In-line Impedance Spectroscopy for Automated Bovine Milk Analysis

C. Beyer, M. Höchemer, T. Volden, J. Goldowsky, R. Limacher

Early screening of somatic cell count (SCC) in incoming raw milk samples loads, i.e. before pooling, is necessary for avoiding waste of large amounts of milk and would enable new services in herd management. In collaboration with an industrial and academic partner, CSEM developed a simple, portable and automated milk analysis tool that determines efficiently SCCs within less than three minutes. This project comprised the elaboration of high-level signal processing and auto-clustering algorithms based on various concepts of machine learning methods.

The quality of bovine raw milk depends, among other factors, on the concentration of somatic cells. A high somatic cell count (SCC > 100 cells / µl) indicates a potential infection of the cow udder (mastitis) and decreases the quality of dairy products like cheese. Presently, analyses of the somatic cell load in raw milk are performed manually (by classical microscopy) at laboratories of large dairy facilities or fully automated at centralized analytical service laboratories. A microfluidic measurement system, based on impedance flow cytometry, provides a more cost efficient, label-free and real-time determination of SCC (Figure 1).

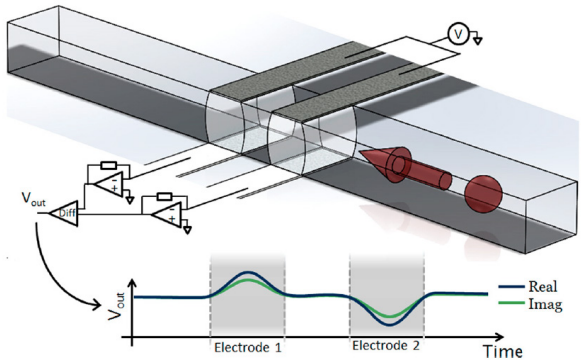

Figure 1: Measurement setup for impedance flow cytometry based on differential impedance spectroscopy – Figure from Spencer et al [1].

In the framework of an InnoSuisse project, CSEM developed in collaboration with a local SME and the University of Applied Sciences Lucerne a novel algorithm for determination of SCC in the range of 10 to 2000 cells per microliter. The project comprised three major steps: data collection, signal pre-processing and SCC determination (Figure 2).

Regarding the first step, a total of over 600 fresh milk samples of various origin including pooled milk from tanks and up to 350 individual cows from 5 different farms were processed over a time period of 18 months. The impedimetric time signal of each sample was recorded at least 4 times for at least 30s. Its somatic cell count was additionally determined by the independent SuisseLab institute.

Major challenge of the analysis is the missing bijective correlation between detected events and somatic cells. Only the somatic cell count, *i.e.*, cell concentration, is known for each measurement sample, thus a correlation between roughly 25 million measurement points and one value has to be found.

An essential focus of the work was directed towards signal pre-processing with the aim to reduce the measured variance of samples with a similar number of somatic cells. Based on the raw time signal, automated algorithms for signal offset correction, event detection, *i.e.*, potential somatic cells, and signal feature extraction were developed and successfully implemented. Thus, reducing the input data for step 3 by a factor 1000 and more. Various measured parameters could be identified and extracted from the time signal allowing to reduce the overall assessed signal variance.

For the final SCC determination, two data analysis approaches have been investigated in parallel, based either on the time signal of detected events or on extracted signal features. Both approaches utilised and explored various concepts of machine learning, especially the deep learning ones. In the recent decade, deep learning concepts became dominant in the machine learning community for its strong feature of self-learning ability. The algorithm extracts multiple level representations of the data through its deep architecture and learns the underlying structures and sophisticated high-level features of the data. [2].

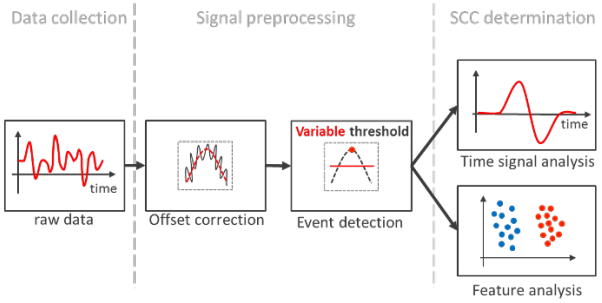

Figure 2: Implemented workflow of signal analysis path.

The first approach explored and optimized a recurrent neuronal network (LSTM) using clipped time signals of detected events from pre-processing. The algorithm estimates the probability for each event to be a somatic cell based on trainings of the assessed fresh milk samples. The second approach focused on cluster algorithms, like DBSCAN, and semi-supervised anomaly detection (AD) methods, such as Gauss-based, Isolation Forest and Robust Covariance, in combination with Support Vector Regression models. The first approach yields more promising results compared to the second, while both approaches provide comparable performance as human experts manually analyzing the measurements.

For final testing, both approaches have been integrated into a lean and cost-efficient embedded system (Raspberry Pi) together with a speed optimized pre-processing algorithm enabling a fully automated determination of SCC. This system is currently being evaluated. The idea is to combine it with the assessment of other milk quality parameters. Hence, the final device integrated as a mobile solution might avoid waste of large amounts of milk enabling novel services in herd management.

[1] Spencer, *et. al.* (2011) "Positional dependence of particles in microfludic impedance cytometry." Lab on a Chip 11.7: 1234-1239.

[2] Bengio, *et al.* (2013) "Representation learning: A review and new perspectives." IEEE transactions on pattern analysis and machine intelligence 35.8 (2013): 1798-1828.