A Machine Learning approach to fault detection in transformers by using vibration data

A. Tavakoli*, L. De Maria**,
B. Valecillos***, D. Bartalesi**, S. Garatti*, S. Bittanti*

* Dipartimento di Elettronica, Informazione e Bioingegneria - Politecnico di Milano, piazza L. da Vinci 32, 20133, Milan, Italy (amirhossein.tavakoli@mail.polimi.it, {simone.garatti,sergio.bittanti}@polimi.it)
**RSE S.p.A, via Rubattino, 20134, Milan, Italy (Letizia.demaria@rse-web.it, Daniele.bartalesi@rse-web.it)
***Trafoexpert GmbH, Switzerland (b.valecillos@trafoexperts.com)

Abstract: Transformer Vibration Technique is considered an effective method to monitor structural elements of transformers, in particular, to detect loose or deformed windings. As it is well known, vibrations vary with the sensor location on the transformer tank, which makes the number and the placement of sensors critical aspects for fault detection. In this paper, we investigate this issue by analyzing vibration spectra collected from various sensors installed on the tank of a typical oil filled power transformer operating under two limit cases, namely absence or presence of clamping looseness on windings. Support Vector Machines (SVM) are employed and an extensive analysis is performed to understand the informativeness of data corresponding to various sensors so as to figure out the appropriate number of sensors and their best location. This way fault detection is eventually achieved with a reduced and optimized number of sensors, resulting in a significant saving of time and costs.

Keywords: oil transformers, winding-looseness fault-detection, machine learning, data analysis.

1. INTRODUCTION

A proper clamping pressure in transformers' windings is essential to withstand high internal electromagnetic forces that arise during external short-circuit events. A transformer with loose or deformed windings is heavily exposed to permanent damage in presence of external short [1]. Recently, there has been a growing interest toward the application of monitoring techniques and devices for early warning on transformers' faults, in view of a more targeted and efficient asset management. Moreover, significant benefits on predictive maintenance of transformers have been obtained by the Industry 4.0 paradigm, where embedded sensors and devices play a key role as effective data sources for predictive analysis. In this framework, techniques based on the vibration of the transformer tank in steady-state have proved to be a promising tool to diagnose windings faults, [1,2]. Specifically, these approaches are based on the acquisitions of vibration fingerprints of the transformer tank before and during its service. These fingerprints are collected by means of sensors, like accelerometers, mounted in different positions of the tank. Various data analysis methodologies can be then exploited to perform fault diagnosis based on these collected vibration data. In [1] a model with electrical current, voltage, and temperature as inputs is reported to identify winding deformation. The malfunctioning is detected by comparing the measured 100Hz vibration signal with the vibration magnitude estimated by the model. To construct a model, the geometry of the transformer and its parameters are required though. When the parameters of the transformer are uncertain or unknown one must resort to different routes, akin to the black-box paradigm. For example, in [3], various indicators, like total harmonic distortions, the sum of harmonic amplitudes and ratio of main harmonics, are used to classify transformers into new, used, and anomalous.

The aim of this paper is to investigate the feasibility of a machine learning classification technique to predict the looseness of windings by vibration spectra. Surprisingly, the usage of machine learning for condition monitoring of power transformers seems to be a quite unexplored research direction. To the best of our knowledge, the only contribution in the same vein is [4], where artificial neural networks are considered.

In this paper, we investigate the usage of Support Vector Machines (SVM), [5,6,7], on vibration spectra experimentally recorded from the tank of a typical distribution transformer, subject to tightening or loosening clamping of its windings pack. As it is well known, vibrations vary with the sensor location on the transformer tank, which makes the number and the placement of sensors critical aspects for fault detection. In particular, the goal is to eventually obtain a classifier for the tight vs. loose condition that requires the final user to use the smallest possible number of sensors and that is resilient to misplacement of sensors as it often happens in real application. The main contribution of this paper is to propose an analysis of the informativeness of the data corresponding to the various sensors locations so as to figure out which is the minimal number of sensors and their best location for the training of the robust SVM classifier. This way fault detection is eventually achieved with a reduced and optimized number of sensors, resulting in a significant saving of time and costs. The effectiveness of the proposed approach is carried out by means of numerous tests. All the results reported in the present paper have been obtained by using a set of real data obtained from laboratory experiments.

The paper is organized as follows. In Section 2, after a preliminary description of the experimental layout to collect vibration data and to reproduce the fault on transformer's windings, the fault detection problem is introduced. The proposed SVM model and analysis is explained in Section 3,

while the experimental results obtained by applying SVM to vibration spectra are discussed in Section 4. Some conclusions are drawn in Section 5.

2. DATA ACQUISITION FRAMEWORK

2.1 Test transformer

A new oil filled three-phase transformer (42kV/580V, 750kVA, cooling with circulation of both oil and air natural (ONAN) cooling) was used as a test transformer (Figure 1) for collecting vibrations. It was tested under two tight (corresponding to no fault) and loose (fault) clamping. The last one has been simulated by loosening the winding clamping rods, which compress the windings at the top of the structure (see the inset of Figure 1). Therefore, the fault condition is represented by the transformer with the three phases simultaneously loosened at the same time.

Vibration measurements were carried out on the load transformer either under fault and no-fault conditions. Actually, unlike no load operation, where tank vibrations are driven by magnetostrictive forces in core sheets, during load operation, the vibrations of the transformer are mainly generated by the windings, due to the presence of high electro-dynamic forces [1]. Therefore, any looseness on windings will be more likely exalted during load tests with respect to no-load conditions.

Fig. 1. The 750kVA test transformer. Insets: a detail of the looseness of winding clamping structure (left) and of measurement points (right).

During load tests, the transformer was energized from the High Voltage (HV) side to nominal current (10A) with secondary shorted. A grid of 22 measurement points was defined on the Low Voltage (LV) sidewall of the tank (Figure 1). Tank's Vibrations were measured in each point of the mesh, at the top and at the bottom of the tank, either under tight and loose clamping.

2.2 Vibration Sensing System

The vibration sensing system used for the tight and loose tests is based on an optical accelerometer, the Electro Optical conversion Unit and a conditioning and recording Unit, as previously described in [8]. Due to the dielectric nature of its elements this optical accelerometer can be safely installed on serviced transformers. The accelerometer's sensitivity is 100mV/g and the frequency bandwidth (flat up to 1000Hz) has been chosen to detect main frequencies of mechanical oscillations of the transformer's windings and core. The Electro-optic Unit (EOU) output signals are recorded as a .wav file by means of a digital audio recorder (ZOOM Corp.H5-Handy Recorder) with a 24-bit resolution and a 96 kHz maximum sampling frequency.

A proprietary software developed under LabView (National Instrument) environment, was exploited for the off-line processing of the recorded time signals.

2.3 Vibrations data

During the experiment, 22 vibration time-series, one time-series for each position on the transformer tank, were recorded both for tight and loose windings.

Two minutes acquisition and 44 kHz frequency sampling were set as acquisition parameters. Each recorded time signals were subdivided in 100 time-subseries, each of which was averaged, filtered and Fast Fourier Transformed into the frequency domain (which seems to be better suited to reveal the information content of vibrations). Only magnitudes were considered. A data normalization was then applied by enforcing the value corresponding to each harmonic to be in the [0,1] range. Apart from fostering numerical stability, normalization has also motivations related to the structure of vibration in transformers. The vibration inside transformer can be transmitted through the liquid (insulating oil) inside the tank and also through some metal joints inside the transformer. Therefore, the possible similar patterns inside the tank can have different magnitudes in the outer body of the tank due to the different passed route. The aim of normalization is achieving meaningful information from patterns rather than relying on the magnitude of just one specific harmonic, [8]. Eventually, to better focus on the real information content of data, we eliminated frequencies which, provably from physical considerations, carry very little information (i.e. harmonics having very small magnitude or showing no significant differences for tight and loose conditions). Accordingly, our spectra were defined over the frequencies 50Hz, 100Hz, 150Hz, …, 500Hz, see also [8].

Thus, altogether, our dataset consists of approximately $22 \cdot 2 \cdot 100 = 4400$ pairs (x, y) where x is a normalized spectrum (which can be represented as a vector of real numbers as we consider magnitudes only) and y is a label which can be either "tight" or "loose". x is called the feature vector. The dataset is balanced since half of the pairs are labelled "tight" and half are "loose". Moreover, the dataset can be decomposed in 22 balanced sub-datasets, one for each sensor position in the considered grid.

Figure 2 shows two examples of averaged vibration spectra recorded in the same point of the tank. These two spectra (Fig. 2) were recorded at different times, after repositioning the sensor approximately in the same point of the tank, that is, with a tolerance less than 5cm. Though similar spectra should be expected, their comparison evidenced relevant harmonic differences (i.e. at 200Hz, 300Hz, and 400Hz).

Figure 3 shows load vibration spectra measured in the upper part of tank for different sensor positions: left (Fig. 3a) middle (Fig. 3b), right (Fig. 3c), respectively. Each figure shows the

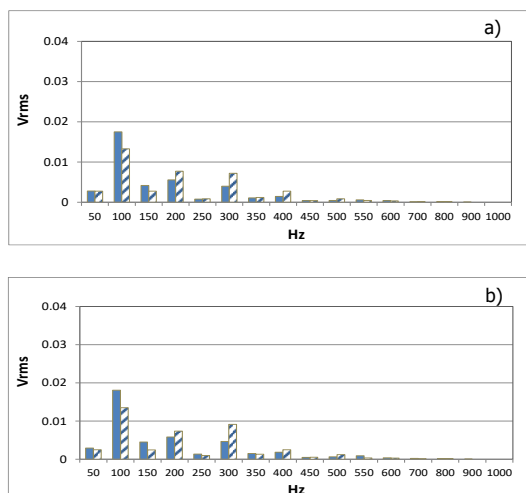vibration harmonics recorded under tight (solid) and loose (striped) windings, respectively.

Fig. 2. Magnitude of load vibration spectra measured in the same position at the top of the tank at two different times: (a) t_1 and b) t_1b, respectively. Tight windings (blue solid) vs. loose ones (blue striped).

A comparative analysis of tight and loose spectra among these three positions (Figure 3) evidenced a different harmonics contents with a different magnitude of the fundamental harmonic (100Hz) depending on the sensor position.
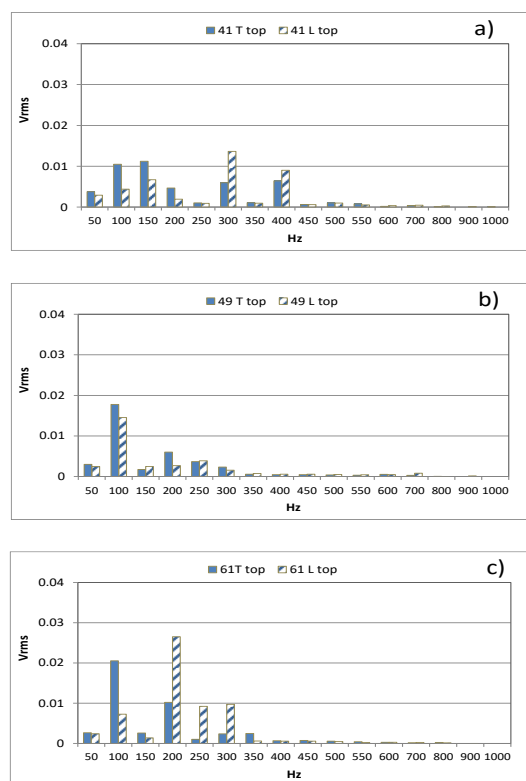

Fig. 3. Magnitude of load vibration spectra measured at the top of the tank: left side (a) middle (b) and right side (c), respectively. Tight windings (blue solid) vs. loose ones (blue striped).

Moreover, spectra corresponding to the top of the tank (Fig.3) and the bottom (Fig. 4) present an opposite mutual ratio

between the 100 Hz harmonic magnitude in tight and loose conditions.

This visual inspection reveals that detection of the tight vs. loose condition based on the assessment of 100 Hz harmonic magnitude as suggested by physical considerations is not possible without any further knowledge on the transformer. Moreover, due to the complexity of the vibration data obtained from all tank positions, extracting quantitative information on the fault of the transformer from simple classification rules based on vibration spectra seems challenging. This motivates the usage of more sophisticated black-box tools like the SVM discussed in the next section.
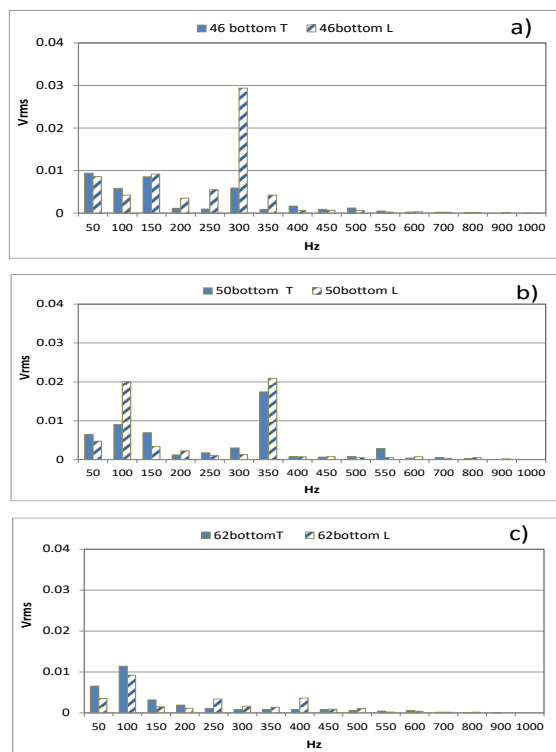

Fig. 4. Magnitude of load vibration spectra measured at the bottom of the tank: left side (a) middle (b) and right side (c), respectively. Tight windings (blue solid) vs. loose ones (blue striped).

3. FAULT DETECTION VIA SVM

The Support Vector Machines (SVM, [5,6,7]) is a supervised learning method whose goal is to learn a mapping (called classifier) from the feature vector x to the label y. The classifier thus is a natural tool to predict the value of the label given the feature vector for a new observation. The classifier is learnt based on a training dataset $\{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$, where N is the number of so-called training examples. SVM is based on separating hyperplanes (linear classifiers) that are selected so as to increase the margin between points corresponding to distinct labels as much as possible (possibly by tolerating that few data examples are misclassified as indicated by a user-chosen misclassification regret parameter ρ). This way, the classifier capability of separating (aka correctly classifying) other, unseen, observations is fostered, [9]. Before finding the separation hyperplane, however, SVM maps data to a higher dimensional feature space so as to augment the separation of the data

points belonging to distinct classes. This is done by specifying a kernel $k(x, x')$, which automatically defines a (possibly infinite-dimensional) transformation through the Mercer relation $k(x, x') = \langle \phi(x), \phi(x') \rangle$. Typically, Gaussian kernels are used, $k(x, x') = \exp(-\|x - x'\|^2/2\sigma^2)$, which enhance the separation of classes. See [9,10,11,12] for further details. The training of the classifier involves also the selection of the so called hyperparameters (ρ and the kernel $k(x, x')$) so as to improve the classifier as much as possible, [9,11]. To avoid overfitting, [13,14], k-fold Cross Validation is used, [15]. That is the training dataset is randomly split into k subsets and repeatedly k-1 subsets are used for training while the remaining dataset is used to validate a given choice of the hyperparameters (in our problem, k is selected equal to 10). Once the hyperparameters are chosen, the final classifier can be learnt over the whole training dataset by re-running SVM for the last time. The final classifier can be used to predict the labels from new measured features and hence to predict whether we are in a tight winding (no fault) or loose winding (fault) condition.

The goal of this paper is to check the potentiality of SVM to produce a tool for the diagnosis of transformers to be sold along with the transformer itself to the final user. In this respect, the following observations arise:

1. The producer cannot in general perform an extensive data acquisition campaign as the one that we performed for the analysis presented in this paper. The reason is that acquiring data from many different sensor positions is time-consuming and expensive. It is thus safe to assume that the classifier is trained based on measurements obtained from a limited number of sensor positions and the less the number of positions the better;

2. Also the final user, who aims at detecting whether the transformer is faulty or not, will use a limited number of sensor positions (for the same reasons as in point 1) and moreover it must be considered that he/she could place the sensors in positions other than those that are used when the classifier is trained. Indeed, keeping memory of the exact position seems unreasonable (loose winding detection becomes relevant when dealing with aged transformers that have operated for years) and errors may always happen.

In order to account for 1 and 2 above, the analysis cannot be limited to the training of an SVM classifier from the available dataset. We instead proceed as follows.

The dataset is partitioned in many different ways according to the sensor positions corresponding to the available data, and repeated experiments are performed where the classifier is trained based on data corresponding to certain sensor positions and then its fault prediction capabilities are tested against new data referring to other positions. In this way, provided that enough combinations are considered, it is possible then to obtain an indication of the safer positions for performing testing for various classifiers and, more ambitiously, to determine which is the least number of sensors and which are the best sensor positions to be used in the training phase so as to obtain a classifier whose prediction capabilities are robust enough with respect to the sensor positions used during testing. In other words, we obtain useful indication of how to use SVM so as to obtain a procedure that is simultaneously feasible and reliable in practice according to the points 1 and 2

above. The results of the analysis here proposed over the data described in Section 2 are presented in the next section.

4. EXPERIMENTAL RESULTS

The experimental result has been divided into 4 main sections. In Section 4.1, the reliability and limiting aspects of the trained classifier on each sensor position will be discussed in details and it will be clarified why this type of classifier is not practical. In Section 4.2, the capability of SVM for more comprehensive classifier will be assessed. In this step the classifier will be extended for all upper and bottom part of transformer. In Section 4.3, by using the classifier proposed in Section 4.2, the robustness of the model will be analysed for different positions. Finally, by exploiting information obtained in previous sections, the least number of sensors for fault detection will be discussed in section 4.4. It will be shown that this set of sensors is sufficient for diagnosis of transformer.

The reliability of the method in each step is expressed by the following parameters: accuracy, sensitivity and specificity, [16]. The accuracy is the ratio of truly detected data (that is true positive, i.e. loose, and true negative, i.e. tight, correctly identified data) to all data; sensitivity is the ratio of the truly positive detected to all positive data; specificity is the ratio of the correctly negative data identified to all negative data.

4.1. Training and testing classifier for each position

Initially, data referring to each position are split into training and test data (respectively, 75% and 25% of the data). The capability of the SVM classifier trained from data referring to one single position is first assessed via test data referring to the same sensor position. This procedure has been implemented for each sensor position and the accuracy of the SVM classifiers for corresponding test data is always 100%. Figure 5 shows an example of the separation of the two classes for sensor position 41.

However, the model built from one sensor position is not in general capable to predict data coming from different sensor positions. Consequently, by repeating this experiment, there is no guarantee that the proposed models are able to predict the fault of the system due to possible misplacements of the exact location of the sensor.

4.2. *Extension of classifier to all positions on the top and bottom*

To solve this problem, it is needed to have a more comprehensive model, which will be able to detect possible faults while positions of the sensors have not been restricted to very small specific areas. Therefore, the next step is to build the classifier by collecting training data coming from all sensor positions of the upper part of transformer and subsequently assessing its reliability for the remaining 25% of test data, again randomly chosen from all upper part sensor positions (Figure. 6).

The accuracy, specificity, and sensitivity of this classification for the model trained from all sensors in the upper part of the tank are all 100%. The same approach has been followed for data referring to bottom part taken from 11 sensors (Fig. 7). The accuracy, sensitivity and specificity of classification for

test data referring to sensors at the bottom part are respectively: 99.8%, 99.6% and 100%.

The accuracy achieved by the obtained classifiers indicate that the selection of the 10-dimensional input feature space and the usage of SVM with Gaussian kernels was appropriate. Moreover, since the two classifiers obtained in this Section 4.2 exploit data coming from sensors in different positions, it is expected that they show a better robustness with respect to a misplacement of sensors in the test phase than classifiers trained over data collected from a single sensor (Section 4.1). The next section aims at assessing this robustness property.
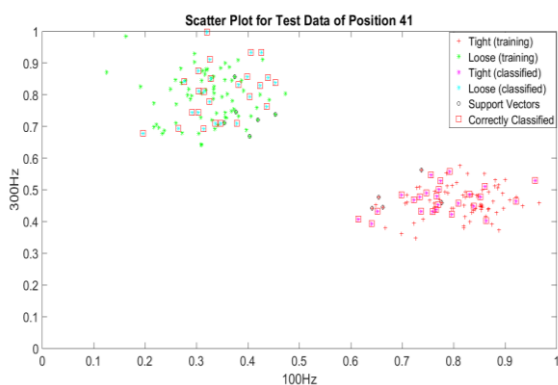
Fig. 5. Scatter Plot projected on two features (100Hz and 300Hz) derived from position 41, test and training data referring to same position, 41.

4.3. Testing classifier against new data from excluded positions

At each step all data referring to one specific sensor will be kept as test data. Then, remaining data referring to other sensors referring to upper part or bottom part positions will be considered as training ones. For instance, all data sets referring to position 41 will be kept as test data while the remaining data related to other sensor positions located in the upper part of transformer (i.e. data referring to the other 10 sensors) are used to train the SVM classifier. In this situation, position 41 can be interpreted as totally unobserved vibrations for the classifier and the reliability of the trained classifier indicates its robustness against this unobserved position (Fig. 8). This procedure has been implemented for every position and Table 1 illustrates the validity of the model for totally unseen data coming from the excluded sensor in the upper part of transformer.
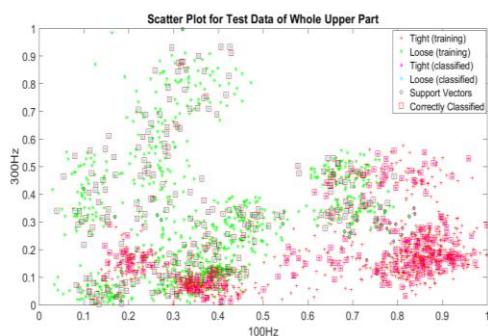
Fig. 6. Scatter Plot projected on two features (100Hz and 300Hz) derived from all upper part positions
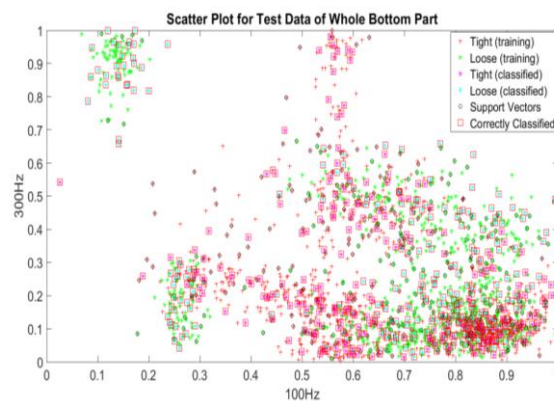
Fig. 7. Scatter Plot projected on two features (100Hz and 300Hz) derived from all bottom part positions
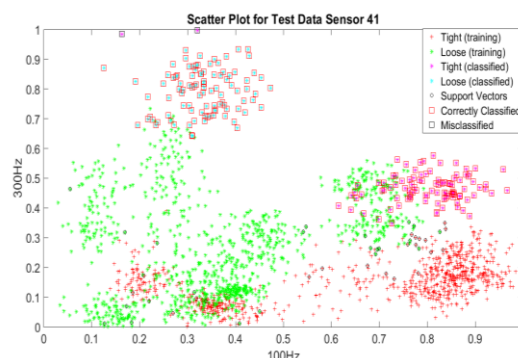
Fig. 8. Scatter Plot projected on two features (100Hz and 300Hz) for classification of data referring to excluded sensor position 41

From Table 1, indications about the safer sensor positions (i.e. enhancing more robustness) for the training of the classifier emerge. While the reliability of classifiers for unseen sensor positions is high for most positions on the upper part of transformer, there are few positions (e.g. 55 and 63) which are not predictable by the classifiers obtained from data collected from the other positions. This indicates that sensor positions 55 and 63 must be used to enforce robustness with respect to the sensor position used during the test.

Table 1. Reliability of classifiers on test data referring to excluded positions in upper part of transformer, expressed via Accuracy, Specificity and Sensitivity

Position	Accuracy (%)	Specificity (%)	Sensitivity (%)
41	99	100	98
43	95.5	91	100
45	89.9	100	85
49	100	100	100
51	100	100	100
53	100	100	100
55	68.5	37	100
57	96.5	100	93
59	99.5	99	100
61	88.9	100	77.8
63	50	0	100

A similar experiment with the dataset referring to the bottom part of the transformer reveals that in the bottom only two positions (60 and 64) can be correctly predicted by the classifiers trained based on data referring to sensors in the other positions in the bottom part. This fact indicates that

models obtained from data collected from the bottom of the transformer are more sensitive to the sensor position used in the test phase. In other words, the SVM classifier trained at the upper part of transformer is more robust with respect to sensor misplacements rather than models trained at the bottom part. Hence, data referring to upper part of transformer is respectively safer to be used. It is perhaps worth mentioning that another experiment reveals also that merging training data from upper and bottom parts do not increase the accuracy of the prediction for neither bottom nor upper part.

4.4. Least number of required sensors

The experimental results in Table 1 on the robustness of the trained SVM by vibration data belonging to upper part can be used to reduce the number of sensors used for training. Specifically, one can progressively add sensor positions, starting from those leading to highest decreases of accuracy when removed from the data set. The procedure is halted when the classifier trained from data coming from the selected set of sensors achieves a high level of reliability also for data referring to the other sensor positions on the upper part of transformer not used for training.

Table 2. Reliability of classifiers trained by least number of sensors on test data referring to excluded positions in upper part of transformer, expressed via Accuracy, Specificity and Sensitivity.

Position	Accuracy (%)	Specificity (%)	Sensitivity (%)
41	100	100	100
45	90.6	86	100
51	100	100	100
53	100	100	100
57	90.5	100	81
59	100	100	100
61	87.9	100	75.8

In our experiment, the least number of sensors which has this capability of classifying with high reliability unseen data from any position in the upper part of transformer is 4 and consists of positions 43, 49, 55, and 63. The reliability of the classifier trained from data referring to these positions against test data referring to all the other positions are reported in Table 2. Based on Table 2, it is possible to claim that the selected sensor positions are enough to obtain a classifier that is robust with respect to any possible misplacement in the positioning of the sensor at the upper part of transformer during the test phase.

5. CONCLUSIONS

In this paper, we have considered the problem of detecting malfunctioning in electrical transformers. To this purpose, we have resorted to the Support Vector Machine classification technique by using vibration data measured with sensors located in various positions on the transformer tank. The obtained result shows the high reliability of SVM when applied to consistent dataset (i.e. taken always from the same position on the tank) and moreover we also devised robust classifiers that allows one to detect the fault even though the test data are collected from a position on the tank different from that corresponding to the data used to train the classifier. This robustness property is significant in two respects. First, it shows that misplacements of sensors location, which is quite likely, have a limited impact on the detection of fault. Second, it reveals that fault detection can be achieved with a reduced number of sensors, resulting in a significant saving of time and costs.

ACKNOWLEDGMENTS

REFERENCES

[1] García, B., Burgos, J. C., & Alonso, Á. M. (2005). Transformer tank vibration modeling as a method of detecting winding deformations-part II: experimental verification. *IEEE Transactions on Power Delivery*, *21*(1): 164-169.

[2] Hu, Y., Zheng, J., & Huang, H. (2019). Experimental Research on Power Transformer Vibration Distribution under Different Winding Defect Conditions. *Electronics*, *8*(842): 1-19.

[3] Bartoletti, C., Desiderio, M., Di Carlo, D., Fazio, G., Muzi, F., Sacerdoti, G., & Salvatori, F. (2004). Vibro-acoustic techniques to diagnose power transformers. *IEEE Transactions on Power Delivery*, *19*(1): 221-229.

[4] Booth, C., & McDonald, J. R. (1998). The use of artificial neural networks for condition monitoring of electrical power transformers. *Neurocomputing*, *23*(1-3): 97-109.

[5] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory (pp. 144-152). ACM.

[6] Müller, K. R., Smola, A. J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., & Vapnik, V. (1997). Predicting time series with support vector machines. In: *International Conference on Artificial Neural Networks* (pp. 999-1004). Springer, Berlin, Heidelberg.

[7] Cristianini, N., & Scholkopf, B. (2002). Support vector machines and kernel methods: the new generation of learning machines. Ai Magazine, 23(3), 31-31.

[8] Tavakoli, A., De Maria, L., Bartalesi, D., Garatti, S., Bittanti, S., Valecillos, B., & Piovan, U. (2019). Diagnosis of transformers based on vibration data. In: *2019 IEEE 20th International Conference on Dielectric Liquids (ICDL)* (pp. 1-4).

[9] Vapnik, V., Golowich, S. E., & Smola, A. J. (1997). Support vector method for function approximation, regression estimation and signal processing. In: *Advances in neural information processing systems* (pp. 281-287).

[10] Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

[11] Smola, A. J., Schölkopf, B., & Müller, K. R. (1998). The connection between regularization operators and support vector kernels. *Neural networks*, 11(4): 637-649.

[12] Scholkopf, B., Sung, K. K., Burges, C. J., Girosi, F., Niyogi, P., Poggio, T., & Vapnik, V. (1997). Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE transactions on Signal Processing*, *45*(11), 2758-2765.

[13] Bittanti, S. (2019). *Model Identification and Data Analysis*. Wiley.

[14] Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2): 83-85.

[15] Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, 4: 40-79.

[16] Sammut, C., & Webb, G. I. (Eds.). (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.