Specification of neck muscle dysfunction through digital image analysis using machine learning

Filip Paskali[1], Angela Dieterich[2] and Matthias Kohl[3]

[1]Institute of Precision Medicine, Medical and Life Sciences, Hochschule Furtwangen, Villingen-Schwenningen
F.Paskali@hs-furtwangen.de
[2] Physiotherapie, Fakultät Gesundheit, Gesellschaft, Hochschule Furtwangen, Studienzentrum Freiburg
Angela.Dieterich@hs-furtwangen.de

Abstract. Everyone has or will have experience some degree of neck pain. Typically, neck pain is associated with the sensation of tense, tight or stiff neck muscles. However, it is unclear whether the neck muscles are objectively stiffer with neck pain. Some investigations documented higher stiffness of the neck muscles with neck pain, while others did not find differences compared to asymptomatic study participants. This is a cross-sectional, observational study that analyses shear wave elastography data obtained from 38 women. In this study, we trained machine learning models that can classify the shear wave elastography images at the level of an expert. Knowledge on a potentially increased objective stiffness of the neck muscles is important when related to diagnosis or therapeutic decisions. Moreover, such an automated approach enables a computed image analysis, which may provide new insights of the physiological properties of the neck muscles in individuals suffering from neck pain.

Keywords: Neck pain 1; Shear Wave Elastography 2; Ultrasound 3; Image Analysis 4; Machine Learning 5.

1 Introduction

Almost everyone has or will have experience some degree of neck pain [1]. The one-year prevalence of disabling neck pain has been estimated between 1.7% and 11.5% [2], and the risk is especially high in middle-aged woman who have already experience with neck pain [3]. Typically, neck pain is associated with the sensation of tense, tight or stiff neck muscles [4]. Therapeutic interventions often include measures to reduce neck muscle tone. However, it is unclear whether the neck muscles are objectively stiffer with neck pain. Some investigations documented higher stiffness of the neck muscles with neck pain [5], [6], while others did not find differences compared to asymptomatic study participants [7], [8]. Knowledge on a potentially increased objective stiffness of the neck muscles is important when related to diagnosis or therapeutic decisions.

Ultrasound shear elastography is the superior and reliable modality for measurement of tissue stiffness [9], [10]. However, the methods to measure muscle stiffness vary between studies. The active and individual positioning of the region of interest has biasing potential. Some studies reported inter-session stiffness variability [11]. An automated image analysis might provide new insights into differences of the mechanical properties of the neck muscles between individuals with and without chronic neck pain.

Recently, there has been a growing number of publications focusing on the usage of the artificial intelligence algorithms for medical diagnostics in ultrasonography imaging. The machine learning algorithms have been employed for classification of chronic liver diseases [12], breast tumors [13], thyroid nodules [14], and prostate cancer [15]. On other hand, deep learning convolutional neural network algorithms have been used for segmentation of muscles, as well as automatic measurement of muscle thickness and muscle fat infiltration [16]–[18]. To the best of our knowledge, there has been no attempts to employ the machine learning algorithms in ultrasound elastography images for classification of neck pain.

The experimental work presented here provides a fully automatic investigation using machine learning techniques to examine in detail group differences in two-dimensional ultrasound elastography images measured in adult woman suffering from neck pain and women without symptoms.

2 Materials and Methods

2.1 Study

This is a cross-sectional, observational study that analyses data obtained from 38 women. The participants belong to two groups, 20 women suffering from chronic neck pain and 18 women without symptoms. The inclusion criteria for the pain group were non-specific pain longer than six months with symptoms duration of at least one week, neck stiffness sensation and Neck Disability Index higher than 10/50. On other hand, for the control group, no history of recurrent neck or low back pain or neck pain that affected the neck function or required treatment. Exclusion criteria were major circulatory, neurological, or respiratory disorders, pregnancy, cervical spinal surgery, participation in neck muscle training in past 6 months, body mass index higher than 30 and intake of medication that can affect muscle stiffness. On the days of the imaging, participants were asked to not use pain medication. As shown in the demographic summary (Table 1), the sample size, age, and body mass index of the two groups is comparable. At the same time, it is noticeable that the range of motion in flexion and extension, and maximal voluntary isometric contraction is limited in the pain group.

Table 1. Demographic summary: mean ± SD. Abbreviations: MVIC, maximal voluntary isometric contraction; NRS, numerical rating scale.

	Neck pain, n = 20	**Control, n = 18**
Age (years)	52.5 (12.0)	48.5 (9.0)
Body Mass Index (kg/m^2)	23.8 ± 3.2	22.2 ± 2.4
Range of Motion neck flexion	50.6° ± 12.4°	69.4° ± 10.8°
Range of Motion neck rotation (left + right)	115.5° ± 14.3°	145.5° ± 26.6°
MVIC neck extension (N)	56.3 ± 18.5	64.0 ± 19.2
Pain today (NRS)	3.6 ± 2.2	0
Pain duration (years)	8.0 (16.3)	-
Neck Disability Index % (scored 0% - 100%)	32.5% ± 12.3%	-

Participants performed four diverse activities including head lift from prone, stressful office work, balancing a weight (1 kg) on the head, and graded isometric neck extension under force levels of 12 N, 24 N, 36 N and 48 N. All activities were repeated and measured in three sessions. More information about the study cohort can be found in Dieterich et al. 2020 [7].

2.2 Image acquisition

Shear wave elastography images were obtained on the neck extensor muscle group 1 cm lateral to the spinous processes, the transducer centered at the C4 level in longitudinal orientation (Acuson S3000; Siemens, Germany, 9L4 linear transducer with 4 cm footprint). Maximal shear wave speed of 10m/s was set. Gain (14-20 dB), dynamic range (45-65 dB), and image depth 3.5-4.5 cm were adjusted for good visualization. In total, 1099 images were recorded. More information about the imaging in Dieterich et al, 2020 [7]

Each measurement produced three images, an ultrasound image; a color coded elastography image, indicating the shear wave velocity in the region of interest with range 0.5 -10 m/s; and a color-coded quality map presenting the quality of each pixel ranging from low to high quality. The images have a resolution of 1024 x 768 and are stored in BMP file format.

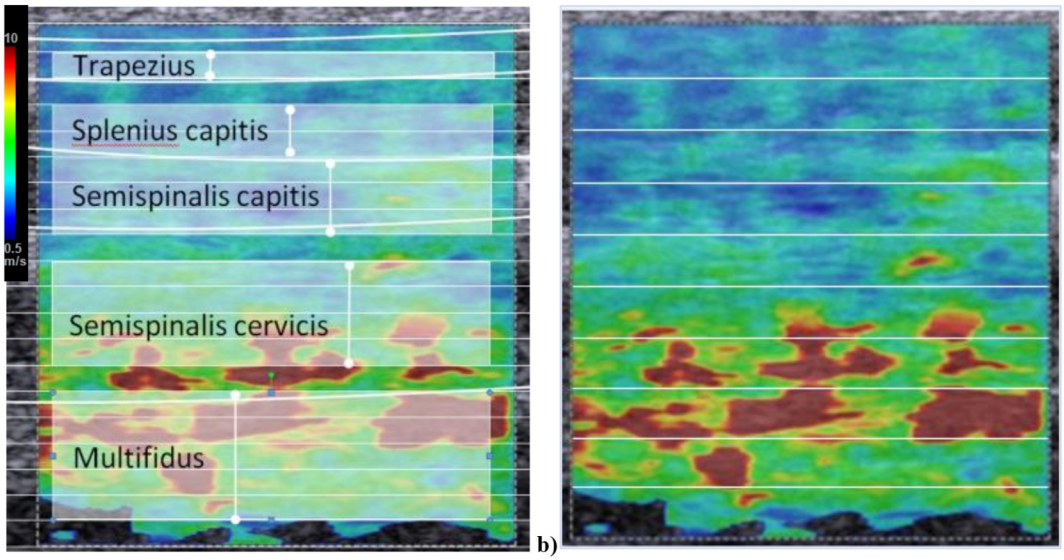

Fig 1. a) Color coded elastography shear wave image with muscle layer organization b) Color coded elastography shear wave image with representation of the horizontal segments.

2.3 Image Analysis

The image analysis and feature extraction was carried out using the free open-source programming language Python [19].

For the machine learning, 30 diverse features were handcrafted, extracting numeric information from the shear wave velocity color-coded images, as described below.

The first set of numerical features was computed by performing various statistical operations to the whole image, such as computing grand mean, median and standard deviation. Furthermore, the images were separated in three color channels, red, green, and blue. The statistical operations and grand sum were computed for each color channel of each image. The images were converted into grayscale color mode and same features were extracted.

For the second set of features, an inverse mapping was performed from RGB color mode to stiffness using the full range shear velocity scale (Figure 1a). The number of pixels in the images was quantified from three stiffness level categories: low level stiffness category with shear velocity between 0,5 – 3,67 m/s, medium level stiffness category with shear velocity between 3,67 – 6,84 m/s, and high level stiffness category with shear velocity between 6,84 – 10 m/s. Additionally, the mean area and the weighted average distance of high level stiffness areas from the top edge of the images was calculated. As weight, the size of the area was used.

The third set of features was obtained with Gabor filtering, which in several studies was used for texture classification [20]–[22]. In this study, we used mean and variance of the resulting convoluted image as features, using the following four different Gabor kernels (Fig. 2).

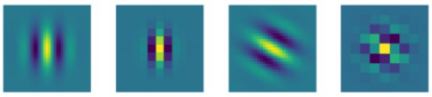

Fig 2. The Gabor kernels used for convolution

Finally, the images were divided in ten equal horizontal segments (Figure 1b) and the aforementioned operations were repeated for each of the segments leading to total number of 330 features. To the 330 features, the 14 activities performed by the participants, together with the imaging session number were added as dummy variables. Overall, this gives 345 features extracted from 1099 images.

2.4 Machine learning

A supervised binary classification was performed, using six classification machine learning algorithms from the package scikit-learn [23]. The main task was a prediction of one of the two target classes: participants with neck pain and participants without neck pain. The class label was assigned to the images before the machine learning. The training was repeated 50 times, with a new random train-test split, where the model was trained with 1049 images 10-fold cross validated and tested with 50 held-out images. Finally, in order to investigate the feature importance, the best performing model was trained with the whole dataset and feature importance were extracted.

3 Results

Table 2 displays the arithmetic mean and the confidence interval of performance metrics, as empirical quantiles (2.5%; 97.5%) computed from 500 cross-validation results.

Table 2. Summary of the cross-validation during training of the classification machine learning algorithms (mean, confidence interval 95%) Abbreviations: KNN – K-Nearest Neighbor; SVM – Support Vector Machine; AUC – Area Under Curve.

Metrics	KNN	Logistic Regression	Naïve Bayes	SVM	Decision Trees	Random Forest
Accuracy	0. 789 (0.785-0.792)	0.703 (0.700-0.707)	0.622 (0.618-0.625)	0.710 (0.706-0.714)	0.707 (0.703-0.711)	**0.810** (0.807-0.813)
Balanced Accuracy	0.789 (0.786-0.792)	0.701 (0.698-0.705)	0.620 (0.617-0.624)	0.709 (0.705-0.713)	0.706 (0.702-0.709)	**0.808** (0.804-0.811)
Precision	0.763 (0.759-0.768)	0.686 (0.682-0.691)	0.593 (0.588-0.597)	0.690 (0.686-0.695)	0.690 (0.685-0.695)	**0.815** (0.811-0.819)
Recall	**0.796** (0.791-0.801)	0.674 (0.668-0.680)	0.605 (0.599-0.612)	0.690 (0.684-0.696)	0.680 (0.674-0.686)	0.769 (0.763-0.775)
AUC Score	0.849 (0.846-0.852)	0.774 (0.771-0.778)	0.683 (0.678-0.687)	0.776 (0.773-0.780)	0.706 (0.702-0.709)	**0.895** (0.892-0.897)
Brier Score	0.157 (0.155-0.159)	0.201 (0.199-0.204)	0.364 (0.360-0.368)	0.196 (0.195-0.198)	0.293 (0.289-0.297)	**0.152** (0.151-0.153)

The trained models with 1049 images were used to predict classes of the 50 held-out images. The result of this evaluation is summarized in Table 3.

Table 3. The summary of the results obtained 50 evaluations with 50 held-out images of the classification machine learning models. Abbreviations: KNN – K-Nearest Neighbor; SVM – Support Vector Machine; AUC – Area Under Curve.

Metrics	KNN	Logistic Regression	Naïve Bayes	SVM	Decision Trees	Random Forest
Accuracy	0.792 (0.779-0.805)	0.695 (0.677-0.713)	0.619 (0.600-0.638)	0.697 (0.679-0.715)	0.724 (0.708-0.739)	**0.811** (0.795-0.827)
Balanced Accuracy	0.791 (0.777-0.805)	0.691 (0.673-0.710)	0.618 (0.599-0.638)	0.694 (0.675-0.713)	0.722 (0.706-0.739)	**0.807** (0.791-0.823)
Precision	0.763 (0.741-0.786)	0.673 (0.646-0.699)	0.590 (0.562-0.617)	0.668 (0.643-0.693)	0.699 (0.677-0.721)	**0.807** (0.786-0.828)
Recall	**0.802** (0.778-0.826)	0.669 (0.635-0.704)	0.612 (0.585-0.639)	0.689 (0.658-0.721)	0.714 (0.685-0.743)	0.780 (0.752-0.808)
AUC Score	0.854 (0.842-0.865)	0.757 (0.737-0.778)	0.677 (0.656-0.698)	0.760 (0.739-0.780)	0.722 (0.706-0.739)	**0.892** (0.879-0.905)
Brier Score	0.154 (0.146-0.162)	0.209 (0.199-0.220)	0.368 (0.349-0.387)	0.201 (0.194-0.208)	0.276 (0.261-0.292)	**0.150** (0.145-0.156)

An inspection of the data in Table 2 and 3 reveals that there is an overlap between confidence intervals of training and testing results, with slightly decrease in the performance during the testing step. Random Forest

model outperforms other models in almost every performance metrics; hence it has the best performance during the training and testing step. The second-best performing algorithm is K-Nearest Neighbor.

The most important features extracted from the Random Forest model trained with the whole dataset are presented in Table 4. What is interesting in this table is that most of the represented features are coming from the deepest horizontal segment, that corresponds to the multifidus muscle closest to the spine.

Table 4. Top 20 most important features extracted from trained Random Forest model

	Features
1	blue_layer_value_sum_hsegment_1/10
2	blue_layer_value_sum_hsegment_2/10
3	red_layer_value_sum_hsegment_1/10
4	blue_pixels
5	gray_value_min_hsegment_10/10
6	gray_value_hsegment_9/10
7	red_layer_value_median_hsegment_10/10
8	gray_value_median_hsegment_10/10
9	red_layer_value_sum_hsegment_2/10
10	gabor_kernel4_mean_hsegment_10/10
11	gabor_kernel1_mean_hsegment_10/10
12	median_hsegment_10/10
13	green_layer_value_median_hsegment_10/10
14	gabor_kernel3_mean_hsegment_10/10
15	mean_hsegment_10/10
16	red_layer_value_mean_hsegment_10/10
17	gray_value_hsegment_3/10
18	gray_value_mean_hsegment_10/10
19	gray_value_hsegment_6/10
20	blue_layer_value_sum_hsegment_3/10

4 Discussion

Previous studies, when analyzing neck muscle pain, mainly focused on the superficial neck muscles [5], [6], [8], [24]. Dieterich et al, 2020 was the first study that included the deep muscles. Xie et al. 2019 reported a problem in assessment of the deep muscles, in terms of high variability [25]. Furthermore, Haldemann et al. 2009 also noted that the measurement in shear wave elastography is very sensitive to the orientation of muscle fibers [1]. Only few studies have found a correlation between neck muscle pain and stiffness levels in shear wave elastography images [5], [6], while other studies could not find significant difference [7], [8]. An initial objective of this research was to successfully train a machine learning model, that can predict with high accuracy the group of an elastography shear wave image. So far, the neck pain diagnostics is a complex task, based on the subjective sensation of the patient and the diagnostic experience of the trained physician [4]. A machine learning model that can classify with an accuracy of or better than a trained diagnostician will improve medical diagnosis and therapy.

In this study we trained and evaluated a machine learning model that can diagnose neck pain with high accuracy by analyzing selected features in shear wave elastography images. Closer inspection of the Table 2 and 3 shows an overlap of the confidence intervals of training and testing results, with tendency of slightly lower performance in the testing results. However, there are a few exceptions where the mean result is slightly higher in the results of testing step compared to the results from training step. This might be a result of heterogeneity of the medical data, and the variance introduced by the pseudo-random selection process, or overfitting due to the high number of features in the training step. This assumption is supported by the fact that the results from

the algorithms that have an embedded feature selection are clearly better than the other algorithms. The possibility to train a classification machine learning algorithm with very good accuracy provides some tentative initial evidence that there is a correlation between subjective sensation of stiffness and tissue properties recorded with shear wave elastography modulus. Our research is still in progress, therefore further work has to be done with focus on feature selection and hyper-parameter optimization in order to attain more generalizable well performing model.

The second objective of this study was to find a correlation between extracted features and physiological tissue properties, with the intention to find reliable biomarkers that can be used to objectify and enhance the future diagnostics. What stands out in the Table 4 is that the most important features chosen from the Random Forest algorithm predominantly are extracted from the lowest horizontal segment which correspond to the deepest neck muscle. This might suggest that the main difference is located in the deep neck muscles, closest to the spine. Contrary to our expectations, the activities that participants performed do not have an important role in the classification. Due to complexity of the features extracted from the images, it is not easy to find a direct link between these features and physiological neck muscle tissue properties. Further investigations will be necessary to gain a better understanding of a possible correlation between information in shear wave elastography images and biological properties, and to better understand the mechanisms underlying neck muscle pain.

5 Conclusion

We successfully trained machine learning algorithms that can classify the shear wave elastography images at the level of an expert. Neck pain is a complex condition/experience and until now the diagnostics are highly subjective, dependent on the participants personal sensation of pain, and the diagnostic experience of the diagnostician. The algorithms used in this study could be regarded as first step to objectify the neck pain diagnosis. Moreover, such an automated approach enables a computed image analysis, which may provide new insights into the differences of the physiological properties of the neck muscles in individuals with neck pain compared to asymptomatic individuals. Further research is required to better understand the discrepancies between both groups and ascertain the possibility of adopting the shear wave elastography as a robust and reliable diagnostic tool for neck pain diagnosis.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Authors' Contribution

F. P. wrote the manuscript and carried out the analysis. A.D. provided the sample datasets and supervised the project. M.K. contributed to the interpretation of the results and supervised the project.

Acknowledgements

The work was supported by an internal grant from Hochschule Furtwangen.

References

1. S. Haldeman, L. J. Carroll, and J. D. Cassidy, "The empowerment of people with neck pain: introduction. The Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders," *J. Manipulative Physiol. Ther.*, vol. 32, no. 2 Suppl, Art. no. 2 Suppl, Feb. 2009, doi: 10.1016/j.jmpt.2008.11.006.
2. S. Hogg-Johnson *et al.*, "The burden and determinants of neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders," *Spine*, vol. 33, no. 4 Suppl, Art. no. 4 Suppl, Feb. 2008, doi: 10.1097/BRS.0b013e31816454c8.
3. P. R. Blanpied *et al.*, "Neck Pain: Revision 2017," *J. Orthop. Sports Phys. Ther.*, vol. 47, no. 7, Art. no. 7, Jul. 2017, doi: 10.2519/jospt.2017.0302.

4. J. C. MacDermid, D. M. Walton, P. Bobos, M. Lomotan, and L. Carlesso, “A Qualitative Description of Chronic Neck Pain has Implications for Outcome Assessment and Classification,” *Open Orthop. J.*, vol. 10, pp. 746–756, 2016, doi: 10.2174/1874325001610010746.
5. S. Taş, F. Korkusuz, and Z. Erden, “Neck Muscle Stiffness in Participants With and Without Chronic Neck Pain: A Shear-Wave Elastography Study,” *J. Manipulative Physiol. Ther.*, vol. 41, no. 7, Art. no. 7, Sep. 2018, doi: 10.1016/j.jmpt.2018.01.007.
6. J. Hvedstrup, L. T. Kolding, M. Ashina, and H. W. Schytz, “Increased neck muscle stiffness in migraine patients with ictal neck pain: A shear wave elastography study,” *Cephalalgia*, vol. 40, no. 6, Art. no. 6, May 2020, doi: 10.1177/0333102420919998.
7. A. V. Dieterich, U. Ş. Yavuz, F. Petzke, A. Nordez, and D. Falla, “Neck Muscle Stiffness Measured With Shear Wave Elastography in Women With Chronic Nonspecific Neck Pain,” *J. Orthop. Sports Phys. Ther.*, vol. 50, no. 4, Art. no. 4, Apr. 2020, doi: 10.2519/jospt.2020.8821.
8. H. Ishikawa *et al.*, “Changes in stiffness of the dorsal scapular muscles before and after computer work: a comparison between individuals with and without neck and shoulder complaints,” *Eur. J. Appl. Physiol.*, vol. 117, no. 1, Art. no. 1, Jan. 2017, doi: 10.1007/s00421-016-3510-z.
9. R. Akagi and S. Kusama, “Comparison Between Neck and Shoulder Stiffness Determined by Shear Wave Ultrasound Elastography and a Muscle Hardness Meter,” *Ultrasound Med. Biol.*, vol. 41, no. 8, Art. no. 8, Aug. 2015, doi: 10.1016/j.ultrasmedbio.2015.04.001.
10. A. V. Dieterich *et al.*, “Shear wave elastography reveals different degrees of passive and active stiffness of the neck extensor muscles,” *Eur. J. Appl. Physiol.*, vol. 117, no. 1, Art. no. 1, Jan. 2017, doi: 10.1007/s00421-016-3509-5.
11. Ž. Kozinc and N. Šarabon, “Shear-wave elastography for assessment of trapezius muscle stiffness: Reliability and association with low-level muscle activity,” *PloS One*, vol. 15, no. 6, Art. no. 6, 2020, doi: 10.1371/journal.pone.0234359.
12. I. Gatos *et al.*, “A Machine-Learning Algorithm Toward Color Analysis for Chronic Liver Disease Classification, Employing Ultrasound Shear Wave Elastography,” *Ultrasound Med. Biol.*, vol. 43, no. 9, pp. 1797–1810, Sep. 2017, doi: 10.1016/j.ultrasmedbio.2017.05.002.
13. Y.-J. Mao, H.-J. Lim, M. Ni, W.-H. Yan, D. W.-C. Wong, and J. C.-W. Cheung, “Breast Tumour Classification Using Ultrasound Elastography with Machine Learning: A Systematic Scoping Review,” *Cancers*, vol. 14, no. 2, Art. no. 2, Jan. 2022, doi: 10.3390/cancers14020367.
14. B. Zhang *et al.*, “Machine Learning–Assisted System for Thyroid Nodule Diagnosis,” *Thyroid*, vol. 29, no. 6, pp. 858–867, Jun. 2019, doi: 10.1089/thy.2018.0380.
15. R. R. Wildeboer *et al.*, “Automated multiparametric localization of prostate cancer based on B-mode, shear-wave elastography, and contrast-enhanced ultrasound radiomics,” *Eur. Radiol.*, vol. 30, no. 2, pp. 806–815, Feb. 2020, doi: 10.1007/s00330-019-06436-w.
16. K. Orhan, G. Yazici, M. E. Kolsuz, N. Kafa, I. S. Bayrakdar, and Ö. Çelik, “An Artificial Intelligence Hypothetical Approach for Masseter Muscle Segmentation on Ultrasonography in Patients With Bruxism,” *J. Adv. Oral Res.*, vol. 12, no. 2, pp. 206–213, Nov. 2021, doi: 10.1177/23202068211005611.
17. K. A. Weber *et al.*, “Multi-muscle deep learning segmentation to automate the quantification of muscle fat infiltration in cervical spine conditions,” *Sci. Rep.*, vol. 11, no. 1, p. 16567, Aug. 2021, doi: 10.1038/s41598-021-95972-x.
18. I. Loram *et al.*, “Objective Analysis of Neck Muscle Boundaries for Cervical Dystonia Using Ultrasound Imaging and Deep Learning,” *IEEE J. Biomed. Health Inform.*, vol. 24, no. 4, pp. 1016–1027, Apr. 2020, doi: 10.1109/JBHI.2020.2964098.
19. G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.
20. C. Palm and T. M. Lehmann, “Classification of color textures by gabor filtering,” vol. 11, no. 2, p. 26, 2002.
21. F. Bianconi and A. Fernández, “Evaluation of the effects of Gabor filter parameters on texture classification,” *Pattern Recognit.*, vol. 40, no. 12, pp. 3325–3335, Dec. 2007, doi: 10.1016/j.patcog.2007.04.023.
22. M. Idrissa and M. Acheroy, “Texture classification using Gabor filters,” *Pattern Recognit. Lett.*, vol. 23, no. 9, pp. 1095–1102, Jul. 2002, doi: 10.1016/S0167-8655(02)00056-9.
23. F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, no. 85, Art. no. 85, 2011.
24. J. Aljinović *et al.*, “Can measuring passive neck muscle stiffness in whiplash injury patients help detect false whiplash claims?,” *Wien. Klin. Wochenschr.*, vol. 132, no. 17–18, Art. no. 17–18, Sep. 2020, doi: 10.1007/s00508-020-01631-y.
25. Y. Xie, L. Thomas, F. Hug, V. Johnston, and B. K. Coombes, “Quantifying cervical and axioscapular muscle stiffness using shear wave elastography,” *J. Electromyogr. Kinesiol. Off. J. Int. Soc. Electrophysiol. Kinesiol.*, vol. 48, pp. 94–102, Oct. 2019, doi: 10.1016/j.jelekin.2019.06.009.