Transformer winding type recognition based on FRA data and a support vector machine model

Xiaozhou Mao¹, Zhongdong Wang¹, Peter Crossley¹, Paul Jarman¹, Andrew Fieldsend-Roxborough², Gordon Wilson²

¹The University of Manchester, Manchester M13 9PL, UK
²National Grid, Warwick CV34 6DA, UK
E-mail: zhongdong.wang@manchester.ac.uk

Abstract: Frequency response analysis (FRA) is regarded as the most effective technique to detect mechanical faults of transformers. Over the years, FRA measurement data have been collected by utilities into transformer asset databases. The characteristic of FRA data is fundamentally determined by the transformer’s equivalent electrical circuit, which consists of inductance and capacitance parameters that are windings’ design and structure dependent. Different winding types tend to have different FRA characteristics, and a transformer’s design information such as winding type, dimension etc. is often not known to the utility but critically important for asset management. This study reviews the state-of-the-art transformer FRA databases and application of machine learning techniques in this field, and proposes to apply a support vector machine (SVM) model onto the FRA data to identify the winding type. The SVM model is first trained by FRA traces of transformers with known winding types, and after testing, the SVM model is then applied to FRA traces with unknown winding information. A set of data from the UK’s National Grid FRA database, was used to demonstrate and verify the SVM model. All transformers used in this study are 400/275/13 kV transmission transformers, which were designed using four different winding types, namely multiple layer, plain disc, interleaved disc and single helical windings. The proposed method can successfully identify the correct winding type.

1 Introduction

Windings are important electrical components of a transformer, and the choice of winding construction type is greatly influenced by the transformer manufacturer's historic experience, the transformer's voltage and power rating. In general, for the same power rating, a higher voltage winding prefers to use the winding construction type which gives a larger winding series capacitance due to the requirement to withstand a stringent BIL level [1]. It is well known that different winding construction types are susceptible to different modes of mechanical deformations, hence knowing the winding design information is helpful for transformer fault diagnosis. However in practice, transformer asset managers in utilities know it too well that a large number of transformers operating in their networks are lack of design information such as winding type, structure and dimension, as this information is the manufacturers’ safe guarded know-how. Effective asset management, especially for those transformers without any technical support from the original equipment manufacturer, calls for the development of non-intrusive winding type recognition techniques.

Frequency response analysis (FRA) has been developed as an effective and sensitive technique to identify winding mechanical movement such as displacement or deformation [2, 3]. An FRA measurement is normally carried out by the transformer manufacturer in factory, as the ‘reference’ for the future diagnostic FRA measurements which are either conducted after the transformer's transportation and installation or after a system short circuit fault current passing through the transformer. The short circuit current could induce an electromagnetic force, several hundred time larger, and lead to the mechanical deformation/displacement/damage of a winding [4]. When measuring the frequency response of a winding, a sinusoidal low-voltage signal is injected at one end and the response signal is received at the other end of the winding. The injected signal covers a large frequency range from a few Hertz to several mega-Hertz. The FRA data consists of the magnitude ratio and phase shift between the injected and received signals against the frequency. More attention is paid to the magnitude frequency spectrum, and when compared with the ‘reference’, any alteration, especially on the resonant frequencies, may indicate a mechanical fault hence needs to be investigated [5]. To interpret the FRA data, the numerical indices, including the statistic indicators [6] and the transfer function expression [7–9], are proposed to objectively describe the differences between the measured and the ‘reference’ FRA data. Another interpretation method is based on the development of an equivalent electrical circuit network model of the transformer. The electrical circuit network models developed can be categorised into three types: the white box model which is constructed using the geometric parameters of the transformer [10–12]; the grey box model which usually shares the identical or similar circuit topology as the white box model, whilst its electrical circuit parameters are estimated from the transformer's terminal measurements [13]; and the black box model which can take any mathematic format to accurately reproduce the measured FRA data and with no physical resemblance to the winding [7–9].

The FRA characteristics at different frequency regions are determined by different parts of a transformer. For the 400/275/13 kV transformers to be investigated, the frequency range can be split into three regions 0.005–2, 2–20 and 20–1000 kHz, which are dominated by the core, inter-winding interaction and the structure of the winding-under-test, respectively. The frequency region >1 MHz is influenced by the FRA measurement setup such as the bushing and the earthing lead. For the frequency region of 20–1000 kHz, the ratio between winding series capacitance and shunt capacitance is the main shaping factor [10]. Different winding types have different typical values of winding series capacitance, and thus different FRA characteristics in this frequency region. This lends itself to the winding type recognition.

Different artificial intelligence methods, especially pattern recognition techniques have been applied for the transformer fault diagnosis [14–25]. Those methods have also been applied to FRA for various purposes, such as estimation of parameters of transformer models [19], identification of winding faults [14, 20–24] and classification of winding types [25]. Machine learning is a...
subset of the artificial intelligence, and it includes the supervised methods and the unsupervised methods. The supervised machine learning methods require observations with known labels as input to build an identification model, and this model could be used to estimate the labels of new observations [14, 20–22]. The unsupervised machine learning methods classify the input observations according to the similarity or dissimilarity among them, without any guidance, i.e. the labels of observations [24, 25].

Artificial neural networks (ANNs) are computation models which can be either supervised or unsupervised, and they are among the most popular artificial intelligence methods employed in the FRA field. The ANN models simulate the way that biological nervous systems analyse and process information. ANN has been successfully applied to identify the type, location and severity of transformer winding faults [20, 21]. However, the learning process of ANN cannot be observed, and its output results are difficult to understand. Hierarchical clustering method is an unsupervised machine learning method which groups similar observations together according to distances between each other [24]. Initially, each observation is viewed as a single cluster. The algorithm repeatedly merges the closest two clusters together until the number of clusters reduces to the desired level. The hierarchical clustering method has been successfully applied to identify the winding fault types [24] and winding structure types [25]. In [25], the hierarchical clustering method can only distinguish between windings with high or low series capacitance. Support vector machine (SVM) method is another supervised machine learning method for classification problem. It finds an optimal hyperplane which separates two types of data in binary problems. Multiple binary SVM classifiers are used when the number of data types to be identified is larger than two. The SVM method has been employed successfully for the recognition of transformer winding fault types and series [14, 22]. The advantage of SVM is that it can deal with small size of samples, non-linearity and high-dimensional issues. In addition, it can overcome the problem of local minimums in neural networks.

Pattern recognition techniques developed so far assumed that the winding type information is known beforehand and the solutions derived focused solely on that type. However, for different winding types, the characteristics of their FRA are different, thus the fault pattern may differ significantly. The identification of winding type is the cornerstone for further study of winding mechanical faults in order to establish a generic interpretation guide. On the other hand, as mentioned, different winding types are susceptible to different types of mechanical faults, thus knowing the design information such as the winding type can be helpful for transformer asset management. Therefore, there is an urgent need to identify winding types through FRA measurement data.

In this study, a novel SVM-based winding type recognition method is proposed. The SVM is trained and tested by FRA traces from the UK’s National Grid FRA database of a group of three winding three phase 400/275/13 kV auto transformers. Four winding types, including multiple layer, plain disc, interleaved disc and single helical are used in these transformers. Then, the SVM model built is applied to FRA traces without winding type information to test its performance in winding type recognition. The test results and sensitivity studies confirm that the proposed method is comprehensive and can be used along with expert experience and forensic information to aid transformer winding fault interpretation and transformer asset management.

2 FRA traces of different winding types

White box model has been developed to investigate the correlation between the equivalent electrical parameters and the characteristics of FRA traces [10–12]. The white box model constructs R–L–C electrical circuit using known geometric structure information of transformers. It is found through simulation studies that for a single winding, having a winding series capacitor and an inductor in parallel causes a trough on the FRA magnitude response, while having a shunt capacitor and an inductor in series results in a peak. The latter refers to the structure that the shunt capacitor is connected to the middle of two halves of inductor, in ‘T’ shape. The resistor only controls the sharpness of the peaks and troughs. For 400/275/13 kV auto transformers, the frequency region of 20 kHz to 1 MHz is mainly controlled by the properties of winding-under-test. In this frequency region, the ratio between the winding series capacitance and shunt capacitance is the most influential factor on the FRA shape. Among the four winding types investigated in this study, the multiple layer winding and interleaved disc winding [26] have a relatively high series capacitance. The plain disc winding has a relatively lower series capacitance and the single helical winding has the smallest series capacitance.

In the UK’s National Grid FRA database, both the magnitude and phase are recorded, and only the magnitude frequency spectrum is used, which proves to be sufficient for winding type recognition. Due to the particular FRA measurement device used, four measurements were taken for 5 Hz to 2 kHz, 50 Hz to 20 kHz, 500 Hz to 200 kHz and 5 kHz to 2 MHz, respectively, and 400 evenly distributed frequency points are recorded for each region. To eliminate these duplicate measurement points at lower frequencies, 40 redundancy data were eliminated for the second, third and fourth measurements, resulting in 1280 data points per FRA trace from 5 Hz to 1 MHz. The ‘typical’ FRA traces of aforementioned four winding types are plotted in Fig. 1. These are the measured FRA data on a winding with its winding type known, selected based on empirical experiences after observing 33-phase 400/275/13 kV transmission transformers with known winding types, including series, common and tertiary windings. In the database, the same windings may have several measurements which were taken at different time, or with different connections of windings.

The single helical winding, used for the tertiary windings, has the highest magnitude roughly from 5 Hz to 100 kHz. This is because the tertiary winding has the lowest voltage and power rating (13 kV, 60 MVA). The plain disc winding has a lower winding series capacitance, and its FRA trace has the oscillatory camel humps characteristics which can be observed between about 20 and 500 kHz. Either an overall rising or flat trend may appear on this frequency region for a plain disc winding. Multiple layer and interleaved disc windings have high winding series capacitance, and their FRA traces have a rising trend from 20 kHz to 1 MHz. The FRA trace of interleaved disc winding has a smoother rising trend, while that of multiple layer winding has some obvious fluctuations. The reason behind this observation could be that the interleaved disc winding has a higher winding series capacitance in proportion to the winding shunt capacitance, as compared to the multiple layer winding, and this higher ratio eliminates the appearance of resonant frequencies in this region [10].
3 Support vector machine

A hard-margin classification model requires all features to be accurately classified, while a soft-margin classification model allows a certain amount of features to be misclassified. Normally for soft-margin model, a regularisation parameter can be used to avoid overfitting, by restricting the norm of the parameters.

SVM is a supervised learning model which was initially proposed for the two-type classification problem [27]. Multiclass classification problems can also be solved by SVMs if using multiple binary classifiers. An SVM is a generalised linear classifier, which can be also applied to the non-linear classification problem in combination with kernel method. In this study, only the linear classifier is discussed. A hard-margin SVM is used in the paper.

The SVM algorithm finds a separation hyperplane from two groups of observations with known types and utilises it to categorise new examples. The hyperplane is also referred to as the decision boundary. The process of devising an SVM can be divided into two stages: training where the optimal hyperplane(s) is obtained from observing training data; and testing where the SVM is validated with a group of observations with known types. Upon completion, the SVM can be used to classify new observations.

3.1 Binary SVM classifier

A binary SVM classifier finds an optimal hyperplane which isolates two groups of observations with known types. The distances from the hyperplane to the nearest observation from the two groups should be equal. For example, in the two-dimensional space of Fig. 2, an optimal hyperplane, represented by the bold line, leans neither to the nearest observation from type C1 (triangles) nor the nearest observation from type C2 (circles). Each observation, either a triangle or circle, is called a feature in SVM algorithm. The boundary features, which are circled in dashed line in Fig. 2, determine the hyperplane of binary classifiers, and they are the nearest observations to the classification hyperplane from each feature type. Such boundary features are called support vectors in SVM. There may be more than one support vector from each type of features.

In the two-dimensional space in Fig. 2, the classification hyperplane \( g(x) = 0 \) is a line and it is defined as

\[
g(x) = w^T \cdot x + b = (w_1 \quad w_2)(x_1 \quad x_2) + b = 0
\]

where \( x \) is a point, a \( 2 \times 1 \) vector, located on the classification line in the two-dimensional space, \( w \) is the vector of fitted linear coefficients and \( b \) is bias. The vector \( w \) and the bias \( b \) determine the slope and the vertical intercept of the classification hyperplane in the two-dimensional space, respectively. The signs of two types of features on two sides of the classification hyperplane are different. In Fig. 2, the sign of \( g(x) = wx + b \) is positive for features from type C1 and negative for features from type C2. Define \( y \) as the classification label

\[
y = \begin{cases} 1, & \text{for type C1} \\ -1, & \text{for type C2} \end{cases}
\]

For the convenience of computation, the whole plane can be scaled such that for the support vector \( x_i \), \( g(x_i) = [w^T \cdot x_i + b] \) equals to 1. To achieve an optimal classifier, the geometrical margin \( 1/\|w\| \) between the support vector and the classification hyperplane, is to be maximised which is equal to minimising \((1/2)\|w\|^2\). Thus, the problem to find the optimal classification hyperplane in the two-dimensional space can be described mathematically as:

\[
\begin{align*}
\min & \frac{1}{2} \|w\|^2 & & \text{(objective function)} \\
\text{s.t.} & y_i(w^T \cdot x_i + b) - 1 \geq 0 & & \text{(constrain function)}
\end{align*}
\]

This is a convex quadratic programming problem, which also applies to finding the optimal classification hyperplane in higher dimensional spaces.

To solve (3), a Lagrange equation can be defined, by combining the constrain function with the objective function with a non-negative Lagrange multiplier \( \alpha \)

\[
L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i(w^T \cdot x_i + b) - 1)
\]

Set \( \theta(w) \) as

\[
\theta(w) = \max \{ L(w, b, \alpha) | \alpha_i \geq 0 \}
\]

Since \( \alpha_i \geq 0 \), when the constrain function in (3) is satisfied, there exists \( \theta(w) = (1/2)\|w\|^2 \). Thus the objective function in (3) can be expressed as

\[
\min_{w, b} \theta(w) = \min_{w, b} L(w, b, \alpha)
\]

Under Karush-Kuhn-Tucker condition (constrain function is satisfied and \( L \) is differentiable regarding \( w \) and \( b \)), the problem in (6) can be converted into its dual problem according to Lagrange duality [28]

\[
\max_{\alpha_i \geq 0} \min_{w, b} L(w, b, \alpha)
\]

To solve (7), \( L \) should be minimised regarding \( w \) and \( b \), by setting their partial derivatives as 0, and hence

\[
w = \sum_{i=1}^n \alpha_i y_i x_i
\]

\[
\sum_{i=1}^n \alpha_i = 0
\]

Now the problem can be expressed by only the Lagrange multiplier \( \alpha \) as

\[
\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\
\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \ldots, n \\
\sum_{i=1}^n \alpha_i y_i = 0
\]

The Lagrange multiplier \( \alpha \) can be solved using the sequential minimal optimisation algorithm [29]. After \( \alpha \) is computed, \( w \) and \( b \) are derived as

\[
w = \sum_{i=1}^n \alpha_i y_i x_i
\]
The purpose of this study is to use SVM algorithm to classify the multiclass classification problem. ‘One-versus-one’, ‘one-versus-all’ and ‘binary tree’ are three commonly adopted multiclass classification strategy [30]. In this study, ‘one-versus-one’ method is used. This means that between every two classes, a decision is made on which class the new feature is more similar to. Finally, the new feature is identified as the class which wins the most votes.

4 Winding type classification

The purpose of this study is to use SVM algorithm to classify the winding type of transformers. To obtain a well-functioned SVM model, the input features should be carefully selected. The statistic indices are not suitable as they are normally used for the comparison between two FRA traces, which means they indicate the similarity or dissimilarity between two sets of data, but not the characteristics of a single set of data. Although the transfer function can accurately describe the FRA traces in a mathematical way, one problem is that the parameters of a transfer function do not have a fixed quantity for both complex poles and zeros. This is because the number of peaks and troughs on different FRA traces from different winding types may vary. Besides, the quantity of real zeros and poles is also hard to be unified. Therefore, it is difficult to find appropriate input features for the SVM model using transfer function expression. As a result, in the study, the measured magnitude responses of different FRA traces, with unified format, are selected as the input features. FRA analysers on the market use different frequency and amplitude resolutions, which lead to the fact that the FRA trace may have different number of frequency points. In case such a scenario arises, pre-processing of data is needed and by applying transfer function estimation method, the measurement data in different frequency resolution can be expressed into a mathematical equation, which can reproduce FRA traces in the same desired format.

The FRA traces used for training and testing the SVM model are obtained from National Grid’s FRA database with the same format, and all together 108 FRA traces are used in this study, including 30 multiple layer winding FRA traces, 36 plain disc winding FRA traces, 27 interleaved disc winding FRA traces and 15 single helical winding FRA traces.

Before training the SVM model, standardisation should be applied to all 108 FRA traces as follows:

\[ x' = \frac{x - x_m}{\sigma} \quad (13) \]

where \( x \) is the original vector consisting of the 1280 FRA amplitudes in dB for the frequency range from 5 Hz to 1000 kHz, \( x_m \) is the mean value of the original vector, \( \sigma \) is the standard deviation of original vector, and \( x' \) is the standardised dimensionless vector. The range of \( x' \) is roughly from −4 to 3. Although the classification difficulty will be increased after standardisation, the focus can be put more on the traces' characteristics, rather than relying on the difference in magnitude.

There are altogether four winding types to be identified. Since the ‘one-versus-one’ strategy is adopted, a binary classifier is needed for every two winding types. Therefore, a total number of \((4 \times 3/2) = 6\) binary classifiers should be built. The input features are the standardised FRA magnitude response \( x' \). If there are altogether \( n \) FRA traces used as training feature, and there are 1280 points in each FRA magnitude response, the input of the SVM model to be built is an \( n \times 1280 \) matrix.

Using the methodology introduced in Section 3, the weight vector \( w \) and bias \( b \) for each binary classifier can be computed with a given set of training data. By doing so, the multiclass classification SVM model can be built. Four labels are used for the four investigated winding types. The output of the SVM model is the label of the identified winding type, and the meaning of each label is as follows:

Label 1: multiple layer winding.
Label 2: plain disc winding.
Label 3: interleaved disc winding.
Label 4: single helical winding.

4.1 Cross validation process

Cross validation and bootstrap are two resampling methods, which can be used to evaluate the effectiveness of classification models when the quantity of available features is limited. Bootstrap resamples randomly with replacement, which means duplicate features may be sampled. In a soft-margin SVM model, the weight of duplicate features changes, thus the parameters of SVM model can be influenced. A hard-margin SVM classification model does not notice the duplicate features.

The cross validation method without replacement is preferred. \( K \)-fold cross validation is therefore used to verify the applicability of SVM model under the small sample size setting. It divides the original data into \( K \) roughly equal-sized folds. One fold is used as the testing data, and the rest \( K - 1 \) folds are used as the training data. Each of the \( K \) folds should be used as the testing data once, which means \( K \) classification models are built. This makes the most of the available information. When the value of \( K \) is selected, the average accuracy can be used to evaluate the performance of the SVM model. With the change of \( K \), the performance of the model may change. The model with the highest accuracy can be chosen.

The FRA traces from the A, B, and C phase of the same transformers should be all grouped into either the training or testing data group due to their similarity, in order to guarantee the credibility of the classification model. Considering there are 15 single helical FRA traces from five different transformers, the maximum amount of folds is five. The 108 FRA traces investigated in this study are divided into 2, 3, 4 and 5 roughly equal-sized folds, respectively. For example, for the two folds cross validation, there are 12 frequency responses from multiple layer winding, 18 frequency responses from plain disc winding, 18 frequency responses from interleaved disc winding and six frequency responses from single helical winding are used as training data, and the rest of the frequency responses are used as testing data. The number of training and testing traces when \( K \) changes are listed in Table 1. For each \( K \)-fold cross validation, different combinations of training traces are selected randomly without repetition for 20,000 times. To guarantee that the sampling size 20,000 is large enough, the cross validation process is conducted twice. The corresponding

| Table 1 | Cross validation accuracy |
|---------|--------------------------|
| Winding type | Number of training traces | Number of testing traces | Accuracy 1, % | Accuracy 2, % |
| K=2      | 12 18 18 21 24 30 12 21 6 18 9 18 6 | 18 18 9 9 18 9 9 12 6 9 9 9 6 | 97.70 | 97.75 |
| K=3      | 18 24 18 21 24 30 12 18 9 18 9 18 | 12 12 9 9 6 6 9 9 9 6 9 9 | 99.54 | 99.56 |
| K=4      | 21 27 21 21 24 30 12 21 9 21 9 9 | 9 9 6 6 9 9 9 9 6 9 9 9 | 99.97 | 99.97 |
| K=5      | 24 30 21 21 24 30 12 21 6 21 6 6 | 6 6 6 6 3 3 6 6 3 6 6 3 | 100.00 | 100.00 |

Once \( w \) and \( b \) are obtained, the classification hyperplane is found. Therefore, the type of new features can be identified according to its sign.

\[ b = -\frac{\text{max}(w^T \cdot x_i | y_i = -1) + \text{min}(w^T \cdot x_j | y_j = 1)}{2} \quad (12) \]
average accuracies are listed in Table 1. It can be seen that the overall accuracy is satisfactory, which proves the feasibility of winding type classification through SVM method. Among all the SVM models built, the lowest accuracy is 38.89% when $K = 2$. Though for this particular model, the training features used may not be appropriate, however the accuracy is still $> 25\%$ which can be resulted from random guess when distinguishing the four winding types. In this cross validation process, with the increase of number of training data, the accuracy of the SVM model increases.

Since the models of two-fold cross validation have the lowest accuracy, one of the models built when $K = 2$ with an accuracy of 100% is of more interest for further investigation in order to increase our understanding on when and why the SVM method can work well.

### 4.2 FRA traces and training process

For the SVM model with an accuracy of 100%, which is built when $K = 2$, altogether 54 FRA traces are used to train the SVM model, and another 54 FRA traces are used to test the model.

As mentioned before, the boundary frequencies of 2 and 20 kHz are empirical; hence the whole range from 5 Hz to 1 MHz should be used in the following study of the SVM model. There exists a great similarity for FRA traces in the frequencies lower than 20 kHz for all the multiple layer, plain disc and interleaved disc windings, and this similarity is reasonable as the frequency region is dominated by the core and the interaction between windings. Inclusion of the low frequency regions would increase the complexity for classification but and enhance the confidence level of the SVM model developed.

For the multiple layer winding type, the training features are 12 FRA traces from two 500 MVA transformers, including the common and series windings from A, B, and C phases, as plotted in Fig. 3a. The testing features are 18 FRA traces from three 750 MVA transformers, including the common and series windings from A, B and C phases, as plotted in Fig. 3b. It can be seen that for the multiple layer windings, training features have more obvious oscillations than the testing features in the frequency region from 20 to 1000 kHz, which means that the training features are more typical.

For the plain disc winding type, the training features are 18 FRA traces from two 750 and one 1000 MVA transformers, including common and series windings from A, B and C phases. The testing features are 18 traces from two 750 MVA and one 1000 MVA transformers, including common and series windings from A, B and C phases. As mentioned earlier in Section 2, plain disc windings’ FRA traces can either have a rising or flat trend in the frequency region from 20–1000 kHz. All the training and testing features have rising trend in Fig. 4, except that three testing FRA traces from the A, B and C phases of one 750 MVA transformer’s series windings have flat trend, as shown in Fig. 4b.

The choice of FRA traces for the interleaved disc winding type is limited because only a small quantity of frequency responses from this type is available in National Grid database. For this winding type, there are 18 training FRA traces from three 750 MVA transformers of the same manufacturer, from A, B and C phases in Fig. 3.
phase of common and series windings, as plotted in Fig. 5a. The testing features are 9 FRA traces from A, B and C phases of the series windings of three 750 MVA transformers, as plotted in Fig. 5b.

For the single helical winding type, 6 FRA traces from two 750 MVA transformers are used as training features while nine FRA traces from two 1000 MVA and one 750 MVA transformers are used as testing features, as plotted in Fig. 6.

Once the parameters of SVM model are computed, the distances from the training vectors to the hyperplanes of each binary classifier can be calculated. For example, the distances from 12 training multiple layer winding FRA traces (Label 1) to the following three binary classifiers, 1 versus 2, 1 versus 3 and 1 versus 4, are tabulated in Table 2. For each binary classifier, the closest training vectors to its hyperplane are the support vectors. The distances between the support vectors and the corresponding hyperplane should ideally be 1. In Tables 2–5, the distances from all 54 training vectors to the hyperplanes of the relevant binary classifiers are listed. The range of these distances is from −1.5 to −1, and 1 to 1.5, which indicates that the distances between every two training vectors from the same winding type are small, due to their high similarity. Considering inevitable calculation errors, all the training vectors, whose distances to any classification hyperplanes range from 0.9990 to 1.0010, are considered as support vectors. All the support vectors are noted with ‘∗’ in Tables 2–5. As stated, there may be more than one support vector for each winding type.

It can be seen in Table 2 that for Multiple Layer VS Interleaved Disc (1 versus 3) classifier, traces 8, 9 and 11 are all close to 1. The three traces are plotted in Fig. 7. It can be seen that trace 9 and trace 11 almost overlap with each other. However, trace 8 differs from the other two traces. This means that not all the support vectors have a high similarity in the shape of FRA responses between each other, though they have the same distance to the classification hyperplane.

4.3 Testing process

Two testing features are selected as an example to show the process of winding type prediction. One testing feature is of the multiple layer winding, measured on A Phase, common winding, and the other testing feature is of the plain disc winding, measured on A Phase, common winding.

For the multiple layer versus plain disc (1 versus 2) classifier, its weight matrix w is plotted against frequency in solid lines in both Figs. 8a and b. Its bias b is −0.1956. The first testing feature xα1 to be identified is plotted in Fig. 8a after standardisation and the second testing feature xα2 is plotted in Fig. 8b after standardisation, both in dashed lines. The values ∑i∈αwi · x+i + b against frequency for the two testing features xα1 and xα2 are plotted in Figs. 8c and d. The final value of g(xα) can be calculated as: g(xα1) = 1.0954 − 0.1956 = 0.8998 and g(xα2) = −8392 − 0.1956 = −8392.1956 = −1.0347. For the first testing feature, positive number 0.8998 suggests that it should be a multiple layer winding. For the second testing feature, the negative value suggests that it should be a plain disc winding. The higher the absolute value of g(xα), the farther the testing feature is from the classification hyperplane, and the higher the prediction confidence is. In Fig. 8c, the value of ∑i∈αwi · x+i + b for the first testing feature starts from a negative number and then becomes positive in the frequency region 2–20 kHz dominated by winding interaction. Its absolute value, i.e. prediction confidence, grows rapidly from 0.1184 to 0.8998, in the frequency domain controlled by winding type. Such behaviour suggests that in this example, weight factor of the magnitude points on the frequency region from 20 to 1000 kHz plays the most important role in the decision process. In Fig. 8d, the value of ∑i∈αwi · x+i + b for the second testing FRA trace remains negative over the whole frequency range. For the frequency region controlled by winding properties, from 20 to 200 kHz, there exists the camel humps characteristics and the absolute value of g(xα2) increases to 1.3070 as well as the prediction confidence. From 200 to 1000 kHz, the camel humps disappear and the rising trend with oscillation becomes similar to the characteristics of the multiple layer winding type, thus the absolute value of g(xα2) decreases to 1.0350, as well as the prediction confidence. The prediction results from all six binary classifiers are listed in Table 6. For the first testing vector, the multiple layer winding type wins the most votes, three, which suggests that the winding corresponding to xα1 has the multiple layer winding type. Similarly, the winding with regard to the second testing feature xα2 is classified as having the plain disc winding type. The proposed method successfully identifies the correct winding type in both cases.

Noticeably, even though the testing feature xα1 belongs neither to the plain disc winding nor the interleaved winding, the plain disc versus interleaved winding classifier determines it having the interleaved disc winding type. It indicates that the testing feature xα1 is more analogous to the feature of an interleaved disc winding than that of a plain disc winding. This conclusion corresponds to the fact that the FRA traces of multiple layer and interleaved disc windings share some similarity in the rising magnitude trend from 20 to 1000 kHz, though their oscillation levels are different.

All the 54 testing features are identified correctly, as shown in Table 7. For plain disc windings, although all the training FRA traces have rising magnitude trend, the three testing FRA traces with flat trend are classified into the right winding type. The classification result is encouraging, in that the overall characteristics of FRA traces can be identified by considering not the magnitude but the trend. The testing process confirms that the developed SVM model can correctly classify FRA traces that are different from training data and is ready for winding type recognition.

4.4 Application

A total of 51 FRA traces, from ten 400/275/13 kV transformers, without winding type information are used to assess the performance of the developed SVM. The 51 FRA traces are chosen to train the SVM. The study is devised to see if the performance of the developed SVM model is improved.

Winding classification by expert experience suggests that six of them are of multiple layer winding type, 36 of them are of plain disc winding type and nine of them are of interleaved disc winding type. The plain disc winding type was widely used for 400/275/13 kV auto transformers before 1960s when the interleaved disc winding type had not been invented, thus it takes up the largest number. The FRA traces of single helical windings are included here because the measurement information indicates they belong to tertiary windings and tertiary windings use only the single helical winding type. However, this does not mean that the FRA traces of single helical windings can be excluded from training data of the SVM model, since the measurement information may sometimes be missing.

Tests show that the SVM model generates the same classification results as the expert experience. As shown in Table 8, according to the two classification methods, 6 traces are both classified as multiple layer winding type, 36 traces are both classified as plain disc winding type and 9 traces are both classified as interleaved disc winding type. The proposed method achieves 100% success rate in winding type recognition.

5 Sensitivity study

Training features determine the structure of SVM model and parameters of SVM classifiers. Therefore, the features chosen for training the SVM play an important role on the performance of SVM. It is crucial to explore the choice of training features does not severely deteriorate the accuracy of winding type recognition. To this end, three sensitivity studies are carried out to investigate the impact of training data selection on the SVM performance. First of all, multiple layer winding type features initially used for testing are chosen to train the SVM. The study is devised to see if the
SVM design can be easily swayed by training data as these features are similar to another winding type. In SVM theory, training features can be classified into two categories: non-support vectors and support vectors based on their distances to the classification hyperplane. The other two sensitivity studies focus on the impact of removing these two types of vectors from training data on the SVM design and performance.

Fig. 10 plots the distances from all training features with both multiple layer winding type and plain disc winding type to the multiple layer versus plain disc classifier as discussed in Section 4. It can be seen from the figure that the features with these two different winding types are separated by a large distance of 2, compared to the distances between features with the same winding type, the largest of which is around 0.5. This indicates that the training features with the same winding type are located closely but not necessarily share similarity in their FRA trace shapes, as confirmed in Figs. 3 and 4.

5.1 Exchange training and testing vectors

As shown in Section 4, the magnitude responses of the training features and testing features for the multiple layer winding type are different, in the frequency region between 20 and 1000 kHz. The testing features shown in Fig. 3b are smoother than the training features shown in Fig. 3a, which means that the testing features are more similar to those of the interleaved disc winding type.

When the training features and testing features are exchanged for the multiple layer winding type, the classification result will change, as shown in Table 9. Taking multiple layer features and plain disc features for example, Figs. 11a and b show the distances from the new training data and the new testing data to the new multiple layer versus plain disc classifier, respectively. The classification result during the testing process remains the same for multiple layer features and plain disc features after the exchange.

SVM design can be easily swayed by training data as these features are similar to another winding type. In SVM theory, training features can be classified into two categories: non-support vectors and support vectors based on their distances to the classification hyperplane. The other two sensitivity studies focus on the impact of removing these two types of vectors from training data on the SVM design and performance.

Table 2 Distance from the multiple layer input features to the hyperplane of binary classifier

| Multiple layer training traces 1 versus 2 1 versus 3 1 versus 4 |
|-----------------------------|-----------------------------|-----------------------------|
| 1  | 1.0243 | 1.0058 | 0.9997* |
| 2  | 1.0173 | 1.1676 | 1.0013 |
| 3  | 1.0001* | 1.0459 | 1.0027 |
| 4  | 1.4132 | 1.1755 | 1.0055 |
| 5  | 1.4006 | 1.5139 | 1.0000* |
| 6  | 1.4301 | 1.2443 | 1.0022 |
| 7  | 1.0465 | 1.0139 | 1.0030 |
| 8  | 1.0986 | 1.0004* | 1.0094 |
| 9  | 1.0151 | 1.0002* | 1.0000* |
| 10 | 1.4159 | 1.1933 | 1.0074 |
| 11 | 1.5086 | 1.0002* | 1.0594 |
| 12 | 1.4092 | 1.2991 | 1.0000* |

High Volt., 2020, Vol. 5 Iss. 6, pp. 704-715
This is an open access article published by the IET and CEPRI under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)
Table 3 Distance from plain disc input features to the hyperplane of binary classifier

| Plain disc training traces | 1 versus 2 | 2 versus 3 | 2 versus 4 |
|---------------------------|------------|------------|------------|
| 1                         | 1.0615     | 1.0506     | 1.0078     |
| 2                         | 1.0658     | 1.1224     | 1.0003*    |
| 3                         | 1.0385     | 1.0362     | 1.0005*    |
| 4                         | 1.1457     | 1.1119     | 1.0213     |
| 5                         | 1.0673     | 1.0813     | 1.0002*    |
| 6                         | 1.1076     | 1.0831     | 1.0189     |
| 7                         | 1.0076     | 1.0148     | 1.0281     |
| 8                         | 1.0357     | 1.0859     | 1.0126     |
| 9                         | -0.9999*   | 0.9999*    | 1.0067     |
| 10                        | -1.1179    | 1.1222     | 1.0002*    |
| 11                        | -1.0548    | 1.0869     | 1.0000*    |
| 12                        | -1.0781    | 1.0566     | 1.0518     |
| 13                        | -1.0358    | 1.0260     | 1.0383     |
| 14                        | -1.0600    | 1.1123     | 1.0218     |
| 15                        | -1.0002*   | 0.9999*    | 1.0311     |
| 16                        | -1.0300    | 1.0003*    | 1.0826     |
| 17                        | -1.0000*   | 1.0216     | 1.0200     |
| 18                        | -1.0358    | 1.0146     | 1.0706     |

Table 4 Distance from interleaved disc input features to the hyperplane of binary classifier

| Interleaved disc training traces | 1 versus 3 | 2 versus 3 | 3 versus 4 |
|---------------------------------|------------|------------|------------|
| 1                               | -1.0943    | -1.0155    | 1.0000*    |
| 2                               | -1.1479    | -1.0431    | 0.9995*    |
| 3                               | -1.0674    | -1.0614    | 1.0033     |
| 4                               | -1.0807    | -1.4572    | 1.0067     |
| 5                               | -1.0006*   | -1.4724    | 1.0000*    |
| 6                               | -1.0003*   | -1.4434    | 1.0058     |
| 7                               | -0.9998*   | -1.0001*   | 1.0055     |
| 8                               | -0.9996*   | -1.0136    | 1.0084     |
| 9                               | -1.0614    | -1.0054    | 1.0046     |
| 10                              | -1.0004*   | -1.4887    | 1.0170     |
| 11                              | -1.0218    | -1.4380    | 1.0110     |
| 12                              | -1.1275    | -1.4767    | 1.0183     |
| 13                              | -1.1843    | -1.0541    | 1.0035     |
| 14                              | -1.2479    | -1.0548    | 1.0000*    |
| 15                              | -1.1119    | -1.0263    | 1.0060     |
| 16                              | -1.0390    | -1.4586    | 1.0077     |
| 17                              | -1.0177    | -1.4654    | 1.0017     |
| 18                              | -1.1013    | -1.4587    | 1.0066     |

However, it can be seen from Fig. 11c that one interleaved disc winding type testing feature is wrongly classified as multiple layer type. The voting results for this interleaved disc winding type feature are shown in Table 10.

Due to the change of multiple layer training features, all three classifiers related to this winding type are affected. When such classifiers are applied to the interleaved disc winding type feature, the values of $g(x)$ change. Most importantly, the sign of $g(x)$ alters for the classifier 1 versus 3, leading to incorrect classifying it as multiple layer winding type. Eventually, the feature is falsely identified as multiple layer type because it wins the most votes according to the classification criterion in a multiclass SVM problem. However, the classification confidence is not high for this specific classifier, being 0.1886 which is $\approx 0.3717$ when the original SVM classifier is used. Consequently, there is a high probability for it to cross the classification hyperplane, once the training features change again. The exchange of training and testing features for multiple layer winding reduces the dissimilarity between the multiple layer training features and interleaved disc training features.

When the training data and testing data from the four winding types are swapped, the accuracy of the SVM prediction model drops to 67%.

5.2 Delete non-support vectors

This section examines the impact of deleting non-support vectors from training data on the performance of the proposed SVM. After the SVM model is built, the support vectors can be identified. All the support vectors are noted by * in Tables 2–5. Instead of using all features, this time only the support vectors are used for training the SVM. The result shows that the newly built SVM model is identical to the original SVM model.

A comparison is done in Fig. 12 on the weight matrixes of multiple layer versus plain disc classifier, before and after the non-support vectors are deleted. It shows that the two weight matrixes exactly overlap with each other. The correlation coefficient and average difference are used to describe the similarity between the parameters of SVM before and after the change of training data. All the correlation coefficients for the six binary classifiers are 1.0000, with 0 mean differences. The biases of the six binary classifiers also remain unchanged, with a difference value of 0. The bias of 1 versus 2 classifier before and after is listed in Table 11.

5.3 Delete support vectors

This section investigates the effect of deletion of support vectors from training data on the SVM performance of winding type recognition. The sensitivity study is carried out on the multiple layer versus plain disc winding type binary classifier.

As tabulated in Table 2, 12 FRA traces of multiple layer type windings are used as training vectors in the original SVM model. These 12 FRA traces are from A, B and C phases of two transformers. It can be found that the support vector for this binary classifier from multiple layer winding type is the third training
vector in Table 2. FRA traces from three phases of the same transformer series/common winding are similar in shape and their features are close to each other. To accelerate and simplify the process, the training features corresponding to the other two phases are removed together with the support vector to be deleted. After deletion the SVM is trained with the rest features and a new support vector can be identified. Performance of the newly built SVM is assessed. Afterwards the new support vector is deleted together with two features from the same transformer series/common winding. Such process to remove support vectors iterates until only three vectors from the same transformer winding are left. The weight matrices in each iteration are plotted in Fig. 13. The values of the corresponding biases and the performance of SVM are tabulated in Tables 12 and 13, respectively.

Table 6 SVM voting result for testing feature $x_{ij}$

| Classifier | $g(x_{ij})$ value | Testing trace 1 | Testing trace 2 |
|------------|------------------|-----------------|-----------------|
| 1 versus 2 | 0.8998           | ✓               | ✕               |
| 1 versus 3 | 1.1701           | ✓               | ✕               |
| 1 versus 4 | 1.0469           | ✓               | ✕               |
| 2 versus 3 | -0.8589          | ✕               | ✓               |
| 2 versus 4 | 1.2167           | ✓               | ✕               |
| 3 versus 4 | 1.0598           | ✓               | ✕               |

Table 7 Testing result of SVM model

| Winding type     | Training set | Testing set | Accurate number identified | Accuracy, % |
|------------------|--------------|-------------|----------------------------|-------------|
| multiple layer   | 12           | 18          | 18                         | 100         |
| plain disc       | 18           | 18          | 18                         | 100         |
| interleaved disc | 18           | 9           | 9                          | 100         |
| single helical   | 6            | 9           | 9                          | 100         |

On the first iteration, the weight matrix deviates slightly from the original. The deleted features are from a common winding. The SVM model functions well, identifying all testing features correctly. This suggests that the deleted support vectors are similar to a number of remaining training vectors and their deletion does affect the performance of the SVM. On the second iteration, the weight matrix differs from the previous weight matrix considerably. The deleted features are also from the common winding and the remaining features are all from series windings.
This leads to the significant change in the weight matrix because it is determined only by the series windings rather than the combination of common and series windings. The FRA trace of multiple layer series windings is similar to that of interleaved windings, so the SVM trained only by the former can have difficulties to distinguish these two winding types. At this stage nine multiple layer windings are wrongly classified as interleaved disc windings and the prediction accuracy drops to 83%. On the third iteration, only three features from series windings are used as training vectors. The classification accuracy remains at 83%. Slight changes can be observed in the weight matrix. This is because FRA traces obtained from series windings share great similarity. For the nine wrongly classified multiple layer windings, the $g(x)$ values of multiple layer versus interleaved disc classifier are listed in Table 14. It can be seen from the table that the sign of $g(x)$ changes from positive to negative after the first iteration. Hence, the SVM wrongly classify the features as interleaved disc type windings. The study demonstrates the importance of the support vectors in the training of SVMs. The performance of the proposed winding type recognition method depends on if the suitable support vectors are included in training. However, non-support vectors are indispensable because the support vector is a relative concept and there always exists a support vector, but not necessarily a suitable one. It is found that completely removing training features from
either common or series winding can deteriorate the performance of the SVM.

6 Conclusion

In this study, a novel SVM-based method is proposed for transformer winding type recognition using FRA data. The SVM model is built with FRA traces of 400/275/13 kV auto transformers from the UK's National Grid Database. This model is trained using FRA traces with design information, and later tested by different FRA traces with a 100% accuracy rate. Examples are given to show the prediction process and to analyse the influence of different frequency regions on the final classification results. The frequency region from 20 to 1000 kHz controlled by the winding structure plays an important role in the winding type recognition. Subsequently, the SVM is applied to 51 FRA traces with unknown winding type, and the prediction result is validated with the classification made by expert experience. The proposed method successfully identifies the correct winding type in all cases, which demonstrates the satisfactory performance of the SVM-based method. Sensitivity studies are carried out to investigate the impact of training data selection on the performance of SVM. It is concluded that the performance is mainly affected by support vectors. When small changes occur to the support vectors, the SVM model might still produce correct prediction results for the original testing vectors. However, the prediction accuracy will drop once the support vectors change significantly. It is important to identify and include the critical FRA traces in the training data for accurate identification. To ensure the suitable support vectors are used to build the classification model, it is suggested to use as many training FRA traces as possible. Meanwhile, expert judgement and practical experience can be exercised and referred to when training the SVM model.

Table 12 Bias when different numbers of training features are used

|                | Original | Iteration 1 | Iteration 2 | Iteration 3 |
|----------------|----------|-------------|-------------|-------------|
| bias           | -0.1956  | -0.2226     | -0.3663     | -0.3235     |

Table 13 SVM performance when different numbers of training features are used

| Number of training vectors | Number of testing vectors | Number of correct identification | Success rate, % |
|----------------------------|---------------------------|---------------------------------|-----------------|
| original                   | 12                        | 54                              | 54              | 100              |
| iteration 1               | 9                         | 54                              | 54              | 100              |
| iteration 2               | 6                         | 54                              | 45              | 83               |
| iteration 3               | 3                         | 54                              | 45              | 83               |

With the changes in the voltage ratio, power rating, winding types, etc. the frequency region dominated by the properties of winding-under-test, may be different. The linear SVM model built in this study would be only applicable with confidence to the 400/275/13 kV autotransformers with different power ratings (500, 750 and 1000 MVA). Further study should be carried out on winding type recognition of transformers with a variety of voltage levels, such as a mixed database of 275/33 kV, 275/132/13 kV and 400/275/13 kV transformers.

7 References

[1] Heathcote, M.: ‘A J & P transformer book’ (Elsevier, Amsterdam, Netherlands, 2011)
[2] Dick, E.P., Erven, C.C.: ‘Transformer diagnostic testing by frequency response analysis’, IEEE Trans. Power Appar. Syst., 1978, PAS-97, (6), pp. 2144–2153
[3] Sofian, D.M., Wang, Z.D., Jayasinghe, S.B.: ‘Frequency response analysis in diagnosing transformer winding movements-fundamental understanding’. 39th Int. Universities Power Engineering Conf., Bristol, UK, 2004, pp. 138–142
[4] Barros, R.M.R., da Costa, E.G., Araujo, J.F., et al.: ‘Contribution of inrush current to mechanical failure of power transformers windings’, High Volt., 2019, 4, (4), pp. 300–307
[5] Ryder, S.A.: ‘Diagnosing transformer faults using frequency response analysis’, IEEE Electr. Insul. Mag., 2003, 19, (2), pp. 16–22
[6] Samimi, M.H, Tenbohlen, S, Shayegani Akmal, A.A., et al.: ‘Evaluation of numerical indices for the assessment of transformer frequency response’, IET Gener. Transm. Distrib., 2017, H, (1), pp. 218–227
[7] Gustavsen, B., Semlyen, A.: ‘Rational approximation of frequency domain responses by vector fitting’, IEEE Trans. Power Deliv., 1999, 14, (3), pp. 1052–1060
[8] Levi, E.C.: ‘Complex-curve fitting’, IRE Trans. Autom. Control, 1959, AC-4, (1), pp. 37–44
[9] Mao, X, Wang, Z.D., Wang, Z.: ‘Accurate estimating algorithm of transfer function for transformer FRA diagnosis’. IEEE Power & Energy Society General Meeting, Portland, OR, 2018, pp. 1–5
[10] Wang, Z.D., Li, J, Sofian, D.M.: ‘Interpretation of transformer FRA responses – part I: influence of winding structure’, IEEE Trans. Power Deliv., 2009, 24, (2), pp. 703–710
[11] Sofian, D.M., Wang, Z.D., Li, J.: ‘Interpretation of transformer FRA responses – part II: influence of transformer structure’, IEEE Trans. Power Deliv., 2010, 25, (4), pp. 2582–2589
[12] Ang, S.P, Li, J., Wang, Z.D.: ‘FRA low frequency characteristic study using duality transformer core modelling’. Int. Conf. on Condition Monitoring and Diagnosis, Beijing, China, 2008, pp. 889–892
[13] Mitchell, S.D, Welsh, J.S.: ‘Initial parameter estimates and constraints to support gray box modelling of power transformers’, IEEE Trans. Power Deliv., 2013, 28, pp. 2411–2418
[14] Bigdeli, M., Vakilian, M., Rahimpour, E.: ‘Transformer winding faults classification based on transfer function analysis by support vector machine’, IET Electr. Power Appl., 2012, 6, (5), pp. 268–276
[15] Bagheri, M., Zollanvari, A., Nezhivenko, S.: ‘Transformer fault condition prognosis using vibration signals over cloud environment’, IEEE Access, 2018, 6, pp. 9862–9874
[16] Duan, L., Hu, J., Zhao, G., et al.: ‘Method of inter-turn fault detection for next-generation smart transformers based on deep learning algorithm’, High Volt., 2019, 4, (4), pp. 282–291
[17] Junani, H, Kordi, B.: ‘Towards automated statistical partial discharge source classification using pattern recognition techniques’, High Volt., 2018, 3, (3), pp. 162–169
[18] Mitchell, S.D, Welsh, J.S.: ‘Methodology to locate and quantify radial winding deformation in power transformers’, High Volt., 2017, 2, (1), pp. 17–24
[19] Zambra, G.M.V., Ferreira, A.C.C., Caloba, L.P.: ‘Power transformer equivalent circuit identification by artificial neural network using frequency
response analysis’. IEEE Power Engineering Society General Meeting, Montreal, Que., 2006, pp. 6–12

[20] Luo, Y., Ye, J., Gao, J., et al.: ‘Recognition technology of winding deformation based on principal components of transfer function characteristics and artificial neural network’, IEEE Trans. Dielectr. Electr. Insul., 2017, 24, (6), pp. 3922–3932

[21] Ghanizadeh, A.J., Gharehpetian, G.B.: ‘ANN and cross-correlation based features for discrimination between electrical and mechanical defects and their localization in transformer winding’, IEEE Trans. Dielectr. Electr. Insul., 2014, 21, (5), pp. 2374–2382

[22] Liu, J., Zhao, Z., Tang, C., et al.: ‘Classifying transformer winding deformation fault types and degrees using FRA based on support vector machine’, IEEE Access, 2019, 7, pp. 112494–112504

[23] Nurmanova, V., Bagherti, M., Zollanvari, A., et al.: ‘A new transformer FRA measurement technique to reach smart interpretation for inter-disk faults’, IEEE Trans. Power Deliv., 2019, 34, (4), pp. 1508–1519

[24] Abbasi, A.R., Mahmoudi, M.R., Avazzadeh, Z.: ‘Diagnosis and clustering of power transformer winding fault types by cross-correlation and clustering analysis of FRA results’, IET Gener. Transm. Distrib., 2018, 12, (19), pp. 4301–4309

[25] Mao, X., Wang, Z.D., Jarman, P., et al.: ‘Applying unsupervised machine learning method on FRA data to classify winding types’, Int. Symp. on High Voltage Engineering, Budapest, Hungary, 2019, pp. 969–981

[26] De, S., De, A., Bandyopadhyay, G.: ‘Self-capacitance grading approach to suppress winding resonance in EHV transformers under oscillatory terminal disturbances’, High Volt., 2018, 3, (4), pp. 310–315

[27] Vapnik, V.: ‘The nature of statistical learning theory’ (Springer Science & Business Media, Berlin, Germany, 2013)

[28] Shevade, S.K., Keerthi, S.S., Bhattacharyya, C., et al.: ‘Improvements to the SMO algorithm for SVM regression’, IEEE Trans. Neural Netw., 2000, 11, (5), pp. 1188–1193

[29] Platt, J.: ‘Sequential minimal optimization: a fast algorithm for training support vector machines’, Microsoft Research Technical Report, 1998

[30] Hsu, C.W., Lin, C.J.: ‘A comparison of methods for multiclass support vector machines’, IEEE Trans. Neural Netw., 2002, 13, (2), pp. 415–425