Miloš BOŽIĆ[1], Miloš STOJANOVIĆ[2], Zoran STAJIĆ[1], Đukan VUKIĆ[3]
Faculty of Electronic Engineering, University of Niš (1); School of higher Technical Professional Education (2);
Faculty of Agriculture, University of Belgrade (3)

Power Transformer Fault Diagnosis based on Dissolved Gas Analysis with Logistic Regression

Abstract. *Logistic regression (LR) approach for power transformer fault diagnosis, based on dissolved gas analysis (DGA) is presented in this paper. DGA methods proposed by actual standard IEC 60599, often identify wrong fault or cannot even recognize fault type. To overcome these problems, in recent years, several artificial intelligence (AI) approaches are proposed. In this paper LR is applied for the first time in multi-layer and multi class configuration models for transformer fault diagnosis. It is shown that the proposed approach gives a very good classification performance.*

Streszczenie. *Przedstawiono metodę diagnostyki transformatora bazująca na analizie rozpuszczonego gazu metodą logistycznej regresji. Metoda ta umożliwia nie tylko wykrywanie uszkodzeń ale także ich klasyfikację. (**Diagnostyki transformatora bazująca na analizie rozpuszczonego gazu metodą logistycznej regresji**)*

Keywords: power transformer, dissolved gas analysis, logistic regression, fault diagnosis.
Słowa kluczowe: diagnostyka transformatora, analiza gazu, logistyczna regresja

Introduction

Power transformers are of prime importance for electrical power systems. The condition of a power transformer is crucial for its successful operation and, as a consequence, for the reliability of the power system as whole. The detection of incipient faults which may be caused by insulation weakness, malfunction, defects or deterioration is of fundamental importance [1]. A set of modern diagnostic methods is available and applied for oil filled power transformers. These methods allow the operators to plan adequate corrective actions at an early stage.

For many years the DGA method has been used as a tool in transformer diagnostic. The main idea behind the use of DGA is based on the fact that during its lifetime, all oil-cellulose insulated systems generate decomposition gases under the influence of various stresses, both normal and abnormal [2]. The method has been used for several purposes: to detect incipient faults, to supervise suspect transformers, to test a hypothesis or explanation for the probable cause of failures or disturbance which have already occurred and to ensure that new transformer are healthy.

Conventional DGA interpretation methods are: Individual and total dissolved key-gas concentration method [3] (not universally accepted), Rogers ratio method [4], IEC Method [5] and Duval triangle method (Graphical representation method) [6]. However, the identification of fault types by the conventional methods is not always an easy task due to the variability of gas data and operational nature. These methods often give different fault diagnosis results for the same input data. In recent years several AI techniques have been used in order to obtain unique and accurate diagnostic results, such as fuzzy logic (FL) [7], expert systems (ES) [8], artificial neural networks (ANNs) [9] and self-organizing maps [10].

FL and ES methods take a human expertise to form decision-making system and they can incorporate DGA standard. On the other hand, both methods use knowledge base which needs to fill in manually and maintain in the future. ANNs methods map complicated relationship among dissolved gas contents in transformer and corresponding fault type and show a good fault detection performance. Support vector machines (SVMs), proposed by Vapnik in [11], are also widely used for fault detections problems. Applications of SVMs in DGA fault diagnosis are presented in several recent papers [12-16]. In [14], DGA fault diagnosis system is developed based on multi-layer SVM classifier. Fault classification is done in several levels where each level use SVM model to identify one transformer state. In proposed approach, *k-1* SVM models are formed where *k* represent number of transformer states (number of faults plus normal state). Similar approach is used in [15], with multi-layer SVM and genetic algorithm. Application of genetic algorithm in collaboration with SVM is used in order to overcome the problem of choosing good SVM model parameters. Method proposed in [16] overcomes the conventional methods shortcomings in the manner of establishing input vector by the combination of ratios and graphical conventional methods. Than SVM is applied to establish the power transformers faults classification and to choose appropriate gas signature between the DGA conventional methods and proposed method.

In this paper, logistic regression (LR) method for DGA transformer fault classification is applied for the first time, with aim to resolve the problem fault type recognition that occurs in the conventional methods. LR is simple to implement and time efficient by avoiding complex computations which exist in SVMs and ANNs, while maintaining most of the classification performance.

The rest of the paper is organized as follows: Section 2 describe DGA transformer fault types and conventional identification methods. Then Section 3 presents the formulation of the LR and describes the simple method used for optimization. Section 4 includes experiments to verify the proposed approach. Finally, Section 5 outlines the conclusions.

Table 1. Interpretation gas dissolved in the oil [5]

Gas detected	Symbol	Interpretation
Hydrogen	H_2	Electric discharge (corona effect, low partial discharge)
Methane	CH_4	Secondary indicator of an arc or serious overheating
Ethylene	C_2H_4	Thermal fault (overheating local)
Ethan	C_2H_6	Secondary indicator of thermal fault
Acetylene	C_2H_2	Electric fault (arc, spark)
Carbon monoxide	CO	Cellulose decomposition
Carbon dioxide	CO_2	Cellulose decomposition
Oxygen	O_2	Transformer seal fault

DGA interpretation methods

IEC 60599 standard establishes an interpretation by which five gases H2, CH4, C2H2, C2H4 and C2H6 can be used to detect different types of faults. Table 1 shows the diagnostic interpretations applying various key gas concentrations.

All these gases except oxygen and nitrogen may be formed during the degradation of the insulation. The amount and the relative distribution of these gasses depend on the type and severity of the degradation and stress. Internal inspection of hundreds of faulty equipment has led to the broad classes of visually detectable faults [5], presented in Table 2.

Table 2. Fault types

Fault type code	Fault type
PD	Partial discharge
D1	Discharges of low energy
D2	Discharges of high energy
T1	Thermal fault, T < 300 °C
T2	Thermal fault, 300 °C < T < 700 °C
T3	Thermal fault, T > 700 °C

Roger ratio method

In the interpreting gas analysis results, relative gas concentrations are found to be more useful than actual concentrations. Only five gas concentrations (H_2, CH_4, C_2H_6, C_2H_4 and C_2H_2) are sufficient for most purposes. According to the scheme developed by Rogers and later simplified by the IEC, three gas ratios define a given condition. This method uses the following rations:

R1 = C_2H_2 / C_2H_4, R2 = CH_4 / H_2, R3 = C_2H_4 / C_2H_6.

It should be noted that the Rogers ratio method is for analyzing faults and not for detecting the presence of faults. Faults presence can be detected using the total amount of gas limit or increased gas generation rates. Also it is advisable to never make a decision based only on a ratio if either of the gases used in that ratio is less 10 times the amount the gas chromatograph can detect. Table 3 shows Rogers ratios for key gases.

Table 3. Rogers ratios for key gases [1]

Code	Range of Ratios	C_2H_2 / C_2H_4	CH_4/ H_2	C_2H_4/ C_2H_6
	< 0.1	0	1	0
	0.1 – 1	1	0	0
	1 – 3	1	2	1
	> 3	0	2	2
Case	Fault type			
0	No fault	0	0	0
1	Low energy partial discharge	1	1	0
2	High energy partial discharge	1	1	0
3	Low energy discharges	1-2	0	1-2
4	High energy discharges	1	0	2
5	Thermal fault less than 150°C	0	0	1
6	Thermal fault temp. range 150-300°C	0	2	0
7	Thermal fault temp. range 300-700°C	0	2	1
8	Thermal fault temp. range over 700°C	0	2	2

IEC Ratio method

The IEC Ratios method utilizes five gases H2, CH4, C2H2, C2H4 and C2H6. These gases are used to produce a three gas ratios: C2H2/C2H4, CH4/H2 and C2H4/C2H6. The method is very similar to the Rogers Ratio method. Table 4 shows the IEC standard for interpreting fault types and gives the values for three key-gases ratios.

Table 4. Diagnosis using the Ratio method [5]

Fault type	C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6
PD	NS*	< 0.1	< 0.2
D1	>1	0.1 – 0.5	> 1
D2	0.6 – 2.5	0.1 - 1	> 2
T1	NS*	>1 but NS*	<1
T2	< 0.1	>1	1 - 4
T3	< 0.2	> 1	> 4

NS* Non-significant whatever the value

Duval triangle – graphical method

The concentration of the three Duval triangle gases, expressed as a percentage of the total sum (CH_2+C_2H_4+C_2H_2) define a point in a coordinate system represented as a triangular diagram (Fig. 1), which is subdivided in different zones. Each zone is related to a certain type of fault. Duval triangle cannot be used to determine whether or not a transformer has a problem. There is no area on the triangle for a transformer that does not have a problem. In Fig. 1 zone DT corresponds to mixture of thermal and electrical faults. Fig. 1 can be translated in a table that gives the limits of each fault which are summarized in Table 5.

Table 5. Limits of zones [5]

PD	98% CH_4			
D1	23% C_2H_4	13% C_2H_2		
D2	23% C_2H_4	13% C_2H_2	38% C_2H_4	29% C_2H_2
T1	4% C_2H_2	10% C_2H_4		
T2	4% C_2H_2	10% C_2H_4	50% C_2H_4	
T3	15% C_2H_2	50% C_2H_4		

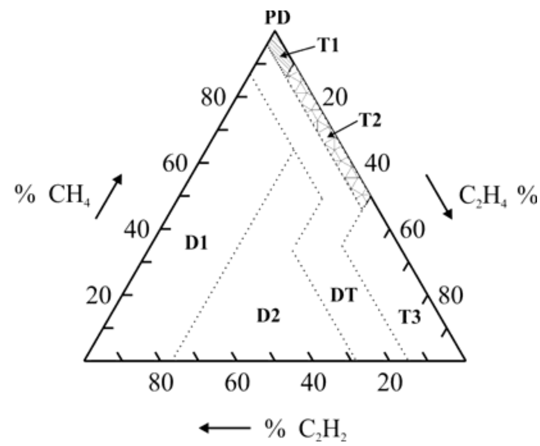

Fig.1. Coordinates and fault zones of the Duval triangle

Logistic Regression

Logistic regression is widely used supervised machine learning technique, mainly applied in solving classification problems [17]. It is a simple and efficient method that provides explicit probabilities of class membership for binary classification tasks.

Let us consider a classification task with *M* training instances $\left\{\left(x^{(i)}, y^{(i)}\right), i=1,\ldots,M\right\}$ where each $x^{(i)} \in R^N$ is an *N* dimensional feature vector and $y^{(i)} \in \{0,1\}$ is a class label. Logistic regression models the probabilities of the class label *y* given a feature vector *x* as in (1):

$$p\left(y=1 \mid x;\theta\right) = h_\theta\left(x\right) = g\left(\theta^T x\right) = \frac{1}{1+e^{-\theta^T x}}, \tag{1}$$

where $0 \le h_\theta\left(x\right) \le 1$ represents hypothesis function, $g\left(\theta^T x\right)$ denotes logistic (sigmoid) function and $\theta = \left[\theta_0, \theta_1, \ldots, \theta_N\right]$ is a vector of learning parameters of the logistic regression model.

Let us assume that:

$$\begin{aligned} P\left(y=1 \mid x;\theta\right) &= h_\theta\left(x\right), \\ P\left(y=0 \mid x;\theta\right) &= 1-h_\theta\left(x\right). \end{aligned} \tag{2}$$

Expression (2) can be written more compactly as:

(3) $$p\left(y \mid x;\theta\right) = (h_\theta(x))^y (1-h_\theta(x))^{1-y}$$

Assuming that the M training examples were generated independently, the likelihood of the parameters can be expressed as:

(4) $$\begin{aligned} L(\theta) &= \prod_{i=1}^{M} p(y^{(i)} \mid x^{(i)};\theta) \\ &= \prod_{i=1}^{M} (h_\theta(x^{(i)}))^{y^{(i)}} (1-h_\theta(x^{(i)}))^{1-y^{(i)}} \end{aligned}$$

Now, given this probabilistic model relating the $y^{(i)}$'s and the $x^{(i)}$'s, it is necessary to obtain optimal parameters θ of logistic regression model. The principal of maximum likelihood says that θ should be chosen to make the data as high probability as possible. i.e., θ should be chosen to maximize $L(\theta)$. Instead of maximizing $L(\theta)$, any strictly increasing function of $L(\theta)$ also can be maximized. In particular, the derivations will be a bit simpler if log likelihood is maximized instead, as in (5):

(5) $$\begin{aligned} l\left(\theta\right) &= \log L(\theta) = \sum_{i=1}^{M} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1-y^{(i)}\right) \\ &\cdot \log\left(1-h_\theta\left(x^{(i)}\right)\right), \end{aligned}$$

For classification problem that involves extremely small sample sizes M (e.g. less than one hundred), regularization term needs to be added into (5) in order to control bias - variance tradeoff [18]. The regularized log likelihood function is defined in (6):

(6) $$\begin{aligned} l\left(\theta\right) &= -\log L(\theta) = -\sum_{i=1}^{M} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1-y^{(i)}\right) \\ &\cdot \log\left(1-h_\theta\left(x^{(i)}\right)\right) + \frac{\lambda}{2}\sum_{j=1}^{N} \theta_j^2, \end{aligned}$$

where λ>0 is regularization parameter, which can be tuned on the training set by some of cross - validation methods. This step is crucial in training LR classifier, because optimal solution for θ can be found only after obtaining optimal λ on the training set. Therefore, performance of LR model directly depends from optimal choice of λ. Accordingly, more details about k - fold cross - validation procedure and grid - search algorithm, which are employed in this study in order to optimize parameter λ will be given in section 4.

The optimization problem given in (6) is convex, so simple gradient descent algorithm always obtains optimal solutions for parameters θ. Note that now because adding minus sign in (6) $l\left(\theta\right)$ needs to be minimized. In vector notation updates of θ according to gradient descent are given by:

(7) $$\theta = \theta - \alpha \nabla_\theta l(\theta)$$

Accordingly, parameters θ_j are simultaneously updated until the $l\left(\theta\right)$ in (6) stops decreasing, by the (8) and (9),

(8) $$\theta_0 = \theta_0 - \alpha \sum_{i=1}^{M}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_0^{(i)},$$

(9) $$\begin{aligned} \theta_j &= \theta_j - \alpha\left[\sum_{i=1}^{M}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j^{(i)} - \lambda\theta_j\right], \\ j &= 1,...,N, \end{aligned}$$

where parameter α represents learning rate that controls speed of convergence and needs to be set manually.

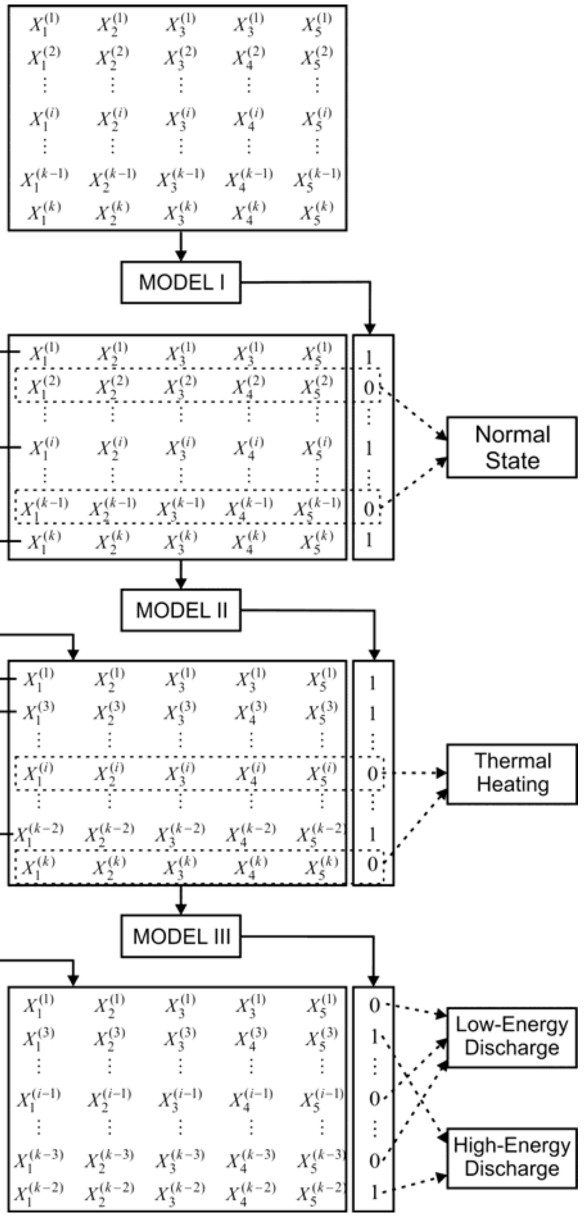

Fig. 2. Multi-layer architecture approach

Once when parameters θ_j are established, classification model can be employed according to (1). Because LR gives us probabilities interpretation of class membership in range [0-1], decision threshold needs to be defined. Every value obtained from (1) which is greater than 0.5 is treated as “1” i.e. instance belongs to class and below 0.5 is treated as “0” i.e. instance does not belong to class.

Logistic regression for binary classification can be easily extended to multi - class problems by employing “one versus all” approach [19]. If $k>2$ is a number of classes, it is needed to train k different binary classifiers between each class and the rest of the classes (which are then treated as a one class). For predicting a new input vector x, the wining class is the i^{th} that maximizes $\max_i h_\theta^{(i)}\left(x\right), i = 1,...,k$.

Transformer faults classification

Proposed approaches

Two architectures are implemented for transformer fault classification in this paper. Multi-layer architecture is proposed for the first time in [14]. As shown in Fig. 2, this approach includes three LR classifier models which are used to identify four transformer states: normal state, thermal heating, high - energy discharge and low - energy discharge. First layer consist of LR model which separate the normal state from the fault states. When fault states are separated from normal state, in second layer LR model separate discharge faults from overheating faults. In third layer, LR model is formed to identify high - energy and low - energy discharge faults.

Fig. 3 shows multi - class approach architecture with “one versus all” coding for multi - class classification. In this approach, number of models is equal to number of classes (transformer states). Every model performs binary classification between one class and all remaining classes. When classification is done for every model in this way, the final class label is determined by finding maximum output from all models, for each of input vectors. This approach overcomes the problem when one class has small number of vectors for model training.

Model formation

For model evaluation data from 500 kV transformers are used, as in [14].

The training data set consist of 25 samples of thermal heating, 15 samples of high - energy discharge, 5 samples of low - energy discharge and 5 samples of normal state. For the testing set 25 history data of the power transformer are used, consisting of 13 thermal heating samples, 2 high - energy discharge samples, 4 normal samples and 6 low - energy discharge samples.

Every instance in training and testing set is composed of 5 features which determine absolute concentration of gases,

$$\left[X_1^{(i)}\, X_2^{(i)}\, X_3^{(i)}\, X_4^{(i)}\, X_5^{(i)}\right]=\left[\mathrm{H_2\, CH_4\, C_2H_6\, C_2H_4\, C_2H_2}\right].$$

Based on chosen architecture described in previous section and shown in Fig. 2 and Fig. 3 and corresponding training sets, 3 LR models are trained for multi-layer and 4 LR models for multi-class approach. When using regularized LR, parameter λ needs to be optimized. For this purpose, *k* - fold cross - validation (*k=5*) procedure is used. The training set is randomly subdivided into *k* disjoint subsets of approximately equal size and the LR classifier is built *k* times with the current λ. Each time, one of the *k* subsets is used as the test set and the other *k-1* subsets are put together to form a training set. After *k* iterations, the average value of cost function is calculated for the current λ. The entire process is repeated with an update of the parameter λ until the given stopping criterion is reached. Parameter λ is updated in the given range using predefined equidistant steps, according to the grid-search procedure. After obtaining the optimal λ, classifier model is trained on the training set.

Experimental results and discussions

Evaluation criterions for binary classification are defined based on all possible outcomes of predicted class labels which are given in Table 6.

Table 6. All possible prediction outcomes for binary classification

	Actual class ($y^{(i)}=1$)	**Actual class** ($y^{(i)}=-1$)
Predicted class ($h_\theta(x)=1$)	TP - True positives	FP - False positives
Predicted class ($h_\theta(x)=-1$)	FN - False negatives	TN - True negatives

Three most commonly used criterions, derived from Table 6 are precision, recall and F - measure, defined by (6), (7) and (8).

$$\text{Precision}=\frac{TP}{TP+FP} \tag{6}$$

$$\text{Recall}=\frac{TP}{TP+FN} \tag{7}$$

$$\text{F}=\frac{2\,P\,R}{P+R} \tag{8}$$

Precision determines if element which is classified in class $h_\theta(x)=1$ actually belongs to that class $y^{(i)}=1$. Recall determines how much examples of a given class classifier can recognize. The F - measure is the harmonic mean of precision and recall and it determines the tradeoff between precision and recall.

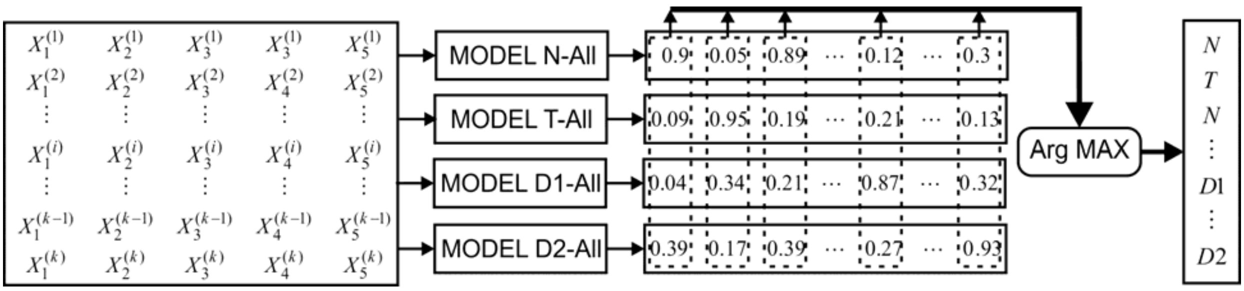

Fig. 3. Multi-class architecture approach

In multi class classification problems additional criterion needs to be defined based on the F - measure, macro-average defined in (9) [20].

$$F^{macro}=\frac{1}{k}\sum_{i=1}^{k}F_i \tag{9}$$

Precision, recall and F measures for all build models in multi-class architecture are shown in Table 7. General architecture behavior is described with macro F measure which has the value of 0.79.

Table 7. Precision, recall and F measures for multi-class architecture

Model	Recall	Precision	F	F^macro
D1-All	1.00	0.67	0.80	0.79
D2-All	0.67	0.80	0.73	
T-All	1.00	0.93	0.96	
N-All	0.50	1.00	0.67	

Table 8 shows precession, recall and F measures for three build models in multi-layer architecture. It is obvious that first and second model identify all fault successfully. Macro F measure has value of 0.93 which is expected with regard of individual model fault identification success.

Table 8. Precision, recall and F measures for multi-layer architecture

Model	Recall	Precision	F	F^macro
N-D+T	1	1	1	0.93
T-D	1	1	1	
D1-D2	1	0.67	0.8	

Table 9 show results of faults identification on test data set with LR and conventional methods. Compared to conventional methods, both LR architectures have proved more successful. Multi-class LR architecture have two incorrectly identified faults on test set while multi-layer architecture have one wrong identified fault. Conventional methods best fault identification result are obtained using Duval method, with the result that is worse than results obtained with both LR architectures. Accuracy on test set for LR approaches and conventional methods are given on the bottom of Table 9.

Conclusion

In this paper, the Logistic regression method is implemented for the faults classification of power transformer, using the dissolved gas analysis. In the LR approach, regularization term is involved in order to overcome difficulties of small sample size and to prevent over fitting. The advantages of using LR instead of other more complex machine learning based techniques lies in its simplicity and time efficiency.

The experimental results show that the diagnostic results of both LR architectures, multi-class and multi-layer, significantly outperform conventional methods. Compared with other AI approaches, the proposed method shows competitive performance for fault diagnosis.

Considering small training/test data sets, obtained results are excellent. Future research should be oriented towards expansion of training/test set with more instances which will definitely lead to better results. Possibility of introduction of new features based on DGA also can be considered. This needs to be examined more, and one unique data set has to be built for evaluation of both AI and standard methods, as well as revision of existing standard.

Table 9. Test data fault diagnosis by logistic regression and conventional methods

No.	Real fault state	Logistic regression		Conventional methods		
		Multi-class	Multi-layer	Rogers	IEC	Duval
1.	D2	D2	D2	D2	D2	D2
2.	D2	D2	D2	D2	D2	D2
3.	D1	**D2**	**D2**	**Un**	D1	**D2**
4.	D1	D1	D1	**Un**	**PD**	**T**
5.	D1	D1	D1	**Un**	**Un**	**D+T**
6.	D1	D1	D1	**Un**	**Un**	D1
7.	D1	D1	D1	**T**	**T**	**T**
8.	D1	D1	D1	**N**	**T**	**T**
9.	T	T	T	T	T	T
10.	T	T	T	T	T	T
11.	T	T	T	T	T	T
12.	T	T	T	T	T	T
13.	T	T	T	T	T	T
14.	T	T	T	T	T	T
15.	T	T	T	T	T	T
16.	T	T	T	T	T	T
17.	T	T	T	T	T	T
18.	T	T	T	T	T	T
19.	T	T	T	T	T	T
20.	T	T	T	**Un**	**Un**	T
21.	T	T	T	T	T	T
22.	N	N	N	**D2**	**Un**	**-**
23.	N	N	N	N	**Un**	**-**
24.	N	**T**	N	**T**	**Un**	**-**
25.	N	N	N	**Un**	**Un**	**-**
Acc. [%]		92	96	60	60	80*

Un – Unknown state, method cannot identify fault state.
* - Result without normal state detection, Duval method cannot identify normal state.

Acknowledgements

This work was supported by the Ministry of Science and Technological Development, Republic of Serbia (Project number: III - 44006).

REFERENCES

[1] ABB**,** Service Handbook for Transformers - ABB Zurich, Switzerland, 2006.
[2] United States Department of the Interior (Bureau of Reclamation), Transformer Maintenance Manual, Facilitates Instructions, Standards and Techniques, 3 (2000), No. 33.
[3] IEEE Standard C57.104-1991 - Guide for the Interpretation of Gasses Generated in Oil-Immersed Transformers, (1991).
[4] R.R. Rogers**,** IEEE and IEC Codes to Interpret Incipient Faults in Transformers Using Gas in Oil Analysis, CEGB Transmission division, (1995).
[5] International Electro Technical Commission (IEC 60599), Mineral Oil-Impregnated Electrical Equipment in Service, Interpretation of Dissolved and Free Gas Analysis - IEC, (2007).
[6] Duval M., A review of faults detectable by gas-in-oil analysis in transformers, IEEE Electrical Insulation Magazine, 18 (2002), 8-17.
[7] Q. Su, C. Mi, L.L. Lai, P. Austin**,** A fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer, IEEE Trans. Power Syst., 15 (2000), No. 2, 593-598.
[8] C.E. Lin, J.M. Ling, C.L. Huang, An expert system for transformer fault diagnosis using dissolved gas analysis, IEEE Trans. Power Del., 8 (1993), No. 1, 231-238.
[9] Y. Zhang, X. Ding, Y. Liu, P.J. Griffin, An artificial neural network approach to transformer fault diagnosis, IEEE Trans. Power Del., 11 (1996), No. 4, 1836-1841.
[10] K. Thang, R. Aggarwal, A. McGrail, D. Esp, Analysis of power transformer dissolved gas data using the self-organizing map, IEEE Trans. Power Del., 18 (2003), No. 4, 1241-1248.
[11] V.N. Vapnik, The Nature of Statistical Learning Theory - Springer-Veriag, New York, (1995).
[12] Piotr S. SZCZEPANIAK, Marcin KŁOSIŃSKI, Maximal margin classifiers applied to DGA-based diagnosis of power transformers, PRZEGLĄD ELEKTROTECHNICZNY (Electrical Review), 2 (2012), 100-104.
[13] Mu ZHANG, Kun LI, Huixin TIAN, Multiple SVMs Modeling Method for Fault Diagnosis of Power Transformers, PRZEGLĄD ELEKTROTECHNICZNY (Electrical Review), 7b (2012), 232-234.
[14] L. Ganyun, H. Cheng, H. Zhai, L. Dong, Fault diagnosis of power transformer based on multi-layer SVM classifier, Electric Power Systems Research, 75 (2005), 9-15.
[15] S.-w. Fei, X.-b. Zhang, Fault diagnosis of power transformer based on support vector machine with genetic algorithm, Expert Systems with Applications, 36 (2009), 11352-11357.
[16] K. Bacha, S. Souahlia, M. Gossa, Power transformer diagnosis based on dissolved gas analysis by support vector machines, Electric Power System Research, 83 (2012), 73-79.
[17] D. Hosmer, S. Lemeshow, Applied Logistic Regression (2nd ed.) - John Wiley & Sons, (2000).
[18] C. Bielza, V. Robles, P. Larrañaga, Regularized logistic regression without a penalty term: An application to cancer classification with microarray data, Expert Systems with Applications, 38 (2011), 5110–5118.
[19] R. Rifkin, A. Klautau, In Defense of One-Vs-All Classification, Journal of Machine Learning Research, 5 (2004), 101-141.
[20] R. E. Fan, C. J. Lin, A study on threshold selection for multi-label classification, Technical Report, National Taiwan University, (2007).

***Authors**: Miloš Božić, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: milos1bozic@gmail.com; Miloš Stojanović, School of higher Technical Professional Education, Aleksandra Medvedeva 20, 18000 Niš, Serbia, E-mail: milosstojanovic10380@yahoo.com; Zoran Stajić, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: zoran.stajic@elfak.ni.ac.rs; Đukan Vukić, Faculty of Agriculture, University of Belgrade, Nemanjina 6, 11080 Beograd-Zemun, Serbia, E-mail: vukicd@agrif.bg.ac.rs.*