ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

From the right part of Fig. 3, we can make the following observations. Similar to the removal of centroids, the achieved mean accuracy values are better when the outliers are defined based on all data samples (MAX), instead of on prototypes ($\overline{MAX}$). The best overall result is achieved by the MIN approach when one single participant is removed from the training set. Moreover, the top three results were obtained by the MIN approach (for the removal of 1, 4 and 10 participants from the training set).
Note that the BVDB has the special property that each participant has exactly the same amount of data samples. For imbalanced data sets, where each participant is represented by a different number of data samples, one has to find an appropriate weighting factor for the sum of distances in the left part of Eq. (1), e.g. one divided by the number of distance values.

6 Conclusion

In this study, we introduced our idea of defining the participant with the lowest sum of distances to all other participants, i.e. the central point, as a candidate for training data clean up. For future work, we define the following research directions. First, one should test our approach, which we denoted by MIN, on other affect related data sets to confirm its effectiveness. Second, in this work we evaluated four different distance based approaches for participant based outlier and centroid detection. It could also be beneficial to combine (the best) two of the proposed methods. And third, instead of compressing the data specific to one participant to one single prototype ($\bar{d}$ from Eq. (1)), one could compute class-specific prototypes for each participant to preserve some information of each participant's data distribution. In general, the reported outcomes could motivate the implementation of novel approaches for data preprocessing/training set selection techniques, based on the removal of centroids.

References

[1] Steffen Walter, Sascha Gruss, Hagen Ehleiter, Jun-Wen Tan, Harald C. Traue, Stephen Clive Crawcour, Philipp Werner, Ayoub Al-Hamadi, and Adriano O. Andrade. The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system. In *CYBCONF*, pages 128–131. IEEE, 2013.

[2] Markus Kächele, Patrick Thiam, Mohammadreza Amirian, Friedhelm Schwenker, and Günther Palm. Methods for person-centered continuous pain intensity assessment from bio-physiological channels. *J. Sel. Topics Signal Processing*, 10(5):854–864, 2016.

[3] Patrick Thiam, Viktor Kessler, and Friedhelm Schwenker. Hierarchical combination of video features for personalised pain level recognition. In *ESANN*, pages 465–470, 2017.

[4] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[5] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[6] Peter Bellmann, Patrick Thiam, and Friedhelm Schwenker. *Multi-classifier-Systems: Architectures, Algorithms and Applications*, pages 83–113. Springer International Publishing, Cham, 2018.

[7] Patrick Thiam, Peter Bellmann, Hans A. Kestler, and Friedhelm Schwenker. Exploring deep physiological models for nociceptive pain recognition. *Sensors*, 19(20):4503, 2019.