One-Class Classifiers for Novelties Detection in Electrical Submersible Pumps

Gabriel Soares Baptista*, Lucas Henrique Sousa Mello*, Thiago Oliveira-Santos*, Flávio Miguel Varejão*, Marcos Pellegrini Ribeiro† and Alexandre Loureiros Rodrigues*
* Universidade Federal do Espírito Santo (UFES), Vitória, ES, Brazil
†CENPES/PDP/TE, Petrobras, Rio de Janeiro, RJ, Brazil

***Abstract*—Detecting anomalies and fault novelties is of high interest in the industry due to the scarcity of fault examples to train classification systems. In this article two algorithms for anomaly detection, One-Class SVM and Isolation Forest, are successfully used as effective methods for detecting fault novelties in problems of electrical submersible pumps. Faults in submersible electric pumps generate an enormous cost for companies in the oil and gas sector, since the cost of stopping production to change the equipment is excessive, which makes it necessary to identify problems before implementation. Empirical evaluation shows that both one-class classifiers performed satisfactorily, obtaining macro f-measure values of approximately 0.86. For comparison purposes, a Random Forest trained in a conventional binary classification manner is tested and achieved a macro f-measure of 0.95. Results show that the proposed solutions can have practical applications in the classification of problems in electrical submersible pumps, changing the way the oil and gas industry addresses this difficulty.**

I. Introduction

An Electrical Submersible Pump (ESP) [1] is an industrial equipment widely used in offshore oil and gas production. It is composed of multiple pumps and electric motors. Increasing the durability of this equipment by avoiding acquiring faulty components is an important task studied in the literature [2]–[5]. Since an Electrical Submersible Pump (ESP) is deployed under deep waters, any maintenance is unfeasible. Failures may lead to significant financial losses due to the high cost of replacing the equipment and especially because of the interruption of production over a long period of time. The machine learning techniques play a major role in this scenario, which may lead to automatic and precise fault diagnosis.

Many works have addressed such a problem by using machine learning techniques on a real-world dataset [2]–[5]. The dataset has data from five different conditions. In addition to the normal condition (without defect), a vibration signal collected in the test can indicate misalignment, unbalance, and mechanical rubbing. In some cases, a faulty sensor may generate abnormal vibration behaviour. This is considered as a faulty pattern, although the abnormal behaviour is not necessarily related to the equipment. In [2], a series of comparisons are made regarding the results of various classifiers in this multiclass dataset. The work in [3] shows a comparative study of classifier architecture, with emphasis on Extreme Learning Machine, showing that the proposed method has an excellent learning capability in this type of problem. The authors in [4] develop a methodology for combining decision templates with several classifiers, seeking improvements in the diagnosis of the ESPs. All these before mentioned works used a set of hand-crafted features extracted from the frequency-domain spectrum of the vibration signal. A recent study [5] takes an approach based on triplet networks, a special deep neural network technique, in which the neural network is trained for automatically extracting relevant features directly from the frequency-domain spectrum raw data, without the intervention of a specialist.

The methodologies covered in previous works are valuable resources in the analysis of ESPs, since specialists who carry out this analysis efficiently are very scarce. Usually, the knowledge needed to perform this task is built over many years of experience and is not easily taught. Experts use computational tools to visually analyze the spectrum of vibration signal data collected from several sensors along different points of an ESP system. Although this procedure is quite effective in reducing the risk of failure after deployment of the system, the lack of trained professionals for the service requires more robust methods capable of dealing with the problem in a simpler way. This also highlights the difficulty of acquiring and labelling new data for the problem.

Although the aforementioned methods have raised the standards for dealing with the problem, they all assume a supervised setup with a finite well-known set of possible faults that an ESP may present. However, this is not always the case in practice since a new unknown type of fault may appear unexpectedly. This fact introduces a challenging problem, since an expert must detect the new fault type and produce new training instances to retrain the classification model. Considering the challenging process of generating and the possibility of these events being rare, it may not be possible to have a representative training dataset for the fault.

The machine learning techniques already employed in this context are not able to identify patterns of novel faults that may eventually occur because they are not trained with these patterns, once they are rare or may not even have appeared yet. Therefore, this work approaches the original problem from a different perspective, trying to indicate whether the new data has normal or abnormal behavior without using abnormal sig-

This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, by National Council for Scientific and Technological Development – CNPq and by CENPES-Petrobras.

nals during training. In this case, a normal equipment indicates a submerged electric pump that, before being installed, does not present any type of failure, as its signal only presents characteristics of the normal type. The importance of clearly identifying whether the equipment fits this type is enormous, as the cost generated by stopping production to change the equipment is much higher than the acquisition cost. The idea is to learn what is the normal behavior and send an alarm anytime something different is measured. This enables detecting any abnormal signal, including those that may appear with an unknown fault type not available in training data. Given the new problem description, this work does not aim at overtaking the state-of-the-art approaches for the supervised task but to provide a viable alternative to detect new faults when they appear.

The aforementioned problem is addressed in the literature as anomaly or outlier detection and is usually referred as one-class classification in the field of machine learning. One-class classification methods are viable options for anomaly detection, since they require only one target class (the majority class, typically the normal class). These methods are able to detect if new incoming data belongs to the normal class or not, without requiring any training examples from the abnormal classes. Researchers have been active in this field and a good overview of the recently proposed methods can be found in [6].

In this context, this work leverages the use of one-class classification techniques for fault detection in ESP. More specifically, it investigates two well-known methods, the One-Class SVM [7] and Isolation forest [8], for detecting abnormal ESP behaviour. The main challenge of this study compared to previous approaches is detecting new untrained faults. Results show that both of them can efficiently detect abnormal signal patterns in ESP approaching the results of state-of-the-art methods trained in a fully supervised manner.

The remainder of this paper is organized as follows. Section II briefly presents the methodology for acquiring vibration signals. Section III discusses how fault detection of ESPs may be performed using the One-class Classifier approach. Section IV shows the adopted experimental methodology. Section V compares empirically the one-class classifiers with a Random Forest binary classifier used as an upper limit. Section VI discusses the results and presents ideas for future works.

II. Signal Processing for Fault Diagnosis Systems

An ESP is an equipment composed of electrical motors, pumps and seals. Accelerometers are attached at strategic positions of an ESP for collecting vibration signals.

Once the sensors are attached, the ESP is tested in a suitable laboratory and under certain frequency and flow operating conditions. Each sensor measures the acceleration of vibration at a certain point in the ESP. The vibration signals are therefore collected from the time domain at a sampling rate of 4096 points per second and converted to the frequency domain as it is an easier format to be analyzed by humans. A specialist is able to analyze the graphics of these signals in the frequency domain and provide the correct diagnosis. However, signals in the frequency domain are still difficult to diagnose by traditional machine learning methods due to the high number of points per signal and due to the complexity of the relationship between these points. Due to this difficulty, a set of features was extracted with the help of a specialist, seeking to represent the signals in a synthesized way by 8 features used in [2].

- **median(3,5)** Median of the amplitudes in the interval (3Hz, 5Hz);
- **median(F-1,F+1)** Median of the amplitudes in the interval (**F**-1Hz, **F**+1Hz);
- **a:** Coefficient a of the exponential regression of type $e^{(a \cdot X + b)}$ where X is an array of equally separated frequencies from 5Hz to 19Hz.
- **b:** Coefficient b of the exponential regression of type $e^{(a \cdot X + b)}$ in the interval (5Hz, 19Hz);
- **rotation1x:** Frequency of the highest amplitude in the interval (**F**-3Hz, **F**-0.2Hz);
- **peak1x:** Amplitude in **rotation1x**;
- **peak2x:** Amplitude in 2·**rotation1x**;
- **rms(F-1,F+1)** Root mean square of the amplitudes in the interval (**F**-1, **F**+1).

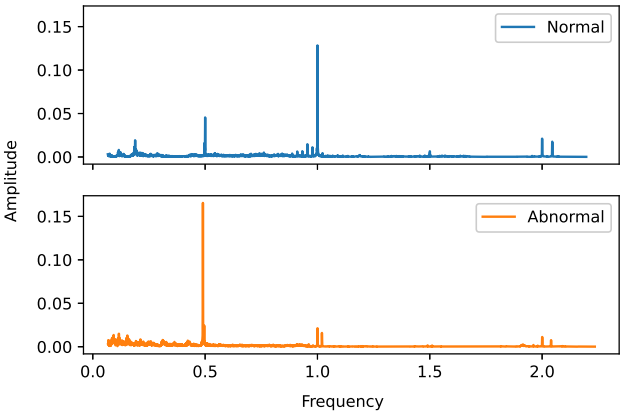


Fig. 1: Sample of a normal and abnormal signal.

III. Fault Detection of ESPs via One-Class Classification

Novelty detection or detection of new fault types is a challenging problem. An ideal solution should not require the intervention of a specialist. In this situation, an ideal solution should work directly on the available data, not assuming any type of special distribution for the abnormal classes. In this paper, we propose to adopt one-class algorithms for solving the fault detection problem in ESPs. Instead of relying on a representative training dataset composed of multiclass training data, our solution uses only signals of normal behaviour.

The proposed system starts with a preprocessing step that transforms signals of ESPs with normal behaviour in the

Fig. 2: A ESP system with six components and attached 36 sensors. Each square or circle represents a sensor, the symbology represents orthogonal placement.

time domain (usually having 400000 points) to the frequency domain and then extracts the eight real-valued features defined in Section II. Subsequently, a one-class classification model is trained on these eight pre-defined features in order to adjust the threshold parameter defining what is normal and what is not based on the normal data only. After trained, the one-class model can be applied to infer whether a new signal of unknown class belongs to the normal or abnormal group (including new and old faults). Figure 3 illustrates the proposed solution. Note that the proposed system is independent of how many fault types exist and how many are currently known since it does not consider the fault during training.

Two one-class classification models are investigated to compose the system, one-class SVM [7] and Isolation Forest [8], therefore they are described in detail in the following subsections.

A. One-Class SVM

The one-class SVM classifier [7] adapts the binary support-vector-machine (SVM) classifier to one-class problems, learning the limits of the target class (normal signal), and being able to classify points outside this limit (outliers).

1) The standard Support Vector Machine: A SVM tries to find a hyperplane for separating data into the two classes. The SVM considers the data set $\Omega = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ as points where $x \in \mathbb{R}^d$ in any space where x_i is the i-th entry point and $y_i \in \{-1, 1\}$ indicates the class the data belongs. A ϕ nonlinear function creates a nonlinear decision boundary by projecting the data to a larger dimension where there may be a "straight" hyperplane that separates the data from the two classes. When this hyperplane is projected back into the original space it assumes a non-linear form. The distance from the hyperplane to the nearest points of each class is ideally equal, seeking the maximum margin between classes. To avoid that noise data provoke classification overfitting, the gap variables ξ_i are introduced. This allows some points to stay within the margin, while a constant $C > 0$ determines the trade-off between maximizing the margin and the number of points of training data within the margin, i.e. training errors. The function that must be minimized in the SVM classifier is defined as:

$$\begin{aligned} &\min_{w,b,\xi_i} \frac{\|w\|^2}{2} + C\sum_{i=1}^{n}\xi_i \\ &\text{subject to:} \\ &y_i\left(w^T\phi\left(x_i\right)+b\right) \geq 1-\xi_i \quad \text{for all } i=1,\ldots,n \\ &\xi_i \geq 0 \quad \text{for all } i=1,\ldots,n \end{aligned} \tag{1}$$

When this quadratic problem is solved using Lagrange multipliers, the rule of the decision function for a data point x becomes:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{n}\alpha_i y_i K\left(x, x_i\right) + b\right)$$

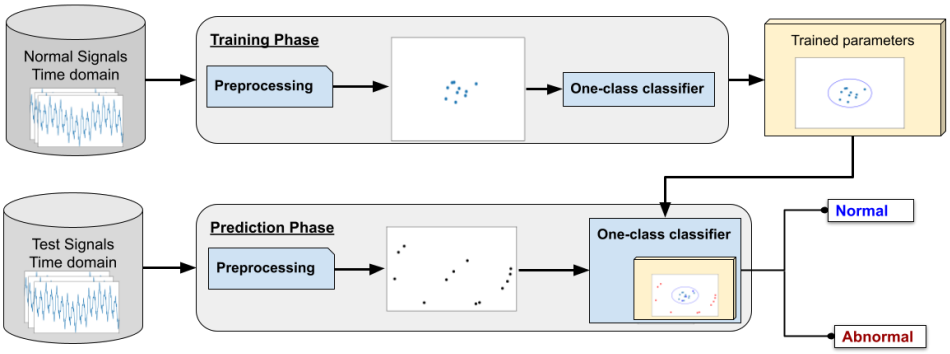
where the function $K(.,.)$ is a kernel function and α_i are the Lagrange multipliers, every $\alpha_i > 0$ is weighted in the decision function and, therefore, "supports" the machine. Considering that SVMs are usually sparse, there will be few Lagrange multipliers with a non-zero value.

2) One-Class SVM: An approach for adapting standard SVM classifiers for one-class problems was proposed by [7]. This method separates the data from the origin and maximizes the distance of the hyperplane from the origin, resulting in a binary function that captures regions of the space where the probability density of the data resides, returning +1 in a "small" region (capturing the data points training) and -1 elsewhere.

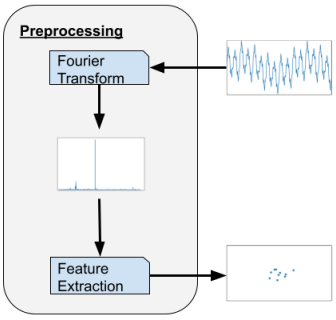
Therefore, the quadratic equation to be minimized is defined as:

$$\begin{aligned} &\min_{w,\xi_i,\rho} \frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho \\ &\text{subject to:} \\ &\left(w\cdot\phi\left(x_i\right)\right) \geq \rho-\xi_i \quad \text{for all } i=1,\ldots,n \\ &\xi_i \geq 0 \quad \text{for all } i=1,\ldots,n \end{aligned} \tag{2}$$

It is possible to visualize the similarity of the objective functions (1) and (2). In the standard SVM, the parameter C is a regularization parameter used to decide the smoothness of the fit. In (2), the parameter ν assumes this role, defining an upper limit in the fraction of outliers (mistaken classified examples in the training set) and is a lower limit of the number of training examples used as a support vector. Unlike the parameter C in (1), the parameter ν always varies from 0 to 1 and represents the proportion of outliers (or mistaken classified examples) in the training set. Therefore, based on this interpretation, setting a value for this parameter is a feasible task for the domain specialist.



(a) Training and Prediction phases for one-class classifiers.



(b) Preprocessing done on signals in the time domain.

Fig. 3: Proposed fault detection process via one-class classification.

Resolving by the Lagrange technique again and using a kernel function for calculations of the scalar product, the new decision function becomes:

$$f(x) = \operatorname{sgn}\left((w \cdot \phi(x_i)) - \rho\right) = \operatorname{sgn}\left(\sum_{i=1}^{n} \alpha_i K(x, x_i) - \rho\right)$$

Therefore, this approach creates a hyperplane characterized by ω and ρ that has maximum distance from the origin in the F resource space and separates all data points from the origin. Another approach [9], not used in this article, creates a circumscribed hypersphere around the data in the resource space.

In this work, the most used choice for the kernel function $K(.,.)$ is adopted, whereas an empirical experiment was conducted to estimate the proportion of outliers ν. The radial basis function (or RBF kernel), which is mathematically defined as in (3), is used as kernel function.

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{n_f \cdot \sigma^2} \|\mathbf{x} - \mathbf{x}'\|^2\right) \tag{3}$$

where n_f is the number of features (8 in our case) and σ is the sample standard deviation.

An empirical study was conducted to set the value for the parameter ν. A sample of 30 normal condition (without defect) examples was selected and their classification was reviewed by the specialist. Miss-classification was detected on 6 examples, therefore we set $\nu = 0.02$. In order to avoid bias of any kind, those 30 examples were discarded from the training process.

B. Isolation Forest

An approach for adapting tree-based methods to one-class problems was proposed in [8]. The method seeks to isolate anomalies, since they are "few and different" and, therefore, more susceptible to isolation. In a random tree induced by the data, the partitioning of the instances is repeated recursively until all instances are isolated, producing shorter paths for anomalies, because by assumption they are easier to isolate. Indeed, fewer number of anomalies results in a smaller number of partitions, since a fewer number of instances requires a smaller number of splits in the dataset for isolating each instance. A data set X = $(x_1, x_2, \ldots, x_n)$ is splitted recursively by randomly selecting an attribute q and a division value p, until the tree reaches a maximum size or data can no longer be split (there is a single instance or all instances have the same features values). Path Length $h(x)$ is defined as the number of edges that traverses from the root node to the leaf node for an instance x in an isolation tree. Since an isolation tree (iTree) has a structure equivalent to a binary search tree (BST), the average estimate of height $h(x)$ for determining path length in the tree is the same as for unsuccessful search in a BST, being defined as:

$$c(n) = 2H(n-1) - (2(n-1)/n)$$

where n is the number of instances in the dataset and H is the harmonic number, which can be estimated by $H(i) \approx ln(i) + 0.5772$ (Euler's constant). As $c(n)$ is the average of $h(x)$ found, this value is used to normalize $h(x)$ and get an estimation of the anomaly score for a give instance x:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}},$$

where $E(h(x))$ is the average value of $h(x)$, taken over all trained trees composing the isolation trees method. It is interesting to note that the values of s, anomaly score, indicate that:

- if s is very close to 1, then x is very likely to be an anomaly;
- if s is much smaller than 0.5, then x is likely to be a normal value;
- if for a sample, all instances have s close to 0.5, it is safe to assume that there are no anomalies in the sample.

The authors in [8] verify, based on empirical tests, that convergence is usually achieved before 1000 trees (estimators), but this quantity was chosen to be sure that the method converges well.

IV. Experimental Methodology

This section describes the experimental setup, the dataset used and the classification models used for solving our problem.

A. Dataset

The dataset described in [5] is used to conduct the experiments of this paper. The dataset is composed of 5617 vibration signals obtained through multiple sensors coupled to strategic points of an ESP (see Section II). Each signal can show evidence of three types of possible faults: misalignment, imbalance and rubbing. There are cases in which the sensor itself is defective, causing abnormal behavior in the collected data. Although not related to the equipment, a faulty sensor is considered a fault of the system. In this paper, the one-class classifiers are trained with signals with a normal behaviour. In testing phase, signals with abnormal behaviour are added for performance evaluation. The dataset is composed of 4493 normal signals (80%) and 1124 abnormal signals (20%). The complete list of distribution of fault types among signals is presented in Table I.

TABLE I: Class distribution of 5617 collected vibration signals.

Class name	Type	A priori distribution [%]
Normal	normal	80
Rubbing	abnormal	4.86
Faulty sensor	abnormal	5.25
Misalignment	abnormal	0.93
Unbalance	abnormal	8.96

The dataset in question is not in the public domain, however the authors intend to make it available through an article, so it would be interesting to follow the authors' website.

B. An upper bound classifier

In order to analyze the effectiveness of the one-class classifiers for solving the fault detection problem, a binary classifier, trained with the normal examples (negative class) and the abnormal classes (positive class) is used as an upper bound. Since the binary classifier is trained with more training data, and deals with an easier problem than one-class approaches, it is expected that one-class approaches achieves an inferior performance. Therefore, the binary classifier can be seen as an upper limit. The classification model adopted for the upper bound classifier is Random forest [10], an ensemble tree-based method commonly used for classification problems. It has been successfully used for fault diagnosing of ESPs [2], being the best one among the tested supervised classifiers models.

The algorithm starts from the construction of a set of independent and identical initialized decision trees. These decision trees are made distinct from each other by randomly selecting subsets of features and instances for training each one. After the model is trained, each decision tree contributes with a single vote for a class. The most voted class gives the final decision.

It is important to emphasize that although the one-class classifier also deals with a binary problem, the problem setup is inherently different from the binary classifier used as upper bound (i.e., Random Forest). The upper bound requires samples from all classes for training, whereas the one-class uses only the normal class. Therefore, if samples of a new type of fault appear, the Random Forest would not be trained to recognize them, as the one-class classifiers would.

C. Model evaluation setup

In order to compare quantitatively all chosen classifiers models, the Macro F-measure is used as a performance metric since there is an evident imbalance in the dataset, making it an appropriate metric for this problem. This metric is defined as follows. Denote c as the number of classes in the classification problem. For each class j, there are individual true positives, false positives and false negatives, defined as tp_j, fp_j and fn_j, respectively. Macro-averaged precision and macro-averaged recall are defined as

$$\mathrm{Precision}_M = \frac{1}{c}\sum_{j=1}^{c}\frac{\mathrm{tp}_j}{\mathrm{tp}_j + \mathrm{fp}_j}$$

and

$$\mathrm{Recall}_M = \frac{1}{c}\sum_{j=1}^{c}\frac{\mathrm{tp}_j}{\mathrm{tp}_j + \mathrm{fn}_j}.$$

Macro-averaged F-measure is the harmonic mean of precision and recall:

$$F_M = \frac{2 \cdot \mathrm{Precision}_M \cdot \mathrm{Recall}_M}{\mathrm{Precision}_M + \mathrm{Recall}_M}. \quad (4)$$

A traditional 10-fold stratified cross validation is used to evaluate the upper bound Random Forest and the one-class solution. The entire dataset was divided in 10 folds of almost equal size and class distribution. For the upper bound Random Forest, the different faulty signals are converted to abnormal examples and the binary classier is evaluated. The same configuration is used for one-class classifiers, but faulty signal (abnormal examples) are removed at training time.

The Random Forest in this work uses 1000 decision tree with no pruning, where each one uses a maximum of 3 features. The Gini criterion was used to measure the quality of the division for each decision tree.

V. Experimental Results

The experiments were conducted in the python programming language using Scikit Learn [11] as a machine learning framework and testing platform. The experiments aim to accept or reject the hypothesis that one-class classifiers are an adequate tool for detecting novelties in vibration signals from electrical submersible pumps. The results are presented in Figures 4, 5 and Table II.

Figure 4 shows a boxplot with the classifiers performance on all folds, while Table II shows the average performance. As expected, the Random Forest achieved the highest performance. However, it can only be used as an upper limit given that it accesses the abnormal data in the training process. Both one-class classifiers have satisfactory and similar performances, achieving an average macro F-measure of approximately 0.86. Clearly, it is observed that there is no significant difference in the F-measure performance between One-Class SVM and Isolation Forest. These results suggest that proposed solution may be usefull in the oil and gas industry to detect abnormal

behaviour in Electrical Submersible Pumps. Indeed, a model that predicts a test example as an outlier with probability of 20% (the priori distribution of abnormal examples in the dataset) leads to a much lower F-measure of 0.5.

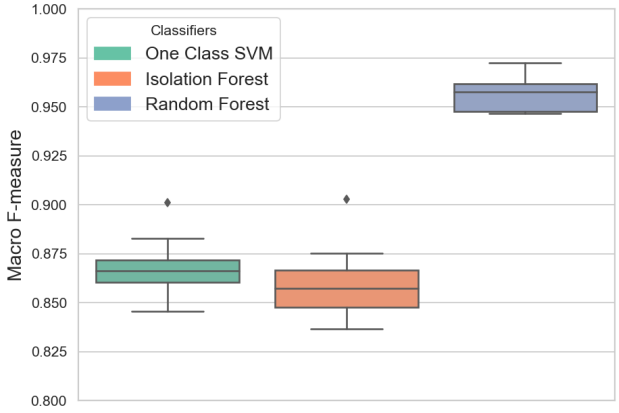


Fig. 4: Boxplot for Macro F-measure for the ten-fold cross-validation for one-class classifiers. The binary classifier (Random Forest) is used for comparison purposes (upper bound).

TABLE II: Macro F-measure average of the classifiers tested over the 10 folds.

Classifier	***Mean***	***Std deviation***
One-Class SVM	0.8685	0.0143
Isolation Forest	0.8602	0.0179
Random Forest (upper bound)	0.9564	0.0084

Figure 5 presents the confusion matrix of the one-class classifiers. One may observe that classifiers are not limited in correctly classifying only normal data. The Isolation Forest detected more anomalies than the One-class SVM but also classified more normal examples as faulty. The Random Forest achieved a slightly better performance than the Isolation Forest when detecting anomalies and also a slightly better performance than the One-class SVM when detecting normal samples. Even though both one-class classifiers have similar overall performances, the ability of the Isolation Forest to detect more anomalies than the One-class SVM makes the former method more appropriate for real application. Classifying a faulty ESP as normal may lead to significant financial losses due to the high maintenance costs, and particularly because of long-term production interruptions.

TABLE III: Recall of each class obtained by classifiers.

Class name	Classifier		
	One-Class SVM	Isolation Forest	Random Forest
Rubbing	39.92	75.46	79.12
Faulty sensor	97.28	98.98	98.64
Misalignment	98.08	84.62	92.31
Unbalance	71.77	93.04	99.64

As it can be seen in Table III, there is indeed a significant difference of performance at detecting rubbing. It seems that

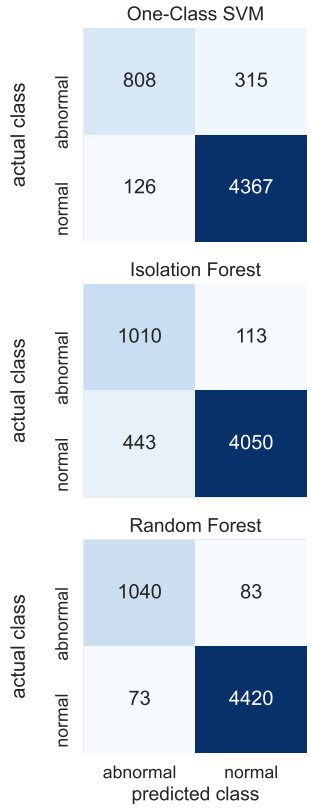


Fig. 5: Confusion Matrix resulting from cross-validation for each classifier. From left to right: One-Class SVM, Isolation Forest and Random Forest.

signals with a rubbing behaviour appear to be quite normal by all three classifiers, specially by one-class SVM. The Isolation Forest and the Random Forest are very consistent in the other 3 classes. It can also be noted that the One-Class SVM presents difficulties in dealing with signals of unbalance, although it is a very consistent classifier for dealing with signals of faulty sensor and misalignment.

Finally, the Table IV shows the average time of each classifier.

TABLE IV: Average time of each classifier for 10 runs on a computer with a AMD Ryzen 7 1800x processor and 16GB of RAM memory.

Classifier	**Average Time (seconds)**
One-Class SVM	0.1798
Isolation Forest	6.3124
Random Forest	11.9425

VI. Conclusions

In this paper, a proposal was presented to use one-class classifiers to detect novelties in electrical submersible pump fault detection problems. The results show a good performance in both evaluated one-class classifiers. The overall F-measure performances are much higher than 0.5 (performance of a random model) and came close to Random Forest, which

performs a binary classification task by using additional data not available to one-class classifiers. The results obtained show that both one-class classifiers behave similarly. They have a great power of generalization in this type of problem of finding novelties, differently from the classifiers studied in the previous works [2]–[5] that do not have this generalization capability.

The one-class classifiers have a similar average macro F-measure. However, their recall ratio (see Table III) are considerable different. In the real application where a low false negative rate is more desired than a low false positive rate, the Isolation Forest is preferable. In addition, the Isolation Forest is known to be less sensitive to changes in the parameters than the one-class SVM, which is very sensitive to the ν parameter.

Based on those results, a possible approach for future studies would be an ensemble of one-class classifiers, giving preference to those classifiers that present better performance in case of divergence of the result. Moreover, finding an efficient and effective deep neural network approach for one-class classification is one of the possible future work of this research. For instance, the deep one-class classification [12] may be investigated.

REFERENCES

[1] G. Takacs, *Electrical submersible pumps manual: design, operations, and maintenance*. Gulf professional publishing, 2017.

[2] T. Oliveira-Santos, T. W. Rauber, F. M. Varejão, L. Martinuzzo, W. Oliveira, M. P. Ribeiro, and A. Rodrigues, "Submersible motor pump fault diagnosis system: A comparative study of classification methods," in *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, Nov 2016, pp. 415–422.

[3] T. W. Rauber, T. Oliveira-Santos, F. de Assis Boldt, A. Rodrigues, F. M. Varejão, and M. P. Ribeiro, "Kernel and random extreme learning machine applied to submersible motor pump fault diagnosis," in *2017 International Joint Conference on Neural Networks, IJCNN 2017, Anchorage, AK, USA*, 2017, pp. 3347–3354.

[4] T. Oliveira-Santos, A. Rodrigues, V. Rocha, T. Rauber, F. Varejão, and M. Ribeiro, "Combining classifiers with decision templates for automatic fault diagnosis of electrical submersible pumps," *Integrated Computer-Aided Engineering*, vol. 25, pp. 1–16, 2018.

[5] L. H. S. Mello, M. P. Ribeiro, T. O. Santos, F. M. Varejão, and A. L. Rodrigues, "Metric learning for electrical submersible pump fault diagnosis," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[6] S. S. Khan and M. G. Madden, "One-class classification: taxonomy of study and review of techniques," *The Knowledge Engineering Review*, vol. 29, no. 3, pp. 345–374, 2014.

[7] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, J. C. Platt *et al.*, "Support vector method for novelty detection." in *NIPS*, vol. 12. Citeseer, 1999, pp. 582–588.

[8] F. T. Liu, K. Ting, and Z.-H. Zhou, "Isolation forest," 01 2009, pp. 413 – 422.

[9] D. Tax and R. Duin, "Support vector data description," *Machine Learning*, vol. 54, pp. 45–66, 01 2004.

[10] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python ," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[12] L. Ruff, R. Vandermeulen, N. Görnitz, L. Deecke, S. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," 07 2018.