Mikko Haikonen*, Hannu Toivonen, and Jerker Björkqvist*

# Stress test evaluation of classifiers for audio based bearing diagnostics in HVAC machines

**Keywords:** Diagnostics, Machine learning, Bearing monitoring, Audio classification, HVAC maintenance

***Corresponding Author: Mikko Haikonen:** Solita Oy, E-mail: mikko.haikonen@solita.fi
**Hannu Toivonen:** Åbo Akademi University, E-mail: hannu.toivonen@abo.fi
***Corresponding Author: Jerker Björkqvist:** Åbo Akademi University, E-mail: jerker.bjorkqvist@abo.fi

## 1 Background

Condition based monitoring attempts to measure the situation of industrial equipment or process. This measurement is important in detecting equipment changes or damages. Servicing mechanical devices has traditionally been relying heavily on aural observations. Accurate audio diagnostics require profound understanding. Those skills are not always available on site, especially for remote location. Therefore automated or assisted diagnostics are very interesting fields in predictive maintenance.

## 2 Objectives

In applying machine learning classifiers in condition based monitoring there are two practical problems. The first one is the overfit of classifiers. That may cause problems whenever the situation changes at monitored process after training data is being collected and the classifier is being trained. The second problem is the cost of audio data storage. Data compression is one of the ways to decrease data storage costs. However, it is not always clear how much information we can afford to loose in order to maintain appropriate classification performance.

The objective of our study was to evaluate classifiers in their capacity to detect the relationship between the selected audio features and bearing changes. Audio samples from HVAC (heating, ventilation, and air conditioning) system were used for classification experiments. Main focus in evaluation was to prevent overfit. Ability to correctly classify the audio signals from before and after overhaul was primary objective. Attention was also paid to the computational cost of the classifiers. An additional aim was to provide additional sanity check mechanisms to avoid selecting a classifier that is being overfit. For this we examined how lossy compression of input data affects the classification performance. This is how we estimated the reliability of K-fold cross validation results.

## 3 Methods

We used the audio samples collected from an HVAC system before and after overhaul. In the overhaul one of the two drive and fan unit was replaced with new one. Audio signals from air- and contact microphone were recorded. To avoid excess computations the amount of data was reduced by random sampling of training and testing audio files.

The audio classification features were selected in a way that attempts to reduce the need for additional feature engineering if the detection task changes. We used a feature space that allows linking the classification results back to the physical process. Therefore we used features like spectrum (FFT), autocorrelation (ACF) and partial autocorrelation (PACF). An additional benefit of ACF and PACF features is their close link to data generating process characteristics [1]. Autocorrelation feature can also bring insights into summative- and modulation effects [2]. Both autocorrelation and partial autocorrelation are also closely linked to linear predictive coding used in human speech processing [3].

For classification we used random forest, support vector machines (SVM), learning vector quantisation and decision tree. We used relatively simple models to maintain the link between results and a physical process. That link is useful when the classification results are taken into operational use.

The process of training and cross-validating classifiers was used as tool to understand how well the selected feature space reflects the mechanics and the class of samples. Our stress test complemented the results from traditional cross validation results.

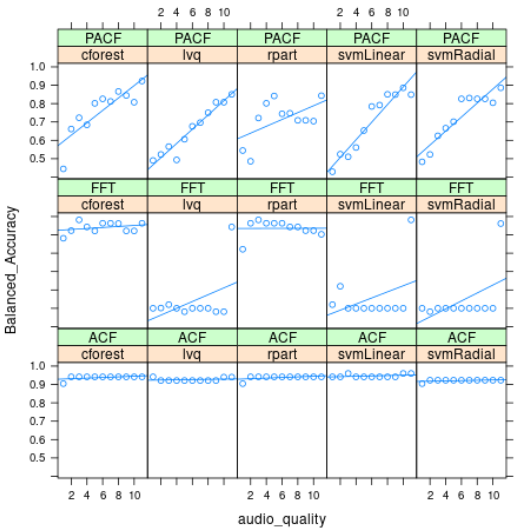


**Fig. 1.** Example stress profiles: Impact of audio quality to balanced accuracy in Stress test results for selected classifiers and feature sets (horizontal axis of individual plots is audio quality indicator where 0 stands for poorest and 11 is original audio recording)

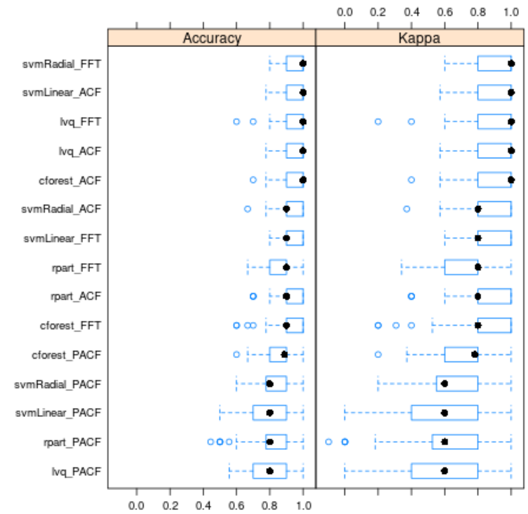


**Fig. 2.** Cross validation results: One can compare these cross validation to find classifiers that perform well in cross validation and very poorly in stress test

Traditional cross validation methods indluding e.g. K-fold [4] cross validation were used [5]. Our key idea was to extend e.g. K-fold cross validation by stress testing the classifiers. Stress test means that the trained classifiers are stressed by testing them with samples of varying audio quality. For each classifier and validation sample, we used several levels of audio quality. We will then examine how the prediction accuracy reacts to this stress. The purpose of the stress test was to examine how sensitive the classifiers are to changes in input data. From this experiment, we obtain a stress profile for each classifier and each type of audio feature. We used these stress profiles to find the best combination of classifier and feature set for the task at hand.

Since analysis of audio samples and speech analysis are similar problems it is very natural to use speech compression as a vehicle for stress testing the classifiers and consider audio samples as "machine speech".

## 4 Results

Stress test results fig. 1 indicate that a good classification performance in K-fold cross valication tests fig. 2 may drop dramatically with slightest reduction of input data quality. Example of such a classifier was support vector machine (SVM) trained and tested using FFT feature set. We also experienced that the classifier performance depends on selection of the feature set even if two feature sets have direct mathematical translations like spectrum and autocorrelation (ACF) features do. These sensitivities are not always seen in cross validation results but regrettably also depends on selection of K-fold parameters.

## References

[1] Box G E, Jenkins G M. *Time Series Analysis, forecasting and control*. Holden-Day Inc. 1976.
[2] Lange H F. *Correlation Techniques*. Iliffe Books Ltd. 1967.
[3] Rabiner Lawrence R, Schafer Ronald W. *Introduction to Digital Speech Processing*. now Publishers Inc. 2007.
[4] T Hastie, R Tibshirani, J Friedman. *The Elements of Statistical learning*. Springer Science Business Media. 2009.
[5] Kuhn M, Johnson K. *Applied Predictive Modeling*. Springer Science+Business Media. 2013.