstly, the image of the original pain map is gray-scaled to get a one-dimensional matrix instead of a

three-dimensional RGB-matrix. This matrix is then converted into a matrix consisting of zeroes and ones, where the pain pixels are symbolized with ones. An original pain map and a pain map consisting of a binary matrix is shown in figure 3.12.

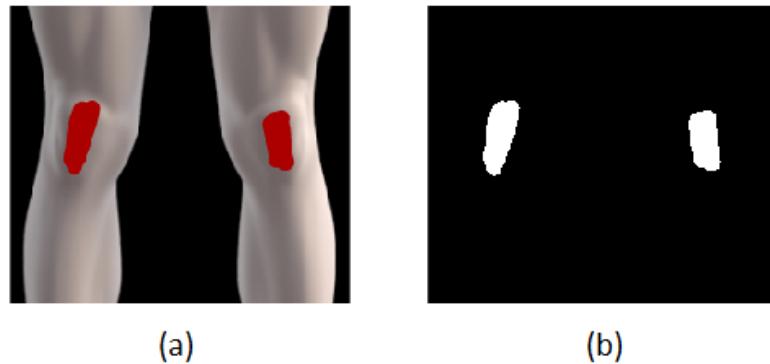


Figure 3.12: (a) Original pain map and (b) image consisting of a binary matrix where white color represents the pain pixels.

An illustration of this data representation is created to convey how the data is assembled and transferred to the model. The illustration is shown in figure 3.13, where a matrix containing image-vectors for all the pain maps and appurtenant gender and either symptom duration or pain intensity.

	Binary image-matrix												Gender Duration/ pain intensity
Image-vector 1	0	0	0	0	1	0	0	1	0	1	2		
Image-vector 2	1	0	0	1	0	1	1	0	1	0	0		
Image-vector 3	1	0	1	0	0	1	1	0	1	0	1		
Image-vector 4	0	1	1	1	0	1	1	1	0	1	0		
Image-vector 5	1	0	1	0	1	1	1	0	0	0	2		
Image-vector 6	0	0	0	1	0	0	1	1	1	1	1		
Image-vector ...	1	0	1	0	0	0	0	1	1	1	1		
Image-vector n	0	0	1	1	1	1	0	1	1	1	0		

Figure 3.13: An illustration of the matrix of the morphology-representation. The matrix consist of image-vectors for each individuals where the two last columns indicate the appurtenant gender (blue column vector) and either symptom duration or pain intensity (green column vector). The image-vectors have a length equal to the number of pixels in the pain maps.

3.3.2 Regions-representation

The second representation of the data is a matrix consisting of vectors with 20 values which indicate pain in relation to the knee regions. An image of the knee regions as shown in figure 2.7 are converted into a matrix consisting of 20 values, which represent each knee regions.

This matrix is superimposed to the binary image of the pain map, which results in a matrix with pain pixels represented in each knee region. In figure 3.14 are the knee regions and the pain associated with the regions illustrated.

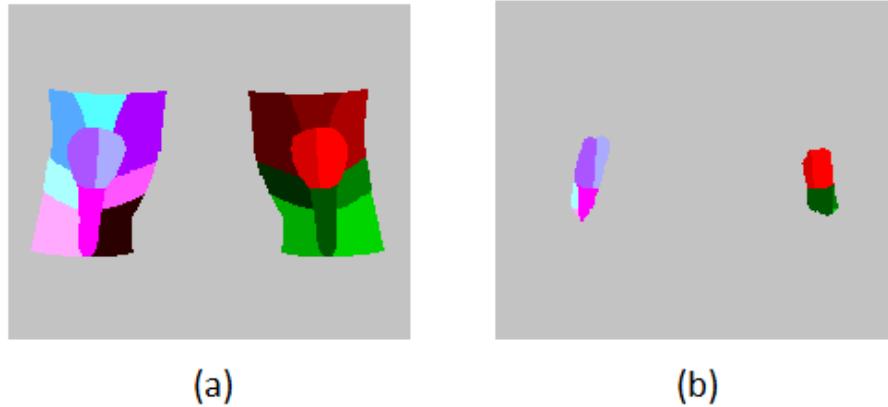


Figure 3.14: (a) Knee regions and (b) pain in the specific regions.

After superimposing the two matrices, knee regions and pain pixels, the number of pixels in each active knee region is found. This number is compared to the total number of pixels that are in each knee region, so knee regions with less than 5 percent pain are excluded. The threshold on 5 percent is chosen based on the analysis in section 3.2.2. As a result a vector with 20 values is created with zeros and ones, where one represent an active region. This data representation is implemented in the deep learning models the same way as the morphology-representation, which is illustrated in figure 3.13. The only difference is that the length of the image-vectors respond to the 20 regions, and therefore are there only 20 values in this data representation.

3.3.3 Combined-representation

The third representation of the data is a matrix consisting of individuals' pain divided into the knee regions. In this representation the superimposed matrix from the region-representation is used. Since the data representation should reflect the morphology of the pain and divide the pain into the different knee regions is one-hot encoding used. One-hot encoding is a way to separate categorical data into binary data [46]. This means that the 20 values for each knee region do not have a correlation. After one-hot encoding, the superimposed matrix consists of 20 layers where each layer represents a knee region. An illustration of this data representation is shown in figure 3.15.

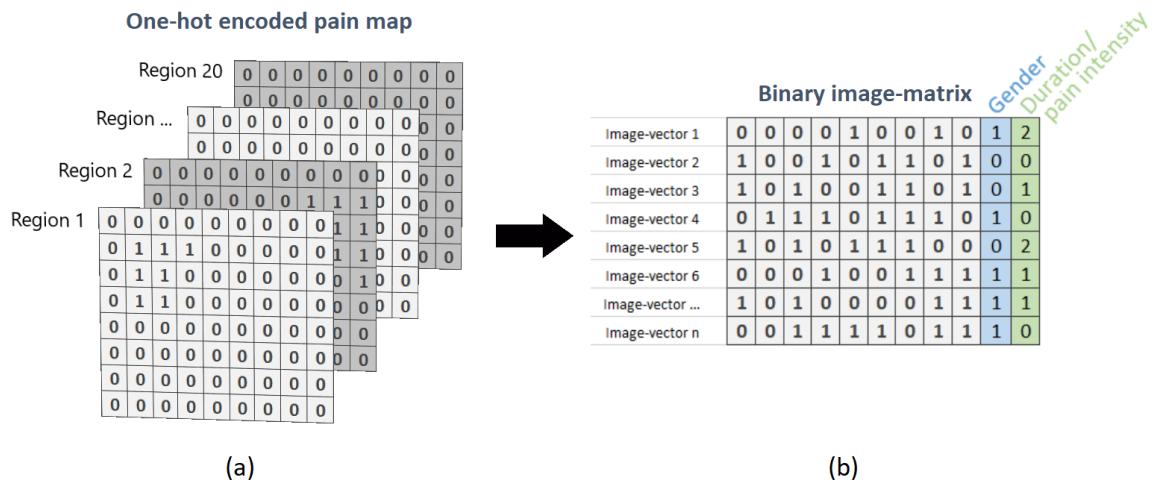


Figure 3.15: (a) An illustration of the one-hot encoded pain map and (b) shows the images-vectors in one assembled matrix with gender and either symptom duration or pain intensity.

3.4 Programs

In this project it is chosen to use Python v3.6.3 for development of the neural network. Python is an object-oriented and general-purpose programming and scripting language. Python is among other things used for programming websites, mobile applications, desktop GUI's, but also used for machine learning programming. When developing a machine learning application, there are different libraries that can be used, where some of the most popular is the Theano and the TensorFlow libraries.[47] In this project the TensorFlow v1.3.0 library has been used. TensorFlow is an open source library for development of machine learning applications, that has been released by Google [47].

maybe something about keras... if we are gonna use it. Keras is a high-level neural network library, that runs on top of either TensorFlow or Theano. Keras is a simplified version of the two libraries, which makes it easier to program in Python, but still allows for building complex models.[47]

3.5 Deep learning (but I would like to think another name for this section)

In this project the deep learning method is used to classify the pain maps and gender by determined outputs. The data is a set of 2 - dimensional images and the outputs are pain intensity and duration. For classification purpose supervised convolutional neural network is used followed by fully-connected layers at the end. This architecture is chosen to the interest of morphology and location of the pain. Model is trained and tested on a single computer with GPU and ran on Python programming language with Keras library and Tensorboard tool for visualizing results. Specifications of hardware and software are described in detail in section 4.2.

The architecture of model is presented in PICTURE(If nesceserry). From the higher point of view, the model can be architecturally splitted into two main parts. Feature extraction

part, where convolutional and pooling layers alternates in order to extract the most important features out of the image. Second part is called classification, where computed feature maps will be assigned to particular class of the output layer. In order to obtain a better accuracy of deep learning classifier, methods like activation function, regulation and optimization will be used within the layers.

Activation function is used to fit nonlinear patterns in the data and to control when the node in the model should be considered as an active or not. In this project "Rectified linear unit" (ReLU) activation function will be used as given since this is the most common for neural network (equation) [31].

Regulation is a method for restricting model from overfitting. The most common which is used in this project is called dropout. It randomly omits the fixed percentage amount of neurons in order to significantly improve the training accuracy. [Dropout]

Training the network requires to minimize the objective function (loss function) in order to get a better overall performance and the optimizer algorithm does that. Unfortunately, there are no easy way to know which one of optimizers fit the model best, so in this project stochastic gradient descent and Adam optimizers were tested and the one which performed better is used in the final model. [optimizers]

3.5.1 Software and hardware

The model was created and tested using Keras library, which is a high-level neural networks API (Application programming interface). Keras is written in Python and capable of running together with TensorFlow library. Using CPU or GPU on this platform the most mathematical operation is computed in a computerized and simple way.[Keras website] The computational part is written in Python script which defines all necessary operations for specific computations. A small example of the concept of using tensorflow could be seen in Appendix B. The plots for this model are generated using Tensorboard visualizing tool. It allows to watch a graph during the training, gives plot quantitative metrics and shows data like images that pass through it. Besides visualizing part of this tool it is also useful for debugging as it allows to observe the graph at a present time.[?]

The classifier was created on a machine with 4x "Intel® Core™ i7" CPU's and one single GPU of type "Geforce GTX 970M" with specifications listed in table 3.16.

CPU specifications	Value
Cores	4
Clock speed	2,6 GHz
Cache	6 MB SmartCache
Memory bandwidth	34,1 GB/s

GPU specifications	Value
CUDA cores	1280
Base Clock speed	924 MHz
Memory	6144 Mb
Memory bandwidth	120 GB/sec

Figure 3.16: Specifications of CPU's and GPU [website],[website].

Chapter 4

Description of models

4.1 Data handling and design choices of the models

This chapter presents how the pre-processed data is loaded and prepared for training of the models in Python, as well as the different architectures of the models, and features related to their implementation. Three separate models have been made fitting to each of the data representations described in section 3.3.

4.1.1 Data handling in python

The pre-processed data is loaded from a *.mat* file into python. Data within the file is pre-shuffled and split into a training and test subsets, as described in ???. In Python the training subset is further divided into a training and validation set. The training set is used to find the "optimal" weights with the backpropagation algorithm [32]. In this project, training set makes the biggest percentage amount of the data up to 75%. The validation set is used for estimating the generalization error of the model, and to examine how adjustments to hyperparameters effect the model performance [33]. Hyperparameters often define many different values that can be adjusted, to control the behavior of the algorithm, to which some parameters may affect the runtime and computational cost when training the model [31]. The validation set makes up 10% of the available data for the different models. The test subset will only be used to evaluate the generalization of the model, and therefore will not contain data that has been used during the training [33]. By keeping the test subset separate it will act as new unseen data, for the model. It contains 15% of the data.

For the binary, and combined pain map representation, the images were reshaped from a row vector, back into a matrix to retain their 2D structure.

Furthermore the classification labels were one-hot encoded, so they were compatible to the number of outputs in the models. This is a result of the loss function *categorical_crossentropy* used for the models, that tries to reduce loss between the categories. This demands that the class label has the same dimension as the number of outputs in the model [43].

4.1.2 Optimization techniques

To try to reduce overfitting and improve generalization of the neural network different techniques are applied network models.

One method used to automatically chose hyperparameters is used, and is known as grid search [31]. By using this method a listing of different hyperparameters were tested on a model, that allowed for choosing the parameter that give the highest improvements in performance in relation to the validation loss and accuracy.

Grid search is mainly practical when there are only a few hyperparameters that needs to be tested, because of the exponentially computational cost follows [31].

Optimizer used for the models

The Adam has been implemented in the three models, based on the findings of comparing it to the performance results with stochastic gradient descent optimizer. Adams optimizers showed a higher accuracy and significant higher training speed. Comparison results can be seen in ??

Kernel-initializer

Kernel-initialization was performed, since it affects the performance of the model. A grid search of kernel-initializer using *uniform*, *lecun_uniform*, *normal*, *glorot_normal* and *glorot_uniform* were tested and results revealed that *glorot_normal* performed the best by 3,2 % compared to the second best. Results can be seen in ??

Learning rate

In order to determine the most optimal *learning_rate* for the adam optimizer a several different values were tested. To fit the model this parameter was set to *0,0001* (lower than default value) meaning that the convergence of gradient descent is reached slower but with more accurate *minima*. It showed the

Batch training

Initial experimentation of the batch size showed little flexibility, because of the available computation power for training of the model. The batch size could not exceed a value of four, by reason of the relative high image resolution (912 x 2315). As a result of this it was chosen reduce the resolution (233 x 251). Resizing of the images, are described in 3.3.

Furthermore, the grid search of batch parameters were performed with values of 5, 10, 15, 20, and the highest accuracy with the value 5 were chosen

Cross-validation

Because the amount of data for this project is limited it is chosen to implement m-fold cross validation, where the training data is divided into m number of subsets. Each of the subsets can function a either a validation set or as a part of a training set e.g. if a classifier is trained m times, then each time a different subset will be used as a validation set, and the rest is used for training. [33] Because of the property of cross validation, it can be used as a way of investigate a general accuracy since all data is included during training, but may not be beneficial for every kind of problem. [33]

Dropout

Dropout is implemented in the models due to the benefits described in section 2.6.3. The dropout is only used within the hidden fully-connected layers, where it's defined to randomly drop 50 % of the nodes.

Activation function

ReLU activation function is used as controlling unit for nodes activation. It defines when the node is considered as active. In the model ReLU is implemented in convolutional and fully-connected layers.

Sigmoid is implemented only in the output layer, because it

It has something to with the fact that the output layer performs the classification and needs sigmoid is more sutible for this task then ReLu...

4.1.3 Training of the networks

Supervised learning is used for training in all the models. The generic input for all of the model is gender, along with the different image representation, which are presented in /refto data. These inputs trained and then compared against their respective category label. WE CAN MENTION HERE THE TYPES OF DATA AND TYPES OF MODELS IF THEY ARE NOT PRESENTED LATER

Temp-placeholder: The process of making the neural network model has been a trial and error process, because there is not an actual “cookbook” for developing NN (This statement is from a not vaild source, but so far it's the only one that i have found.) [31]

4.2 Vector image model

The architecture for this model only contains fully connected layer, since the data representation only contains 21 element vector that reflects the active pain regions and gender as described in ???. It is evaluated that there would not be any gains in making the model more complex e.g. adding of convolution, based on the information available from vector, since the level of detail in relation to morphology is very simple. The architecture of the model is illustrated in Figure 4.1.

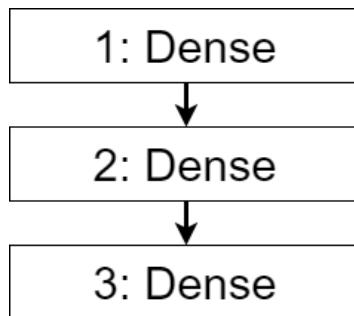


Figure 4.1: Arcitechture of the neural network using knee region representation.

The model consists of three layers, where the input and output layer..(WE ADD MORE)

4.3 Raw binary representation model

The architecture of this model is based on the typical structure of a convolutional network, where the first layers alternate between convolutional layers and max pooling layers [11]. This defines the first part of the model. The following layers consists of three fully connected layers, and output layer, and defined the second part of the model. An overview of the architecture

is shown in Figure 4.2. Convolution layer are implemented for this pain map representation, because of their ability to extract morphology features from images, as written in ??

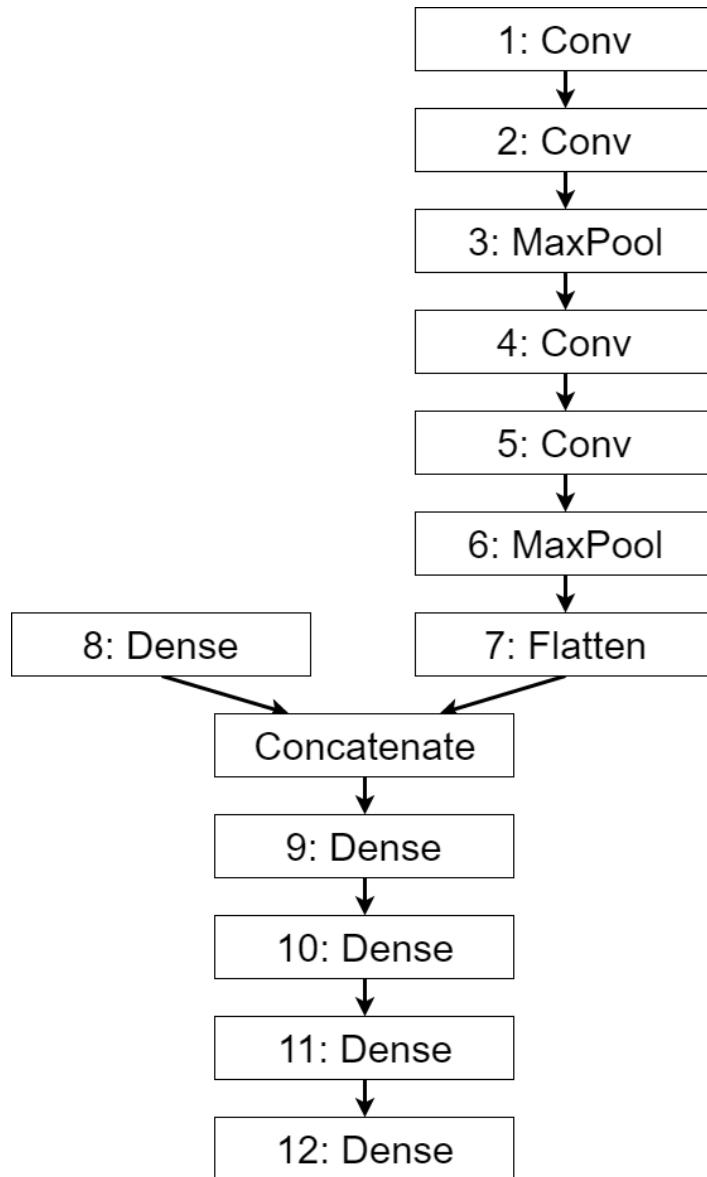


Figure 4.2: Architecture for the neural network model using binary pain map representation.
REMEMBER TO INCLUDE THE INPUT IN THE FIGURE

Rather than feeding both pain map representation and gender into the same input layer, they are separated and used as input at two different locations. The binary pain map representation works as input in first part of the model, where the input layer is a convolution layer. This layer is setup to receive a input shape to that of the dimension of the pain map, that is defined during re-scaleing of the pain maps in ?. Gender works as secondary input in the second section of the model, along with the pain maps features extracted through the convolution layers. Before the pain maps features reach the fully connected part of the network it is flattened from a matrix to a single row in order to merge the features with gender. The merged data passes the fully connected layers and reaches the output layer where it is given a percentage value according to which class it fits the most. The second part of the model

resembles the simple representation model, described in 4.2

THIS NEEDS TO BE REWRITTEN: The reason for separating gender and binary images is given as separate inputs is because of that there is no benefit in feeding gender through several convolutional layers, since these layer are use for looking at the shapes of the pain. The reason for using gender as input this far into the model, is a result of the way that convolution works

4.4 Combined representation model

The architecture of this model is nearly identical to that of the binary representation model as described in 4.3. The main difference can only be seen in the input layer for the pain map representation, where the input shape is altered to contain 20 layers per pain maps instead of one. This is the result of the one hot encoding done to the images as described in subsection 3.3.3.

Chapter 5

Results

Bibliography

- [1] Liam R. MacLachlan, Natalie J. Collins, and Et.al. The psychological features of patellofemoral pain: a systematic review. 2017. doi: 10.1136/bjsports-2016-096705.
- [2] T.O. Smith, B.T. Drew, and Et.al. Knee orthoses for treating patellofemoral pain syndrome (review). 2015. doi: 10.1002/14651858.CD010513.pub2.
- [3] M. S. Rathleff, B. Vicenzino, and Et.al. Patellofemoral Pain in Adolescents and Adulthood: same same, but different? 2015. doi: 10.1007/s40279-015-0364-1.
- [4] Wolf Petersen, Andree Ellermann, and Et.al. Patellofemoral Pain Syndrome. *Clinical Orthopaedics and Related Research*, 2013. doi: 10.1097/01.blo.0000229284.45485.6c.
- [5] Erik Witvrouw, Michael J. Callaghan, and Et.al. Patellofemoral Pain: consensus statement from the 3rd International Patellofemoral Pain Research Retreat held in Vancouver, September 2013. 2014. doi: 10.1136/bjsports-2014-093450.
- [6] Kay M. Crossley, Michael J. Callaghan, and Et.al. Patellofemoral pain. 2016. doi: 10.1136/bjsports-2015-h3939rep.
- [7] Scott F Dye. Patellofemoral Pain Current Concepts: An Overview. *Sports Medicine and Arthroscopy Review*, 2001.
- [8] Shellie A. Boudreau and Susanne et.al Badsberg. Digital pain drawings: Assessing Touch-Screen Technology and 3D Body Schemas. 2016. doi: 10.1097/AJP.0000000000000230.
- [9] Shellie A. Boudreau, E. N. Kamavuako, and Et.al. Distribution and symmetrical patellofemoral pain patterns as revealed by high-resolution 3D body mapping: a cross-sectional study. 2017. doi: 10.1186/s12891-017-1521-5.
- [10] Mathias Haefeli and Achim Elfering. Pain assessment. 2005. doi: 10.1007/s00586-005-1044-x.
- [11] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep Learning. *Nature Insight Review*, pages 436–444, 2015. doi: 10.1038/nature14539. URL <https://www.nature.com/nature/journal/v521/n7553/pdf/nature14539.pdf>.
- [12] Frederic H. et. al. Martini. *Anatomy & Physiology*. 2012.
- [13] IASP. IASP Taxonomy, 2012.
- [14] R. F Schmidt. Nociception and Pain. In *Human Physiology*. Springer Berlin Heidelberg, 1989. ISBN 978-3-642-73831-9. doi: 10.1007/978-3-642-73831-9_10.
- [15] R. F Schmidt. Nociception and Pain. In *Fundamentals of Sensory Physiology*. Springer Science & Business Media, 2 edition, 2013. ISBN 9783662011287.

- [16] Emma Briggs. Understanding the experience and physiology of pain. 2010.
- [17] Kay M. Crossley, Michael J. Callaghan, and Et.al. Patellofemoral pain. 2015. doi: 10.1136/bmj.h3939.
- [18] Katherine Stabenow Dahab. Q angle. *Encyclopedia of Sports Medicine*, 2011. doi: 10.4135/9781412961165.n415.
- [19] Jarred Younger and Sean Mackey. Pain outcomes: A brief review of instruments and techniques. 2009.
- [20] Kanishka M. Ghosh and David J. Deehan. Soft tissue knee injuries. In Mark Wilkinson, editor, *Orthopaedic surgery: lower limb*. Elsevier, 2010.
- [21] Maria Ferreira Valente, José Ribeiro Pais, and Et. Al. Validity of four pain intensity rating scales. 2011. doi: 10.1016/j.pain.2011.07.005.
- [22] Ewa M. Roos and L. Stefan Lohmander. The knee injury and osteoarthritis outcome score (KOOS): from joint injury to osteoarthritis. *Bio Med Central*, 2003.
- [23] Marie Grunnesjö. The course of pain drawings during a 10-week treatment period in patients with acute and sub-acute low back pain. 2006. doi: 10.1186/1471-2474-7-65.
- [24] Geoffrey D. Schott. The cartography of pain: The evolving contribution of pain maps. 2010. doi: 10.1016/j.ejpain.2009.12.005.
- [25] D. W. Elson, S. Jones, and Et.al. The photographic knee pain map: Locating knee pain with an instrument developed for diagnostic, communication and research purposes. 2010. doi: 10.1016/j.knee.2010.08.012.
- [26] Mads Nielsen. *Den digitale revolution – fortællinger fra datalogiens verden*. Datalogisk Institut, 2010. ISBN 978-87-981270-5-5.
- [27] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015. ISSN 0036-8075. doi: 10.1126/science.aaa8415. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.aaa8415>.
- [28] Jürgen Schmidhuber. Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61:85–117, 2015. ISSN 18792782. doi: 10.1016/j.neunet.2014.09.003. URL <http://www.sciencedirect.com/science/article/pii/S0893608014002135>.
- [29] Jacopo Acquarelli, Twan van Laarhoven, Jan Gerretzen, Thanh N. Tran, Lutgarde M.C. Buydens, and Elena Marchiori. Convolutional neural networks for vibrational spectroscopic data analysis. *Analytica Chimica Acta*, 954:22–31, 2017. ISSN 18734324. doi: 10.1016/j.aca.2016.12.010. URL <http://linkinghub.elsevier.com/retrieve/pii/S0003267016314842>.
- [30] Mohri Mehryar, Rostamizadeh Afshin, and Talwalka Ameet. *Foundations of Machine Learning*. 2012. ISBN 9780262018258. URL <https://ebookcentral.proquest.com/lib/aalborguniv-ebooks/reader.action?docID=3339482{&}ppg=17{#}>.

- [31] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
URL <http://www.deeplearningbook.org>.
- [32] Yoshua Bengio. *Neural Networks: Tricks of the Trade: Second Edition*. 2012. ISBN 978-3-642-35289-8. doi: 10.1007/978-3-642-35289-8_26. URL http://dx.doi.org/10.1007/978-3-642-35289-8{_}26.
- [33] Richard Duda, Peter Hart, and David Stork. *Pattern Classification*. Second edi edition, 2000. ISBN 9780471056690.
- [34] Alaa Ali Hameed, Bekir Karlik, and Mohammad Shukri Salman. Back-propagation algorithm with variable adaptive momentum. *Knowledge-Based Systems*, 114:79–87, 2016. ISSN 09507051. doi: 10.1016/j.knosys.2016.10.001. URL <http://www.sciencedirect.com/science/article/pii/S0950705116303811?via%203Dihub>.
- [35] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2323, 1998. ISSN 00189219. doi: 10.1109/5.726791. URL <http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf>.
- [36] Sebastian Ruder. An overview of gradient descent optimization algorithms. 2016. ISSN 0006341X. doi: 10.1111/j.0006-341X.1999.00591.x. URL <http://arxiv.org/abs/1609.04747>.
- [37] Sebastian Raschka. Single-Layer Neural Networks and Gradient Descent. 2016. URL http://sebastianraschka.com/Articles/2015_singlelayer_neurons.html.
- [38] Sixin Zhang, Anna Choromanska, and Yann LeCun. Deep learning with Elastic Averaging SGD. 2014. ISSN 10495258. URL <http://arxiv.org/abs/1412.6651>.
- [39] Ning Qian. On the Momentum Term in Gradient Descent Learning Algorithms The Momentum Term in Gradient Descent. *Neural Networks: The Official Journal of the International Neural Network Society*, 5213(12(1)):145–151, 1999. URL [#aep-bibliography-id16](https://www.sciencedirect.com/science/article/pii/S0893608098001166).
- [40] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*, 2015. ISSN 09252312. doi: <http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503>. URL <http://arxiv.org/abs/1412.6980>.
- [41] M. Patacchiola and A. Cangelosi. Head pose estimation in the wild using Convolutional Neural Networks and adaptive gradient methods. *Pattern Recognition*, 71, 2017. ISSN 00313203. doi: 10.1016/j.patcog.2017.06.009. URL https://ac.els-cdn.com/S0031320317302327/1-s2.0-S0031320317302327-main.pdf?{_}tid=a09e6e8a-d356-11e7-9b28-00000aacb360&{#}acdnat=1511775647{_}f91c81864b46ddf0becfc12de21bce48.
- [42] Int8. Optimization techniques comparison in Julia: SGD, Momentum, Adagrad, Adadelta, Adam. *Www*, 2016. URL <http://int8.io/comparison-of-optimization-techniques-stochastic-gradient-descent-momentum-adagrad-and-adadelta>

- [43] François Chollet. Keras Documentation, 2015. URL <https://keras.io>.
- [44] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. ISSN 15337928. doi: 10.1214/12-AOS1000. URL <https://dl.acm.org/citation.cfm?id=2670313&CFID=818407627&CFTOKEN=74532044>.
- [45] Aglance Solutions. Visual insight for clinical reasoning – Navigate Pain, 2015. URL <http://www.navigatepain.com/>.
- [46] David Money Harris and Sarah L. Harris. Sequential Logic Design. In *Digital design and computer architecture*. Elsevier, 2012. ISBN 9780123978165.
- [47] Manohar Swamynathan. *Mastering Machine Learning with Python in Six Steps*. 2017. ISBN 978-1-4842-2865-4. doi: 10.1007/978-1-4842-2866-1. URL <http://link.springer.com/10.1007/978-1-4842-2866-1>.

Appendix A

Knee injury and Osteoarthritis Outcome Score (KOOS) [?]

KOOS KNEE SURVEY

Today's date: ____ / ____ / ____ Date of birth: ____ / ____ / ____

Name: _____

INSTRUCTIONS: This survey asks for your view about your knee. This information will help us keep track of how you feel about your knee and how well you are able to perform your usual activities.

Answer every question by ticking the appropriate box, only one box for each question. If you are unsure about how to answer a question, please give the best answer you can.

Symptoms

These questions should be answered thinking of your knee symptoms during the **last week**.

S1. Do you have swelling in your knee?

Never	Rarely	Sometimes	Often	Always
<input type="checkbox"/>				

S2. Do you feel grinding, hear clicking or any other type of noise when your knee moves?

Never	Rarely	Sometimes	Often	Always
<input type="checkbox"/>				

S3. Does your knee catch or hang up when moving?

Never	Rarely	Sometimes	Often	Always
<input type="checkbox"/>				

S4. Can you straighten your knee fully?

Always	Often	Sometimes	Rarely	Never
<input type="checkbox"/>				

S5. Can you bend your knee fully?

Always	Often	Sometimes	Rarely	Never
<input type="checkbox"/>				

Stiffness

The following questions concern the amount of joint stiffness you have experienced during the **last week** in your knee. Stiffness is a sensation of restriction or slowness in the ease with which you move your knee joint.

S6. How severe is your knee joint stiffness after first wakening in the morning?

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

S7. How severe is your knee stiffness after sitting, lying or resting **later in the day**?

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

Knee injury and Osteoarthritis Outcome Score (KOOS), English version LK1.0

Pain

P1. How often do you experience knee pain?

Never <input type="checkbox"/>	Monthly <input type="checkbox"/>	Weekly <input type="checkbox"/>	Daily <input type="checkbox"/>	Always <input type="checkbox"/>
-----------------------------------	-------------------------------------	------------------------------------	-----------------------------------	------------------------------------

What amount of knee pain have you experienced the **last week** during the following activities?

P2. Twisting/pivoting on your knee

None <input type="checkbox"/>	Mild <input type="checkbox"/>	Moderate <input type="checkbox"/>	Severe <input type="checkbox"/>	Extreme <input type="checkbox"/>
----------------------------------	----------------------------------	--------------------------------------	------------------------------------	-------------------------------------

P3. Straightening knee fully

None <input type="checkbox"/>	Mild <input type="checkbox"/>	Moderate <input type="checkbox"/>	Severe <input type="checkbox"/>	Extreme <input type="checkbox"/>
----------------------------------	----------------------------------	--------------------------------------	------------------------------------	-------------------------------------

P4. Bending knee fully

None <input type="checkbox"/>	Mild <input type="checkbox"/>	Moderate <input type="checkbox"/>	Severe <input type="checkbox"/>	Extreme <input type="checkbox"/>
----------------------------------	----------------------------------	--------------------------------------	------------------------------------	-------------------------------------

P5. Walking on flat surface

None <input type="checkbox"/>	Mild <input type="checkbox"/>	Moderate <input type="checkbox"/>	Severe <input type="checkbox"/>	Extreme <input type="checkbox"/>
----------------------------------	----------------------------------	--------------------------------------	------------------------------------	-------------------------------------

P6. Going up or down stairs

None <input type="checkbox"/>	Mild <input type="checkbox"/>	Moderate <input type="checkbox"/>	Severe <input type="checkbox"/>	Extreme <input type="checkbox"/>
----------------------------------	----------------------------------	--------------------------------------	------------------------------------	-------------------------------------

P7. At night while in bed

None <input type="checkbox"/>	Mild <input type="checkbox"/>	Moderate <input type="checkbox"/>	Severe <input type="checkbox"/>	Extreme <input type="checkbox"/>
----------------------------------	----------------------------------	--------------------------------------	------------------------------------	-------------------------------------

P8. Sitting or lying

None <input type="checkbox"/>	Mild <input type="checkbox"/>	Moderate <input type="checkbox"/>	Severe <input type="checkbox"/>	Extreme <input type="checkbox"/>
----------------------------------	----------------------------------	--------------------------------------	------------------------------------	-------------------------------------

P9. Standing upright

None <input type="checkbox"/>	Mild <input type="checkbox"/>	Moderate <input type="checkbox"/>	Severe <input type="checkbox"/>	Extreme <input type="checkbox"/>
----------------------------------	----------------------------------	--------------------------------------	------------------------------------	-------------------------------------

Function, daily living

The following questions concern your physical function. By this we mean your ability to move around and to look after yourself. For each of the following activities please indicate the degree of difficulty you have experienced in the **last week** due to your knee.

A1. Descending stairs

None <input type="checkbox"/>	Mild <input type="checkbox"/>	Moderate <input type="checkbox"/>	Severe <input type="checkbox"/>	Extreme <input type="checkbox"/>
----------------------------------	----------------------------------	--------------------------------------	------------------------------------	-------------------------------------

A2. Ascending stairs

None <input type="checkbox"/>	Mild <input type="checkbox"/>	Moderate <input type="checkbox"/>	Severe <input type="checkbox"/>	Extreme <input type="checkbox"/>
----------------------------------	----------------------------------	--------------------------------------	------------------------------------	-------------------------------------

Knee injury and Osteoarthritis Outcome Score (KOOS), English version LK1.0

For each of the following activities please indicate the degree of difficulty you have experienced in the **last week** due to your knee.

A3. Rising from sitting

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A4. Standing

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A5. Bending to floor/pick up an object

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A6. Walking on flat surface

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A7. Getting in/out of car

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A8. Going shopping

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A9. Putting on socks/stockings

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A10. Rising from bed

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A11. Taking off socks/stockings

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A12. Lying in bed (turning over, maintaining knee position)

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A13. Getting in/out of bath

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A14. Sitting

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A15. Getting on/off toilet

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

Knee injury and Osteoarthritis Outcome Score (KOOS), English version LK1.0

For each of the following activities please indicate the degree of difficulty you have experienced in the **last week** due to your knee.

A16. Heavy domestic duties (moving heavy boxes, scrubbing floors, etc)

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

A17. Light domestic duties (cooking, dusting, etc)

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

Function, sports and recreational activities

The following questions concern your physical function when being active on a higher level. The questions should be answered thinking of what degree of difficulty you have experienced during the **last week** due to your knee.

SP1. Squatting

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

SP2. Running

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

SP3. Jumping

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

SP4. Twisting/pivoting on your injured knee

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

SP5. Kneeling

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

Quality of Life

Q1. How often are you aware of your knee problem?

Never	Monthly	Weekly	Daily	Constantly
<input type="checkbox"/>				

Q2. Have you modified your life style to avoid potentially damaging activities to your knee?

Not at all	Mildly	Moderately	Severely	Totally
<input type="checkbox"/>				

Q3. How much are you troubled with lack of confidence in your knee?

Not at all	Mildly	Moderately	Severely	Extremely
<input type="checkbox"/>				

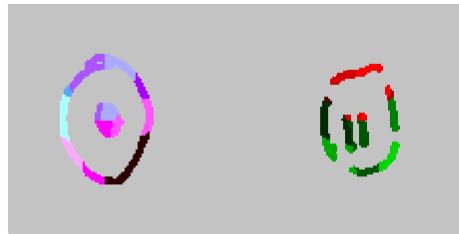
Q4. In general, how much difficulty do you have with your knee?

None	Mild	Moderate	Severe	Extreme
<input type="checkbox"/>				

Thank you very much for completing all the questions in this questionnaire.

Appendix B

Threshold analysis



(a) Threshold 0 percent



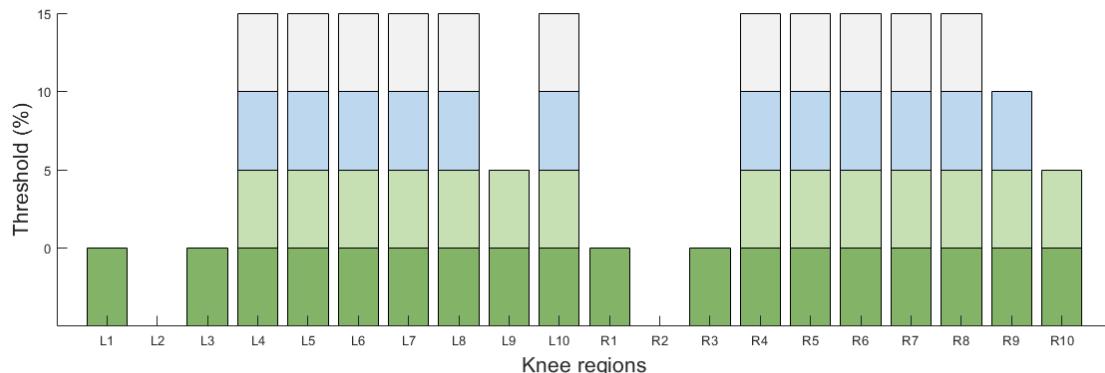
(b) Threshold 5 percent



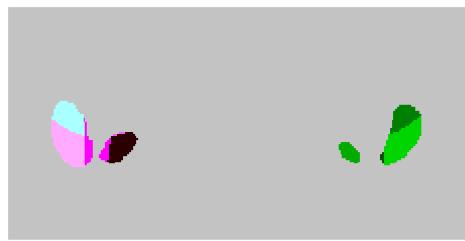
(c) Threshold 10 percent



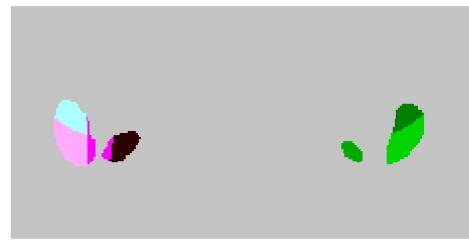
(d) Threshold 15 percent



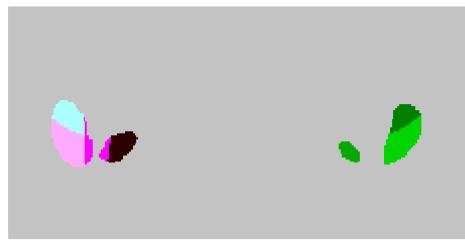
(e) Active regions



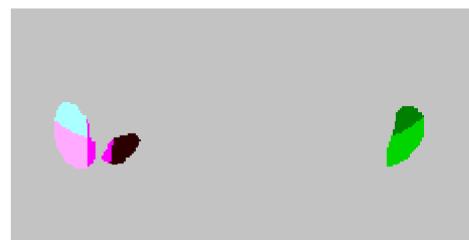
(a) Threshold 0 percent



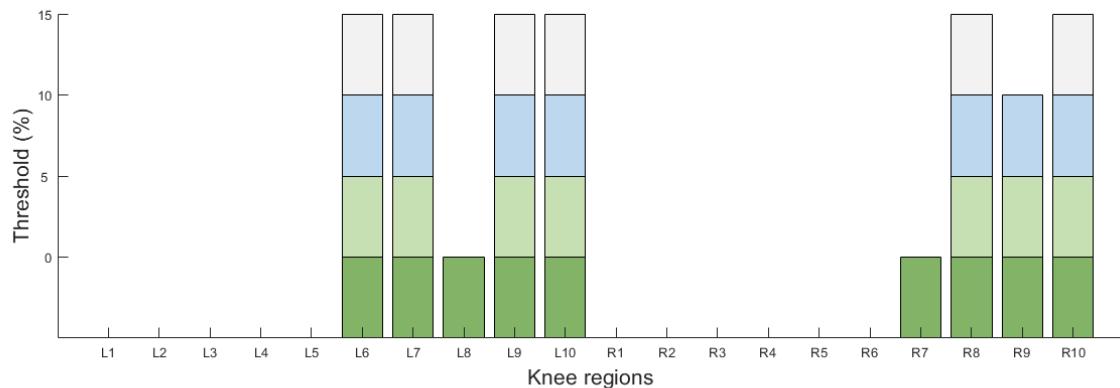
(b) Threshold 5 percent



(c) Threshold 10 percent



(d) Threshold 15 percent



(e) Active regions



(a) Threshold 0 percent



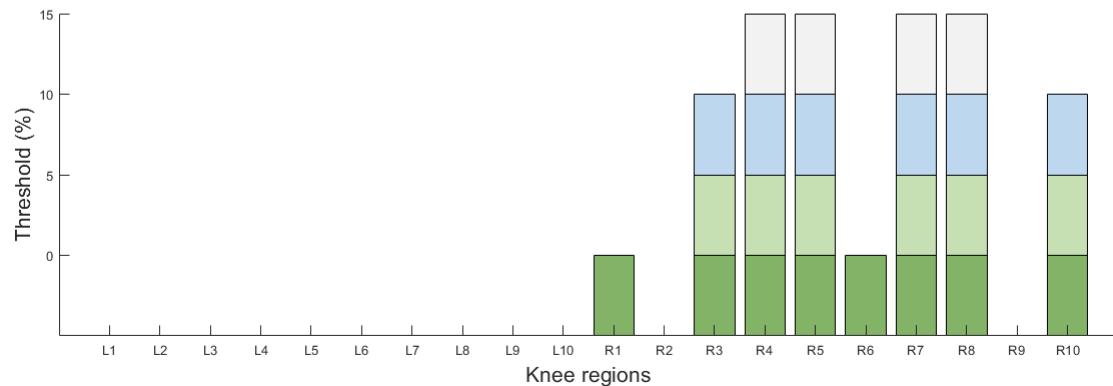
(b) Threshold 5 percent



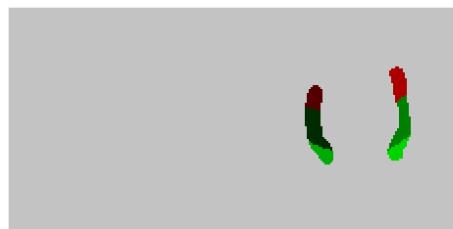
(c) Threshold 10 percent



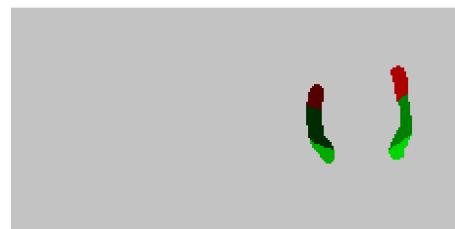
(d) Threshold 15 percent



(e) Active regions



(a) Threshold 0 percent



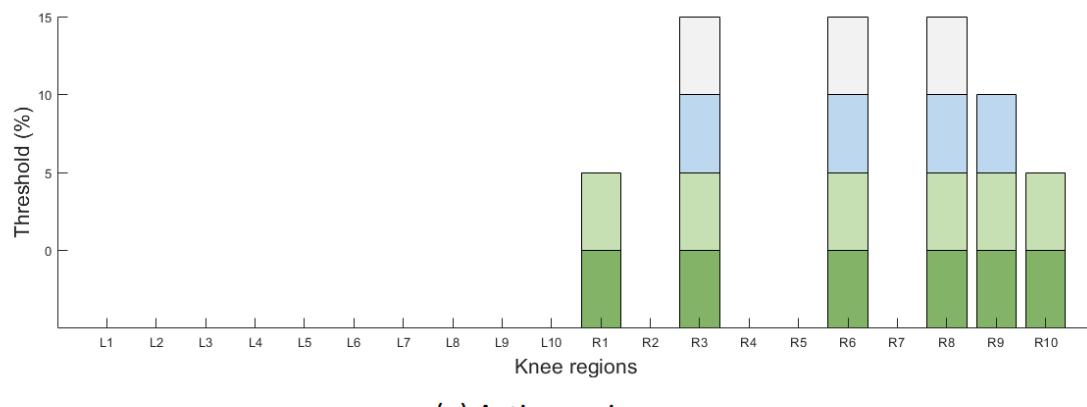
(b) Threshold 5 percent



(c) Threshold 10 percent



(d) Threshold 15 percent



(e) Active regions