How spikes in the preceding activity are related to the response time of a nerve fiber in microneurography?

Alina Troglio

November 17, 2020

1 Introduction

About $6 - 10\%$ of the world population suffers from chronic neuropathic pain [1]. In particular, the complexity of the disease and the limited treatment, which often has side effects, makes pain relief difficult and points to the importance for improvement of neuropathic pain [2]. It has been discovered that spontaneous activity of afferent nociceptive C-fibers is an important peripheral mechanism contributing to neuropathic pain [3]. These findings were achieved by using the examination technique of microneurography. Microneurography is a technique for measuring and recording nerve impulses and C-fiber activities in humans and in animals. There have been studies investigating the dependencies between induced activities, i.e. number and frequency of action potentials and latency changes of the following fiber action potentials, e.g Schmelz [4]. The preceding activity is changing the latency. To the best of our knowledge, a model that describes these dependencies has not yet been developed. Such a model could help to analyze spontaneous activity via reverse engineering. Up to now it is not possible due to the bad signal-to-noise ratio to analyze and describe spontaneous activity and therefore the number of action potentials. Solely latency shifts of the electrically induced action potential following spontaneous activity indicate that there has been spontaneous activity and can give a rough estimation of its magnitude and the corresponding action potential number.
In this project, the relationship between the preceding activity and the latency is exploited by using the number of action potentials to define features for these activities. It starts with an introduction explaining which data is used to build the model and how the feature space is defined to describe the preceding activity. Afterwards, the results of modeling the relationship between activity and response

latency are presented. The discussion highlights the challenges with the chosen approach and data.

2 Methods

The following chapter describes the steps for defining the feature space and building the model. First, an overview of the workflow is given, followed by the data and the software, that provided the fiber recording and the template matching to separate the fibers. Finally, the feature space is defined and the input and output values for the model are determined.

2.1 Workflow

One of the challenges in machine learning is to collect enough data and extract suitable features to describe the preceding activities. To extract features several steps are required before the model can be built. Figure 1 shows an overview of the workflow. As mentioned in the introduction, this model should help with spontaneous activities, as these may occur in patients with chronic pain. Spontaneous activities are not predictable and cannot be observed in all patients. This requires a proxy, which must be treated as spontaneous activity, for example mechanical stimulation by scratching or touching the skin area of interest. The recording is examined for fibers and a single fiber is selected that responds to electrical stimulation and mechanical stimulation. Then the fiber must be extracted using template matching. After the data has been exported, the response latency needs to be calculated and the action potentials are counted in a defined time interval. Therefore, a feature space is created based on the spike count and the placement of the action potentials. With the defined features and the response time a model can be created and the goodness of fit can be measured.

2.2 Data

The data for this work is a pressure-clamped fiber recording of a rat. The software Spike2[1] developed by Cambridge Electronics Design Limited (CED) was used to record the data and to extract the features of previous activities. Spike2 is a software for recording continuous life science data and analyzing it. Since animal data sets have a high signal-to-noise-ratio and are limited in the number of fibers, it is preferable to use animal data for first approaches in the search for the best features. There are $2 - 3$ fibers in the record. Regular electrical stimuli are always applied. The regularity varies and there are frequencies of $0.25 - 4$ Hz.

[1] http://ced.co.uk/us/products/spkovin

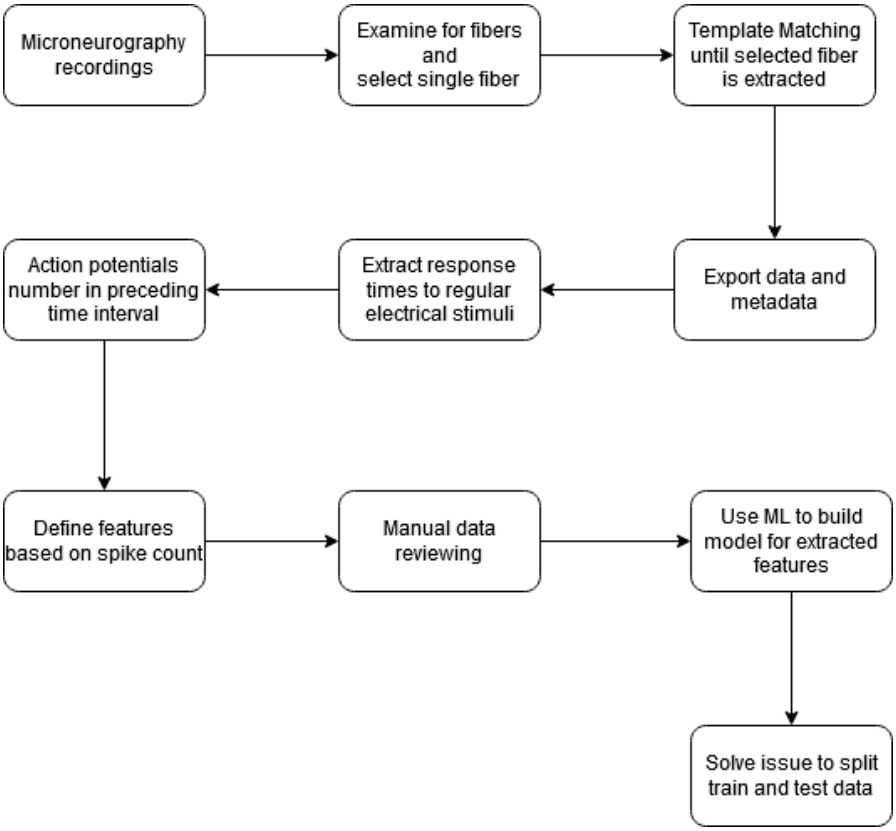

Figure 1: Overview of the workflow for this project. The most time consuming parts were extracting the data, finding suitable features and extracting them to create a model.

Additional mechanical stimuli are applied irregularly. The time between a regular electrical stimulus and the response of the selected fiber of this stimulus is measured as the response time and later used as the output value for the model. The number of action potentials of this fiber in a predefined time is the input value. Fiber separation is required to extract the selected fiber. A problem in human microneurography is the simultaneous detection of several fibers [5]. Therefore, the fiber separation is essential to investigate the change of latency for the fiber of interest. In Spike2 it is based on template matching [6]. The user can choose between automatic and manual template generation and a variety of parameters. The templates are created based on the amplitude and shape of the chosen action potential and the matched action potentials are marked. Because it is an ideal setup with low noise and distinguishable fibers, the template for the fiber can be created and manually checked to ensure that the markings are correct even without medical expertise. Considering that this method should work with human data

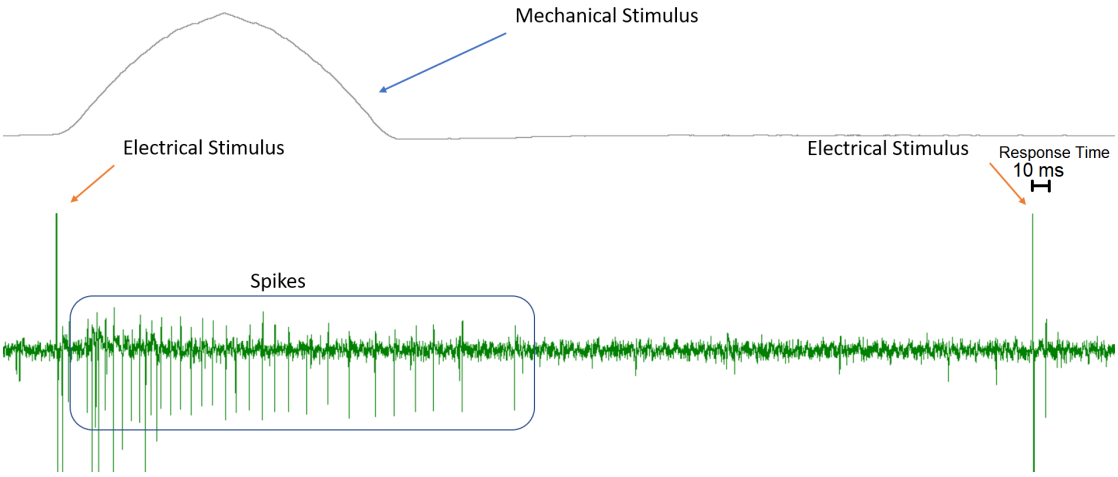

Figure 2: Overview of data extracted from Spike2 and showing the spikes as a result of the electrical and mechanical stimulation as input to the model and the response time to the next regular electrical stimulus as output.

and fibers that are not distinguishable by vision, a different approach must be taken, but for now this separation is sufficient. The data can be directly extracted into a CSV format for further processing.

2.3 Feature Space and Model

For the implementation and the modeling, the Python[2] package scikit-learn[3] was used. The model aims to establish the relationship between the number of spikes in a time interval and the latency change to an electrical regular pulse after this interval. The input features are the number of action potentials in a time interval. Using the numbers of action potentials is one way to quantify the nerve activities. As output the response time must be calculated. Since the data provide regular pulses, the response time is the difference between one of the regular electrical action potentials and the action potential of the selected fiber (see Figure 2). In the future the influence of spontaneous activity should be considered, but for this work the mechanical activity serves as a proxy. In these data, the mechanical stimulation lies between two regular electrical stimuli.
After quantifying the activities, a time interval must be specified, for example 4 seconds, 2 seconds, 12 seconds, 100 seconds, 400 seconds, . . . , to count the fiber action potentials during this time before the regular electrical stimulus, which is used

[2]https://www.python.org/
[3]https://scikit-learn.org/stable/index.html

to measure the latency. In the determined time intervals before the electrically induced AP, the action potentials of the fiber of interest are counted. This results in the input and output features. With this one-dimensional input and output vector a regression model can be fitted. To further investigate the behavior and use the placement of the action potentials as well, the feature space is extended. The time interval is cut into different pieces and the action potentials in each cut are counted. Levels are defined (see Figure 3). In Level 0 there is an input value of $2^0 = 1$. In Level 3 there are $2^3 = 8$ input values. On a given level the input and output values can be used to create a regression model with multiple input values. 5-Fold[4] is used to split the training and test data.
The average of the coefficient of determination (R^2) and the mean squared error are calculated to measure the goodness-of-fit of the model for each fold. In microneurography it is a problem to find the best method to split the training and the test data, because the stimuli can influence the latency of a fiber for minutes [4] and are not independent from each other. In this work 20% of the data points in each fold are randomly selected, but in the future it has to be investigated whether there are better approaches to select the train and test split.

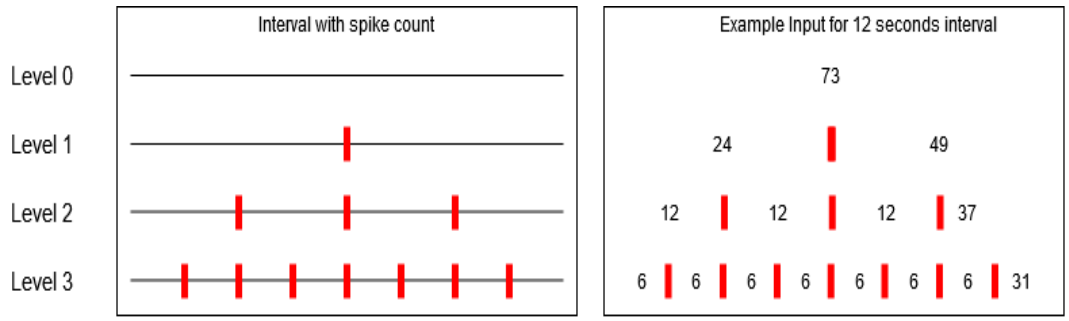

Figure 3: Level definition and example for a time interval of 12 seconds with a total number of 73 spikes.

3 Results

In this section the results of the work are presented. A plot has been created to show the counted number of action potentials in previous times interval with mechanical stimuli by frequency. Afterwards, two data sets were extracted from the fiber recording. The first data set contains all intervals divided by regular

[4]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

electrical stimulation with additional mechanical stimulation (29 intervals). The second data sets includes all intervals (approximately 6850 intervals).
The results for the first data set using linear regression are shown. Different time ranges and levels are considered and a plot of the one dimensional input values and predicted output values is given for the most promising time interval. Finally, the second data set is used. Therefore, all intervals are considered and a linear regression model and a support vector regression model are fitted to describe the relationship between preceding activities and response time.

3.1 Relationship Number of Spike and Response Time

The first attempt to extract features was to cut the recordings into intervals by the regular electrical stimuli and count the spikes in these intervals with additional mechanical stimuli. The relationship between frequency and response time was investigated. The lower the frequency, i.e. the greater the distance to the next regular electrical stimulus, the more spikes are in the interval (see Figure 4).

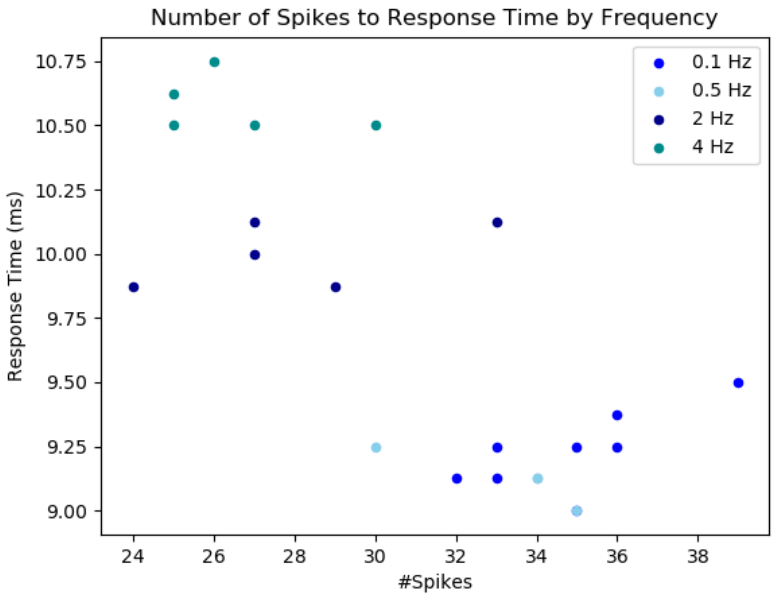

Figure 4: Spike number to response time separated by frequency. A lower frequency results in more spikes per interval and a lower response time. Higher frequencies result in more action potentials and higher response time.

If the frequency is high ($2 - 4$ Hz), the number of spikes in the interval is lower compared to the intervals with lower frequencies ($0.1 - 0.5$ Hz). This gives

an indication that the previous stimulation is still affecting the fiber for a longer period of time. High frequencies indicate that the fiber has no time to recover and therefore the latency is increased. The idea is to use different periods of time as preceding activity that go beyond the limits of the regular electrical stimulation intervals.

3.2 Prediction of Response Time

The first model that predicted the response time considered the intervals with additional mechanical stimulation, but this yields only 29 data pairs of spike count and latency to train and test the model. Linear regression was used to make a first prediction for the response time and the results are presented in chapter 3.2.1. Since the 29 data pairs exclude most of the fiber activity history and the data set has very small size, all intervals were considered in chapter 3.2.2 and linear and support vector regression were performed.

3.2.1 Intervals with mechanical stimulation

The data set provides a total of 29 intervals with additional mechanical stimulation. Due to the different frequencies the number of spikes can be significantly increased especially at high frequency and for a long time interval.

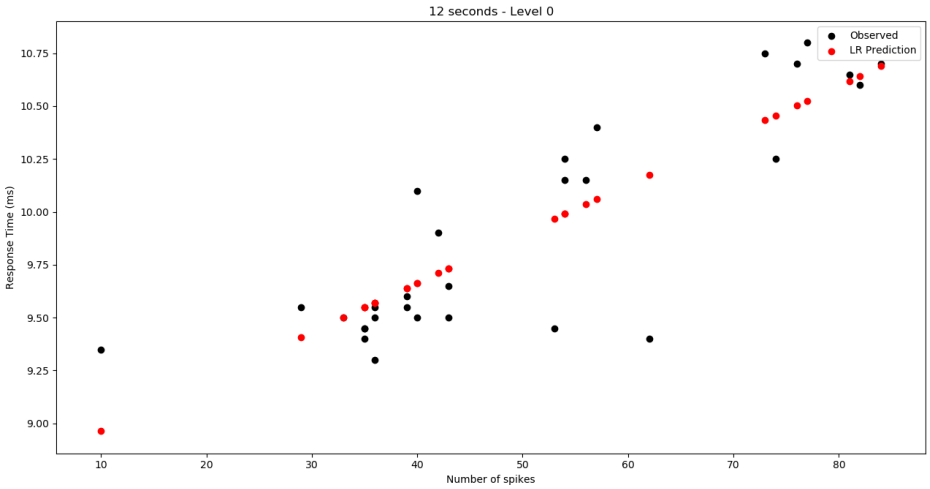

Figure 5: Linear Regression for 12 seconds on level 0 for 29 data points. The number of spikes is plotted against the response time in ms.

Due to the small data size, the 5-fold method was used, i.e. $23-24$ data points are data points for training and $5-6$ data points for testing. The linear regressor is fitted with the training data, tested with the test set and scored with the R^2 score and the maximal error averaged over all 5 folds. In Figure 5 the line for the predicted values is drawn for 12 seconds at level 0. For the preceding activity $1, 2, 8, 12, 16$ and 100 seconds are tested with the levels $0-3$ as shown in Figure 3. The highest score of R^2 occurred for 1 second at level 2 (input is 4-dimensional), 2 seconds at level 1 (input is 2-dimensional) and 12 seconds at level 2. All results can be found in Table 1. The seconds column is the time for the considered activity.

Seconds	Level	R^2	Max Error
1	0	0.8046	0.3765
1	1	0.8726	0.3300
1	2	0.8837	0.3006
1	3	0.4286	0.6446
2	0	0.8670	0.3181
2	1	0.8892	0.2901
2	2	0.8618	0.3024
2	3	0.8682	0.2953
8	0	0.8494	0.2989
8	1	0.8608	0.3323
8	2	0.6317	0.5033
8	3	0.8318	0.3691
12	0	0.6746	0.5175
12	1	0.8524	0.3469
12	2	0.8823	0.3173
12	3	0.5328	0.6099
100	0	0.7126	0.4865
100	1	0.3365	0.6955
100	2	0.2329	0.7749
100	3	0.4654	0.5908

Table 1: R^2 and maximal error for linear regression model, that uses the number of spikes as input and the response time as output.

3.2.2 All intervals

The microneurography record contains about 6850 regular electrical stimuli.

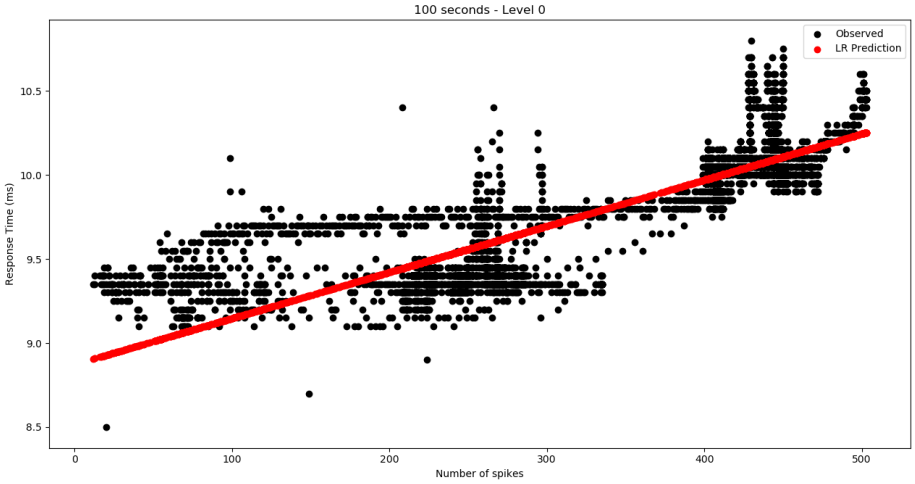

Figure 6: Linear regression for 100 seconds on level 0 for 6850 data points. The number of spikes is plotted against the response time in ms.

Seconds	Level	R^2	Mean squared error
4	0	0.495	0.067
4	1	0.494	0.067
4	2	0.493	0.067
12	0	0.668	0.044
12	1	0.671	0.043
12	2	0.674	0.043
100	0	0.736	0.035
100	1	0.773	0.030
100	2	0.802	0.026
100	3	0.809	0.025
100	4	0.812	0.025
400	0	0.509	0.065
400	1	0.681	0.042
400	2	0.749	0.033

Table 2: R^2 and mean squared error for linear regression model, that uses the number of spikes as input and the response time as output.

Considering all intervals with regular electrical stimulation and using 5-fold, about 5500 data points are used for training and about 1375 data points for testing.

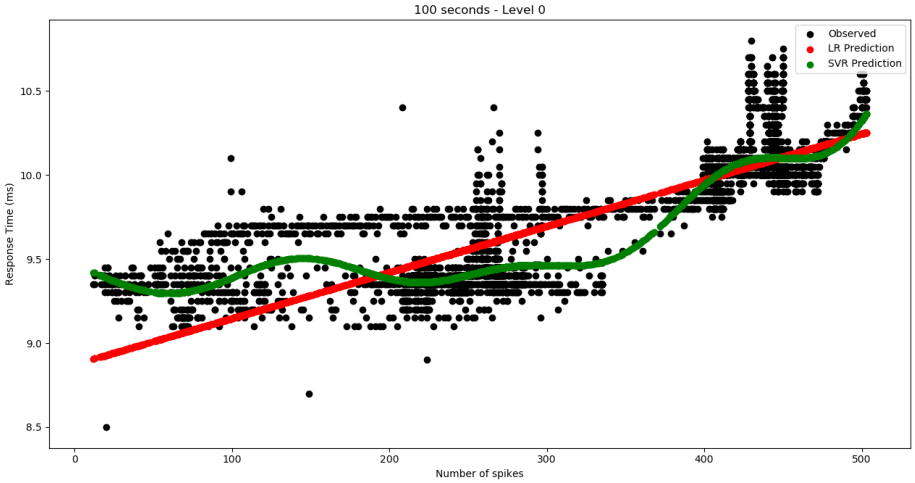

Figure 7: Linear regression and support vector regression for 100 seconds on level 0 for 6850 data points. The number of spikes is plotted against the response time in ms.

Seconds	Level	R^2	Mean squared error
4	0	0.787	0.028
4	1	0.806	0.025
4	2	0.804	0.026
12	0	0.818	0.024
12	1	0.859	0.017
12	2	0.871	0.017
100	0	0.836	0.021
100	1	0.889	0.014
100	2	0.929	0.009
100	3	0.946	0.007
100	4	0.953	0.006
400	0	0.607	0.052
400	1	0.873	0.016
400	2	0.900	0.013

Table 3: R^2 and mean squared error for support vector regression model, that uses the number of spikes as input and the response time as output.

The results for linear regression are shown in Table 2 and the plot for a 400 second interval on level 0 is shown in Figure 6. The same results for support vector regression are shown in Table 3 and Figure 7. The best result were obtained for 100 seconds. The results are averaged over all 5-folds.

4 Discussion

The aim of this work was to investigate and model the dependence between preceding activities and latency changes of a fiber. For the modeling, a feature space was defined to describe the activities based on the number of action potentials in a time interval. The activities are the input values for the model and the response times of the fiber are the output values.
Since the fiber recording data is limited, the feature space focuses on the timing and placement of the action potentials to keep the dimension small as possible and to avoid overfitting. In this work, only one single fiber recording was used and two data sets were extracted from it.
The first data set considered intervals with regular electrical and additional mechanical stimulation. The recording contains 29 intervals with both stimuli. It was the first approach to use intervals with additional mechanical stimuli to consider evoked activity only. The best results for these 29 data pairs and these splits of training and test set were obtained for 1 second at level 2, 2 seconds at level 1 and 12 seconds at level 2 using linear regression, indicating that the selected features are indeed useful for modeling and machine learning. However, since it is known that stimulation of a fiber can affect the latency over minutes [4], the next step was to extend the data set and add all activities, including those with only regular electrical stimuli.
This second data set contains about 6850 data points and achieved the best results for 100 seconds on an increasing level using support vector regression. At a point, where the number of input features is further increased, there is no significant improvement anymore and for 100 seconds it starts at level 3, i.e. 8 input values. The mean squared error is .006 and a value of 0 indicates that the predictions fit the data points perfectly. This shows that the defined feature space is promising and the results, measured by the mean squared error are satisfactory for the second data set and encourages to continue this approach of modeling with the created feature space for future work.

Before this approach can be further pursued, the following problems must be solved.
One problem is the division of training and test data, which is a very important step in machine learning to evaluate the performance of the model. In this work,

the train and test data are randomly split. Another option could be to use the first 70% of the recording for training and the last 30% for testing. The next step would be to consider multiple data sets, i.e. recordings from different subjects and to split them into train and test sets after they have been calibrated and made comparable.

The R^2 score is not a valid method for evaluating nonlinear regression [7]. It has only been used to compare the performance of the linear and the support vector regression model. In the future, this should be replaced by one of the various other methods for calculating the goodness of fit suitable for nonlinear regression.

And although the results for the support vector regression are good, the calculated function, for example the green curve in Figure 7, does not necessarily describe the actual relationship between the number of spikes and the latency, but only the best fit. Since it is known that the latency increases almost linearly with additional electrical spikes between regular stimulation [4], the question arises how the mechanically induced spikes and its placement in different preceding interval parts influences the latency and whether linear or even nonlinear behavior can be found. The defined feature space has to be tested with regular data, i.e. the same regular electrical pulses throughout the whole experiment in order to be able to make a statement about how the actual dependencies of spikes and response time might look like.

Finally, the data sets obtained from animal fiber recordings have ideal properties and quality for modeling because the number of fibers was limited and only low noise was present. However, the main goal is to be able to test the features with microneurographic human data containing multiple fibers and noise. Therefore, the next step is it to solve the problem of fiber separation and filter out the noise to prove that the features work for this purpose.

5 Conclusion

The aim of this work was to investigate the feasibility of machine learning and to model the relationship between the number of action potentials and the response time for a pressure-clamped fiber recording of a rat. To the best of our knowledge machine learning was applied for the first time to make predictions for the latency based on the defined feature space. The goal to define a feature space by quantifying the spikes to describe the preceding activity and to predict the latency of a fiber was achieved. After solving the discussed problems with human microneurographic data, this approach will be continued in order to gain useful insights into neuropathic pain and the associated spontaneous activities of C-fibers.

References

[1] van Hecke O, Austin SK, Khan RA, Smith BH, Torrance N. Neuropathic pain in the general population: a systematic review of epidemiological studies. Pain. 2014;155(4):654–62.

[2] Brooks KG, Kessler TL. Treatments for neuropathic pain. Clinical Pharmacist. 2017 12;9.

[3] Serra J, Bostock H, Sol R, Aleu J, Garca E, Cokic B, et al. Microneurographic identification of spontaneous activity in C-nociceptors in neuropathic pain states in humans and rats. Pain. 2012 01;153(1).

[4] Schmelz M, Forster C, Schmidt R, Ringkamp M, Handwerker H O , Torebjörk H E . Delayed responses to electrical stimuli reflect C-fiber responsiveness in human microneurography. Experimental Brain Research. 1995 May;104(2):331–336.

[5] Forster C, Handwerker HO. Automatic classification and analysis of microneurographic spike data using a PC/AT. Journal of Neuroscience Methods. 1990;31(2):109 – 118. Available from: `http://www.sciencedirect.com/science/article/pii/0165027090901559`.

[6] Spike2 - Template Matching;. `http://ced.co.uk/us/products/spksstm#`.

[7] Spiess AN, Neumeyer N. An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A Monte Carlo approach. BMC pharmacology. 2010 06;10:6.