

Automatic Learning Techniques for on-line Control and Optimization of Transformer Core Manufacturing Process

P. Georgilakis⁽¹⁾ N. Hatziaargyriou⁽²⁾ D. Paparigas⁽¹⁾ J. Bakopoulos⁽¹⁾ S. Elefsiniotis⁽¹⁾

⁽¹⁾ Schneider Electric AE, Elvim Plant, P.O. Box 59, GR-32011, Inofyta, Viotia, Greece

⁽²⁾ Electric Energy Systems Laboratory, National Technical University of Athens, Greece

Abstract - In this paper, a novel computer based learning framework that has been developed and applied for the on-line control and optimization of transformer core manufacturing process is presented. The proposed framework aims at predicting core losses of wound core distribution transformers at the early stages of transformer construction. Moreover, it is used to improve the grouping process of the individual cores by reducing iron losses of assembled transformers. Three different automatic learning techniques (namely decision trees, artificial neural networks and genetic algorithms) are combined and their relevant features are exploited.

Index Terms - Decision trees, neural networks, genetic algorithms, intelligent core loss modeling.

I. INTRODUCTION

In today's competitive market environment there is an urgent need for the transformer manufacturing industry to improve transformer efficiency and to reduce costs, since high quality, low cost products and processes have become the key to survival. Transformer efficiency is improved by reducing *load* and *no load (iron)* losses. Methods used by transformer designers in order to decrease the iron losses tend to increase the load losses and vice versa. The decision on the best design is based on loading and other specifications of each individual transformer. In most cases however, it is required that the transformer is designed with minimum iron losses [1]. These are particularly important, considering the fact that under normal operating conditions the transformer is continuously energized, i.e., 24 hours per day, every day, and therefore considerable energy is consumed in the core, while load losses occur only when the transformer is on load. Thus, iron losses constitute one of the main parameters of transformer quality.

Accurate prediction of transformer iron losses is an important task in transformer manufacturing, since it protects the manufacturer from paying loss penalties. One possible method to avoid this risk, is to design the transformer at a lower magnetic induction, resulting in an increase of the transformer cost since more magnetic material is required [2]. Satisfactory prediction of iron losses can be achieved if various parameters involved in the process, both qualitative

and quantitative, are taken into consideration. In the current practice the calculations are based on graphs (e.g., the loss curve) and tables obtained from past measurements on actual transformers. This is dictated by the fact that there is no analytical relationship expressing the effect of the various parameters on transformer iron losses. Concerning the loss curve for example, only the influence of the rated magnetic induction on iron losses for each specific magnetic material is considered.

For these reasons, iron loss prediction during design of transformers with stacked cores has been investigated in [3,4]. More specifically, in [3] the effects of a number of core production attributes on transformer core loss performance have been investigated. In [4] spatial distribution components and total core losses are calculated using a generic 2D finite difference method.

This paper deals with iron loss reduction of wound core distribution transformers. The origin of our work is the effective use of measurements taken at the first stages of transformer construction, in order to minimize iron losses of transformer (final product), while previous works were concentrated on the more accurate calculation of transformer iron losses during its design phase. The measurements taken at the first stages of core construction are collected and stored in databases. Each database corresponds to different conditions (environment), i.e. to a certain supplier, grade and thickness of magnetic material. When a satisfactory number of measurements has been collected, automatic learning is used in order to extract the information included in the databases [5]. More specifically, decision trees (DTs) [6,7] are used to select the most relevant attributes among a large set of candidate ones and to produce "if-then-else" decision rules. These rules are applicable at the early stages of core production, and allow possible corrective actions during the manufacturing process.

Neural networks (NNs) [8] as well as a hybrid decision tree-neural network approach [9,10] are used to predict iron losses at the early stages of transformer manufacturing. Each of neural networks is suited to a different environment. Selection of the most appropriate network (or equivalently environment) is based on the satisfaction of customers' requirements and several technical and economical criteria

[11]. The attributes selected by the decision trees are used as inputs to the neural networks.

The intelligent core loss model (i.e., the model of iron losses obtained through the neural network) is applied on-line to optimally combine the individual cores, in order to reduce the iron losses of assembled transformers. Unfortunately, for any practical number of cores, the complexity of an exhaustive search for their optimum grouping is such that a direct implementation would be practically unfeasible. For this reason, a genetic algorithm (GA) approach [12] is adopted in this paper.

In conclusion, in this paper an efficient computer-based environment is presented, providing on-line control of transformer core manufacturing process. The computer environment consists of data acquisition systems, automatic learning software, and statistical processing and graphical visualization toolbox. Results from the application of the proposed techniques on a transformer industry demonstrate the feasibility and practicality of this approach. Significant reduction of transformer iron losses is observed in comparison to the current practice.

II. OVER VIEW OF DECISION TREES

The Decision Tree methodology [13] is a non-parametric learning technique able to produce classifiers about a given problem in order to reduce information for new, unobserved cases. The DT is a tree structured upside down, built on the basis of a Learning Set (LS). The LS comprises a number of preclassified measurement sets (MS) defined by a list of candidate attributes.

The construction of a DT starts at the root node with the whole MS of the LS. These MS are analyzed in order to select the test that splits them "optimally" into a number of most "purified" subsets, which correspond to the successor nodes. For the sake of simplicity, a 2-class partition is considered in the following analysis, i.e. each MS is characterized as acceptable or as non-acceptable. The selection of the optimal test is based on maximizing the additional information gained through that test. A measure of the information provided by a test is based on the entropy of the examined subset. This procedure, known as the *optimal splitting rule*, is applied recursively to each new node, to build the corresponding subtrees.

In the following, the basic terms involved in the tree-building procedure are introduced and the algorithm is explained in detail.

Each node possesses a subset of MS with the following characteristics:

E_n : the MS subset of node n of the DT.

N : size (number of MS) of E_n .

n_A : number of acceptable MS in E_n .

n_{NA} : number of non-acceptable MS in E_n .

The relative frequencies of acceptable (f_A) and non-acceptable (f_{NA}) MS for node n will be:

$$f_A = \frac{n_A}{n_A + n_{NA}} = \frac{n_A}{N}, \quad (1a)$$

$$f_{NA} = \frac{n_{NA}}{n_A + n_{NA}} = \frac{n_{NA}}{N}. \quad (1b)$$

The entropy of E_n with respect to the class partition of its elements, is defined as:

$$H_C(E_n) = -(f_A \log_2 f_A + f_{NA} \log_2 f_{NA}). \quad (2)$$

$H_C(E_n)$ is a measure of the class-purity of the node subset E_n and, consequently, of the uncertainty of the classification of a state by this node. The following relations hold for $H_C(E_n)$:

$$0 \leq H_C(E_n) \leq 1, \quad (3a)$$

$$H_C(E_n) = 0 \Leftrightarrow (f_A = 1 \text{ or } f_{NA} = 1), \quad (3b)$$

$$H_C(E_n) = 1 \Leftrightarrow f_A = f_{NA} = 0.5. \quad (3c)$$

A test T is defined at node n as:

$$T: A_i \leq t, \quad (4)$$

where A_i is the value of attribute i of a particular MS and t is a threshold value.

By applying the test T to all MS of node n , E_n is split into two subsets E_{n1} and E_{n2} :

$$E_{n1} = \{MS \in E_n : A_i \leq t\}, \quad (5a)$$

$$E_{n2} = \{MS \in E_n : A_i > t\}. \quad (5b)$$

If n_i is the number of MS in E_{ni} , with $i=1,2$ then the corresponding frequencies are given by:

$$f_1 = \frac{n_1}{n_1 + n_2} = \frac{n_1}{N}, \quad (6a)$$

$$f_2 = \frac{n_2}{n_1 + n_2} = \frac{n_2}{N}. \quad (6b)$$

The entropy of E_n with respect to the partition induced by T is:

$$H_T(E_n) = -(f_1 \log_2 f_1 + f_2 \log_2 f_2), \quad (7)$$

where $H_T(E_n)$ is a measure of the uncertainty of the outcome of test T .

The mean conditional entropy of E_n , given the outcome of test, corresponds to the residual entropy after the application of T and is defined as:

$$H_C(E_n | T) = f_1 H_C(E_{n1}) + f_2 H_C(E_{n2}). \quad (8)$$

The information gained from the application of test T is expressed by the achieved reduction of the learning subset entropy:

$$I(E_n; T) = H_C(E_n) - H_C(E_n | T). \quad (9)$$

A more objective (less biased) estimator of the merit of test T is provided by the normalized information gain, defined as:

$$C(E_n; T) = \frac{2I(E_n; T)}{H_C(E_n) + H_T(E_n)} \in [0, 1]. \quad (10)$$

For each node of the DT, the *optimal splitting rule* consists of selecting the test which minimizes the classification uncertainty, i.e. the one with the highest information gain $C(E_n; T)$. The best test is obtained by sequential testing of all attributes and candidate thresholds and comparing their information gain.

In order to detect whether the calculated information gain for each test reflects an actual increase in information from the application of test T to unforeseen MS or it is simply apparent (i.e. a random effect, due to the limited size of the sample), a suitable statistical hypothesis test is used [5]. More specifically, under the hypothesis of no correlation between the test T and the class partition in the Universe U of the MS (resp. U_n), that is for zero actual increase in information, the random variable $N^*I(E_n; T)$, which is an estimator of the total actual information gain, is χ^2 -distributed with 1 degree of freedom and its expected value is positive and inversely proportional to the size of the subset E_n :

$$N^*I(E_n; T) \sim \chi^2(1). \quad (11)$$

If α is the risk level of not detecting situations of only apparent information gain and X_{cr} the value such that $P(X > X_{cr}) = \alpha$, where X a random variable following an χ^2 distribution with 1 degree of freedom, then the following statistical test can be formulated: "The node splitting test T is rejected as uncorrelated with the class partition if $QI = N^*I(E_n; T) < X_{cr}$ ", where N is the number of learning states in E_n .

In order to detect if a node is terminal, i.e. sufficiently "class-pure", the *stop splitting rule* is used, which compares the classification entropy of the node with a minimum preset value H_{min} . If it is lower than H_{min} , then the node is sufficiently class-pure and it is not further split. Such nodes are labeled LEAVES:

$$\text{If } H_C(E_n) < H_{min} \Rightarrow \text{LEAF}. \quad (12)$$

Otherwise (if $H_C(E_n) \geq H_{min}$), a suitable test is sought to divide the node, by applying the optimal splitting rule. In the case that no test can be found with a *statistically significant* information gain, the node is declared a DEADEND and it is not split.

The steps followed by the tree-building algorithm are summarized in the following:

1. The procedure starts from the top node with the whole LS.
2. It is examined if the node should be better split by applying the stop splitting rule:
 - If the node subset is sufficiently class-pure \Rightarrow LEAF
 - If not \Rightarrow proceed to Step 3.
3. Selection of the optimal test (optimal splitting rule).
4. Statistical significance testing of the optimal test.
 - If the measured information gain is statistically significant \Rightarrow Test Node. Proceed to Step 5.

- If there is no statistically significant way to expand the node \Rightarrow DEADEND.

5. Two successor nodes are created, using the optimal splitting test.

6. Steps 2-5 are recursively applied to the successor nodes.

DTs are tested using test sets (TS), comprising a number of similar, preclassified, but independent MS. The class of each of these MS is compared to the class of the terminal node finally led to by applying the tests of the various non-terminal nodes. This comparison provides the DT classification error rate (C_e).

III. FROM DECISION TREES TO NEURAL NETWORKS

A binary DT induces a hierarchical partitioning over the decision space. Starting with the root node, each internal (test) node partitions its associated decision region into two half spaces. It is obvious that all the conditions along any particular path from the root to the terminal node of the DT must be satisfied in order to reach the particular terminal node. Thus, each path of a DT implements an AND operation on a set of half spaces. If two or more terminal nodes result in the same class, then the corresponding paths are in an OR relationship.

Multi layer perceptrons are feedforward neural networks consisting of one input layer, one or more hidden layers and one output layer. A MLP with two hidden layers used for classification performs the following functions. The first hidden layer is the partitioning layer that divides the entire feature space into several regions. The second hidden layer is the ANDing layer that performs ANDing of partitioned regions to yield convex decision regions for each class. The output layer is the ORing layer that combines the results of the previous layer to produce disjoint regions of arbitrary shape.

It is concluded from the above that a DT and a four-layer perceptron are equivalent in terms of input-output mapping. In addition, a DT can be reformulated as a neural network, called entropy network (EN), by following the rules proposed in [9]:

- a. The *Input Layer (IL)* consists of one neuron per attribute selected and tested by the DT.
- b. The 1st hidden layer called *Partitioning or Test Layer (TL)* consists of one neuron per DT test node.
- c. The 2nd hidden layer called *ANDing Layer (AL)* consists of one neuron per DT terminal node.
- d. The *output layer or ORing Layer (OL)* consists of one neuron per DT class.

The connections between the neurons of the above four layers implement the hierarchy of the DT. In particular, each neuron of the TL is connected to the neuron of the IL corresponding to the tested attribute. In addition, each neuron of the AL is linked to the neurons of TL corresponding to the test nodes located on the path from the top node towards the

terminal node. Finally, each neuron of the output layer is connected to the neurons of AL corresponding to the DT terminal nodes. In comparison to the standard MLPs that are fully connected, the entropy network has fewer connections, or equivalently less number of parameters, reducing the time needed for training.

The entropy network can be used for classification, with some modifications however it can be also used for prediction [10]. In this case the OL is replaced by a single output neuron, fully connected to all neurons of the AL and the resulted network is retrained. This methodology is called hybrid DT - NN (HDTNN) approach. After HDTNN convergence, the network is used to predict the test states and after that to classify them accordingly, providing the so-called hybrid DT - NN classifier (HDTNNC).

IV. OVERVIEW OF GENETIC ALGORITHMS

GA is a robust optimization method that works above a set of candidate solutions named population and performs a number of operations based on genetics.

An initial population of k chromosomes $X(0)=(x_1, \dots, x_k)$ is randomly generated. The initial population $X(0)$ is used for the creation of new generation populations $X(n)$, $n>0$. The creation of $X(n)$ at generation (or GA cycle) n is performed by applying a set of operations on population $X(n-1)$, described below. This procedure is repeated in an iterative way, until $X(n)$ converges to an optimal solution of the problem.

For the transformer iron loss minimization problem, the total predicted iron losses P_i of the $L/2$ transformers are used as an objective function to estimate the performance of all chromosomes x_i , $i=1, \dots, k$ in a given population. However, a *fitness function* is used to map objective values to fitness values, following a *linear normalization scheme*. In particular, chromosomes x_i are ranked in ascending order of $P_i(x_i)$, since the objective function is to be minimized. Let $r(x_i) \in \{1, \dots, k\}$ be the rank of chromosome x_i , $i=1, \dots, k$. Defining an arbitrary fitness value F_B for the best chromosome, the fitness of the i -th chromosome is given by the linear function:

$$F(x_i) = F_B - [r(x_i) - 1]D, \quad i = 1, \dots, k, \quad (13)$$

where D is a decrement rate. Thus, the average objective value of the population is mapped into the average fitness [14]. After fitness values, $F(x_i)$, $i=1, \dots, k$, have been calculated for all members of the current population, *parent selection* is then applied so that a fitter chromosome gives a higher number of offspring and thus has a higher chance of survival in the next generation. A *proportionate scheme*, implemented by the *roulette wheel selection* procedure [15], is used for parent selection, ensuring that each chromosome has a growth rate proportional to its fitness value.

A set of new chromosomes (offspring) is then produced by mating the selected parent chromosomes and applying a

crossover operator. The genetic material of the parents is combined in a random way in order to produce the genetic material of offspring. *Mutation* is then applied to the newly created chromosomes, introducing random gene variations that are useful for restoring lost magnetic material, or for producing new material that corresponds to new search areas. A small mutation probability ensures that only a small gene proportion is altered in each generation.

Once new chromosomes have been generated for a given population $X(n)$, the next generation population, $X(n+1)$, is formed by inserting those new chromosomes into $X(n)$ and deleting an appropriate number of older chromosomes, so that each population consists of k members. The exact number, C , of old chromosomes to be replaced by new ones defines the *replacement strategy* of the GA and greatly affects its convergence rate. All of the above description refers to simple GA cycle. Several cycles need to take place, that is, several generations $X(n)$ need to be produced until the population converges to an optimal solution. For this reason, the procedures of fitness evaluation, parent selection, crossover and mutation are repeated until a termination criterion is satisfied. Usually the GA terminates when the best chromosome fitness remains constant for a large number of generations, indicating that further optimization is unlikely.

V. AUTOMATIC LEARNING TECHNIQUES AND TRANSFORMER CORE LOSSES

A. Problem Description

Fig. 1 shows the assembled active part of a wound core distribution transformer. It can be seen that two small individual cores (width of core window equal to $F1$) and two large individual cores (width of core window equal to $F2$) need to be assembled. In general, the width $F2$ is twice that of $F1$.

As it is observed from Fig. 1 the four cores are arranged in space as (from left to right): a small core, followed by a large, followed by another large and finally followed by a small one. Let us denote as "11" and "12" the left small and large core respectively, while as "13" and "14" the other two cores. Thus, the core arrangement from left to right will be "11"- "12"- "13"- "14", as depicted in Fig. 1.

Typically, transformer iron losses depend upon the grade of steel, its thickness, current frequency, magnetic flux density and weight. These factors are taken into account during the transformer design stage. A number of additional factors affect iron losses during manufacturing, such as the kind of lamination insulation, annealing, core construction, quality of assembly, etc. However, it is not possible to consider all these factors analytically and, therefore, the calculations are based on graphs and tables obtained from past measurements on actual transformers. The basic data taken from these tables are updated by coefficients that

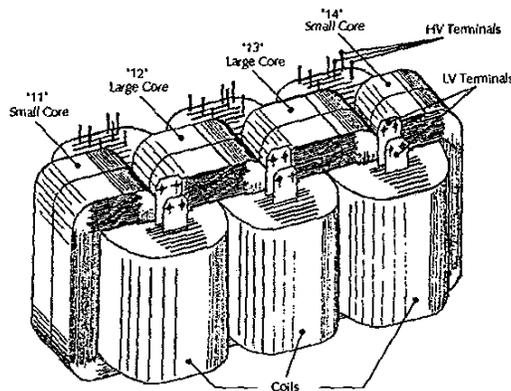


Fig. 1. Assembled active part of wound core distribution transformer.

account for the specific features of the magnetic core design and technology of core production.

Indicative loss curves traditionally used to estimate iron losses of individual cores and of assembled transformers are shown in Fig. 2. Using these loss curves, only the influence of the rated magnetic induction on iron losses for each specific magnetic material is considered.

It should be noted that during transformer construction actual weights and losses of individual cores diverge from the theoretical ones. Although these deviations are within predictable statistical limits, they cause variations in the iron losses of assembled transformers.

The conventional technique used to reduce the variation in iron losses of assembled transformers is to pre-measure and assign a grade (quality category) to each individual core and then combine higher and lower graded individual cores to achieve an "average" value for the entire transformer. This is referred to as conventional grouping process.

In our method, automatic learning techniques are applied in order to obtain an iron loss prediction model. The intelligent model is used on-line, aiming at the minimization of iron losses of assembled transformers.

B. Representation

The *representation* is the first step in the application of automatic learning to a given practical problem [5]. It consists of (a) choosing appropriate input attributes to represent the problem, (b) defining the output information, and (c) choosing the suitable model to represent input/output relationships.

For the transformer iron loss problem, a large set of candidate attributes is initially chosen. The attributes, i.e. the parameters affecting transformer iron losses, have been selected based on extensive research and transformer designers' experience. These attributes are calculated as a combination between measurements taken during the first

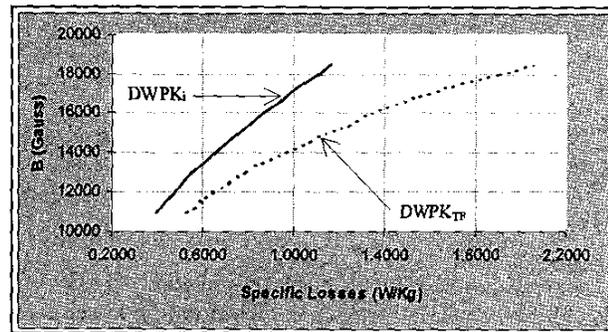


Fig. 2. Typical loss curve.

stages of core production and values theoretically evaluated during the transformer design phase. To define the selected attributes, a brief presentation of the involved parameters is follows.

The transformer designer calculates the following parameters:

$DWPK_i$: Theoretical specific iron losses (W/Kg) of the i -th individual core, $i="11", \dots, "14"$.

$DWPK_{TF}$: Theoretical specific iron losses (W/Kg) of the assembled transformer.

DKg_i : Theoretical weight of the i -th core.

The transformer designer uses the loss curve in order to calculate the theoretical specific iron losses of the individual core and of the assembled transformer. Each loss curve corresponds to a specific supplier, grade and thickness of magnetic material.

The theoretical iron losses, say $DSFL_i$, of the i -th core are given by:

$$DSFL_i = DWPK_i * DKg_i, \quad i="11", \dots, "14". \quad (14)$$

The theoretical total iron losses of the four individual cores of the transformer are:

$$DSFL_{TF} = DSFL_{11} + DSFL_{12} + DSFL_{13} + DSFL_{14}. \quad (15)$$

The theoretical weight of the transformer is equal to the sum of the theoretical weights of its four individual cores:

$$DKg_{TF} = DKg_{11} + DKg_{12} + DKg_{13} + DKg_{14}. \quad (16)$$

The theoretical iron losses of the assembled transformer are calculated as follows:

$$DNLL_{TF} = DWPK_{TF} * DKg_{TF}. \quad (17)$$

After the production and quality control of individual cores the following parameters are known from measurements:

$WPK_{i,mat,a}$: Specific iron losses (W/Kg at 15000 Gauss) of magnetic material of the i -th core.

$WPK_{i,mat,b}$: Specific iron losses (W/Kg at 17000 Gauss) of magnetic material of the i -th core.

AKg_i : Actual weight of the i -th core.

$ASFL_i$: Actual iron losses of the i -th core.

The actual total iron losses, say $ASFL_{TF}$, of the four individual cores are:

$$ASFL_{TF} = ASFL_{11''} + ASFL_{12''} + ASFL_{13''} + ASFL_{14''}. \quad (18)$$

The actual weight, say AKg_{TF} , of the transformer is:

$$AKg_{TF} = AKg_{11''} + AKg_{12''} + AKg_{13''} + AKg_{14''}. \quad (19)$$

The actual specific iron losses of the i -th individual core are, say $AWPK_i$, are given by:

$$AWPK_i = \frac{ASFL_i}{AKg_i}, \quad i = "11", \dots, "14" \quad (20)$$

Table I shows the candidate attributes, initially chosen. As output information, the transformer specific iron losses (i.e., iron losses per weight unit) are selected. The learning sets are composed of sets of input/output pairs corresponding to all possible transformer constructions. Each of the learning sets is suited to a different condition (environment), i.e. to a certain supplier, grade and thickness of magnetic material. In the following, three different environments are considered, each defined in Table II. For example, the environment #1 is characterized by magnetic material of grade M3, according to USA AISI 1983, thickness 0.23 mm, while the supplier of material was SUP_A (Supplier A).

The final step in the *representation* of the transformer iron losses consists of the choice of an appropriate model (or models). The selection is based on trial and error. In the following, decision trees, neural networks, and combined use of decision trees and neural networks are examined and compared.

TABLE I
INITIALLY SELECTED CANDIDATE ATTRIBUTES

Symbol	Expression
ATTR ₁	Rated magnetic induction
ATTR ₂	$(WPK_{11'',mat,a} + WPK_{12'',mat,a} + WPK_{13'',mat,a} + WPK_{14'',mat,a})/4$
ATTR ₃	$(WPK_{11'',mat,b} + WPK_{12'',mat,b} + WPK_{13'',mat,b} + WPK_{14'',mat,b})/4$
ATTR ₄	AKg_{TF}/DKg_{TF}
ATTR ₅	$ASFL_{TF}/DSFL_{TF}$
ATTR ₆	$AWPK_{11''}/DWPK_{11''}$
ATTR ₇	$AWPK_{12''}/DWPK_{12''}$
ATTR ₈	$AWPK_{13''}/DWPK_{13''}$
ATTR ₉	$AWPK_{14''}/DWPK_{14''}$
ATTR ₁₀	$AKg_{11''}/DKg_{11''}$
ATTR ₁₁	$AKg_{12''}/DKg_{12''}$
ATTR ₁₂	$AKg_{13''}/DKg_{13''}$
ATTR ₁₃	$AKg_{14''}/DKg_{14''}$
ATTR ₁₄	$(AWPK_{11''} + AWPK_{12''})/(DWPK_{11''} + DWPK_{12''})$
ATTR ₁₅	$(AWPK_{12''} + AWPK_{13''})/(DWPK_{12''} + DWPK_{13''})$
ATTR ₁₆	$(AWPK_{13''} + AWPK_{14''})/(DWPK_{13''} + DWPK_{14''})$
ATTR ₁₇	$(AWPK_{11''} + AWPK_{12''})/(AWPK_{12''} + AWPK_{13''})$
ATTR ₁₈	$(AWPK_{12''} + AWPK_{13''})/(AWPK_{13''} + AWPK_{14''})$
ATTR ₁₉	$(AWPK_{12''} + AWPK_{13''})/(AWPK_{11''} + AWPK_{14''})$

TABLE II
ENVIRONMENTS CONSIDERED

Characteristic	Environment		
	#1	#2	#3
Supplier	SUP_A	SUP_B	SUP_A
Steel grade	M3	M4	Hi-B
Thickness (mm)	0.23	0.27	0.23

VI. APPLICATION OF DECISION TREES

Decision trees are applied for attribute selection as well as for classification of specific iron losses into acceptable or non-acceptable.

A. Iron Loss Classification

The criterion for classifying transformer iron losses as non-acceptable is based on the comparison of the actual specific iron losses to the theoretical expected (designed) specific iron losses.

In Fig. 3 a characteristic decision tree is illustrated, developed with the 19-attribute list (ATTR₁ + ATTR₁₉ of Table I) and 0.999 confidence level. The DT is constructed from the 1590 MS of the learning set of the environment #2. Its classification error rate, tested with the 760 MS of the test set, is 4.2%.

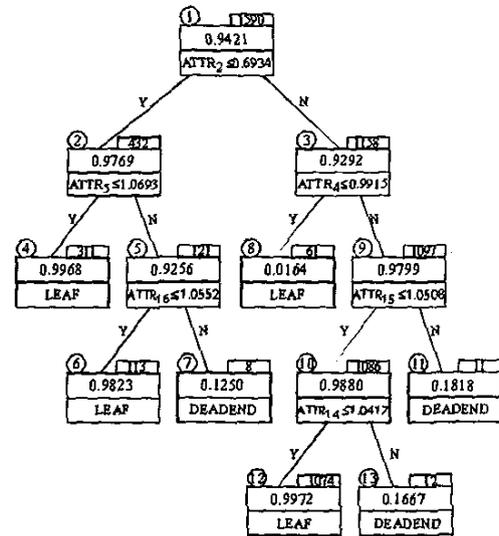


Fig. 3. DT developed using the 19-attribute set.

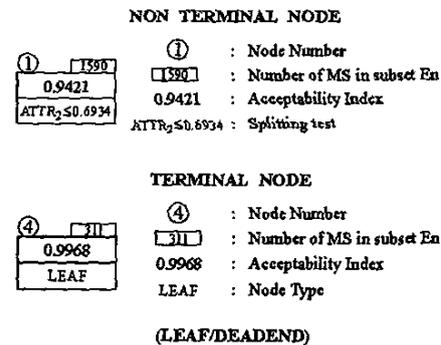


Fig. 4. Notation of the DT's nodes.

The notation used for the DT nodes is explained in Fig. 4. The acceptability index of a node is defined as the ratio of the acceptable MS in the subset E_n of node n to the total number of MS in E_n . For example, among the 1590 MS of node 1, the 1498 MS (94.21%, i.e. acceptability index of 0.9421) are acceptable, while the rest 92 are non-acceptable.

The terminal nodes correspond to a class label (acceptable or non-acceptable) that can be used to classify any measurement set, either belonging to the learning set or completely new one. The class of a terminal node is assigned, using its acceptability index. For example, since node 4 has an acceptability index of 0.9968, then the MS "falling" to this node have a 99.68% probability of being acceptable.

The DT of Fig. 3 consists of 6 test and 7 terminal nodes, and has automatically selected only 6 attributes among the 19 candidate ones of Table I. These attributes in decreasing order of significance are $ATTR_2$, $ATTR_5$, $ATTR_4$, $ATTR_{16}$, $ATTR_{15}$ and $ATTR_{14}$. $ATTR_2$ corresponds to the average specific iron losses of magnetic material of the four individual cores, $ATTR_5$ is the ratio of actual over theoretical total iron losses of the four individual cores, and $ATTR_4$ represents the ratio of actual over theoretical weight of the four individual cores. $ATTR_{16}$ is the ratio of actual over theoretical sum of iron losses of individual cores "13" and "14". Finally, $ATTR_{15}$ and $ATTR_{14}$ are defined in the same way as $ATTR_{16}$, but they correspond to the sum of iron losses of cores "12" and "13", and "11" and "12", respectively. The selection of these attributes is reasonable and expected, since they are all important to the transformer iron losses.

Each terminal node produces one decision rule, on the basis of its acceptability index. For example, from terminal node 6 the following rule is derived: *if $ATTR_2 \leq 0.6934$ and $ATTR_5 > 1.0693$ and $ATTR_{16} \leq 1.0552$, then transformer specific iron losses are of acceptable quality.* Consequently, based on the decision tree of Fig. 3, rules useful for the core production can be derived.

It is desirable to construct transformers leading to nodes 4, 6, and 12, if it technically and economically feasible. These nodes have acceptability indices greater than 98%. On the other hand, construction of transformers with MS "falling" to nodes 7, 8, 11, and 13 must be avoided, since these nodes correspond to non-acceptable quality.

The measurement sets following the rule $ATTR_2 > 0.6934$ and $ATTR_4 \leq 0.9915$ are lead to node 8, and characterized as non-acceptable. In order to avoid this, one way is not to permit $ATTR_4$ to be lower than or equal 0.9915. The method is to produce cores with actual weight not smaller than 0.85% in relation with their theoretical weight.

The measurement sets following the rule $ATTR_2 \leq 0.6934$ and $ATTR_5 \leq 1.0693$ are lead to node 4, and characterized as acceptable. Assuming that the quality of the magnetic material used for core construction satisfies the test $ATTR_2 \leq 0.6934$, then in order to lead to node 4 we must keep $ATTR_5 \leq 1.0693$. This is equivalent to take care that the sum of

the actual iron losses of the four individual cores not to be greater than 6.93% in relation with the sum of their theoretical iron losses.

B. Feature Selection

Feature selection aims at reducing the dimensionality of the input space by dismissing attributes, which do not carry useful information to predict transformer iron losses.

The task of deciding which of the candidate attributes are the most important is an arduous task. The DT technique provides not only attribute selection, but also attribute ranking in the sense that it assigns an information quantity to each attribute [5].

Given a transformer acceptability criterion (i.e., type and threshold value), a DT is built to classify samples and to automatically identify attributes relevant for classification, with respect to this particular criterion.

Two different acceptability criteria are used for the classification of iron losses. According to the first criterion, one transformer is acceptable, if its actual specific iron losses are not greater than $Lim1\%$ of the theoretical specific iron losses (given by the loss curve). According to the second, one transformer is acceptable, if its actual specific iron losses are in the range $\pm Lim2\%$ of the theoretical losses. Parameters $Lim1\%$ and $Lim2\%$ (i.e., threshold values of the transformer acceptability criteria) are defined by the transformer designer in accordance with customer requirements. The above acceptability criteria correspond to alternative customer requirements.

Several DTs were built for various scenarios: two different suppliers of magnetic material, three different grades of magnetic material, two acceptability criteria, and 10 different values for parameters $Lim1\%$ and $Lim2\%$. The DTs were evaluated on the basis of independent test sets. In our investigation only DTs with high acceptability success rates (more than 90%) are taken into consideration. Table III shows the 8 attributes that appear in the test nodes of the resulting decision trees, i.e. automatically selected among the 19 candidate attributes of Table I. The quantitative improvements achieved through the attribute selection process are proved by the reduction of the average absolute relative error in the prediction of iron losses and are described in the following sections.

TABLE III
LIST OF THE MOST SIGNIFICANT ATTRIBUTES

Symbol	Expression
I_1	Rated magnetic induction
I_2	$(WPK_{11,mat,a} + WPK_{12,mat,a} + WPK_{13,mat,a} + WPK_{14,mat,a})/4$
I_3	$(WPK_{11,mat,b} + WPK_{12,mat,b} + WPK_{13,mat,b} + WPK_{14,mat,b})/4$
I_4	AK_{gTF}/DK_{gTF}
I_5	$ASFL_{TF}/DSFL_{TF}$
I_6	$(AWPK_{11} + AWPK_{12})/(DWPk_{11} + DWPk_{12})$
I_7	$(AWPK_{12} + AWPK_{13})/(DWPk_{12} + DWPk_{13})$
I_8	$(AWPK_{13} + AWPK_{14})/(DWPk_{13} + DWPk_{14})$

VII. APPLICATION OF NEURAL NETWORKS

In this section results from the application of different NN structures for classifying and predicting transformer specific iron losses are presented. Prediction aims at estimating the actual specific iron losses, while classification at categorizing the iron losses into acceptable or non-acceptable.

A. Classification Problem

For the classification problem several neural network architectures are tested and compared. More specifically, the fully connected MLP is compared with the entropy network that has fewer connections. Moreover, architectures with one or two hidden layers, with one or two output neurons, and with a variable number of input neurons are tested.

Based on the decision tree of Fig. 3 and the methodology described in Section III, the EN of Fig. 5 can be derived. The EN is composed of 6 input, 6 test, 7 ANDing and 2 ORing (output) neurons. The output information is a two-class classification, i.e., acceptable (A) and non-acceptable (NA) transformers with respect to the DT acceptability criterion considered. The correspondence between the DT nodes and the EN neurons is described in Table IV.

After adaptation of its weights, the EN provides a classification error rate, C_e , of 3.4%. Furthermore, the output layer of the EN is replaced by a single neuron representing transformer specific iron losses and the HDTNN approach is applied. After training and convergence, the NN is used to predict the transformer specific iron losses of the TS and classify them accordingly to the criterion used for DT building. The HDTNNC significantly decreases the C_e to 2.2%.

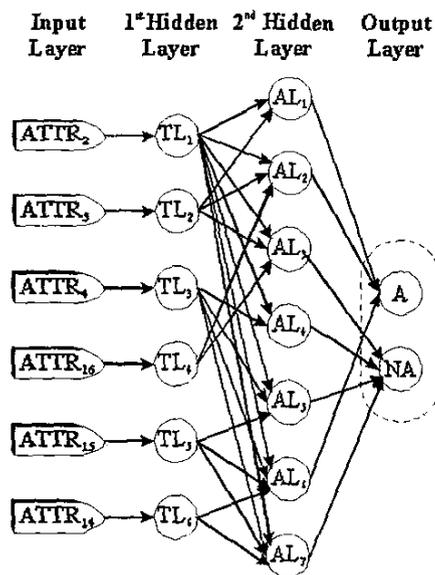


Fig. 5. Entropy network for the DT of Fig. 3.

TABLE IV. CORRESPONDENCE BETWEEN DT NODES AND EN NEURONS

DT	1	2	3	4	5	6	7
EN	TL ₁	TL ₂	TL ₃	AL ₁	TL ₄	AL ₂	AL ₃
DT	8	9	10	11	12	13	
EN	AL ₄	TL ₅	TL ₆	AL ₅	AL ₆	AL ₇	

TABLE V. COMPARING METHODS FOR CLASSIFICATION OF TRANSFORMER IRON LOSSES

Method	Structure	C_e (%)
DT	-	4.2
EN	6-6-7-2	3.4
HDTNNC	6-6-7-1	2.2
MLP (19 attributes)	19-5-2	1.4
MLP (6 DT attributes)	6-9-2	3.2

Moreover, two fully connected MLPs were constructed for the same classification problem. The first MLP comprises 19 input neurons corresponding to the candidate attributes of Table I, while the second comprises the 6 attributes selected by the decision tree of Fig. 3. Both MLPs have only one single hidden layer and two output neurons corresponding to the acceptable and non-acceptable transformers. The first MLP comprises 5 hidden neurons, i.e. a 19-5-2 structure, and presents a C_e of 1.4%. The second MLP has a 6-9-2 structure and a 3.2% classification error rate.

Table V summarizes the results of classification of transformer iron losses. The EN, which is derived by translating the decision tree structure and the second MLP with the 6 attributes identified by the tree, provide very similar classification results. The HDTNNC is more accurate than the EN and the second MLP. The first MLP with the 19 attributes provides the best classification results. Concerning training computational performance, decision trees are by far the fastest method, while among the different neural network approaches the slowest method corresponds to the fully connected MLP.

B. Prediction Problem

In order to predict transformer iron losses, multilayer feedforward neural networks with one output (i.e., transformer specific iron losses) are used. In all cases examined, the fully connected MLP is adopted. The activation functions of all neurons are the sigmoid function.

The Average Absolute Relative Error (AARE), used to evaluate the network performance, is defined as:

$$AARE = \frac{1}{N} \sum_{\text{forall } N \text{ sample}} \frac{|\text{actual losses} - \text{predicted losses}|}{\text{actual losses}} * 100\% \quad (21)$$

Initially a fully connected MLP with the same number of neurons per layer (i.e., 6-6-7-1) with the EN of Fig. 5 is used. The trained MLP presents an AARE of 2.7%. Similar results are obtained with other tested MLPs with two hidden layers. On the other hand, in case of one hidden layer the results are better. The best number of neurons of the hidden layer is

TABLE VI. COMPARING THE PERFORMANCE OF DIFFERENT NEURAL NETWORK STRUCTURES

Structure	AARE (%)
6-6-7-1	2.7
19-5-1	1.8
6-5-1	1.9
8-5-1	1.7

TABLE VII. COMPARING THE LOSS CURVE AND THE MLP IN PREDICTING TRANSFORMER IRON LOSSES

Method	Environment		
	#1	#2	#3
Loss curve	2.9	3.1	3.3
MLP (8-5-1)	1.5	1.7	1.8

selected by trial and error, for the given environment. In particular, for the #2 environment, one hidden layer consisting of a small number of neurons (5 neurons) is found to be completely adequate.

In addition to the 6-6-7-1 MLP, the prediction performance of three different MLPs with one hidden layer and one with two hidden layers is presented in Table VI. The MLPs correspond to the #2 environment. In case of one hidden layer, the number of input neurons is selected to be 19 (corresponding to the attributes of Table I), or 8 (corresponding to the attributes of Table III), or 6 (i.e., the attributes selected by the DT of Fig. 3). The best results are obtained if the 8-5-1 MLP structure is followed. Similar results are obtained for the other environments. This result shows the advantages of the attribute selection process using decision trees.

Moreover, the performance of the 8-5-1 MLP structure is compared with that obtained by the current practice of using the typical loss curve. Table VII presents the average absolute relative error on test set, for the three environments considered, following the current practice (loss curve) as well as the proposed method (neural network). In all cases, the neural network method provides an improved accuracy by more than 45%.

VIII. APPLICATION OF GENETIC ALGORITHMS

A. Genetic Algorithm for Iron Loss Reduction

The intelligent core loss model is exploited on-line in order to reduce iron losses of assembled transformers. More specifically, the accurate prediction of iron losses, provided by the neural networks, is used to improve the grouping process. Thus, assuming that an even number of L small cores and L large cores is available, then $L/2$ transformers can be assembled. Each transformer has four positions where cores can be placed. The two outer positions are occupied by small cores while the other two middle positions are occupied by

large cores (Fig. 1). Each small core can be put to any of the two positions and to any of the $L/2$ transformers. The same assumption exists for each large core. From all possible combinations of grouping $L/2$ transformers, only one combination, providing the optimum iron loss performance, should be selected.

The neural networks can be used in order to minimize (reduce) transformer iron losses, as follows:

- An initial arrangement of $L/2$ transformers is generated by randomly selecting the position of L small and L large cores. For each of the $L/2$ transformers the neural network inputs (attributes) are calculated. Using the neural network weights and thresholds the network output (i.e., the specific iron losses) is estimated, for each of the $L/2$ transformers. After that, the iron losses of each transformer are predicted by multiplying the neural network output with the respective actual transformer weight. Next, the total predicted iron losses, say P_t , of the $L/2$ transformers are evaluated by adding their predicted iron losses. Finally, the arrangement of cores for all the $L/2$ transformers together with the total predicted iron losses, P_t , are stored in a solution table.
- Step a) is repeated for all the different combinations of the $L/2$ transformers.
- From the solution table, the combination having the smaller total predicted iron loss, P_t , is selected. For the selected solution, the arrangement of cores, for all the $L/2$ transformers, is retrieved. This is the solution of the optimization problem.

The effectiveness of this optimization process is strongly reduced, as the number of individual cores increases. For example for $L=6$, the combinations of grouping the $L/2$ transformers are 1800, while for $L=48$ the combinations are approximately $4 \cdot 10^{13}$. For this reason a genetic algorithm approach is adopted for the grouping task. Fig. 6 shows the representation scheme for the arrangement of cores in a transformer. According to this scheme, integer numbers are used for the representation of genes (cores). For example, assuming that 6 small cores and 6 large cores are going to be grouped (Fig. 6), then the integer numbers between 1 and 6

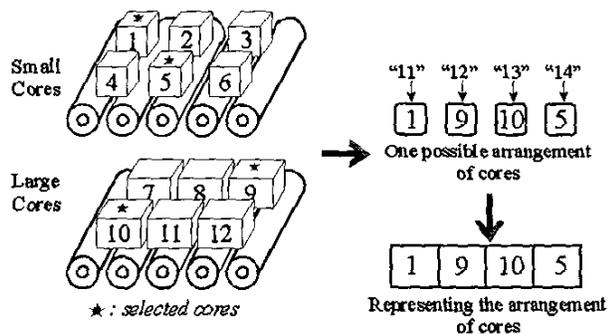


Fig. 6. Representing the arrangement of cores in a transformer.

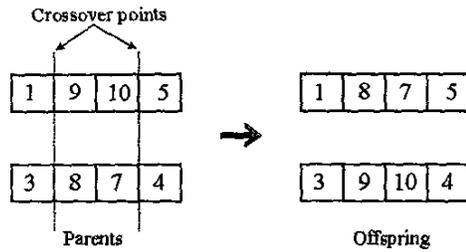


Fig. 7. Example of the two-point crossover operator.

correspond to the small cores, while the consecutive numbers between 7 and 12 represent the large cores. Four genes are used in order to represent the arrangement of cores in one transformer. If the small cores 1 and 4, and the large cores 9 and 10 are selected for the first of the 3 transformers, then one possible arrangement of cores, from left to right, is core 1 is followed by core 9, followed by core 10, and followed by core 5. Then, the string {1,9,10,5} represents the arrangement of cores of the first transformer. The arrangement of cores for the second and the third transformer are generated accordingly. Joining the 3 strings produces the chromosome, i.e. one possible solution of the optimization problem. In general, one chromosome contains $2*L$ genes.

An initial population of k chromosomes $X(0)=(x_1, \dots, x_k)$ is generated by selecting sets of arrangements whose genes correspond to "average" iron losses, according to the conventional grouping process. Traditionally, initial populations are randomly generated, but the above approach increases the possibility of locating sets of arrangements with low losses within the first few GA cycles.

Fig. 7 depicts an example of the crossover operator with two crossover points used for exchanging genes.

B. Experimental Results

The proposed genetic algorithm based grouping process was used in order to group 100 small and 100 large cores of the same production batch of 50 transformers, 160 kVA, 50 Hz.

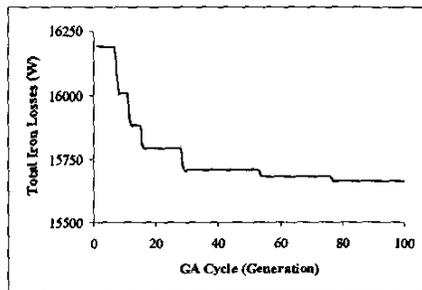


Fig. 8. Genetic algorithm convergence: total iron losses versus the GA cycle.

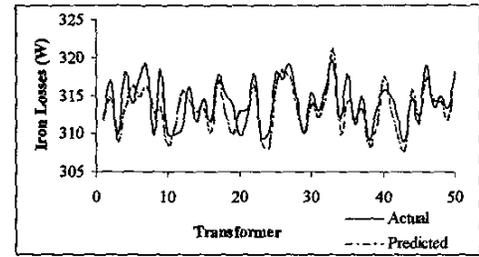


Fig. 9. Evaluating the performance of the genetic algorithm.

Fig. 8 shows the minimum value, over the whole population, of the objective function (or total predicted iron losses $P_i(x_i)$) versus the cycle (or generation) of the genetic algorithm. $P_i(x_i)$ decreases as the GA cycle increases, until it reaches a minimum value of 15664 W at generation 85. Since in the specific experiment half chromosomes are replaced by new ones at each generation ($k=80$ and $C=k/2=40$ have been used), there are cases where all generated offspring have lower fitness than their parents. In these cases the value of the $P_i(x_i)$ remains at the same level, hence the "stepwise" appearance of the curve in Fig. 8. Note that the step "width" increases with the GA cycle, since it is directly related to the probability of further optimization.

The output of the genetic algorithm based grouping process is not only the minimum value of the total predicted iron losses (i.e., $P_i(x_i)=15664$ W for the example of Fig. 8) of the $L/2$ transformers. It also provides for each one of the $L/2$ transformers: (a) the arrangement of the four cores and (b) the predicted iron losses.

Fig. 9 evaluates the performance of the genetic algorithm comparing the predicted to the actual iron losses (measured after transformer construction) for each one of the 50 transformers, for the example of Fig. 8. For the specific example, the *AARE* is 0.57%. The proposed grouping process is tested for all the environments in various production batches providing an *AARE*, smaller than 1.60% for all the produced transformers. This is compared with an *AARE* of 3.15% in prediction of transformer iron losses, usually observed by the conventional grouping process.

The genetic algorithm based grouping process has been implemented through software (GA toolbox) and is currently used in the considered industrial environment. Using appropriate data acquisition systems, measurements are collected and fed to the GA toolbox as well as to a statistical processing and graphical visualization toolbox. The combination of these two systems provides on-line control and optimization of transformer core manufacturing process.

IX. CONCLUSIONS

In this article, automatic learning techniques are used for the prediction of iron losses at the transformer production phase as well as for the reduction of iron losses of the assembled transformers. In particular, three different methods are combined: decision trees, neural networks and genetic algorithms. The goal of decision trees is to investigate the more important attributes and their effect on iron losses as well as to produce decision rules useful for the production of transformers. The neural networks are used to provide satisfactory prediction of iron losses for all the environments considered. Finally, the genetic algorithms are used in combination with neural networks in order to reduce the transformer iron losses. Application of the proposed artificial intelligence framework to transformer manufacturing industry has verified the accurate prediction of iron losses in all the examined environments. Moreover, reduction of the transformer losses, given the individual cores, is achieved.

ACKNOWLEDGMENTS

The authors would like to thank the General Secretariat of Research and Technology of Greece for financing the research programme entitled: "Development of a Quality Control System of Distribution Transformer Magnetic Cores Using Artificial Intelligence Techniques" within the YPER '94 Research Programme. Help from Schneider Electric AE staff in collecting industrial measurements is gratefully acknowledged.

REFERENCES

- [1] S. Austen Stigant, and A.C. Franklin, *The J&P Transformer Book*. Boston: Newnes-Butterworths, 1973.
- [2] B.W. McConnell, "Increasing Distribution Transformer Efficiency: Potential for Energy Savings," *IEEE Power Engineering Review*, vol. 18, no. 7, pp. 8-10, July 1998.
- [3] R.S. Girgis, E.G. teNijenhuis, K. Gramm, and J.E. Wrethag, "Experimental Investigations on Effect of Core Production Attributes on Transformer Core Loss Performance," *IEEE Trans. Power Delivery*, vol. 13, no. 2, pp. 526-531, April 1998.
- [4] G.F. Mechler, and R.S. Girgis, "Calculation of Spatial Loss Distribution in Stacked Power and Distribution Transformer Cores," *IEEE Trans. Power Delivery*, vol. 13, no. 2, pp. 532-537, April 1998.
- [5] L. Wehenkel, *Automatic learning techniques in power systems*. Boston: Kluwer Academic, 1998.
- [6] L. Wehenkel, Th. Van Cutsem, and M. Ribbens-Pavella, "An Artificial Intelligence Framework for On-Line Transient Stability Assessment of Power Systems," *IEEE Trans. Power Systems*, vol. 4, no.2, pp. 789-800, May 1989.
- [7] N.D. Hatziaargyriou, G.C. Contaxis, and N.C. Sideris, "A Decision Tree Method for On-Line Steady State Security Assessment," *IEEE Trans. Power Systems*, vol. 9, no. 2, pp. 1052-1061, May 1994.
- [8] S. Mark Halpin, and Reuben F. Burch, IV, "Applicability of Neural Networks to Industrial and Commercial Power Systems: A Tutorial Overview," *IEEE Trans. Ind. Applications*, vol. 33, no. 5, pp. 1355-1361, Sep./Oct. 1997.
- [9] I. K. Sethi, "Entropy Nets: From Decision Trees to Neural Networks", *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1605-1613, October 1990.
- [10] L. Wehenkel, and V.B. Akella, "A hybrid decision tree - neural network approach for power system dynamic security assessment," *In Proc. of the 4th Int. Symp. on Expert Systems Application to Power Systems*, pp. 397-411, Elsevier-North Holland, 1993.
- [11] P.S. Georgilakis, N.D. Hatziaargyriou, S.S. Elefsiniotis, D.G. Pappas, and J.A. Bakopoulos, "Creation of an Efficient Computer-Based Environment for the Reduction of Transformer Industrial Cycle," *Proc. of the 1st IEE Med. Power '98 Conference*, Nicosia, Cyprus, 1998.
- [12] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley, 1989.
- [13] L. Wehenkel, and M. Pavella, "Decision Trees and Transient Stability of Electric Power Systems", *Automatica*, vol. 27, no.1, pp. 115-134, 1991.
- [14] K.S. Tang, K.F. Man, S. Kwong, and Q. He, "Genetic Algorithms and their Applications," *IEEE Signal Processing Magazine*, pp. 22-37, Nov. 1996.
- [15] H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor: The University of Michigan Press, 1975.



Paul S. Georgilakis was born in Chania, Greece in 1967. He received the Diploma in Electrical and Computer Engineering from the National Technical University of Athens (NTUA), Greece in 1990. In 1994 he joined the Schneider Electric AE, Greece. He has worked in the Development and also the Quality Control Departments of the Industrial Division of the company. He is currently an associate of Technical Division of Schneider Electric AE. He is also working towards his Ph.D. thesis at NTUA. His research deals with Transformer Quality Improvement and application of Artificial Intelligence Techniques to Transformer Design. He is a student member of IEEE and member of CIGRE and the Technical Chamber of Greece.



Nikos D. Hatziaargyriou was born in Athens, Greece, in 1954. He received the Diploma in Electrical and Mechanical Engineering from the National Technical University of Athens (NTUA), Greece in 1976 and M.Sc. and Ph.D. degrees from the University of Manchester Institute of Science and Technology (UMIST), Manchester, England in 1979 and 1982, respectively. He is currently Professor at the Power Division of the Electrical and Computer Engineering Department of NTUA. His research interests include Modeling and Digital Techniques for Power System Analysis and Control. Dr. Hatziaargyriou is a senior member of IEEE and member of CIGRE SC38 and the Technical Chamber of Greece.



Dimitrios G. Pappas was born in Komotini, Greece in 1945. He received the Diploma in Electric Machines & Instruments from the Moscow Energy Institute, U.S.S.R in 1972. From 1974 to 1976 he worked in the Engineering Department of Masina Xrisolouris, Greece. Since 1976, he has been with Schneider Electric AE (former ELVIM), Greece. He has been engaged in several positions, such as Quality Control Manager, Production Manager, Transformer Design Manager and Quality Assurance Director. At present, he is Technical Division Director and Industrial Division Director.



John A. Bakopoulos was born in Athens, Greece, in 1966. He received the Diploma in Electrical and Computer Engineering from the National Technical University of Athens (NTUA), Greece in 1988. He joined Schneider Electric AE in 1989 and has been engaged in different positions such as Technical, Commercial, Development and Projects Departments. From 1994 to 1997 he was Quality Assurance Manager. At present he is Production Manager of Elvim factory producing

Distribution Transformers, LV and MV Panels.



Spiros S. Elefsiniotis was born in Piraeus, Greece in 1967. He received the Diploma in Production & Management Engineering from the Technical University of Crete (TUC), Greece in 1990. He joined Schneider Electric AE in 1990. He has been engaged in several positions, such as Computer Aided Design Engineer (one year), Development Engineer (one year), Inventory Control Manager (three years) and Technical Methods Manager (one year). He is currently Quality Assurance Manager. He is

member of Technical Chamber of Greece (TEE) and European Federation of National Engineering Associations (FEANI).