Estimation of biological variance in coherent Raman microscopy data of two cell lines using chemometrics

Rajendhar Junjuri,[a,b] Matteo Calvarese,[a] MohammadSadegh Vafaeinezhad,[a,b,c] Federico Vernuccio,[d] Marco Ventura,[d,e] Tobias Meyer-Zedler,[a,b] Benedetta Gavazzoni,[d] Dario Polli,[d,e] Renzo Vanna,[e] Italia Bongarzone,[f] Silvia Ghislanzoni[f], Matteo Negro[g], Juergen Popp[a,b,c] and Thomas Bocklitz[*,a,b]

[a]Leibniz Institute of Photonic Technology, Member of Leibniz Health Technologies, Member of the Leibniz Centre for Photonics in Infection Research (LPI), Albert-Einstein-Strasse 9, 07745 Jena, Germany
[b]Institute of Physical Chemistry (IPC) and Abbe Center of Photonics (ACP), Friedrich Schiller University Jena, Member of the Leibniz Centre for Photonics in Infection Research (LPI), Helmholtzweg 4, 07743 Jena, Germany
[c]Max Planck School of Photonics, Jena, Germany
[d]Department of Physics - Politecnico di Milano, P.za L. da Vinci 32, 20133 Milano, Italy
[e]Istituto di Fotonica e Nanotecnologie – CNR, P.za L. da Vinci 32, 20133 Milano, Italy
[f]Department of Diagnostic Innovation, Fondazione IRCCS Istituto Nazionale dei Tumori, Via Giacomo Venezian 1, 20133, Milano, Italy
[g]Cambridge Raman Imaging Ltd, Cambridge, UK

[*]Corresponding author e-mail address: Thomas.bocklitz@uni-jena.de

Supplementary data

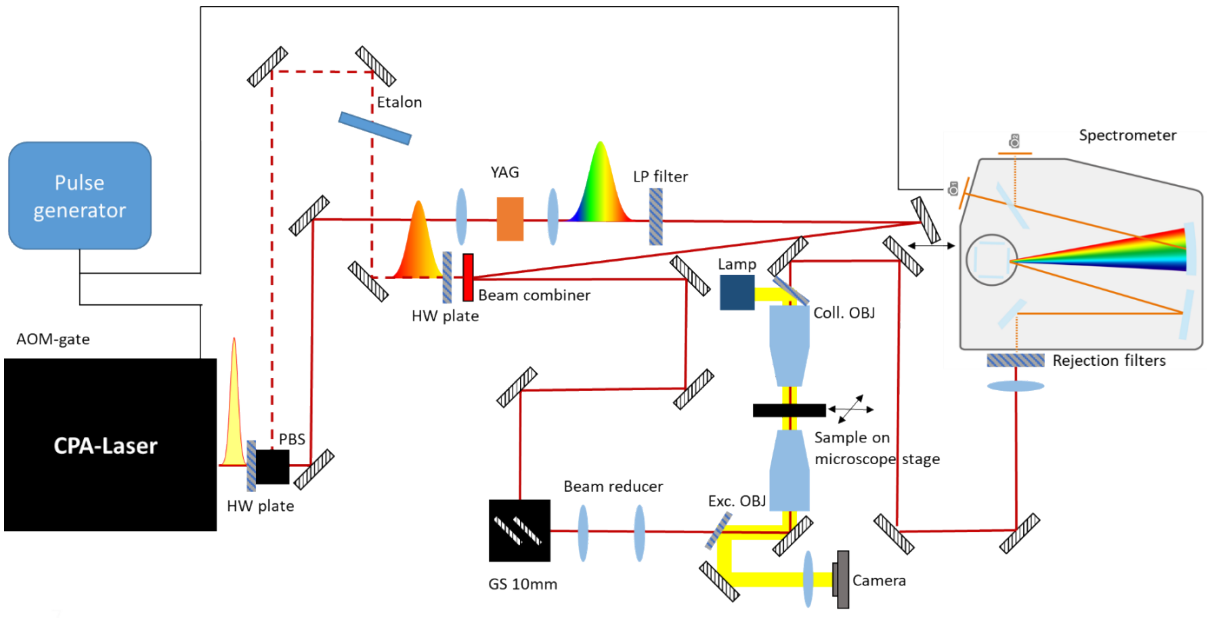

Fig. 1. Broadband CARS setup at the IPHT in Jena.

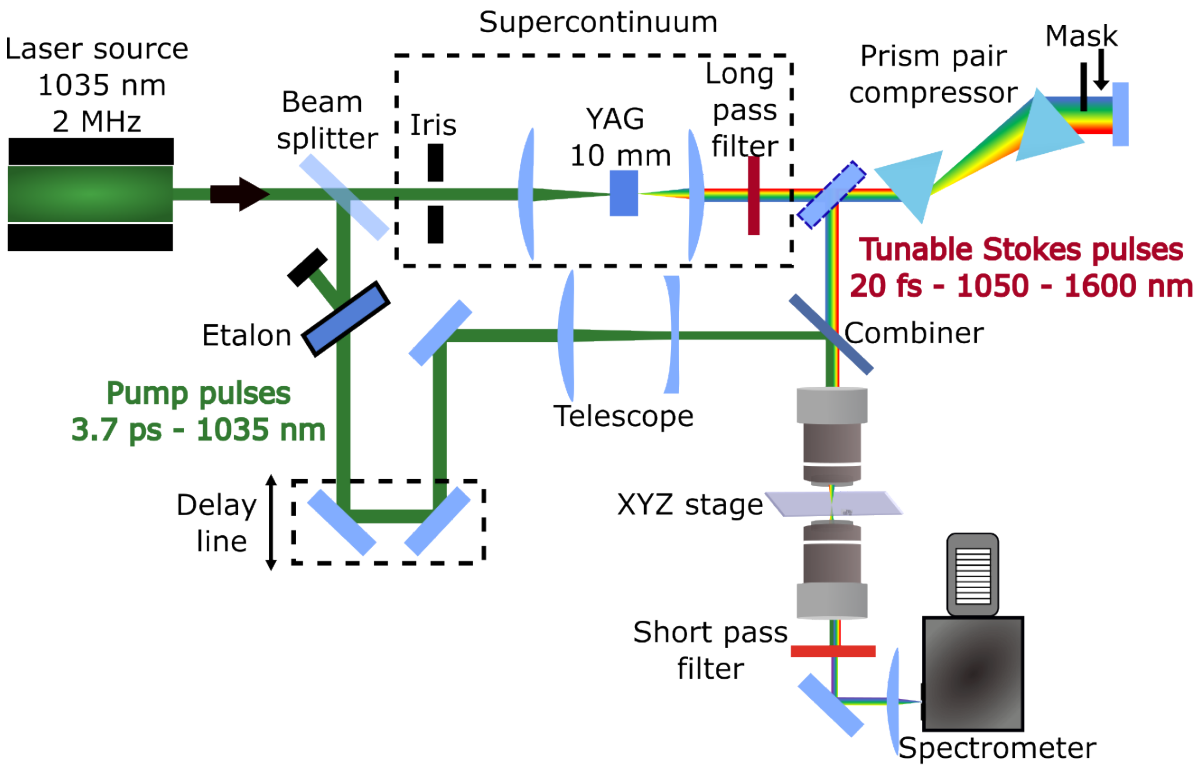

Fig. 2. Broadband CARS setup at the POLIMI in Milano.

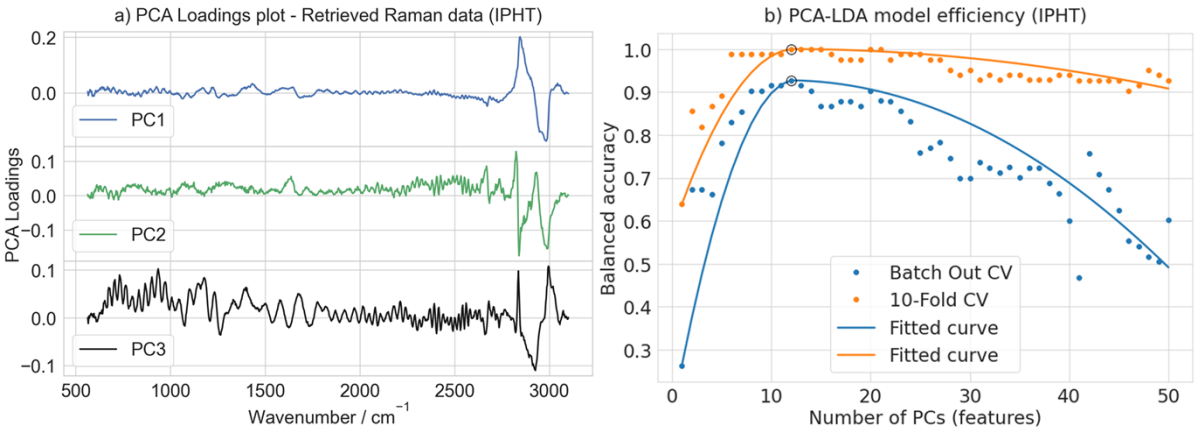

Fig. 3. a) PCA loadings obtained from the preprocessed Raman spectra retrieved from IPHT BACRS spectra. b) Variation of balanced accuracy with the number of PCs used as input to construct the LDA classifier. The black circle on the curve represents the optimal number of components used to create the final LDA classifier. Interestingly, this optimal number is 12 PCs for both CV methods. However, the resulting accuracy differs; it is 100% for the 10-fold CV and 92.7% for the batch-out CV.

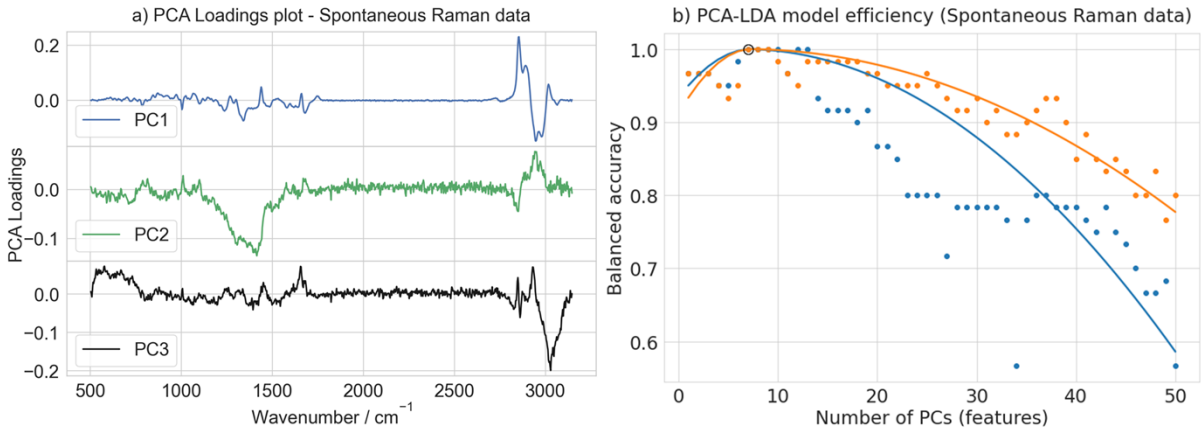

Fig. 4. a) PCA loadings plot obtained from the Spontaneous Raman data. b) Variation of balanced accuracy with the number of PCs given as the input to construct the LDA classifier. The black circle on the curve represents the optimal number of components used to create the final LDA classifier. It is noticed that the optimal number of PCs is the same for both CV methods, and the corresponding accuracy is 100%.