ENHANCING FAULT DETECTION AND DIAGNOSIS IN ORGANIC RANKINE CYCLE SYSTEMS THROUGH PROPER FEATURE SELECTION IN MACHINE LEARNING

Andrés Hernández[1]*, Aitor Cendoya[1], Aldo Serafino[2], Michel Beaughon[2], Vincent Lemort[1]

[1] Université de Liège, Thermodynamics Laboratory, Liège, Belgium

[2] Enertime, R&D division, Courbevoie, France

* *Corresponding Author: jahernandez@uliege.be*

Abstract. In the current energy scenario, enhancing energy efficiency is crucial for sustainable development. Organic Rankine Cycle (ORC) systems play a significant role by recovering waste heat and converting it into useful energy. However, the efficiency and robustness of ORC systems are often compromised by various types of faults, leading to increased maintenance costs. This study focuses on fault detection and diagnosis (FDD) to improve system reliability and reduce maintenance expenses in ORC systems. The proposed methodology utilizes machine learning, specifically support vector machines (SVM), to detect and classify faults such as evaporator fouling and operational deficiencies in the expander and pump. A critical aspect of this approach is the careful choice of features for classification, which can either significantly enhance SVM's effectiveness or lead to misclassification of certain faults. Simulation results explore various feature sets to illustrate their impact on the FDD methodology's efficacy. These scenarios are evaluated in terms of prediction accuracy, demonstrating that proper feature selection improves the long-term efficiency and reliability of ORC systems.

Keywords. Fault detection and diagnosis, Organic Rankine cycle, Support vector machine.

Nomenclature (if relevant)

P	Pressure (Pa)
T	Temperature (°C)
h	Specific enthalpy (J/kg)
$\dot{Q}$	Heat transfer rate (J/s)
$\dot{m}$	Mass flow rate (kg/s)

Special characters

η	Efficiency (-)
η_s	Isentropic efficiency

Subscripts

su	supply
ex	exhaust
hf	Hot fluid
cf	Cold fluid
ev	Evaporator
cd	Condenser
exp	Expander

1 Introduction

The increasing global demand for energy efficiency and sustainability has intensified interest in technologies such as Organic Rankine Cycles (ORCs) for waste-heat recovery and renewable energy generation [1]. The fluctuating nature of the waste heat source make the ORC system to operate frequently at off-design conditions [2], therefore during long-term operation it becomes particularly difficult to identify from the sensor data available, if there exists a fault in a component or sensor, or if it is due to the specific operating conditions. Thus, given these vulnerabilities developing fault detection and diagnosis (FDD) techniques to maintain optimal performance and reliability for these complex energy systems is currently of high interest [3]. Note that effective FDD in ORCs not only reduces downtime and operational costs but also contributes to extending the lifecycle of these systems, enhancing their overall economic viability.

FDD in energy systems has evolved significantly, with various approaches. Model-based approaches, for instance, leverage the physical dynamics of an industrial heat exchanger to predict normal behaviour and detect deviations that may indicate faults [4]. In the context of ORC systems for geothermal

applications, dynamic modelling has proven useful for detecting performance anomalies, helping to support control system design and robust fault diagnosis using residual analysis [5].

Data-driven methods, including machine learning, have been increasingly applied to FDD tasks across complex systems, where they offer adaptability to system variations and the potential for high accuracy. Most scientific contributions of FDD are in the field of refrigeration systems and for heating, ventilation and air conditioning HVAC, as for example [6], where predictive frameworks are explored to automatically assess the system's state of degradation from unlabelled data. Whilst in [7] it is highlighted for the case of Air Conditioning how soft faults decrease the system performance (i.e., increases energy consumption) and are more difficult to detect, thus requiring more sensors.

Hybrid approaches that combine model-base-driven techniques are particularly promising for complex applications, as they can accommodate both the complex dynamics and diverse fault types present in these systems. Methods such as digital twin frameworks and advanced filtering algorithms, including extended Kalman filters and one-class Support Vector Machine (SVM), have been explored for real-time FDD applications in energy systems, including air handling and chiller units [8,9]. Where these frameworks not only enhance fault detection but support decision-making by providing actionable insights into system health and performance.

At the best author knowledge literature on FDD for ORC systems is scarce. For instance, [10] demonstrated the application of machine learning in fault detection for ORC systems dedicated to waste-heat recovery, illustrating the power of supervised learning models to detect four different faults using six sensor data.

This study aims to identify the strengths and limitations of Support Vector Machine for FDD and propose a methodology to optimally define the number and type of features (i.e., sensors or variables) used to train the machine learning algorithm. The findings will contribute to a deeper understanding of fault management strategies in ORC systems, promoting the development of more resilient and efficient energy recovery solutions.

2 Methodology

In this section the description of the organic Rankine cycle architecture is presented, as well as a general analysis of the faults considered in this study and the simulation approach.

2.1 ORC process architecture

The ORC power unit considered corresponds to a 50kWe expander, using refrigerant R1234yf as working fluid and with architecture as depicted in Figure 1.

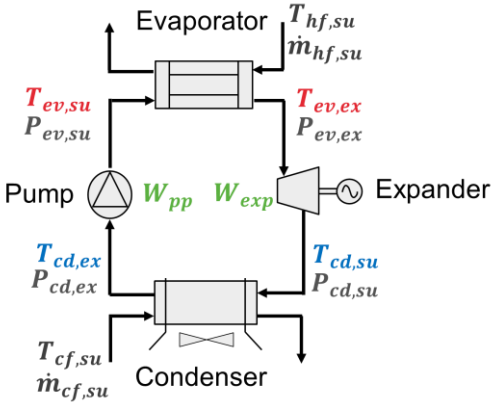

Figure 1: Process architecture.

The refrigerant in liquid state is pumped up, increasing its pressure at the moment of reaching the evaporator inlet, then heat is exchanged between the refrigerant and the secondary fluid which is in this case water at an average temperature $T_{hf,su} = 90$°C and mass flow rate $\dot{m}_{hf,su} = 12.5$ kg/s. The refrigerant boils and is heated up until a superheating state at the expander inlet. The vapour passes through the expander, making it rotate thus producing electrical power when connected to a generator, leading to a drop in temperature and pressure in the refrigerant at condenser input. The condenser ensures that the refrigerant will be back in a liquid state thus closing the loop. Water is used as secondary fluid in the condenser, at an average temperature $T_{cf,su} =$ 10°C and mass flow rate $\dot{m}_{cf,su} = 22.5$ kg/s.

The simulator is built considering refrigerant R1234yf as a working fluid, while its properties are computed using CoolProp [11] in Python. Braze Plate Heat exchanger for both, condenser and evaporator are selected due to their high efficiency for liquid applications. The methodology followed for these heat exchangers corresponds to a discretization according to the different refrigerant zones [12]. The condenser and evaporator are discretized in three zones (zi); superheating (sh), two-phase (tp) and subcooled (sc) zone. Those models are based on physical principles, and the solving process is carried out based on iteration of the saturation pressure at vapour quality. The problem is considered solved

when the difference between the real area and the area computed is lower than a tolerance value of 1 x 10 -6 [12].

The energy balance of the condenser is presented in Equation 1 for the refrigerant side. Additionally, the heat transfer equation presented in Equation 2 is considered to represent any change in the heat exchanger performance by a change in the operational conditions or geometry. For this equation a counter arrangement is considered between the refrigerant and the water loops.

$$\dot{Q}_{cd} = \sum_{i}^{sh,tp,sc} \dot{m}_r \cdot \left(h_{r,su,zi,cd} - h_{r,ex,zi,cd}\right) \quad (1)$$

$$\dot{Q}_{cd} = \sum_{zi}^{sh,tp,cd} \varepsilon_{zi,cd} \cdot \dot{C}_{min,zi,cd} \cdot \left(T_{r,su,zi,cd} - T_{w,su,zi,cd}\right) \quad (2)$$

The effectiveness of the heat transfer process can be computed by Equation 3 and 4, for the single phase and the two-phase zone, respectively.

$$\varepsilon_{zi,cd} = \frac{1 - e^{\left(-NTU_{zi,cd}\cdot\left(1-C_{r,zi,cd}\right)\right)}}{1 - C_{r,zi,cd} \cdot e^{\left(-NTU_{zi,cd}\cdot\left(1-C_{r,zi,cd}\right)\right)}} \quad (3)$$

$$\varepsilon_{tp,cd} = 1 - e^{\left(-NTU_{tp,cd}\right)} \quad (4)$$

The overall heat transfer coefficient in each zone (UA_(zi,cd)) considers the possibility of fouling in the heat exchanger which basically represent the thermal resistance of the fouling thickness in the heat transfer. The UA_(zi,cd)is computed as the sum of the thermal resistance of the water and refrigerant, beside is necessary to consider the thermal resistance of the plates (R_(zi,cd)). To relates the overall heat transfer within the effectiveness of the process Equation 5 is employed, which allows to compute the area, which will the comparison value for the iteration problem.

$$NTU_{zi,cd} = U_{zi,cd} \cdot A_{zi,cd} / \dot{C}_{min,zi,cd} \quad (5)$$

The BPHXs are validated against SWEP manufacturer data for the design condition. The model considered in the validation corresponds to:

• Model B633MH, within 278, 250 and 160 plates.

• Model B439H, within 220, 140 and 120 plates.

The developed model presents an error below 5% for the heat power capacity, 4 K for the exhaust refrigerant temperature and below 10% for the pressure drops.

A centrifugal pump is employed in this configuration due to its high efficiency, where the isentropic efficiency is extracted from the manufacturer curves for the operational range. The pump Movitec V F010/11 from KSB is chosen.

The expander is modelled by thermodynamic equations, assuming that the expansion is an isentropic process. For that an isentropic efficiency of 0.7 [21] is considered, and a volumetric efficiency of 0.95. The Equation 6 is employed to compute the expander power.

$$\dot{W}_{exp} = \dot{M}_r \cdot \varepsilon_{exp} \cdot \left(h_{r,e,ex,exp} - h_{r,su,exp}\right) \quad (6)$$

The nominal conditions of the ORC system are presented in Table 1, where the secondary fluid temperatures and mass flow rates can be analysed. Moreover, the parameters for the expander and pump are also presented, with their respective operational speed.

Table 1: Nominal simulation conditions.

Simulation condition	value
Heat source temperature ($T_{hf,su}$)	90 °C
Heat source mass flow rate ($\dot{m}_{hf,su}$)	12.5 kg/s
Heat sink temperature ($T_{cf,su}$)	10 °C
Heat sink mass flow rate ($\dot{m}_{cf,su}$)	22.5 kg/s
Expander isentropic efficiency	0.7
Pump isentropic efficiency	0.7
Expander speed (N_{exp})	3000 rpm
Pump speed (N_{pp})	2400 rpm

2.2 Faults description

In the present study 4 conditions are considered: the healthy state, heat exchanger fault, expander fault and pump fault. First an attempt for physical interpretation is provided and then the form how the simulator emulates the fault is described.

2.2.1 Heat exchanger fault: some of the most common faults in heat exchangers include fouling, leaking and blocking. Over extended operating periods, fouling faults arise due to crystallization, sedimentation, coking, corrosion, and similar processes. This results in a layer of fouling that gradually builds up on the surface of the heat exchanger. Because the heat conductivity of this

fouling layer is very low, it directly reduces the heat exchanger's efficiency, leading to significant economic losses [15]. For instance, when the evaporator operates at low efficiency, the input temperature to the expander drops, reducing the ORC's work capacity. In this study fouling is emulated by proportionally decreasing the number of plates as fouling occurs. Other types of faults are not considered at the moment.

2.2.2 Expander fault: The expander generates mechanical work by expanding high-temperature, high-pressure organic fluid, directly influencing the system's power output. Expanders are classified as either speed or volume type, with faults typically categorized into thermal performance and mechanical functional faults. Thermal performance faults often show as irregular discharge volume, abnormal pressure differences, and elevated exhaust temperatures, usually due to issues with parts like the air valve, piston ring, or cooling channels. Mechanical functional faults, on the other hand, manifest as unusual vibration, noise, or overheating, stemming from loosened components, wear-induced gaps, or structural cracks. These faults significantly decrease expander efficiency and, if left unaddressed, may lead to severe system failures. Due to the critical impact of mechanical faults, this paper focuses on diagnosing a specific type of mechanical fault, which is emulated as a decrease in the expander isentropic efficiency.

2.2.3 Pump fault: The pump functions to transport and pressurize the working fluid, with types generally classified as positive displacement, dynamic, or electromagnetic, based on how pressure is applied. Pumps are prone to various faults, including reduced speed, blockage, and leakage. When such issues arise, the pump cannot achieve the required pressure, resulting in a reduced mass flow rate of the working fluid. In this work a centrifugal pump is considered with fault emulated by a drop in the isentropic efficiency.

Finally, 4 sets of training data are collected from the ORC thermodynamic model, including a no-fault data set and three data sets with different typical faults, to detect whether a fault occurs in the system as summarized in Table 2.

Table 2: Summary of simulation fault conditions.

Data set	Simulation condition	Class label
No fault	Healthy state	0
Fault 1	Heat exchanger fouling	1
Fault 2	Mechanical fault in expander	2
Fault 3	Mechanical fault in pump	3

2.3 Data generation

In order to stablish during the simulation study conditions similar as the ones encountered on a real application, closed loop control for superheating and subcooling are employed with setpoint in $\Delta T_{sh,ref} = 7$ K and $\Delta T_{sc,ref} = 2$ K.

Additionally, for each simulation condition from Table 2, a random Gaussian distribution is employed for $T_{hf,su}$ and $T_{cf,su}$, with standard deviation 6 and 2, respectively.

3 FDD through Support Vector Machine

In this section the mathematical description of the machine learning classification algorithm implemented as FDD solution is presented, as well as the features selection and the tuning methodology.

3.1 Support Vector Machine (SVM)

The support vector machine, developed in [16], is a highly powerful and flexible model rooted in statistical learning theory. It is effective for both linear and nonlinear classification, regression, and outlier detection, especially with small sample sizes and high-dimensional data [17].

Support vector machines are supervised learning algorithms used for classification that work by finding an optimal hyperplane that maximizes the margin between two classes. In mathematical terms, for a set of training points (x_i, y_i) where $x_i \in \mathbb{R}^n$ and $y_i \in \{-1,1\}$, the objective of SVM is to maximize the margin $\frac{2}{\|w\|}$ between the classes.

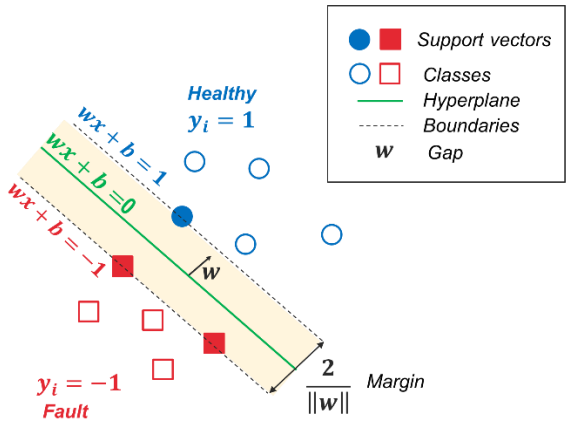

Figure 2: Principle of Support Vector Machine.

This is achieved by solving the optimization problem:

$$\textbf{min} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \tag{7}$$

$$subject\ to: \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i \tag{8}$$
$$for\ i = 1,2,\dots,N \ \text{ and } \ \xi_i \geq 0$$

Where w is the weight vector, and b is the bias term. Note that in order to prevent overfitting and achieving a proper trade-off between achieving a low error on the training data and minimizing the model complexity the regularization parameter C is used, with ξ_i the slack variables representing the degree of misclassification of each data point.

A small value of C allows the optimizer to increase the margin, even if it means misclassifying more points. It makes the decision surface smoother (high bias, low variance). A large value of C tries to classify all training examples correctly by giving the model the freedom to select more complex decision boundaries (low bias, high variance).

Additionally, when the data is not linearly separable, SVM introduces a kernel function $\mathbf{K}(x_i, y_i)$ to transform the data into a higher-dimensional space, enabling linear separation. In this study the Radial Basis Function (RBF) kernel is used as transformation function:

$$\mathbf{K}(x_i, y_i) = \exp(-\gamma\|x_i - x_j\|^2) \tag{9}$$

The parameter γ determines the spread of the kernel, a larger value makes the kernel narrower, meaning each training point has a more localized influence. This can lead to a model that fits the training data closely (risking overfitting). Whilst a smaller value makes the kernel function broader, causing each point to have a more significant influence on the decision boundary, which can help the model generalize better.

In conclusion the optimization problem from Equation 7, Equation 8 and Equation 9, together generate a solution called in machine learning a soft-margin SVM algorithm. Where C controls the regularization and margin, allowing for a trade-off between margin size and classification error, while γ adjusts the influence range of each data point in nonlinear SVMs with kernels.

3.2 Optimal trade-off tuning

A first step before tuning the algorithm consists in normalising the data set between {-1,1}. This can be achieved by applying the following relation

$$\mathbf{X} = \frac{x_i - \mu}{\sigma} \tag{10}$$

with μ the media and σ the standard deviation of each data set. Subsequently, in order to find the optimal C and γ parameters defined in previous section, one can perform an exhaustive search over a specified parameter grid, e.g., C: [0.01, …, 1000] and γ: [0.001, 0.01, 0.1, 100], by evaluating each combination of C and Gamma values using cross-validation.

3.3 Feature selection

In the context of support vector machines and machine learning in general, a feature is an individual measurable property or characteristic of the data being analysed. Features are the input variables used by the SVM model to distinguish between different classes in classification tasks. In the case of FDD for ORC systems using SVM, it is necessary to select between the different sensors available the ones more relevant to correctly detect the different faults.

Due to the limited amount of literature on FDD for ORC systems, there is no clear reference on which sensors or sensor combinations should be used for different fault types. However, considering similar systems, one can use industrial refrigeration systems as a reference. For example, in [6], nine features are selected, namely evaporating temperature, condensing temperature, temperature differences at various cycle components, and superheat. An additional interesting study applied for chillers is encountered in [18], where thirteen different features are employed among them the instantaneous input power, temperature and pressures at different components inputs and outputs, subcooling level, chiller efficiency, and superheat.

In SVM for classification, there is a key trade-off between the number of features (dimensions of the data) and the number of classes (distinct categories being classified). This trade-off affects both the computational complexity of the model and its generalization to new data. In a simple approach one can employ all sensor data available to train a SVM classifier, or use knowledge from the system to interpret from the physics of the system the more suitable features, additionally, one can use algorithms such as the Relief method to optimally select the most important features [19] and [20].

The Relief algorithm is a feature selection method that identifies important features by evaluating their ability to distinguish between classes. For each randomly chosen instance in the dataset, Relief finds

the nearest neighbour from the same class (near-hit) and from a different class (near-miss). It then adjusts the weight of each feature based on how similar its value is in the target instance compared to the near-hit and near-miss instances. If a feature's values are similar within the same class and different across classes, it receives a higher weight, indicating its importance for classification. After multiple iterations, features with the highest weights are selected as the most relevant.

SVM is inherently a binary classifier, therefore, to extend Support Vector Machines (SVM) for multi-class classification (e.g., with 4 classes for the defined healthy and faults to be studied), one can use the One-vs-One (OvO) approach, in which a separate SVM classifier it is trained for every possible pair of classes. For a problem with k classes, this results in $k(k-1)2^{-1}$ classifiers. In the case of 4 classes, it would create 6 binary classifiers (one for each pair of classes). During classification, each classifier makes a vote, and the class that receives the majority of votes across all classifiers is chosen as the predicted class. This method although can be computationally intensive it is effective, especially with smaller numbers of classes as in the present study.

4 Results and analysis

Datasets are generated using the simulation model based on the descriptions in Table 2, with additional severity levels for each fault set at 10%, 30%, and 50%. The algorithms are trained using known data and subsequently tested on unknown data across different scenarios, each representing faults with varying degrees of severity.

The testing procedure is as follows: the algorithm is trained based on the healthy scenario and for each fault type at 10% of severity level. Afterwards, the classification algorithm is tested for the other severity cases i.e., 30% and 50%.

4.1 Scenario 1: All sensor data

In the first scenario, all available sensor data is used as features to evaluate the algorithm's performance. This scenario assumes that the user believes having more information available can enhance the machine learning (ML) predictions. The sensor values to be used as features are listed in Table 3. Note that in this case, raw measured data (RMD) is used to create the features. However, in other cases, calculated data (CD) or secondary variables, which are derived from algorithms or combinations of primary variables, may be used.

Table 3: Variables used as features in scenario 1.

Feature	Type	Feature	Type
$T_{hf,su}$	RMD	$P_{ev,ex}$	RMD
$\dot{m}_{cf,su}$	RMD	$T_{cd,su}$	RMD
$\dot{m}_{hf,su}$	RMD	$P_{cd,su}$	RMD
$T_{ev,su}$	RMD	$T_{cd,ex}$	RMD
$P_{ev,su}$	RMD	$P_{cd,ex}$	RMD
$T_{ev,ex}$	RMD	-	-

The next step is to perform a grid search to tune the C and γ parameters, where the result of each combination is a mean score that reflects how well the SVM model performs on unseen data, as determined by cross-validation. This procedure is then applied to the training datasets (healthy and fault datasets with 10% severity). The results are presented as reference for the nonlinear region encountered, as depicted in Figure 3, leading to $C = 10$ and $\gamma = 0.1$.

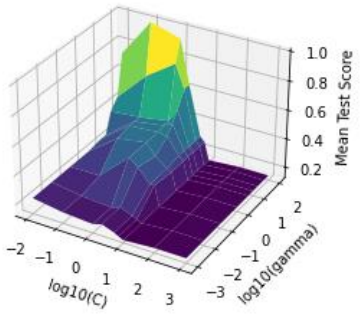

Figure 3: Optimal SVM tuning by cross-validation.

The prediction capabilities of the tuned SVM classifier are then tested on unseen datasets, specifically for severity cases of 30% and 50%. As an example of its performance, the prediction results of the SVM for fault detection and diagnosis (FDD) using features from scenario 1 are presented in Figure 4. In this case, the algorithm fails to predict 8 out of 56 data samples. Notably, it mainly struggles to distinguish between the healthy (class 0) and evaporating fouling (class 1), as outlined in the classification presented in Table 2, whilst for faults in the expander and the pump are correctly detected.

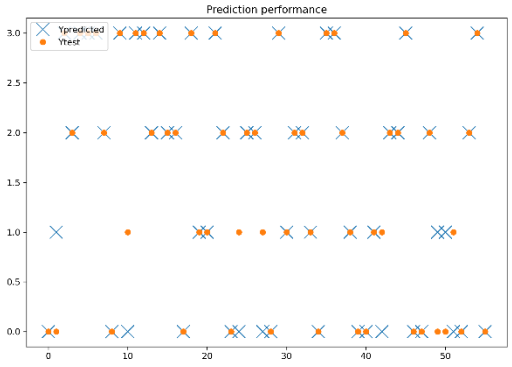

Figure 4: Scenario 1: Prediction performance of FDD with 30% fault severity (error: 8/56).

4.2 Scenario 2: All sensor data + combinations

In the second scenario, all sensor data is used during the training phase, along with additional variables derived from combinations of sensor data. In scenario 1, the main errors occurred because the classifier struggled to distinguish between healthy data points and evaporator fouling. To address this issue, additional information related to system performance should be included. Examples of such variables include the rate of heat transfer $\dot{Q}$, cycle efficiency η, and the power consumed W by the pump or produced by the expander. These additional variables could provide valuable insights, potentially enabling the SVM algorithm to achieve greater accuracy. Table 4 lists the variables used as features, indicating whether each is a raw measured data (RMD) variable or a calculated data (CD) variable. The heat transfer and cycle efficiency are defined as follows:

$$\dot{Q}_{ev} = \dot{m}_{wf}(h_{ev,su} - h_{ev,ex}) \tag{11}$$

$$\eta_{ORC} = \frac{(W_{exp} - W_{pp})}{\dot{Q}_{ev}} * 100 \tag{12}$$

Table 4: Variables used as features in scenario 2.

Feature	Type	Feature	Type
$T_{hf,su}$	RMD	$T_{cd,ex}$	RMD
$\dot{m}_{cf,su}$	RMD	$P_{cd,ex}$	RMD
$\dot{m}_{hf,su}$	RMD	W_{exp}	RMD
$T_{ev,su}$	RMD	W_{pp}	RMD
$P_{ev,su}$	RMD	η_{ORC}	CD
$T_{ev,ex}$	RMD	$\dot{Q}_{cd}$	CD
$P_{ev,ex}$	RMD	$\dot{Q}_{ev}$	CD
$T_{cd,su}$	RMD	$\dot{m}_{wf}$	RMD
$P_{cd,su}$	RMD	-	-

After defining the features, a grid search is performed to tune the C and γ parameters, resulting in $C = 1$ and $\gamma = 0.1$. The training procedure is then applied to 70% of the samples from the healthy and fault datasets with 10% severity, while the remaining 30% is reserved for testing.

The prediction performance for the dataset with 30% severity is shown in Figure 5. It is observed that the FDD algorithm successfully distinguishes between the different faults without any errors. Particularly, it accurately differentiates between healthy data and the evaporator fouling condition without issues. The algorithm is able to classify both datasets i.e., 30% and 50% severity with no error.

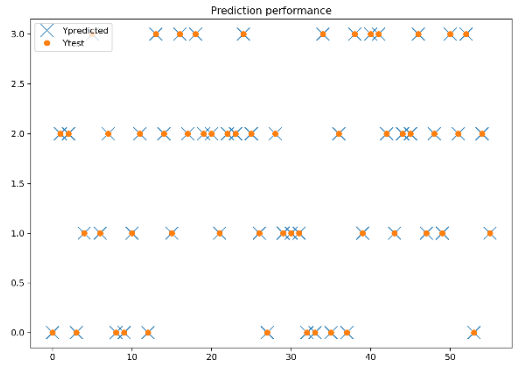

Figure 5: Scenario 2: Prediction performance of FDD with 30% fault severity.

4.3 Scenario 3: Optimizing features

Following the improved results from scenario 1 achieved in scenario 2 by including additional information, an investigation is conducted to determine whether all features used are truly necessary. The Relief algorithm, as described in Section 3.3, is applied to the healthy and 10% severity datasets during training to identify whether fewer features can be used without compromising accuracy in fault detection and diagnosis. Scenario 3 initially considers the same features as in scenario 2, where the weights assigned by the Relief algorithm to each feature represent its relevance or importance in contributing to the classification task. These weights are particularly useful for determining whether a feature can be eliminated, as they specifically indicate how well a feature helps distinguish between classes while considering similarities between samples. The results of this analysis are illustrated in Figure 6.

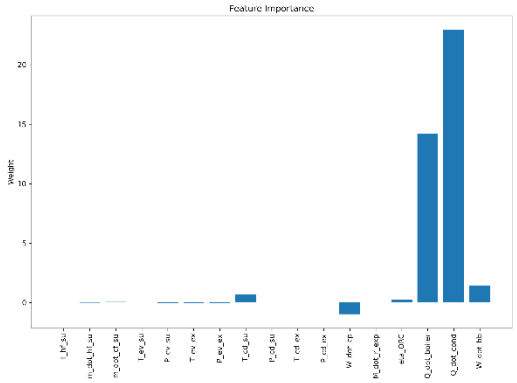

Figure 6: Scenario 2: Prediction performance of FDD with 30% fault severity.

The results shown in Figure 6 suggest that not all features are relevant for accurately classifying fault conditions. Various methods for reducing the number of features can be considered, such as applying a threshold value, selecting a specific percentage, or defining a minimum number of features. In this study, the features with the highest numerical weight values were selected, as shown in Table 5.

Table 5: Variables used as features in scenario 3.

Feature	Type
W_{pp}	RMD
$\dot{Q}_{cd}$	CD
$\dot{Q}_{ev}$	CD

After considering the three selected features, the grid search is performed to tune the C and γ parameters, resulting in C=0.1 and γ=1. The training procedure is then applied to 70% of the samples from the healthy and fault datasets with 10% severity, while once more the remaining 30% is reserved for testing. The trained FDD SVM algorithm is assessed in terms of prediction performance on the 30% and 50% severity fault datasets. The FDD successfully classifies the faults with no errors, achieving the same performance as in scenario 2. However, in scenario 2, 17 features were required, but in this case, only 3 features were necessary. The prediction capabilities for the case of the scenario 3 are depicted in Figure 7.

A final summary of the results obtained for the different proposed scenarios is presented in Table 6, which lists the total accuracy as well as the scores based on fault severity.

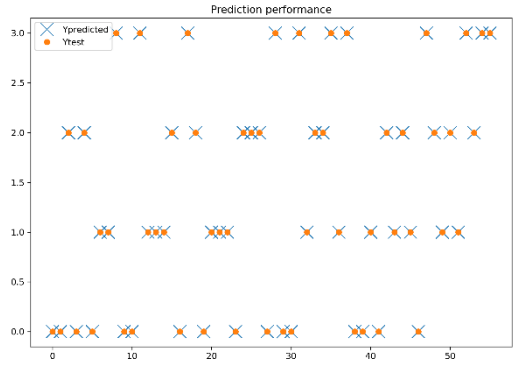

Figure 7: Scenario 3: Prediction performance of FDD with 30% fault severity.

Table 6: Accuracy score of the diagnosis using the proposed SVM-FDD for different fault severities.

Scenario	Number features	Severity	Total Accuracy	Score
1	11	10%	43.75%	7/16
	11	30%	85.71%	48/56
	11	50%	85.71%	48/56
2	17	10%	100%	16/16
	17	30%	100%	56/56
	17	50%	100%	56/56
3	3	10%	100%	16/16
	3	30%	100%	56/56
	3	50%	100%	56/56

5 Conclusions

This paper presents a fault detection and diagnosis (FDD) algorithm for organic Rankine cycle (ORC) power systems, which utilizes a support vector machine (SVM) to classify various operational conditions. Four conditions are considered in this study: the healthy state, evaporator fouling fault, and mechanical faults in the expander or pump. Additionally, the severity of faults is explored at three levels: 10%, 30%, and 50%.

The SVM algorithm is trained using the healthy state and 10% severity fault data, and tested on datasets with 30% and 50% fault severity. The results demonstrate strong predictive performance, with the model effectively generalizing to unseen data.

Regarding model tuning, a grid search approach is employed to identify the optimal tuning parameters for the SVM. Furthermore, the study emphasizes the importance of proper feature selection across three scenarios, ranging from raw sensor data to more

complex variables derived from the measurements. These additional features provide valuable insights to improve the algorithm's performance.

Using the Relief algorithm for feature selection, the number of features is reduced from 17 to just 3, while maintaining excellent predictive performance with no errors.

A limitation of the present study is the reliance on labelled data during the training process. Consequently, future work will focus on the development of self-learning algorithms capable of working with datasets where only a small portion of the data is labelled, and the rest remains unlabelled. This is particularly important for fault detection and diagnosis (FDD), as it can be challenging to obtain sufficient data representing various fault conditions, or to simulate such conditions effectively. For example, emulating the presence of non-condensable gases in Organic Rankine Cycles (ORCs) can be difficult. Self-learning algorithms can address this by using the labelled data to train an initial model, then iteratively improving the model by predicting labels for the unlabelled data and incorporating those predictions into the training process

Acknowledgement

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069740.

The information presented on this document reflects only the authors' view. The Agency is not responsible for any use that may be made of the information it contains.

References

[1] C. Wieland, C. Schifflechner, K. Braimakis, F. Kaufmann, F. Dawo, S. Karellas, G. Besagni, C. Markides, "*Innovations for organic Rankine cycle power systems: Current trends and future perspectives"*, Applied Thermal Engineering, Volume 225, p. 120201, ISSN 1359-4311, 2023.

[2] Dickes, R., *Charge-sensitive methods for the off-design performance characterization of organic Rankine cycle (ORC) power systems*, PhD Thesis, University of Liege, 2019.

[3] Wang, J., Zuo, Q., Liao, G., Luo, F., Zhao, P., Wu, W., He, Z., & Dai, Y. *"Machine Learning–Based Fault Detection and Diagnosis of Organic Rankine Cycle System for Waste-Heat Recovery".* Journal of Energy Engineering, 2021.

[4] D. Dragan, "Fault Detection of an Industrial Heat-Exchanger: A Model-Based Approach", Journal of Mechanical Engineering, Vol. 57 (6), p. 477-484, 2011

[5] S. Choi, S. Krumdieck, "*Dynamic modeling of organic Rankine cycle (ORC) system for fault diagnosis and control system design*", Proceedings 38th New Zealand Geothermal Workshop, 23 - 25, Auckland, New Zealand, November 2016.

[6] R. van de Sand, "A Predictive Maintenance Model for Heterogeneous Industrial Refrigeration Systems", PhD thesis, Tor Vergata University of Rome, Italy, 2021

[7] A. Elmouatamid, B. Fricke, J. Sun, P. Pong, "*Air Conditioning Systems Fault Detection and Diagnosis-Based Sensing and Data-Driven Approaches*", Energies, Vol. 16, p. 4721, 2023.

[8] H. Hosamo, P. Ragnar, K. Svidt, D. Han, H. Nielsen, "*A Digital Twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics*", Energy and Buildings, Vol. 261, p. 111988, ISSN 0378-7788, 2022

[9] K. Yan, Z. Ji, W. Shen, "*Online fault detection methods for chillers combining extended Kalman filter and recursive one-class SVM*", Neurocomputing, Vol. 228, p. 205-212, ISSN 0925-2312, 2017

[10] J. Wang, Q. Zuo, G. Liao, F. Luo, P. Zhao, W. Wu, Z. He, Y. Dai, *"Machine Learning-Based Fault Detection and Diagnosis of Organic Rankine Cycle System for Waste-Heat Recovery"*, Journal of Energy Engineering, Vol. 147(4), p. 04021016, 2021

[11] I. H. Bell, J. Wronski, S. Quoilin, and V. Lemort, '*Pure and Pseudo-pure Fluid Thermophysical Property Evaluation and the Open-Source Thermophysical Property Library CoolProp'*, Ind. Eng. Chem. Res., vol. 53, no. 6, pp. 2498–2508, Feb. 2014

[12] C. Cuevas, J. Lebrun, V. Lemort, and P. Ngendakumana, '*Development and validation of a condenser three zones model'*, Applied Thermal Engineering, vol. 29, no. 17–18, pp. 3542–3551, Dec. 2009.

[13] B. Eck and B. Eck, *Fans: design and operation of centrifugal, axial-flow and cross-flow fans*, 1. engl. ed. Oxford: Pergamon Pr, 1973

[14] V. Lemort, S. Declaye, and S. Quoilin, '*Experimental characterization of a hermetic*

scroll expander for use in a micro-scale Rankine cycle', Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, vol. 226, no. 1, pp. 126–136, Feb. 2012

[15] Markowski, M., M. Trafczynski, and K. Urbaniec. "Identification of the influence of fouling on the heat recovery in a network of shell and tube heat exchangers." Appl. Energy 102 (Feb): 755–764, 2013.

[16] V. Vapnik, "The nature of statistical learning theory". Berlin: Springer, 2000.

[17] L. Qiao, and Q. Chen. "Forecasting models for hydropower unit stability using LS-SVM." Math. Probl. Eng. 2015 (1): 1–9, 2015

[18] K. Yan, J. Zhiwei, S. Wen, *"Online fault detection methods for chillers combining extended kalman filter and recursive one-class SVM"*, Neurocomputing, Volume 228, p. 205-212, ISSN 0925-2312, 2017

[19] M. Robnik-S˘ikonja, I. Kononenko, *"Theoretical and empirical analysis of ReliefF and RReliefF"*, Mach. Learn. 53 (1–2) (2003) 23–69.

[20] Y. Sun, J.Li, *"Iterative Relief for feature weighting"*, in: Proceedings of the 23rd International Conference on Machine Learning, ACM, 913–920, 2006.

[21] M. Pezo, C. Cuevas, E. Wagemann, A. Cendoya, "Net Zero Energy Building technologies - Reversible Heat Pump/Organic Rankine Cycle coupled with Solar Collectors and combined Heat Pump/Photovoltaics - Case study of a Chilean mid-rise residential building", Applied Thermal Engineering, 252, 123683, 2024