Prediction of pain duration and intensity from patellofemoral pain maps using deep learning

Birgithe Kleemann Rasmussen, Ignas Kupcikevičius,
Linette Helena Poulsen, Mads Kristensen

Aalborg University

December 20, 2017

Abstract

Introduction: *Patellofemoral pain (PFP) syndrome is a musculoskeletal condition that presents as pain behind or around the patella without known structural changes [1]. Partial correlations between perceived size of PFP from pain maps and pain duration along with intensity has been indicated in previous studies [2], however morphology and location of PFP remains unexplored in terms of correlation. Based on the objects detection capabilities of deep learning, convolution methods can be used to detect image-features related to morphology. The aim of this study was to determine the performance of deep learning classification according to pain duration and intensity, based on morphology and location of perceived PFP from pain maps.*
Methods and materials: *PFP drawings were collected on lower extremities body-schema and encoded into three different data representations in respect to morphology of pain and location and a combination of the two. The distribution of the outputs were analysed and used for defining the classification intervals for pain duration, below 12 months and above 36 months, and pain intensity, below 4 and above 8 on VAS. Estimation of generalization performance of the models was calculated during the training.*
Results: *The results during training showed a higher accuracy for pain intensity classification than pain duration classification using morphology-representation. Pain intensity had an accuracy on 65.04%, and pain duration had an accuracy on 59.51%. Furthermore, the combined-representation performed with the highest accuracy on 65.14%. The location and morphology-representation scored 63.33%, and 65.04%, respectively, based on pain intensity.*
Discussion: *Despite pain intensity being defined as multidimensional and subjective, the performance accuracy were higher than that of pain duration. The results may indicate that a combination of the morphology and the location of the pain had a higher classification performance in relation to pain duration or intensity. Currently, it is unclear if deep learning methods may be a suitable approach for classifying PFP syndrome to work as support in clinical settings, to which further investigation is necessary. Improvements could be found when more data become available to better reflect generalization patterns in PFP drawings.*

I. Introduction

Patellofemoral pain (PFP) syndrome is a painful musculoskeletal condition that is presented as pain behind or around the patella [1, 2]. PFP syndrome 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]. PFP syndrome may be present over a longer period of time where a high number of individuals experience a recurrent or chronic pain [5]. Chronic pain may be maintained by the phenomenon central sensitization, which may result widespread pain over longer periods of time. Furthermore, PFP syndrome may 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. Because PFP is not caused by structural changes, no definitive clinical test may be used to diagnose PFP syndrome and thereby often diagnosed based on exclusion criterias [4] to which PFP syndrome is also described as an orthopaedic enigma, and is one of the most challenging pathologies to manage [7]. To assist diagnosis of PFP syndrome, 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].

A study by Boudreau et al. [9] indicates, through the

use of pain maps, that there is a correlation between the size of the pain (number of pain pixels) and the pain duration as well as intensity for individuals with PFP longer than five years.[9] However, it is unknown whether pain duration has an influence on morphology of the pain and location, as well as whether morphology of pain and location have an influence on pain intensity. The relation between pain maps and pain duration or pain intensity may be complex, because the perceived PFP is subjective, and considered as multifactorial [10]. Additionally the study by Boudreau et al. [9] did not find a fully correlation between 35 pain maps and pain duration or pain intensity for individuals with a pain duration below 5 years. To investigate the potential nonlinear correlation, a deep learning method was used, which is a method that previously has not been applied on this type of data.
The goals of this study is to explore how accurate a deep learning model can classify pain maps according to pain duration or pain intensity. It is assumed that pain duration is a better predictor than pain intensity, because the perceived pain is subjective, and may be affected by multidimensional factors. The pain maps are encoded into multiple data representations to investigate whether morphology and location are correlated to pain duration or intensity.
The data representations are encoded into three representations, which reflect either morphology of pain or location. It is assumed that a deep learning model will perform better with more information, thus a combination of morphology and location of the pain constitute a data representation. The data representations are refereed to as morphology-, location- and combined-representation. There may be a difference in how gender reports pain intensity, where females reports more intense and frequent pain [11]. Furthermore, there is an imbalance in prevalence between females and males, thus gender is included as a feature in the deep learning model.
The aim of this study was to explore classification performance of a deep learning model, using PFP maps as input to classify according either pain duration or intensity. Furthermore, a secondary aim was to compare the performance accuracy with different pain map representations (morphology-, location- and combined-representation), when predicting pain duration or pain intensity.

II. Methods

This section discuss the pain maps, and how they are processed before using them as input in deep learning models. Furthermore, the multiple pain maps representations are described, whereafter the complexity of the pain maps was investigated by using linear regressions. Finally, the deep learning architectures were presented.

Pain maps

Data used in this study were collected from an ongoing clinical trial (FOXH) which is conducted in collaboration with Danish and Australian universities. The pain maps were drawn by individuals with PFP syndrome through the use of an application, Navigate Pain, in a clinical setting.
Navigate Pain is a software application that is used to visualise the location, morphology and spatial distribution of pain from individuals to healthcare personnel. The application permits individuals to draw their pain with different colors and line thickness onto a body outline, an example is shown in fig. 1. Navigate Pain android was developed at Aalborg University.[12]

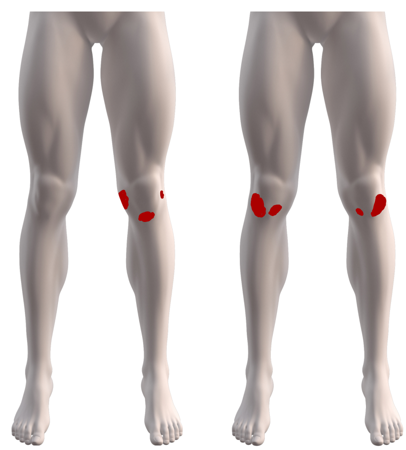

Fig. 1: *Pain maps from individuals with uni- and bilateral PFP. The red markings indicate the area of pain perceived by the individuals.*

The total number of pain maps available was 217, but only 205 pain maps with associated pain duration, and 197 pain maps with associated pain intensity was available. The gender was included as an input, because females may report a more intense and frequent pain than males.

Preprocessing

The pain maps were processed in MatLab version R2017b, where the images were resized to a pixel-size on 252×118, since they were collected at different resolutions (screen sizes) and cropped to only include the knees. To create more pain maps is a split body approach used, where pain maps are divided into two knees. Furthermore, pain was mirrored to represent pain on right knees to minimize the variance in the images. By using split body approach it was assumable that the pain duration and pain intensity were identical for both knees if PFP was bilateral. The total number of pain maps with gender and pain duration was 333, and pain maps with gender and pain intensity was 319, of which 15% was used as test data, and therefore not used to optimize and train the models.
The models should classify pain maps according to pain duration or pain intensity divided into intervals. These intervals were created based on the extremes, which were 0 to 12 months and 36 to 300 months for pain duration, and 0 to 4 and 8 to 10 for pain intensity. It was chosen to divide into extremes, since it was assumed that if the models predicted badly with the extremes, the models would not have a higher predictive value with multiple classifications of the outputs.

Morphology-representation

The original pain maps reflect the morphology of the pain, and do not require further processing than converting the pain maps to a matrix including gender and the output, pain duration or pain intensity. As a result of using the extremes for classifying the number of pain maps decrease to 236 for pain duration, and 196 for pain intensity.

Pain location

The knee is divided into regions based on the underlying anatomical structures, which may have a correlation to pain duration or pain intensity. The locations are divided into 10 regions, which are inspired by Photographic Knee Pain Map (PKPM). The divisions are designed to categorise location of knee pain for diagnostic and research purposes.[13] The knee regions are illustrated in fig. 2.

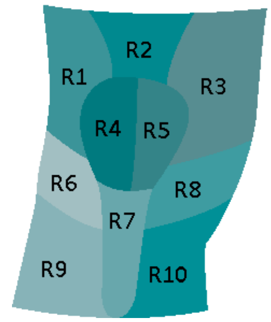

Fig. 2: *The regions of the right knee (R1-R10).*

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.[13]

Location-representation

To investigate whether the location alone have a correlation to the outputs, a simplified representation of the pain maps was created. The location of the pain was reflected by the use of the defined knee regions (fig. 2), where each region represented a value of 0 (not active) or 1 (active) in a vector. The values were defined by using a threshold to determine whether a region was considered active in relation the amount of pain. A threshold was 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 should not be too large so that pain areas was excluded. The threshold was decided based on an analysis on five random pain maps, where threshold values of 0, 5, 10 and 15% was compared. The threshold represent which minimal percentage of pain should be present in a specific region before it is considered active. Based on the analysis a 5% threshold was chosen. As a result of using the extremes for classification, and adding the threshold value the number of pain maps with pain duration decrease to 223, and number of pain maps with pain intensity decrease to 186.

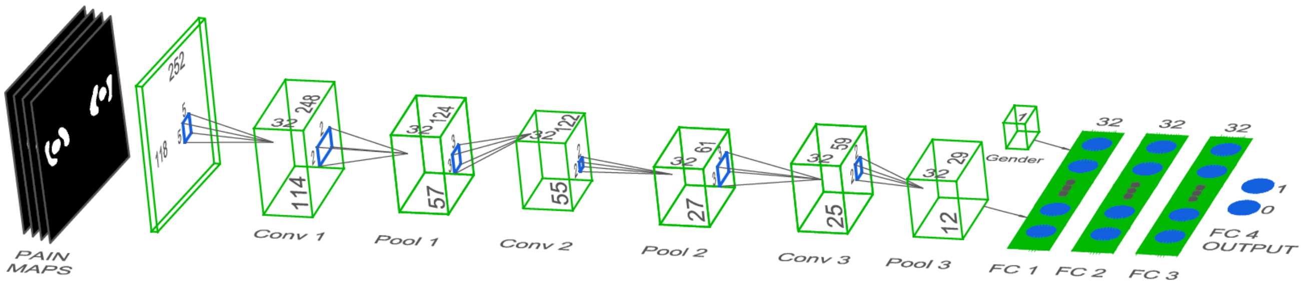

Fig. 3: *The architecture of the deep learning models including the morphology-representation consist of three convolutional, three max pooling, and four fully connected layers.*

Combined-representation

A combination of morphology and location of the pain is created based on components from morphology- and location-representations. The original pain maps are superimposed on the regions, which result in pain pixels reflecting the location with a number from 1 to 10. Before using the representation as input, one-hot encoding approach was used, which made it possible to separate categorical data into binary data [14]. This means that the 10 values do not have a correlation when analysed in the deep learning model. The number of pain maps with pain duration was 331, and number of pain maps with pain intensity was 317. The number of pain maps increased according to the location-representation, because no threshold was applied in this data-representation. By classifying according to the extremes, the number of pain maps decrease to 234 for pain duration, and 194 for pain intensity.

Nonlinearity in pain maps

Given that PFP is subjective and multifactorial it is unlikely that the pain maps and pain duration or pain intensity are linearly correlated. In order to determine if there was a linear relationship, linear regressions were done on simple features reflecting the size of the pain and number of active pain regions. The linear regressions were made in MatLab, and composed a correlation between number of pain pixels and pain duration, number of pain pixels and pain intensity, number of active pain regions and pain duration, and number of active pain regions and pain intensity.

Architecture of the deep learning models

Deep learning models were developed on a computer with 4x "Intel® Core^{TM} i7" CPUs and one single GPU of type "Geforce GTX 970M", using the programming language Python v3.6.3. Libraries used was Keras with a TensorFlow backend.
Multiple deep learning models suitable to the three data representation were created. The models used supervised learning, which is defined as a network learning to classify a given input corresponding to a specific output [15]. The models were designed differently according to its pain maps-representations. The architecture of the models, before optimization, including the morphology-representation is illustrated on fig. 3. These models consisted of three convolutional layers, to which a max pooling layer was designed after each convolutional layer, and ends with four fully connected layers, whereas gender was implemented. The models including the location-representation only use the four fully connected layers of the architecture shown in fig. 3. The last models using the combined-representation were design similar to the fig. 3, only with a difference according to the input image, which in this case included images with 10 layers. The models classified the input, pain maps and gender, in relation to the determined outputs, pain duration or pain intensity.

The convolutional layers

Convolutional Neural Networks (CNNs) is a type of special neural network for processing data with a grid-like topology [15]. CNNs were used to the morphology- and combined-representation because of its capability to perform highly according to image classification. The purpose of the convolutional layer was to recognize the features in the input by taking the image and scan it, then split it up into the feature maps.[15, 16] The architecture of the first convolutional layer consisted of a kernel size on 5×5, and 32 filters. The two following convolutional layers consisted of kernel sizes on 3×3, and 32 filters.

ReLU activation function

The activation function chosen for the hidden nodes in all three models were Rectified Linear Unit (ReLu), which transforms the linear output to nonlinear function by making all negative values zero. ReLu function still remains nearly linear, which means it can easily be optimized with gradient descent based methods [15]. In modern neural networks, ReLU is recommended to use as a default activation function and could be defined as $g(x) = max\{0, x\}$.

Max pooling layers

For the models containing convolutional layers, the convolution layers are followed by max pooling layers, which is a typical structure of a convolutional network [15, 17]. Max pooling layers are used to reduce the size of the dataset, while maintaining features from the feature map. Given a reduction in the data, the computation speed may increase.[15, 16] Max pooling layers are defined after each convolutional layer, to which all have a kernel size of 2×2 with a stride of 2. From the kernel window the highest of the 4 values is extracted to next layer, and used further through the network.

Fully connected layer and output layer

The models consist of four fully connected layers, whereas the 32 feature maps from the previous layer were flattened, and the notation for gender was included in the end of the string, which was used as input in the first fully connected layer with 32 nodes. Additionally, the second and third layers consisted of 32 nodes. The fourth fully connected layer, which also was the output layer, includes a sigmoid activation function. This function operates with a single output, that saturates when its input is either extremely negative or extremely positive [15]. It refers to the output corresponded to the number of classification intervals, pain duration below 12 month, and above 36 month, or pain intensity below 4, and above 8 on VAS.

Dropout algorithm

A dropout algorithm was implemented for the models in the last three hidden fully-connected layers to reduce overfitting while training. The algorithm works by randomly drop a specified faction of the nodes in the given layer, to which the nodes that drop changes during training [18]. Dropout reduce the nodes' ability for co-adaptation, where multiple nodes compute the same features. For the three models the dropout fraction was set to 0.5 (50%) based on a previous study by Srivastava et al. [18], which considered 0.5 as optimal for a multiple range of networks.

Back-propagation algorithm

Back-propagation is a learning process where the weights of the models are adjusted in order to reduce the error calculated between the predicted output, and the correct output.[19] Back-propagation use a method called gradient descents, which computed gradients from the output to the input, in order to minimize the overall output error as much as possible during the learning stage. After each pass of a minibatch, the inputs and weights were multiplied of separate node summed with additional coefficient called bias.[16, 20] Afterwards, a loss was calculated based on a loss function for every input that passed through the network to make the adjustments on the parameters to reduce the loss. As training progressed, the loss should decrease as a result of the parameter adjustments, and improve the performance of the neural network.[15, 17, 19]. This learning process continued until optimal parameters with minimum error was reached.[20]

Training

For training the models, a supervised learning approach was used, to which six models were trained, one for each pain map representation, and for each type of classification. This process was further combined with a structured grid search on the hyperparameters to help set the initial parameters for the models. These hyperparameters referrers to learning rate, kernel initializer, number of filters and nodes, and number of epochs with different batch sizes. Accuracy was used to determine the improvement of performance when testing the multiple parameters. Further manual optimization was performed by evaluating the development in loss, and accuracy during training, and the general performance estimated from an accuracy, sensitivity and specificity. After optimization, the models were trained anew using all of the training data, using the optimal hyperparameters from training and tested with a separate test subset.

III. Results

This section visualize the result from the linear regressions, and performance of the deep learning models using multiple pain maps representation, and different outputs.

Linear correlations

The linear regression between simple features, number of pain pixels or active pain regions, and outputs, pain duration or pain intensity, resulted in the plots shown in fig. 4. The R^2-values support the nonlinearity, shown in the plots, where correlation fig. 4a resulted in a $R^2 = 0.018$, fig. 4b resulted in $R^2 = 0.008$, fig. 4c resulted in $R^2 = 0.011$ and fig. 4d resulted in $R^2 = 0.011$.

Optimization of the models

During the optimization, a structured grid search resulted in different hyperparameters according to each model. This resulted in different optimizations in terms of the models with the morphology-representation. The learning rate was different according to pain duration (0.02), and pain intensity (0.1). A optimization in kernel initializer was defined as glorot_uniform for pain duration, and glorot_normal

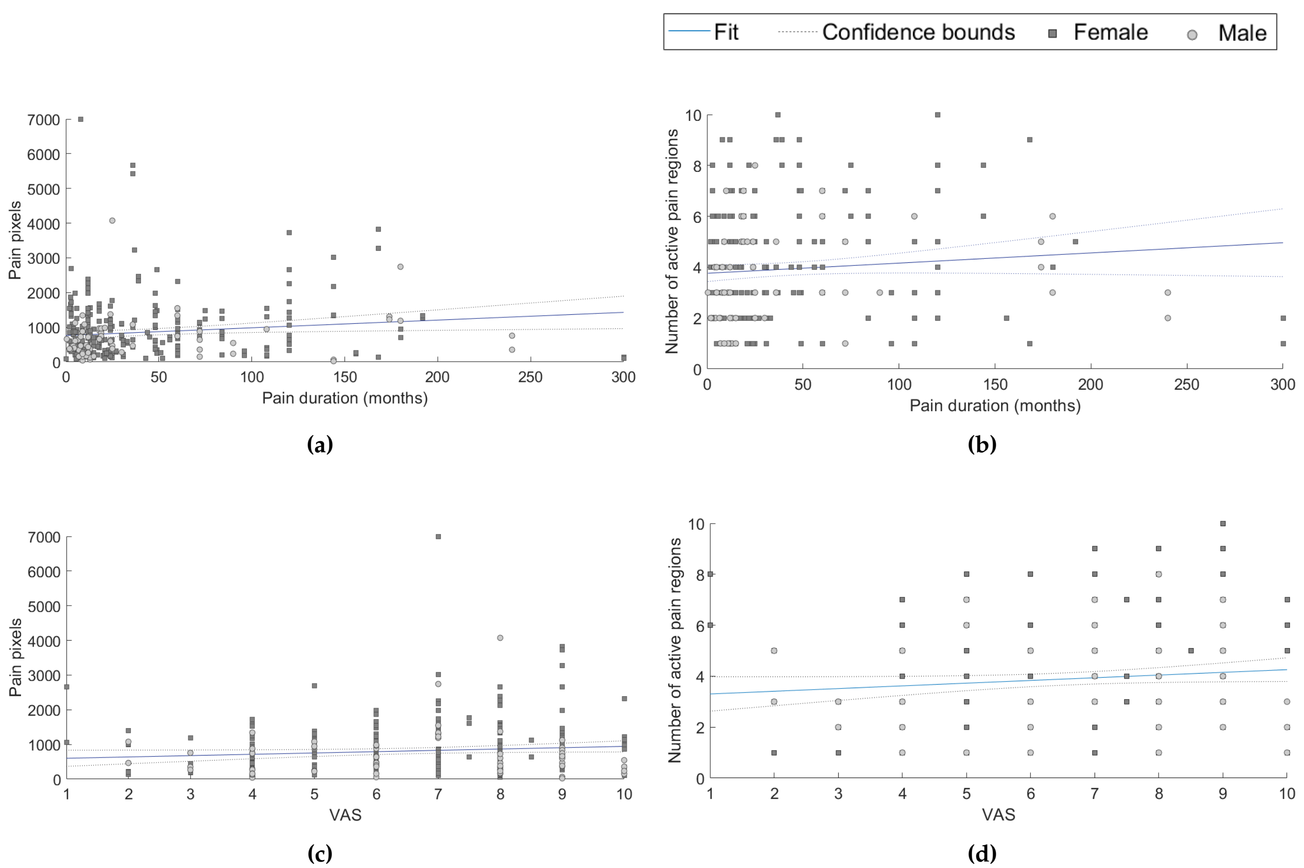

Fig. 4: *Linear correlations of pain pixels and pain duration (a), active pain regions and pain duration (b), pain pixels and pain intensity indicated in VAS (c), and active pain regions and pain intensity indicated in VAS (d).*

for pain intensity. The nodes was changed from 32 in the fully connected layers to 64 for pain duration, and 16 for pain intensity. Lastly, an optimization was found in number of epochs, and batch size, which was changed to 120 and 20 for pain duration, and 140 and 10 for pain intensity. The models including the location-representation had similar results from the optimization. A learning rate on 0.01, a glorot_uniform kernel initializer, 16 nodes, and number of epochs and batch size on 120 and 20. Results of optimization on the combined-representation was almost identical for the two classifications. Both gave best performance with a glorot_uniform kernel initializer, 16 nodes, and with a number of epochs and batch size of 120 and 30, to which the only difference was in the learning rate that for pain duration was 0.1, and 0.001 for pain intensity.

Performance of the models

The average performance accuracy, sensitivity, and specificity of the models during test with new pain maps in different representations are shown in tab. 1. For the pain map representations that resulted in the highest accuracy for either pain duration or pain intensity, confusion matrices were created, and is shown in fig. 5.

	Avg. accuracy (%)	Avg. sensitivity (%)	Avg. specificity (%)
	Morphology-representation		
Pain duration	69.44%	69.23%	69.57%
Pain intensity	60.00%	40.00%	70.00%
	Location-representation		
Pain duration	35.29%	0.00%	35.29%
Pain intensity	60.71%	0.00%	60.71%
	Combined-representation		
Pain duration	55.56%	61.11%	50.00%
Pain intensity	73.33%	0.00%	73.33%

Table 1: *Generalization performance of the models, which use the morphology-, location- and combined-representation when classifying according to pain duration or pain intensity.*

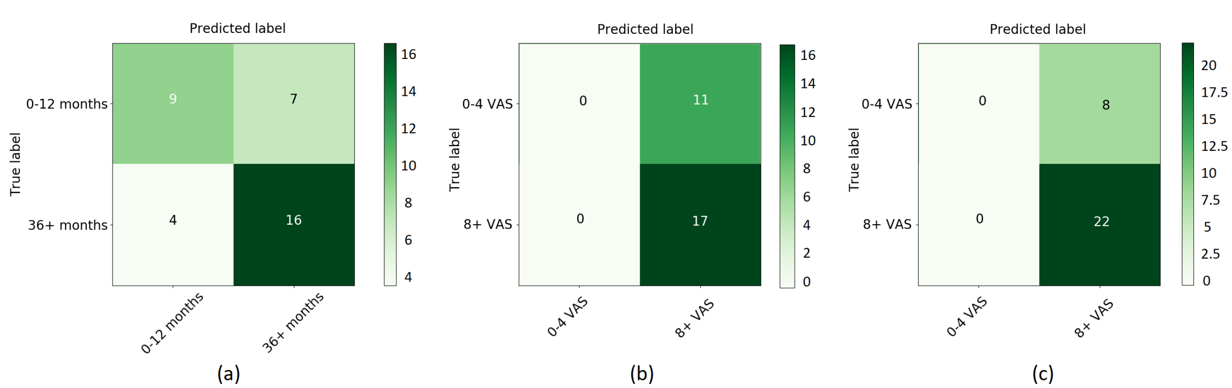

Fig. 5: *Confusion matrices of (a) morphology-representation classified according pain duration, (b) location-representation classified according to pain intensity, and (c) combined-representation classified according to pain intensity.*

IV. Discussion

This section discuss complexity of pain maps, and what may optimize the performance of the models. Furthermore, the results are discussed, whereas the highest performance value of the pain maps representations is evaluated. Finally, the performance according to the output, pain duration or pain intensity, is discussed.

Amount of pain maps

A supervised deep learning model should use five thousand labeled data per category to obtain an acceptable performance [15], whereas this study had a total number of pain maps (n=217) from uni- and bilateral PFP individuals. Based on the limited amount of pain maps, it was chosen to create more images by using a split body approach, where bilateral pain maps were split in two different pain maps. By using the split body approach and mirroring the pain to the right knee, it was assumed that pain duration and pain intensity were identical for both knees. Theoretically, the bilateral PFP may have occurred on one leg first, and afterwards have spreaded to the other knee, which could affect the pain duration. Furthermore, individuals with bilateral pain may feel more pain on one of the knees. This may have resulted in incorrected labeled pain maps, which could have an influence on the performance accuracy of the models.

Classification of pain maps

The R^2-values of the linear correlations were close to zero represents nonlinearity, meaning that there are no simple correlation with the outputs, pain duration and pain intensity. Thus, a more complex model, such as deep learning model, is used to investigate the complexity of pain maps. The deep learning models could classify both pain duration and pain intensity classes in the morphology-representation. Location-representation, despite of the given input, classified only according to the higher intervals (pain duration above 36 months and pain intensity above 8 VAS), and it could not classify according to the lower classes (pain duration below 12 month and pain intensity below 4 VAS). It can be observed in the confusion matrices of location-representation, where sensitivity of both inputs were equal to 0% (thought of putting the image of conf matrix). Combined-representation, classified according to the pain duration, found the patterns in both classes, but overall accuracy was 55.56%, which could be caused by including the knee regions as one of the features. Based on the results, the accuracy of the representations containing the morphology, scored higher compared to location-representation. It can be discussed, that the morphology can be considered to be a better classification feature. Morphology scored 69.44% and 60% while location-representation scored 35.29% and 60.71%, reflecting pain duration and pain intensity. Low score in location-representation is considered to be reflected as a result of imbalance between the duration intervals, as it could be also influenced with the deeper optimization. The low score could also be the result of the simplification of the location of the pain, that leads to the models not being able to find a pattern between the input and target output.

Threshold

The location-representation had a 5% threshold that defined when a pain region was considered active according to the amount of pain. It can be discussed whether this threshold was suitable, since adding the threshold resulted in loss of pain maps that had a very small amount of pain. However, a smaller threshold or no threshold would give active pain regions that might only contain very few pain pixels. Since PFP is described as hard to localize, it is unknown how precise the individuals have drawn their pain, thus every pixel should maybe not be taken into account. The combined-representation did not have a threshold for defining active pain regions, because the morphology of the pain would be affected when discarding small pain regions. This representation is thereby not a complete combination of the morphology- and location-representations.

Classification according to output

Generalization performance of the six models showed that pain duration was a more robust classifier compared to pain intensity. As a result, both classes of pain duration can be predicted in the morphology- and combined-representations, while classes representing pain intensity can be classified only with the morphology-representation. It could be discussed that the lower results of pain intensity against the pain duration, is a reason of that, the pain intensity is a subjective statement and is considered as multifactorial, while pain duration is an objective parameter, and it is thereby expected that the models have a higher performance when classifying pain maps according to pain duration. Given the fact that the sensitivity for multiple models was 0%, the accuracy becomes a reflection of the imbalance between classes in the test set, e.g. if the test subset contained 75% of the class with high pain intensity, the accuracy would also be 75%. This may also be a result of the relative small dataset used in this study. It is unknown whether the reason for the sensitivity to be 0% is caused by that the models cannot find patterns according to the low valued pain intensity, or if it is a "fault" in the model that classifies all inputs as high pain intensity.

Optimization of deep learning models

Optimization of the models is often a time-consuming process based on the picks of the multiple hyperparameters and different algorithms which could be implemented during the development of the models. Activation functions were chosen based on the literature, where ReLU should be used for convolutional and fully connected layers in neural network models, and sigmoid should be picked for binary output layer. However, there was always a way for additional testing with softmax or linear activation function to increase the generalization performance. The dropout algorithm was set to the default 0.5 and used in all models between the fully connected layers to turn off the amount of nodes and prevent the models from overfitting. Additional values could have been tested in order to find most optimal for the every model. Unfortunately, the lack of the time and time-consuming reruns during every optimization cycles, lead to use the most popular hyperparameters as there were many other which were tested with grid search 10-fold cross validation. A further optimization of the models may be found according to the input parameters, to which more physical, and psychological features may increase the performance accuracy. Physical features, such as age, height, weight, physical activity niveau, and sport activity, may influence pain duration or pain intensity. Age could be a relevant feature since the perceived pain is dependent on the individual's personality and character. Younger individuals may feel more pain because of a new and strange pain, than older individuals which have had PFP for a longer period of time. In addition older individuals may feel more pain because of the phenomenon central sentization, which in some cases result in widespread pain. The physical activity niveau, and sport may increase the pain intensity for some individuals because of the patellofemoral loaded activity. Psychological factors are an important feature to consider, because of its influence on pain intensity. Pain is multifactorial and can be influenced of psychosocial factors [21]. Furthermore, other pain as hip pain may influence the PFP. It may be considered to include either pain duration or pain intensity as an input to classify according to either pain intensity or pain duration, because of the possibility that there is a correlation between the two. A limitation for this study was the available computational power for training of the model to which an improvement in performance may be found through more powerful systems or services.

V. Conclusion

During this study the deep learning models were presented for classifying pain maps according to the pain duration and pain intensity. The overall performance of the models were calculated, to which the combined-representation performed the highest accuracy, but the models using the morphology-representation resulted as the most robust classifier. There may be an indication of a pattern to be found between pain maps and pain duration or pain intensity, however further optimization or additional studies is needed to support whether location or morphology contains positive predictive value in terms of classifying pain duration or pain intensity.

References

[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] Amir Haim, Moshe Yaniv, and Et.al. Patellofemoral Pain Syndrome. *Knee Surg sports traumatol arthrose*, 451:223–228, 2006. ISSN 0009-921X. doi: 10.1007/s00167-013-2759-6. URL

http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage{&}an=00003086-200610000-00041.

[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, Susanne Badsberg, and Et.al. 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] E. J. Dansie and D. C. Turk. Assessment of patients with chronic pain. 2013. doi: 10.1093/bja/aet124.

[11] Christoph Pieh, Jurgen Altmeppen, and Et.al. Gender differences in outcomes of a multimodal pain management program. *Elsevier*, 2012. doi: 10.1016/j.pain.2011.10.016.

[12] Aglance Solutions. Visual insight for clinical reasoning – Navigate Pain, 2015. URL http://www.navigatepain.com/.

[13] D. W. Elson, S. Jones, and Et.al. The photographic knee pain map: Locating knee pain with an instrument developed for diagnostic, communcation and research purposes. 2010. doi: 10.1016/j.knee.2010.08.012.

[14] David Money Harris and Sarah L. Harris. Sequential Logic Design. In *Digital design and computer architecture*. Elsevier, 2012. ISBN 9780123978165.

[15] Ian Goodfellow, Yoshua Bengio, and Et.al. *Deep Learning*. MIT Press, 2016. URL http://www.deeplearningbook.org.

[16] Yann LeCun, Léon Bottou, and Et.al. 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.

[17] Yann LeCun, Yoshua Bengio, and Et.al. 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.

[18] Nitish Srivastava, Geoffrey Hinton, and El.al. 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.

[19] Richard Duda, Peter Hart, and Et.al. *Pattern Classification*. Second edi edition, 2000. ISBN 9780471056690.

[20] Alaa Ali Hameed, Bekir Karlik, and Et.al. 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{%}3Dihub.

[21] Ewa M. Roos and L. Stefan Lohmander. The knee injury and osteoarthritis outcome score (KOOS): from joint injury to osteoarthritis. *Bio Med Central*, 2003.