

# Positive-Unlabelled Learning-Based Novelty Detection for Industrial Chillers

## A Data-Driven Approach to Avoid Energy Wastage

Ron van de Sand<sup>1</sup> [<https://orcid.org/0000-0002-8975-4030>],  
Sandra Corasaniti<sup>2</sup> [<https://orcid.org/0000-0002-0434-501X>],  
and Jörg Reiff-Stephan<sup>1</sup> [<https://orcid.org/0000-0003-4176-6371>]

<sup>1</sup>Technical University of Applied Sciences Wildau, Germany

<sup>2</sup>University of Rome "Tor Vergata", Italy

**Abstract.** Chiller systems are used in many different applications in both the industrial and the commercial sector. They are considered major energy consumers and thus contribute a non-negligible factor to environmental pollution as well as to the overall operating cost. In addition, chillers, especially in industrial applications, are often associated with high reliability requirements, as unplanned system downtimes are usually costly. As many studies over the past decades have shown, the presence of faults can lead to significant performance degradation and thus higher energy consumption of these systems. Thus, data-driven fault detection plays an ever-increasing role in terms of energy efficient control strategies. However, labelled data to train associated algorithms are often only available to a limited extent, which consequently inhibits the broad application of such technologies. Therefore, this paper presents an approach that exploits only a small amount of labelled and large amounts of unlabelled data in the training phase in order to detect fault related anomalies. For this, the model utilizes the residual space of the data transformed through principal component analyses in conjunction with a biased support vector machine, which can be ascribed to the concept of semi-supervised learning, or more specifically, positive-unlabelled learning.

**Keywords:** Chiller CBM, Machine Learning, Energy Efficiency

## Introduction

Chillers are applied across many different fields of both industrial and commercial applications. As being large energy consumers with accounting up to 40% of a buildings total energy demand [1], these systems surely open up great potential for optimization. One promising approach to avoid energy losses while operating these systems is known to be condition based maintenance (CBM) [2] which constitutes several steps to asses a systems degradation state, of which fault detection (FD) [3] plays a key-role as it allows to detect novelties caused by one or multiple faults being present. This is particularly important as chillers can loose up to 30% of their coefficient of performance without any obvious sign of degradation being present to the operator [4]. Besides avoiding energy wastage, FD also contributes to saving operation as well as service costs. This can help to predict system malfunctions and therefore prevents unplanned downtimes, which is especially important for industrial applications as a chiller breakdown may cause an entire process line to stop [5].

This topic has attracted the attention of researchers for many decades and a great variety of models have been proposed to date. Generally, these approaches can be divided into model-based and data-driven [6]–[8], whereas the former utilizes a reference model to detect abnormal behaviour. The latter, however, relies on self-adapting algorithms that only require historical machine data [3] making it possible to develop FD models without any prior assumptions about the systems structure or the expected fault characteristics. Another limitation of model-based approaches is that finding a mathematical formulation to model a dynamical process can be difficult as well as time consuming due to changing operation conditions [9], and may sometimes even be infeasible when the system complexity is too high. Data-driven FD models can overcome these issues [3] and are, thus, considered to be more cost effective [10]. On the other hand, most existing data-driven models require labelled data of the dedicated system containing both positive (fault-free) and negative (fault) samples, which are often unavailable. The reason for this is twofold: first, as faults are rare events [11], certain faults may not have occurred in the past and therefore associated data cannot be obtained, and second, labelling is conducted by humans and consequently constitutes a high cost and time factor. However, in most application scenarios one comes across a situation where a few positive labelled samples and a lot of unlabelled data containing positive and negative patterns might be available. In this case, the unlabelled data can be exploited in the training phase in order to increase the model accuracy, which is commonly referred to as positive-unlabelled (PU) learning. Therefore, the work at hand proposes a PU learning based fault detection model which utilizes principal component analysis (PCA) in conjunction with the biased support vector machine (BSVM) [12]. Furthermore, we evaluate the model performance on two datasets stemming from the experimental investigation of different chiller types and demonstrate how the proposed model can be more effective compared to classical novelty detection approaches.

The paper is structured as follows: After this introduction, the related work in this field of research is briefly described. Then, the model principles and the description of the datasets used to train and validate the model are presented. In the following section, the model evaluation is then performed and discussed. Finally, the paper is summarized and an outlook is given.

## Related Works

Data-driven novelty detection has been the subject of research for many years and a great variety of supervised machine learning were presented. For example, Han et al. [13] proposed to first reduce the dataset dimensionality using PCA and then train a support vector machine (SVM) classifier in a multi-class fashion to perform fault detection and fault diagnosis in a single task. This idea was later adopted by Beghi et al. [14] and Li et al. [15], who trained their model in the PCA residual space rather than the principal component space and showed that this approach could significantly improve the fault detection accuracy. Besides, both work considered the problem as a single-class learning task [16], where in [14] the one-class support vector machine (OCSVM) [17] and in [15] the support vector data description (SVDD) [18] is applied. The difference between the two is that the OCSVM classifier tries to solve for a hyperplane that best separates a positive labelled pattern from the rest of the space, whereas the SVDD classifier aims to find a hypersphere enclosing most of the data. It is worth noting that both algorithms yield equal results when used in conjunction with invariant kernels, such as the radial basis function (RBF) kernel [16]. Feature extraction has also been one of the major concerns of other papers such as, for example, in [19] and [20], where the authors applied genetic algorithms in different variations in order to select characteristic features (CF). In both cases, the SVM has been adapted for the final classification task. Other approaches considered the application of linear discriminant analysis [21], Bayesian networks [22] or statistical models [11]. Although many promising approaches were proposed, only few

contributions exist in this field of research considering the case when only a minor number of positive labelled and high number of unlabelled samples are available. Compared to labelled fault samples, labelled fault-free data is easier to obtain as, for instance, after commissioning the chiller or after performing maintainable actions. One example is given in [20], where the authors combined a Kalman filter with a recursive SVM to exploit unlabelled samples. Similarly, the model in [11] is developed in a semi-supervised fashion. Another approach is given by Fan et al. [9], who applied a transfer learning model to incorporate prior-knowledge from another chiller type. Even though some work in this field has already been conducted, a full understanding of PU learning for FD tasks as well as the effect of the number of available positive samples onto the model performance is still lacking. Therefore, this paper proposes a novel approach, namely the PCA-R-BSVM algorithm, for PU learning based novelty detection for industrial chillers and validates its performance based on two datasets. Furthermore, we compare our approach with the existing PCA-R-SVDD [15], or more specifically with the PCA-R-OCSVM [14] model, under the assumption that only a limited number of positive labelled samples is available.

## Model

This section describes the model principles and describes PCA as well as the BSVM algorithm in more detail. Furthermore, a classification metric is introduced that can be employed for evaluating the classification performance based on positive and unlabelled data. This is especially crucial for optimizing the model's parameter as most metrics require the presence of positive and negative labelled instances in the validation phase.

### Principal Component Analysis

PCA is an unsupervised technique that has been widely applied for many dimensionality reduction tasks. In general, it is a linear transformation algorithm which aims to find a new data representation of uncorrelated features, namely the principal components (PC) [14], such that most of the data variance is represented by the first transformed features. Let  $\mathcal{X} \in \mathbb{R}^{n \times k}$  be data sampled from the target chiller, with  $n$  being the number of observations and  $k$  the number of dimensions. The entire dataset can then be expressed as  $\mathcal{D} = \{(x_1, y_1), \dots, (x_n, y_n)\}$  with  $y_i \in \{1, -1\}$  for positive and unlabelled instances respectively. After processing the data  $\mathcal{X}$ , one receives a transformed data representation that can be decomposed into the modelled PC space  $\hat{\mathcal{X}} \in \mathbb{R}^{n \times (k - k_R)}$  and the un-modelled residual subspace  $\tilde{\mathcal{X}} \in \mathbb{R}^{n \times k_R}$  to whose features we will refer to as the residual components (RC) throughout this paper ( $k_R$  is the number of RC). It is worth noting that while the PC space contains most of the data variance, the RC subspace contains noise as well as abnormal information [23]. As has been shown in previous studies [11], [14], [15], training a FD model in the RC space rather than the PC space can significantly improve its classification accuracy as the former is more sensitive to novelties.

### Biased Support Vector Machine

The BSVM algorithm has been proposed by Liu et al. [12] and its working principle is similar to the idea of training the original SVM with imbalanced data, i.e. when instances from one class appear more frequently than from the other class. However, its motivation is slightly different as it assigns high weights to positive and low weights to unlabelled samples in order to bias the data to the favour of the known positive pattern. On the other hand, low weights are assigned to unlabelled patterns as these may also contain positive instances whose labels remain unknown during the training phase. As a consequence, this approach can be assigned to the field of semi-supervised machine learning. Following this idea, the primal form of the optimization problem can be formulated as

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} \|w\|^2 + C^+ \sum_{i=1}^{n_p} \xi_i + C^- \sum_{i=n_p+1}^n \xi_i \\ \text{subject to} \quad & y_i(w^T \tilde{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \end{aligned} \quad (1)$$

where  $w$  is a vector orthogonal to the hyperplane,  $C^+$  and  $C^-$  the regularisation parameters assigned to positive and unlabelled samples accordingly,  $\xi_i$  a slack variable used to solve for a soft-margin decision boundary [16], and  $b$  the bias term. Furthermore, we denote  $\tilde{x}_i$  as any observation from the dataset using only its residual information after PCA transformation and  $n_p$  the number of positive labelled samples in the dataset. Substituting (1) into its Lagrangian dual form, one may use kernels in order to solve for a non-linear decision boundary in the input space. The decision function for an unknown observation  $\tilde{x}$  is defined as

$$h(\tilde{x}) = \sum_{i=1}^n \alpha_i y_i k(\tilde{x}, \tilde{x}_i) + b \quad (2)$$

where  $\alpha_i$  are the Lagrangian multiplies and  $k(\cdot, \cdot)$  the kernel function. Similar to previous studies [14], [19], [20], [24], [25], we apply the RBF kernel function, which is given as

$$k(\tilde{x}_i, \tilde{x}_j) = \langle \varphi(\tilde{x}_i), \varphi(\tilde{x}_j) \rangle = \exp(-\gamma \|\tilde{x}_i - \tilde{x}_j\|^2) \quad (3)$$

where  $\varphi(\cdot)$  is a non-linear mapping function and  $\gamma$  the kernel width parameter that must be tuned.

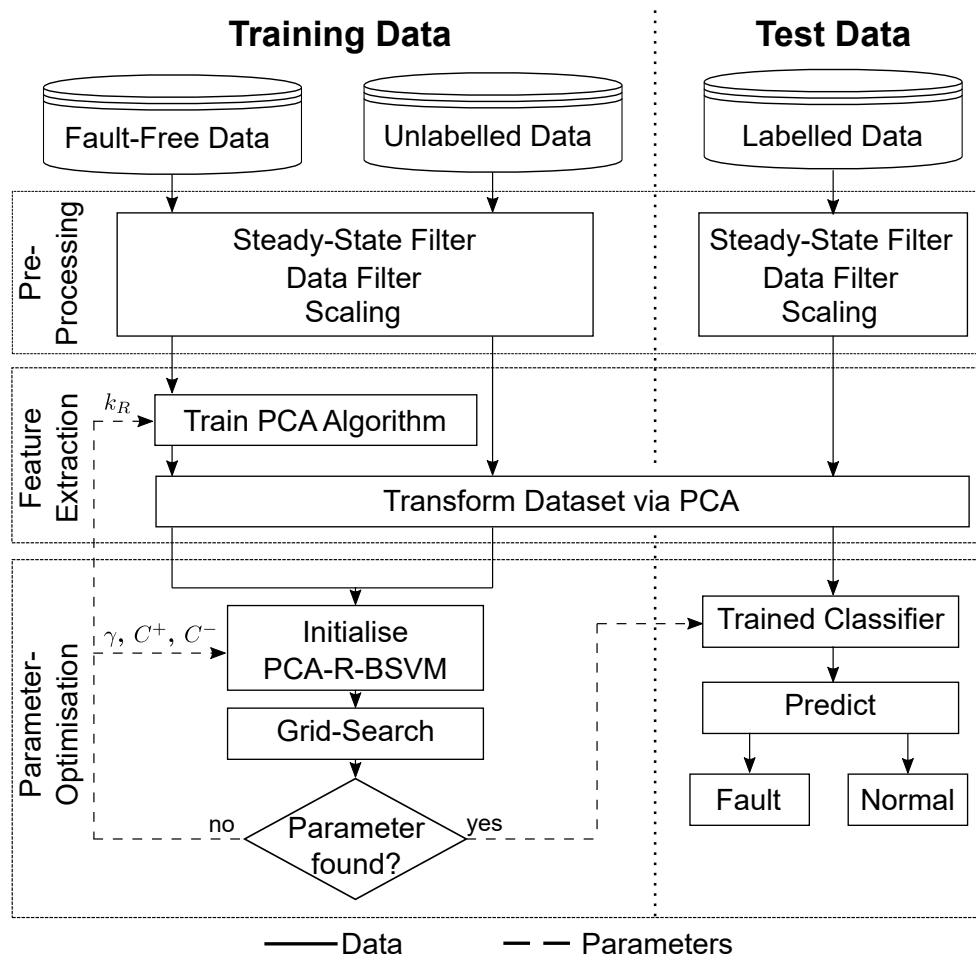
### Classification Metric

One major aspect of PU learning based classification approaches is how to evaluate the classification performance as most metrics require positive as well as negative samples to be available in the dataset. It goes without saying that this is particularly crucial for any parameter optimisation process, such as grid-search, because for any subset of chosen parameters, the algorithm's performance must be evaluated on a given test dataset. Throughout this paper we apply a metric proposed by Lee and Liu [26] which was also applied in the original BSVM paper [12]. Similar to the well-known F-Score, the chosen accuracy metric is formulated as  $r^2 / Pr[h(\tilde{X}) = 1]$ , where  $r$  is the recall and  $Pr[h(\tilde{X}) = 1]$  is the probability that a pattern is classified as positive. As stated in [12], this metric behaves similar to the F-Score metric as it is low when either recall or precision is low and high when both values are high.

### Overview

The model development can generally be distinguished in the preprocessing phase, PCA based feature extraction and parameter optimisation. In general, preprocessing combines three steps, namely: (a) steady-state detection, which is conducted to exclude transient system states from the dataset; (b) data-filtering to increase the signal-to-noise ratio; and (c) data scaling to exclude the negative influence of the different units on the classification result. It is worth noting that we applied the steady-state detector proposed by Beghi et al. [11] in this paper. In the following step, PCA is conducted on the available labelled fault-free samples, which represents the basis for mapping all further data samples into the PC and RC space. The reason for this is that with the PCA model trained on the fault-free data, the RC become more sensitive to faults as these features are subject to changes when novelties are present. As shown in Figure 1, grid-search is introduced for parameter optimisation. In particular, the model depends on the parameters  $\gamma$ ,  $C^+$  and  $C^-$ , which are determined through this searching strategy. In addition, the number of residual components  $k_R$  is adjusted in a similar way. Although solely positive labelled and unlabelled instances are utilized in the training

phase, a fully labelled dataset containing actual positive as well as negative samples is used for validation purposes.



**Figure 1.** Training and testing procedure

## Datasets

In this study, two datasets of two different chiller types is used for training and validating the model. The first datasets was collected during the experimental investigation of a 316kW water-cooled centrifugal chiller. In project ASHRAE RP-1043, carried out by Comstock et al. [27], seven typical chiller faults at four levels of severity were experimentally investigated. The other dataset has its roots in one of our previous works where various faults were investigated using a 100kW ammonia based screw compressor chiller [28], as shown in Figure 2. The data acquisition procedure was similar to that of the ASHRAE RP-1043 project, in which each fault was experimentally investigated at four levels of severity, starting from the lowest level and increasing to the highest. In both projects a test-sequence was defined consisting of several operating states, with each state meeting a steady-state criterion before moving to the next one. As listed in Table 1, 23 features were selected that are shared between both dataset. In the following, these datasets are utilized to train and validate the proposed model. Moreover, it is demonstrated how the model performs with only a minor number of positive samples being available in the training phase.

**Table 1.** Chosen features

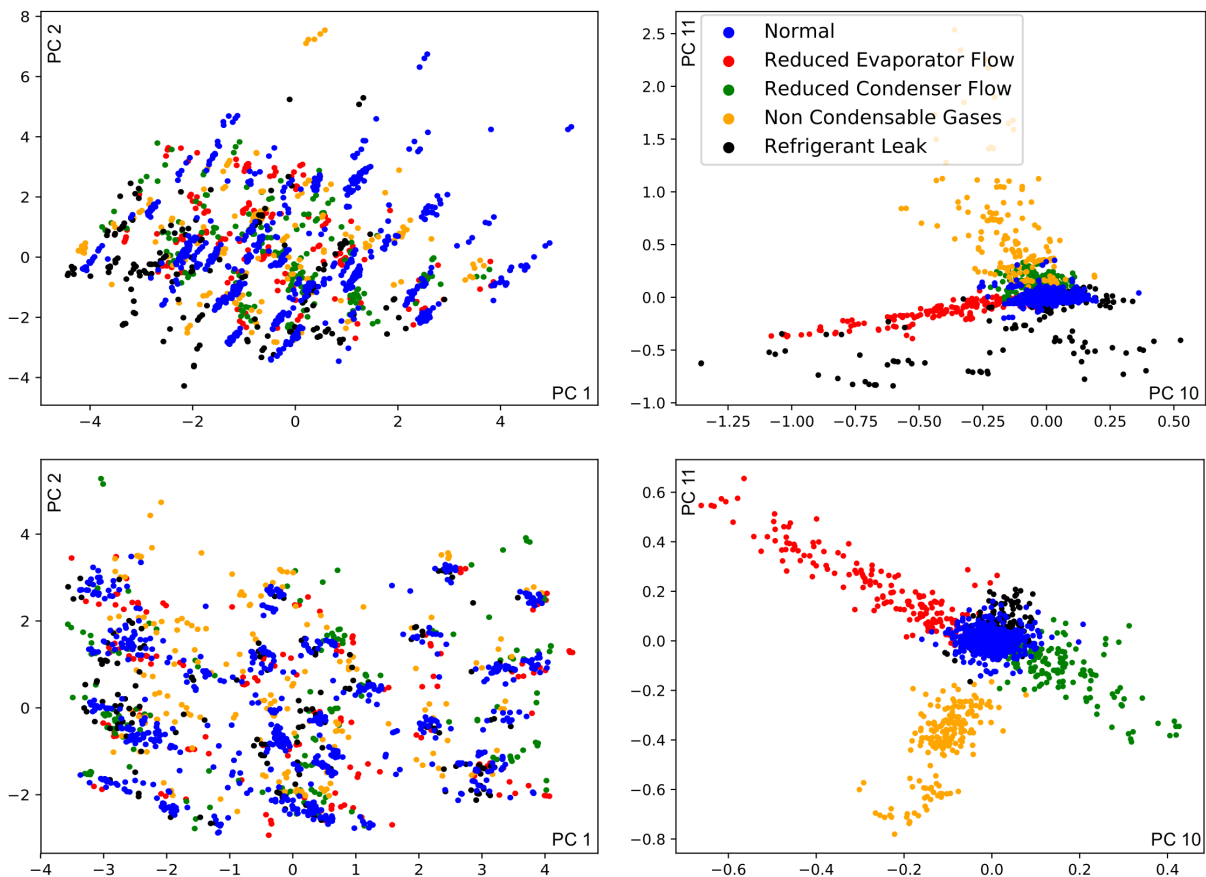
Symbol	Description
$f_1$	Compressor pressure lift
$f_2$	Coefficient of performance
$f_3$	Isentropic compressor efficiency
$f_4$	Polytropic compressor efficiency
$f_5$	Instantaneous compressor power
$f_6$	Instantaneous compressor current
$f_7$	Evaporator cooling rate
$f_8$	Condenser heat rejection rate
$f_9$	Energy balance
$f_{10}$	Evaporator water inlet temperature
$f_{11}$	Evaporator water outlet temperature
$f_{12}$	Condenser water inlet temperature
$f_{13}$	Condenser water outlet temperature
$f_{14}$	Refrigerant temperature in Evaporator
$f_{15}$	Refrigerant discharge temperature
$f_{16}$	Refrigerant suction temperature
$f_{17}$	Refrigerant discharge superheat temperature
$f_{18}$	Refrigerant suction superheat temperature
$f_{19}$	Evaporator approach temperature
$f_{20}$	Evaporator water temperature difference
$f_{21}$	Condenser water temperature difference
$f_{22}$	Refrigerant pressure in evaporator
$f_{23}$	Refrigerant pressure in condenser



**Figure 2.** Test facility

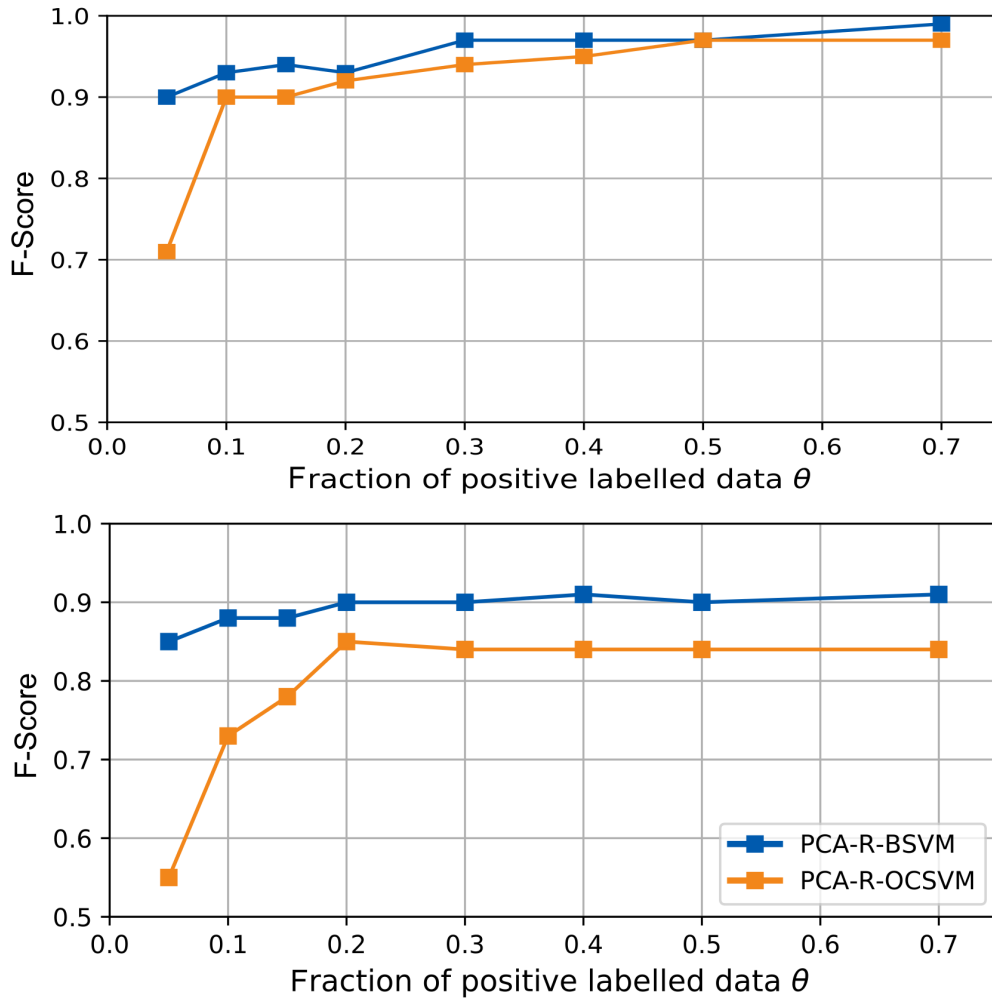
## Results

The model is characterized by two essential aspects; first, it is based on the RC space after processing the data using PCA, and second, it is trained in a semi-supervised fashion as unlabelled data is exploited in the training phase. As can be seen in Figure 3, the fault patterns become more distinct in the RC space. This is because most of the data variance is represented by the first principal components and is induced due to changing operating conditions rather than novelties. Therefore, by removing the PC from the dataset and exploiting the RC space instead, one may yield higher fault detection accuracy.



**Figure 3.** Scatter plot of the principal component and residual space after PCA transformation: upper left: the two first PC (screw compressor), upper right: residual subspace (screw compressor), lower left: the two first PC (centrifugal compressor), lower right: residual subspace (centrifugal compressor)

One crucial aspect of the proposed model is its performance in dependence on the available number of positive labelled instances in the training phase. To show how the model performs on real-world data, we select 1000 observations from the normal and 250 the fault dataset for each chiller type, of which 2/3 is used for training and parameter optimisation and 1/3 for testing. It should be noted that all faults are treated as one class as they represent novelties in this study. We then define a fraction of positive labelled samples  $\theta$  and mark them as positive, whereas the rest, i.e. positive and negative patterns, are defined as unlabelled (-1). As shown in Figure 1, we repeat the model training process described previously for an increasing  $\theta$  and compare the model's classification accuracy with the PCA-R-OCSVM baseline model. For comparison purposes, we train the baseline model in a similar fashion using the classification metric introduced previously. In general, the PCA-R-OCSVM algorithm requires three parameters to be optimized for of which the parameter  $\nu$  replaces  $C^+$  and  $C^-$ , whereas  $\gamma$  and  $k_R$  are equally utilized by both algorithms. Note that in the following, the test dataset is



**Figure 4.** Accuracy plot over fraction of positive labelled samples: models trained on - upper: screw chiller dataset, lower: centrifugal chiller dataset

exploited for comparison purposes using the F-Score metric to evaluate the classification performance. As shown in Figure 4, one can see that the proposed model outperforms the baseline algorithm. This becomes even more distinct when  $\theta$  is low, which certainly proves that by incorporating unlabelled data in the training process, the model's classification accuracy can be significantly improved. The most significant difference between both models is at  $\theta=0.05$ , where the baseline model deviates by 20% considering the screw-chiller dataset and more than 35% when applied on the centrifugal-chiller one. Throughout the experiments, the proposed PCA-R-BSVM model proved to yield better classification results compared to the baseline algorithm by reaching scores as high as 0.97. Furthermore, its accuracy converges faster with increasing values of  $\theta$ , which shows that it can be successfully applied for FD tasks, even when only a minor number of positive labelled samples is available.

## Summary

In this paper a PU-Learning based novelty detection model was introduced for fault detection tasks. While relying on PCA or, more specifically, on the extracted residual components, it was shown that the model was highly sensitive to faults. Besides, it was shown how the model performs with only a limited number of positive samples being available in the training phase while the rest, including actual positive and negative observations, are treated as unlabelled data samples. The key-points of the proposed model can be summarized as follows:



1. Applying PCA onto the dataset can significantly improve the model performance when the first principal components, i.e. the ones associated with the largest values of eigenvalues, are removed from the dataset and the residual components are used for training the classifier
2. The proposed PCA-R-BSVM could outperform an existing novelty detection algorithm and performs well, even with only few labelled data being available
3. The model performance was validated on two datasets stemming from different chiller types

## Acknowledgements

This research was partly supported by the German Federal Ministry for Economic Affairs and Energy (No: ZF4012225RE7) as well as the Ministry for Economic Affairs, Labour and Energy, State of Brandenburg (No: 80237615). Our special thanks go to our partner Potsdamer Anlagenbau und Kältetechnik GmbH for their great support in this research project.

## References

- [1] A. Afshari and L. Friedrich, "A proposal to introduce tradable energy savings certificates in the emirate of abu dhabi," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 1342–1351, 2016.
- [2] ISO, *Condition monitoring and diagnostics of machines–vocabulary*, 2012.
- [3] G. J. Vachtsevanos, *Intelligent fault diagnosis and prognosis for engineering systems*. Wiley Hoboken, 2006, vol. 456.
- [4] S. Wang, Q. Zhou, and F. Xiao, "A system-level fault detection and diagnosis strategy for HVAC systems involving sensor faults," *Energy and Buildings*, vol. 42, no. 4, pp. 477–490, 2010.
- [5] D.-W. Sun, *Handbook of frozen food processing and packaging*. CRC Press, 2016.
- [6] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part ii: Qualitative models and search strategies," *Computers & chemical engineering*, vol. 27, no. 3, pp. 313–326, 2003.
- [7] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis: Part iii: Process history based methods," *Computers & chemical engineering*, vol. 27, no. 3, pp. 327–346, 2003.
- [8] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part i: Quantitative model-based methods," *Computers & chemical engineering*, vol. 27, no. 3, pp. 293–311, 2003.
- [9] Y. Fan, X. Cui, H. Han, and H. Lu, "Chiller fault diagnosis with field sensors using the technology of imbalanced data," *Applied Thermal Engineering*, vol. 159, p. 113933, 2019.
- [10] C. Anger, "Hidden semi-markov models for predictive maintenance of rotating elements," Ph.D. dissertation, Technische Universität, 2018.
- [11] A. Beghi, R. Brignoli, L. Cecchinato, G. Menegazzo, M. Rampazzo, and F. Simmini, "Data-driven fault detection and diagnosis for HVAC water chillers," *Control Engineering Practice*, vol. 53, pp. 79–91, 2016.
- [12] B. Liu, Y. Dai, X. Li, W. S. Lee, and P. S. Yu, "Building text classifiers using positive and unlabeled examples," in *Third IEEE International Conference on Data Mining*, 2003, pp. 179–186.

- [13] H. Han, Z. Cao, B. Gu, and N. Ren, "PCA-SVM-based automated fault detection and diagnosis (AFDD) for vapor-compression refrigeration systems," *HVAC&R Research*, vol. 16, no. 3, pp. 295–313, 2010.
- [14] A. Beghi, L. Cecchinato, C. Corazzol, M. Rampazzo, F. Simmini, and G. A. Susto, "A One-Class SVM based tool for machine learning novelty detection in HVAC chiller systems," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 1953–1958, 2014.
- [15] G. Li, Y. Hu, H. Chen, L. Shen, H. Li, M. Hu, J. Liu, and K. Sun, "An improved fault detection method for incipient centrifugal chiller faults using the PCA-R-SVDD algorithm," *Energy and Buildings*, vol. 116, pp. 104–113, 2016.
- [16] B. Schölkopf, A. J. Smola, F. Bach, *et al.*, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
- [17] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," *Advances in neural information processing systems*, vol. 12, pp. 582–588, 1999.
- [18] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine learning*, vol. 54, no. 1, pp. 45–66, 2004.
- [19] H. Han, B. Gu, J. Kang, and Z. R. Li, "Study on a hybrid SVM model for chiller FDD applications," *Applied Thermal Engineering*, vol. 31, no. 4, pp. 582–592, 2011.
- [20] K. Yan, Z. Ji, and W. Shen, "Online fault detection methods for chillers combining extended kalman filter and recursive One-Class SVM," *Neurocomputing*, vol. 228, pp. 205–212, 2017.
- [21] D. Li, G. Hu, and C. J. Spanos, "A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis," *Energy and Buildings*, vol. 128, pp. 519–529, 2016.
- [22] Z. Wang, Z. Wang, S. He, X. Gu, and Z. F. Yan, "Fault detection and diagnosis of chillers using bayesian network merged distance rejection and multi-source non-sensor information," *Applied Energy*, vol. 188, pp. 200–214, 2017.
- [23] S. Wang and Y. Chen, "Sensor validation and reconstruction for building central chilling systems based on principal component analysis," *Energy Conversion and management*, vol. 45, no. 5, pp. 673–695, 2004.
- [24] H. Han, X. Cui, Y. Fan, and H. Qing, "Least squares support vector machine (LS-SVM)-based chiller fault diagnosis using fault indicative features," *Applied Thermal Engineering*, vol. 154, pp. 540–547, 2019.
- [25] Y. Zhao, S. Wang, and F. Xiao, "Pattern recognition-based chillers fault detection method using support vector data description (SVDD)," *Applied Energy*, vol. 112, pp. 1041–1048, 2013.
- [26] W. S. Lee and B. Liu, "Learning with positive and unlabeled examples using weighted logistic regression," in *ICML*, vol. 3, 2003, pp. 448–455.
- [27] M. C. Comstock, J. E. Braun, and R. Bernhard, *Experimental data from fault detection and diagnostic studies on a centrifugal chiller*. Purdue University, 1999.
- [28] R. van de Sand, C. Falk, S. Corasaniti, and J. Reiff-Stephan, "A data-driven fault diagnosis approach towards oil retention in vapour compression refrigeration systems," in *2019 International IEEE Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE)*, 2019, pp. 197–202.