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

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 pain intensity has been indicated in previous studies [2], however morphology and location of PFP remains unexplored in terms of correlation. Based on the object detection capabilities of deep learning methods can be used to detect image-features related to morphology and location. The aim of this study was to determine the performance of deep learning models for predicting pain duration and pain intensity, based on morphology and location of perceived PFP from pain maps.*

Methods and materials: *PFP drawings were collected on a body-schema and encoded into three different pain map representations in respect of morphology of pain, location of pain, and a combination of morphology and location. The distribution of the outputs were analysed and used for defining the classification intervals for pain duration, to which the intervals were 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 the test showed a higher accuracy for pain intensity classification than pain duration classification using the combined-representation. Pain intensity had an accuracy on 86.67%, and pain duration had an accuracy on 55.56%. However, the morphology-representation resulted in the highest predictive value classifying according to pain duration. The models using the location-representation had the lowest classification performance according to both pain duration and pain intensity, compared to the two other representations.*

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 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 optimization of the models is necessary. Improvements could be found when more pain maps become available to better reflect generalization patterns in PFP drawings.*

1. 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 is typically 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 resulted in widespread pain over time. Ultimately, PFP syndrome may lead to osteoarthritis [4, 6].

PFP is often described as diffuse knee pain, that can be hard for individuals to explain and localize [5]. Despite that individuals feel pain in the knee, there is no underlying structural changes in the knee such as significant chondral damage. There is no definitive clinical test to diagnose PFP syndrome and is 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 man-

age [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 (n=35), that there is a correlation between the size of the pain areas (total number of pain pixels) and the pain duration as well as pain intensity for individuals with PFP duration 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 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 previously has not been applied on this type of data.
The aim of this study was to explore how accurate a deep learning model can classify pain maps according to pain duration or pain intensity. It was assumed that the models classifying according to pain duration would have a higher predictive value than pain intensity, because of the subjectivity of pain, and its possibility of being multifactorial. The pain maps were encoded into multiple pain map representations to investigate the performance of the models using different representations. Representations reflecting the morphology of the pain and the location of the pain were created. It was assumed that a deep learning model would perform better with more features, thus a combined-representation containing morphology and location of the pain was made. The representations are referred to as morphology- (MR), location- (LR), and combined-representation (CR).

2. Methods

This section presents the pain maps and the pre-processing. Furthermore, the multiple pain map representations are described, whereafter the linearity of the pain maps is investigated. Finally, the deep learning architectures are presented.

2.1. Pain maps

Pain maps used in this study were collected from an on-going 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 visualize 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 was developed at Aalborg University.[12]

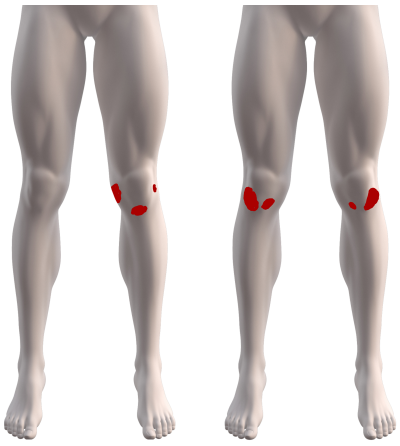

Fig. 1: *Pain maps from individuals with uni- and bilateral PFP. The red markings indicate the area of pain perceived by the individuals.*

The number of pain maps associated with pain duration was 205, and 197 associated to pain intensity measured according the Visual Analog Scale (VAS). The gender was included as an input, because females may report a more intense and frequent pain than males [11], with the intentions of making the deep learning models better at classifying the pain maps.

2.2. Pre-processing

The pain maps were processed in MatLab v. R2017b, where the images were resized, since they were collected at different resolutions (screen sizes) and cropped to only include the knees. To create more pain maps a split body approach was used, where the two knees in the pain maps were separated into

individual images. The left knee was then mirrored to resemble the right knee to minimize the variance in the images. The final pain maps had a pixelsize of 252×118. By using split body approach it was assumable that the pain duration and pain intensity were identical for both knees if PFP was bilateral. This resulted in an increased total number of pain maps with gender and pain duration to 333, and pain maps with gender and pain intensity to 319, of which 15% was used as test data, and therefore not used to optimize and train the deep learning models.
The models were designed to classify pain maps according to pain duration or pain intensity divided into two classes with intervals based on the extremes. The classification intervals were 0 to 12 months and 36 to 300 months for pain duration, and 0 to 4 and 8 to 10 for pain intensity measured on VAS. The reason for choosing extreme intervals, was to separate closely related patterns between the two classes, which resulted in the amount of pain maps as shown in tab. 1.

	Pain duration (months)		Pain intensity (VAS)	
	0-12	36-300	0-4	8-10
MR	114	122	72	124
LR	105	118	68	118
CR	113	121	71	123

Table 1: *The amount of pain maps for the MR, LR and CR for both pain duration and pain intensity using the extremes.*

2.2.1 Morphology-representation

The original pain maps reflect the morphology of the pain, and did not require further processing than converting the pain maps to a matrix including gender and the output, pain duration or pain intensity.

2.2.2 Pain location

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

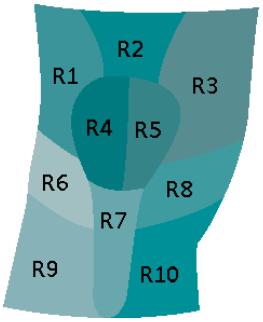

Fig. 2: *The regions of the right knee (R1-R10).*

There are ten regions, where regions 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 regions 4 and 5. Regions 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]

2.2.3 Location-representation

A simplified representation of the pain maps was created to investigate whether the location have patterns related to the pain duration or pain intensity. The location of the pain was reflected by the use of the defined knee regions (fig. 2), where a 10 elements vector were created. Each value represented either an active region (1) or a not active region (0). The values were defined by using a threshold to determine whether a region was considered active in relation the amount of pain in the specific region. Threshold was used to accommodate for the diffuse pain, and the difficulty in localizing PFP. The threshold increased the confidence of an active pain region by avoiding minimal contributions. Simultaneously, the threshold should not be too high to avoid excluding too many or big areas. The threshold indicated which minimal percentage of pain pixels, that should be present in a specific region before it was considered as active. The threshold was decided based on an analysis on five random pain maps, where threshold values of 0, 5, 10 and 15% was compared. Based on the analysis, a 5% threshold was chosen.

2.2.4 Combined-representation

A combination of morphology and location of the pain was created based on components from MR and LR.

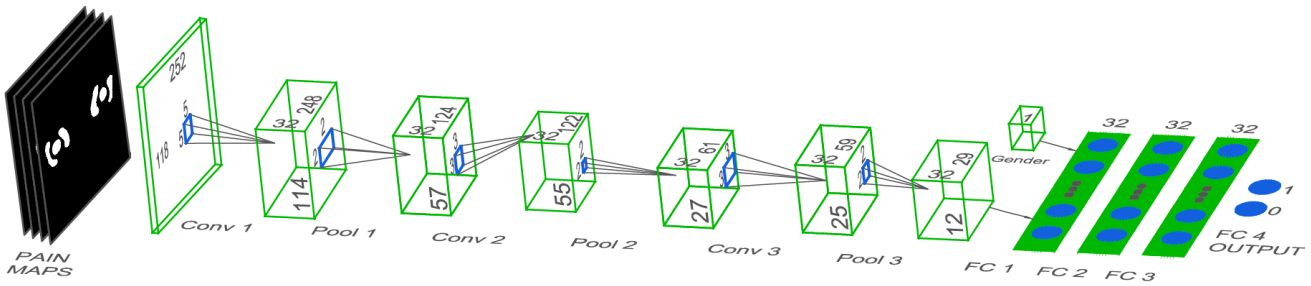

Fig. 3: *The architecture of the deep learning models including the MR before optimization. The models consisted of three convolutional, three max pooling, and four fully connected layers. The gender was included in the first fully connected layer. The boxes represented the output after each layer.*

The original pain maps were superimposed on the knee regions, which resulted in pain pixels reflecting the location with a number from 1 to 10. Before using the representation as input in the deep learning models, a one-hot encoding approach was used, which made it possible to separate categorical data into binary data [14]. As a result, the 10 values did not have a relation to each other when analysed in the deep learning models.

2.3. 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. Linear regressions were made in MatLab, and composed correlations 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.

2.4. Architecture of the deep learning models

Deep learning models were developed on a computer with 4x "Intel® Core™ i7" CPUs and one single GPU of type "Geforce GTX 970M", using the programming language Python v3.6.3. Libraries used was Keras v2.0.8 with a TensorFlow v1.3.0 backend.
Six deep learning models suitable to the three pain map representations and pain duration or pain intensity 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 each pain map representation and type of classification. The architecture of the models, before optimization, including the MR is illustrated on fig. 3. The models classified MR and CR consisted of three convolutional layers, to which a max pooling layer was added after each convolutional layer, and followed by four fully connected layers, where gender was inserted as a secondary input for the models. The only difference in the models using the CR was the input image, which was represented a matrix with 10 layers. The models including the LR consisted of only four fully connected layers of the architecture shown in fig. 3.

2.4.1 The convolutional layers

Convolutional Neural Networks (CNNs) is a type of neural network for processing data with a grid-like topology [15]. CNNs were used to the MR and CR because of its capability to perform highly according to image classification. The purpose of the convolutional layers was to recognize the features in the pain maps by taking the image and scan it, then split it up into 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.

2.4.2 Max pooling layers

For the models containing convolutional layers, each convolution layer was followed by a max pooling layer, 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 maps. Given a reduction in the data, the computation speed may increase.[15, 16] Max pooling layers were 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 was extracted to next layer, and used further through the network.

2.4.3 Fully connected layer and output layer

The models consisted of four fully connected layers, where the first layer received a flattened version of the input. The notation for gender was included in the end of the array, 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, included a sigmoid activation function. This function operates with a single output, that saturates when its input is either extremely negative or extremely positive [15]. The single output refers 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.

2.4.4 ReLU activation function

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

2.4.5 Dropout algorithm

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

2.4.6 Back-propagation algorithm

Back-propagation was used for the learning process where the weights of the models were adjusted in order to reduce the error calculated between the predicted output, and the correct output [19]. Back-propagation is based on gradient descents, which computes 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 for each node summed with additional coefficient 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 updates, and improve the performance of the neural network.[15, 17, 19] This learning process continued until optimal parameters with minimum error were reached.[20]

2.5. Training

The models were trained and optimized with a structured grid search on the hyperparameters to help set the initial parameters for the models. These hyperparameters refer to learning rate, 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 the pain maps was divided into three subsets, a training set consisting of 75%, a validation set consisting of 10%, and a test set of the remaining 15%. A manual optimization was performed, using the validation set, by evaluating the development in loss, accuracy during training, and the general performance estimated from accuracy. After optimization, the models were tested with the test set, using the hyperparameters from the optimization.

3. Results

This section visualizes the results from the linear regressions, and performance of the deep learning models using multiple pain map representations, and different outputs.

3.1. 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 $R^2 = 0.011$.

3.2. Optimization of the models

During the optimization, a structured grid search resulted in different hyperparameters according to each model. For the deep learning models including MR and classification type pain duration, the highest performance was obtained with a learning rate of 0.01, the number of nodes of 64, and the epochs and batch size setting of 120 and 20. The models using the MR and pain intensity had the highest performance when using a learning rate of 0.1, 16 nodes, epochs and batch size of 140 and 10. The models including the LR had similar results from the optimization. Learning rate of 0.01, number of nodes at 16, and a number of epochs and batch size of 120 and 20. Results of optimization on the models including the CR were almost identical. Both resulted in the best performance with 16 nodes,

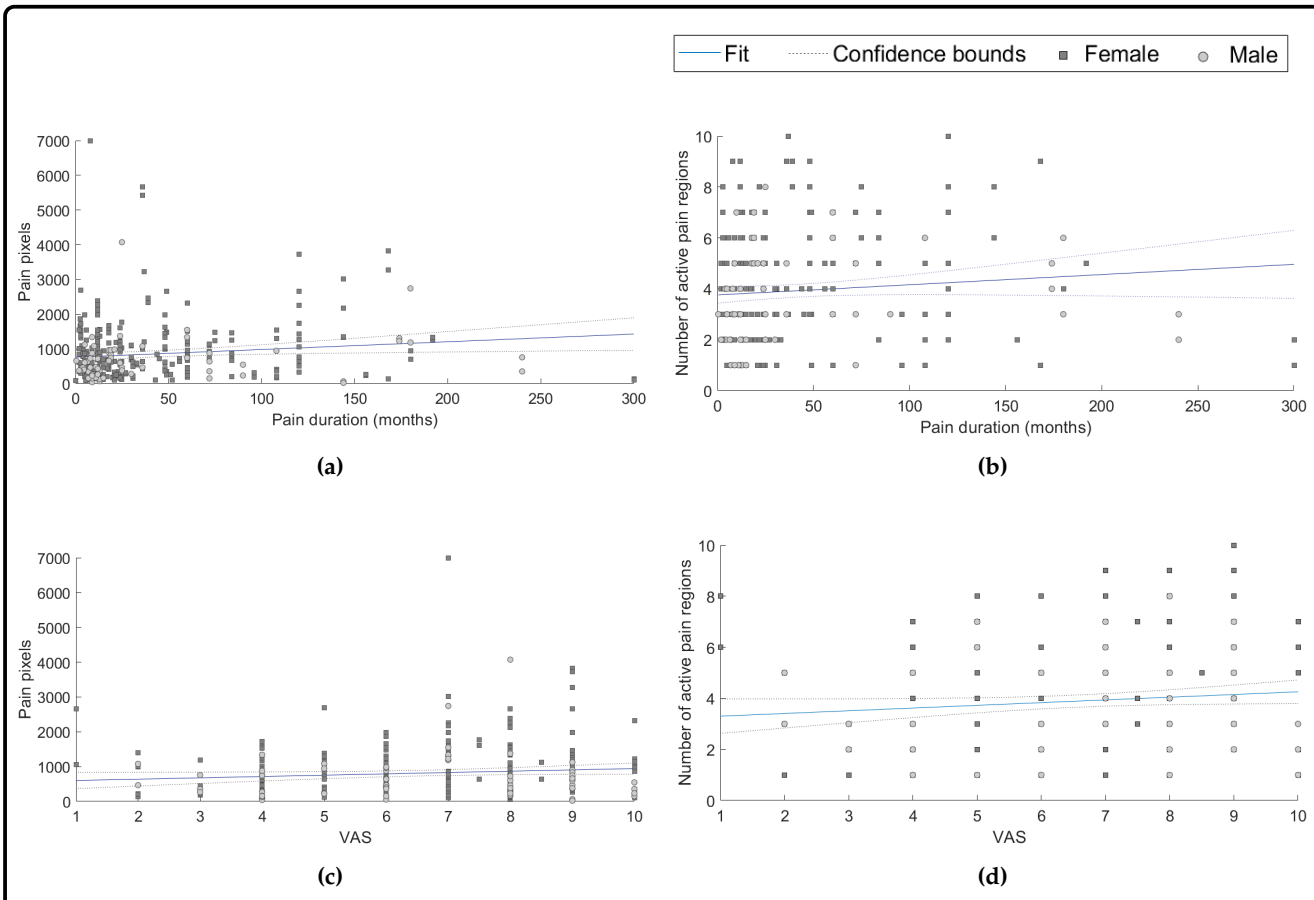

Fig. 4: *Linear correlations of pain pixels and pain duration (a), active pain regions and pain duration (b), pain pixels and pain intensity indicated in VAS (c), and active pain regions and pain intensity indicated in VAS (d).*

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.

3.3. 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. 2. For the pain map representations, confusion matrices were created according to the pain duration as is shown in fig. 5, and confusion matrices according to pain intensity as shown in fig. 6.

4. Discussion

This section discusses 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 map representations is evaluated. Finally, the performance according to the output, pain duration or pain intensity, is discussed.

4.1. Amount of pain maps

In this study the total number of pain maps was 217 from individual with uni- and bilateral PFP. Comparing to the literature, a supervised deep learning model

	Avg. accuracy (%)	Avg. sensitivity (%)	Avg. specificity (%)
		Morphology-representation	
Pain duration	69.44%	56.25%	80.00%
Pain intensity	60.00%	40.00%	75.00%
		Location-representation	
Pain duration	35.29%	0.00%	100%
Pain intensity	60.71%	0.00%	100%
		Combined-representation	
Pain duration	55.56%	55.00%	43.75%
Pain intensity	86.67%	75.00%	90.91%

Table 2: *Generalization performance of the models, which use the MR, LR, and CR when classifying according to pain duration or pain intensity.*

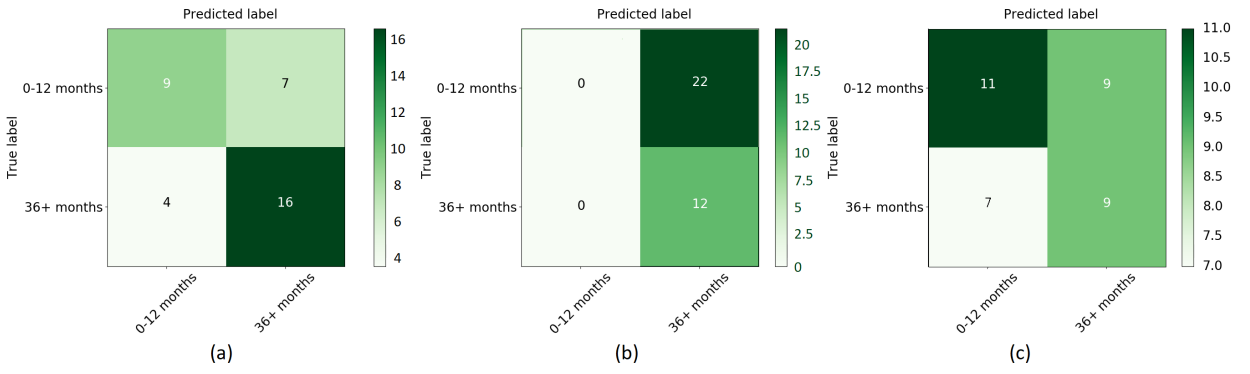

Fig. 5: *Confusion matrices of (a) MR classified according to pain duration, (b) classified LR according pain duration, and (c) classified CR according to pain duration.*

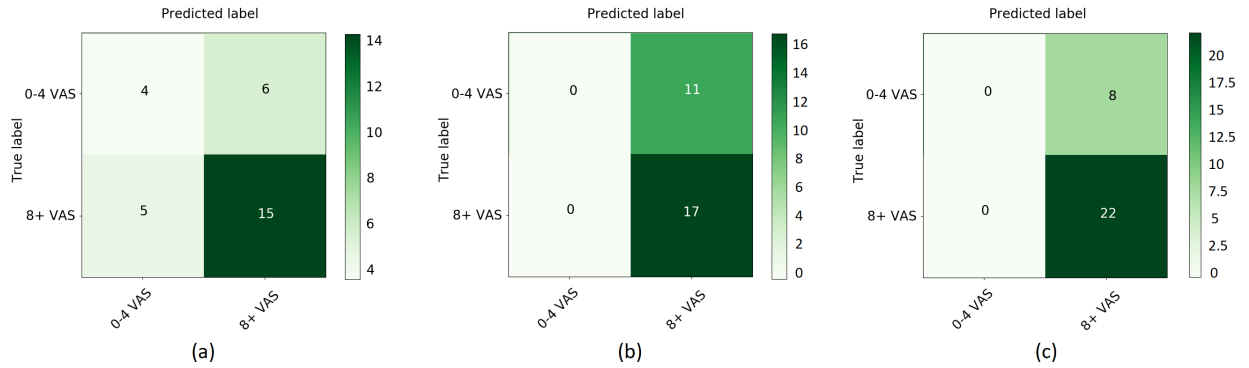

Fig. 6: *Confusion matrices of (a) MR classified according to pain intensity, (b) classified LR according pain intensity, and (c) classified CR according to pain intensity.*

should use five thousand labeled data per category to obtain an acceptable performance [15]. The amount of pain maps was not optimal, to which a split body approach was used to compensate this. During this approach combined with the mirroring 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 incorrect labeled pain maps, which could influence the performance accuracy of the models.

4.2. Classification of pain maps

The R^2-values of the linear correlations were close to zero, which represents nonlinearity. Thus, a deep learning model used to investigate the complexity of pain maps. The models that used the MR resulted in a higher predictive value for pain duration and pain intensity in the higher extremes (36+ months and 8-10 VAS), compared to the lower extremes (0-12 months and 0-4 VAS). By comparing the sensitivity and specificity of the model using MR to predict pain duration or pain intensity, the model predicting the pain duration was better at predicting according to the true low pain duration (0-12 months) and true high pain duration (36-300 months). The models representing LR could not distinguish between the pain maps according to the extremes for both pain duration and pain intensity. The models simply classified all pain maps as being in the higher classes. The model using CR classified better according to pain duration when predicting pain maps with a short duration, where the model that classified according to pain intensity was better to predict the higher pain intensities. Overall the accuracy indicates whether a pattern may be present in the pain maps according to the given classification, however this is not the case when the sensitivity or specificity is zero valued, to which the accuracy only reflects the amount data there is for one class. Bases on the accuracy, sensitivity and specificity for the models, the model using the CR classified according to pain intensity had the highest predictive value.

4.3. Threshold

The LR 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 CR 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 not a complete combination of the MR and LR.

4.4. Optimization of deep learning models

Optimization of the deep learning 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, additional testing could be made by using 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 time and time-consuming reruns during every optimization cycle, lead to use the common hyperparameters as there were many other which were tested with grid search 10-fold cross validation. 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.
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 level, 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 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 sensitization, which in some cases results in widespread pain. The physical activity level, and sport may increase the pain intensity for some individuals because of the patellofemoral loaded activity. Psychological factor is 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 areas, such as hip pain, may influence the pain intensity of PFP.

5. Conclusion

During this study the deep learning models were presented for classifying pain maps according to the pain duration or pain intensity. The overall performance of the models were calculated, to which the CR performed with the highest accuracy according to pain intensity, despite of the subjectivity of pain and the influence of multidimensional factors. There may be an indication of patterns 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.

6. Acknowledgement

Thanks is given to the supervisors Shellie Boudreau and Lasse Riis Østergaard for their feedback and availability during this study.

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.