

*This paper reports the construction of an effective mechanism for synthesizing classification trees according to the fixed initial information in the form of a training sample for the task of recognizing the current state, as well as flood phenomena, of river basins. The built algorithmic classification tree could unmistakably categorize the entire training sample underlying the constructed classification scheme. Moreover, it would demonstrate minimal structural complexity by including components such as the algorithms for autonomous classification and recognition to serve the structure's vertices. The devised method for building the models of algorithms' trees makes it possible to operate training samples composed of a large amount of diverse information of discrete type. It ensures high model accuracy, the rational utilization of the system's hardware resources in the process that generates the final classification scheme, thereby making it possible to build models with predetermined accuracy. The proposed approach to synthesizing the new recognition algorithms is based on a library of already known algorithms and methods. Based on the proposed concept of algorithmic classification trees, a set of models was built that ensured effective categorization and prediction of flood-related events across the Tisza river basin. The proposed indicators of data generalization and quality of the classification tree model make it possible to effectively represent the general characteristics of the model allowing their application to select the optimal algorithm tree from a set of random classification tree methods. The classification trees built have ensured the absence of errors on the data of the training and test sample and have confirmed the efficiency of the approach of algorithm trees*

**Keywords:** *classification model, discrete object, algorithmic classification tree, generalized feature*

UDC 004.8: 004.89: 519.7

DOI: 10.15587/1729-4061.2020.219525

# CLASSIFICATION MODELS OF FLOOD-RELATED EVENTS BASED ON ALGORITHM TREES

**I. Povkhan**

PhD, Associate Professor  
Department of Software Systems  
Uzhhorod National University  
Narodna sq., 3, Uzhhorod,  
Ukraine, 88000  
E-mail: igor.povkhan@uzhnu.edu.ua

Received date 10.10.2020

Accepted date 20.11.2020

Published date 25.12.2020

Copyright © 2020, I. Povkhan

This is an open access article under the CC BY license  
(<http://creativecommons.org/licenses/by/4.0>)

## 1. Introduction

The tasks that are related to classification are very diverse and are tackled in the modern world in all sectors of economics and the social context of human activity. This necessitates building and investigating mathematical models of relevant systems. Currently, there is no universal approach to solving them; several general theories and approaches have been proposed to solve different types of problems. However, a large number of applied tasks, in different areas of natural science where the categorization problems are addressed using software and hardware systems, determine the intensity and relevance of this area of research [1–3]. The concept of classification trees makes it possible to effectively tackle the problems involving data based on arbitrary scales when information is assigned in a natural form. To date, different approaches to the construction of recognition systems (RS) in the form of logical classification trees (LCT) are relevant, with interest in recognition methods involving LCT caused by a series of useful properties that they possess. Work [4] examines the basic aspects of decision tree diagrams; paper [5] introduces methods of branched selection of features based on a statistical branching criterion. The case of classification tree structures based on fixed criteria is an evolutionary development of LCT structures based on a direct transfer scheme [6]. Moreover, an important feature of the concept of classification trees is their high versatility, the possibility of applying to a wide class of applied tasks [7, 8]. One possible direction of application of the concept of classification trees involves tasks associated with the prediction and categori-

zation of flood-related phenomena. Flood phenomena are known to cause significant damage to both the economic and environmental sectors of the public economy and business. They flood and exclude from economic turnover significant agricultural land, settlements, destroy the residential areas and industrial enterprises, dams, transport infrastructure, there are even human casualties. Therefore, a comprehensive study of the conditions that form flood-related events in terms of the categorization of meteorological, hydrographic phenomena is a necessary stage for the further justification of new procedures aimed at forecasting ways to overcome critical environmental situations.

## 2. Literature review and problem statement

Study [9] proposes a scheme of generating the LCT structure based on a stepwise selection of elementary attributes, the disadvantage of which is the heavy dependence of model complexity on the effectiveness of the final minimization, the procedure of tree pruning. Papers [10, 11] suggest a modular scheme to build classifiers in the form of classification tree structures, which makes it possible to circumvent the limitations of conventional decision tree methods. Work [12] proposes an effective scheme for generating generalized features based on constructing the sets of hyperparallelepipeds. The disadvantage of such a scheme is the limitations on the structure of the initial training sample and the non-universality in applied terms. The issues of assessing the structural complexity of LCT models at the minimization stage are

considered in work [13]. The authors of [14, 15] report that the resultant classification rule, which is built by an arbitrary method or by an algorithm of the branched choice of features, has a tree-like logical structure. They emphasize the issue of choosing a qualitative criterion of branching. A logical tree consists of vertices that are grouped into tiers, established at a certain step when building a recognition tree [16]. There is a debatable issue related to the effective minimization of the structure of the built model of the classification tree. An important task tackled in work [10] is the issue of synthesis of recognition trees, which would actually be represented by the tree of algorithms. This approach could make it possible to construct new classifiers based on a modular principle. A potential disadvantage of this structure is greater resource costs and processor time during the model synthesis phase. Unlike existing methods, the main feature of tree recognition systems is that the significance of individual attributes, groups of attributes, or algorithms is defined relative to the function that sets the parsing of objects into classes [17, 18]. Work [19] addresses the basic issue of choosing a branching criterion in the structure of a classification tree scheme whose selection could be influenced by the specificity of the problem. The issues of generating and interpreting classification rules in the LCT structures are tackled in [20]. The task of assessing the informativeness of the attributes [21] in the construction of classification tree vertices remains principal and requires further research into the pre-processing and analysis of the structure of initial data [22]. Moreover, the disadvantage of the functionality of assessing the quality of attributes in the cited papers is the existence of restrictions in terms of generating the LCT structure. Thus, study [23] considers fundamental issues related to the generation of decision trees for the case of low-informative attributes. A potential to improve the cited study could be the use of combinations and sets of attributes in order to generate informative vertices of LCT structures. The ability of LCT/ACT structures (logical and algorithmic classification trees) to perform one-dimensional branching to analyze the impact, importance, quality of individual variables makes it possible to treat variables of different types in the form of predicates. For the case of ACT models, the issue of assessing the quality of relevant branching structures, autonomous classification algorithms, remains relevant [11]. To improve the methods of ACT structures, there is a search for effective criteria for branching the structure of trees. This concept of logical trees is actively used in data intelligent analysis where the goal is to synthesize a model that predicts the value of an objective variable based on a set of initial data at the system input. Paper [24] considers an important issue of analyzing the quality of classification of sets of decision trees. A possible way to improve the overall categorization quality is the use of ensembles of decision trees, the use of bugging and boosting mechanisms [25, 26]. Note that these schemes would provide the necessary accuracy of a classification model only if there is an effective branching criterion [27]. In this case, checking the built LCT models might involve a cross-validation scheme reported in work [28]. The disadvantage of this approach is the resulting complexity of classification models and the need for the procedure of final pruning of the model. As regards the applied aspect, there are a significant number of methods and algorithms that implement the concept of decision trees; however, among them, the following two have become common (C4.5/C5.0, a Ross Quinlan scheme; and the CART scheme, Classification and Regression Tree). The

C4.5/C5.0 scheme uses the so-called theoretic information criterion as a criterion for node selection; the CART algorithm is based on the calculation of the Gini index (a statistical indicator of the discrepancy of an attribute), which takes into consideration the relative distances within the metric between class distributions. The main examples, parameters, mechanisms of operation of a given scheme of classification trees can be found in open resources [29–32]. There, the disadvantage is relatively weak efficiency in terms of vertex selection compared to other modern methods and schemes of LCT structures. The main idea underlying the methods and algorithms for branching the selection of features, and vertices of ACT algorithms can be defined as the optimal approximation of some initial TS by a set of ranked classification algorithms. Then there is a central issue from [10]: the task to choose an effective branching criterion, the selection of vertices, attributes, signs of discrete objects for LCT schemes and ACT algorithms. These fundamental tasks are considered by the authors of [33] when addressing the issues related to the qualitative assessment of individual discrete features, their sets, and fixed combinations, which makes it possible to introduce an effective mechanism for the implementation of branching in the structure of a logical tree. The further advancement of the cited work is the transition from discrete attributes to independent classification algorithms. Important issues that remain include the convergence of the process of building classification trees according to the methods of branched selection of features and the issue of choosing a criterion for terminating the process of logical tree synthesis. The concept of classification trees makes it possible to use the combination of attributes and their sets as the signs, the vertices of a classification tree structure. If one dives deeper and disregards the attributes of objects as branching while selecting individual independent recognition algorithms, evaluated according to TS, a new structure of the ACT would be obtained at the output [10, 11]. Thus, there is a fundamental need to construct an effective method based on the concept of decision trees for recognizing large data sets while rationally utilizing the already accumulated potential of methods and classification schemes.

---

### 3. The aim and objectives of the study

---

The aim of this study is to construct effective models for categorizing flood-related events (the current state) across the Tisza river basin based on the concept of ACT structures.

To accomplish the aim, the following tasks have been set:

- to devise a method and a scheme for constructing ACT structures for the classification tasks of a wide spectrum;
- to build a set of classification models based on a set of independent recognition algorithms to categorize the phenomena of the situational status.

---

### 4. The study materials and methods

---

State the classification problem within a given task. Let some set  $G$  of objects  $x$  is assigned with  $R$ , the partition into the finite number  $k$  of subsets (classes, representations)  $H_i$  ( $i=1, \dots, k$ ),  $G = \bigcup_{i=1}^k H_i$ . The corresponding sets  $H_1, \dots, H_k$  would be called representations, and the elements of the set  $G$  – images or  $H_1, \dots, H_k$  representation representatives.

Objects (images)  $x$  are assigned by the sets of values of some attributes  $x_j$  ( $j=1, \dots, n$ ). If  $x \in H_i$ , it would denote that a given object belongs to the representation  $H_i$ . In a general case, representations  $H_1, \dots, H_k$  can be assigned by the probabilistic distributions  $p(H_1/x), \dots, p(H_k/x)$ , where  $p(H_i/x)$  is the probability (or, in a continuous case, a probability density) of the  $x$  ( $x \in G$ ) belonging to the  $H_i$  representation. Let the condition of the problem sets some initial training sample (TS) in the form of a sequence of training pairs in the following form:

$$(x_1, f_R(x_1)), \dots, (x_m, f_R(x_m)). \tag{1}$$

Here,  $x$  is an element of the initial TS;  $f_R$  determines the value of a recognition function for a given object. Moreover, in addition to the initial TS, a test sample (TeS) (a set of objects of known class affiliation) is also assigned, as some part of the initial TS. Therefore, on the initial condition, TS is the totality (fixed sequence) of some sets (discrete objects), and each set is the totality of values of some features (attributes) and the values of some recognition functions (RF) on this set. Then the totality of attribute values represents some image (a discrete object); the recognition function value attributes this image to the corresponding representation [34, 35]. Thus, this work's task implies building an ACT model with parameters  $p$ , whose structure  $L$  would be optimal  $F(L(p, x_i), f_R(x_i)) \rightarrow \text{opt}$  in relation to the initial data of TS.

At the beginning, let the initial TS in a general form (1) be assigned as a sequence of training pairs of known power classification  $m$ , and some system of length  $n$  of independent and autonomous recognition algorithms for the initial TS  $a_1(x), a_2(x), \dots, a_n(x)$ . Next, it is required to introduce the following sets that represent the parsing of TS data by the corresponding algorithms of classification  $a_i$ :

$$G_{a_1, \dots, a_i} = \{x \in G | a_i(x) = 1\}, \quad (i = 1, \dots, n). \tag{2}$$

Note that here, to simplify explanations, each autonomous classification algorithm  $a_i(x)$  generates an output value only within a binary set  $\{0, 1\}$ , that is, here  $a_i(x) = 1$  in the case of the successful classification of object  $x$ , and  $a_i(x) = 0$  otherwise. Note that the system of sets  $G_{a_1, \dots, a_n}$  would actually represent a complete phased partitioning of the set  $G$  (with the growth in the quantity  $i$  – that is, the classification algorithms involved), which is implemented by independent algorithms  $a_1, \dots, a_n$ . Note that depending on the initial selection of a set of classification algorithms  $a_1, \dots, a_n$ , some of the sets  $G_{a_1, \dots, a_n}$  may be empty – a case of the unsuitability of one particular or more algorithms for the approximation of the current TS. Thus, given the above, and similar to the methods of selection of sets of elementary features [9], it is possible to introduce the following values, which should be considered as a certain criterion of branching in the ACT structure:

$$\begin{aligned} \delta_{a_1, \dots, a_i} &= \frac{S_{a_1, \dots, a_i}}{m}; \quad \psi_{a_1, \dots, a_i}^j = \frac{S_{a_1, \dots, a_i}^j}{S_{a_1, \dots, a_i}}; \\ \rho_{a_1, \dots, a_i} &= \max_j \psi_{a_1, \dots, a_i}^j. \end{aligned} \tag{3}$$

Here, the value  $S_{a_1, \dots, a_n}$  determines the number of occurrences in the initial TS of those training pairs  $(x_s, f_R(x_s))$ , ( $1 \leq s \leq m$ ) that meet the base condition to belong to  $x_s \in G_{a_1, \dots, a_n}$ . Accordingly, the value  $S_{a_1, \dots, a_n}^j$ , ( $j=0, 1, \dots, k-1$ ) determines the number of occurrences in TS of those pairs  $(x_s, f_R(x_s))$ , ( $s=1, \dots, m$ ) that meet the conditions  $x_i \in G_{a_1, \dots, a_n}$  and  $f_R(x_s) = j$ .

Note that if the object is  $x_s \notin G_{a_1, \dots, a_i}$  for all  $s=1, \dots, m$ , then it is clear that the values  $\delta=0$  and  $\psi=0$ , at ( $j=0, \dots, k-1$ ). Thus, here the magnitude  $\delta$  characterizes the frequency of occurrences of the terms of the sequence  $x_1, \dots, x_m$  of discrete objects in the set  $G_{a_1, \dots, a_n}$ , and, accordingly, the value of  $\psi$  characterizes the frequency of belonging of some object  $x$  to the  $H_j$  class provided  $x \in G_{a_1, \dots, a_i}$ . Note that the condition  $x \in G_{a_1, \dots, a_i}$  is equivalent to the condition that in the  $a_1, \dots, a_i$  algorithm sequences one finds such an algorithm  $a_y$  that  $a_y(x) = 1$ . Then the  $\delta$  magnitude characterizes the information efficiency of recognizing the belonging of some object  $x$  to one of the  $H_0, \dots, H_{k-1}$ , classes, of course, if  $x \in G_{a_1, \dots, a_i}$ .

In the next step, again there is a fundamental issue about the belonging of an object  $x$  to the  $H_0, \dots, H_{k-1}$  classes, that is, the issue of forming a classification rule. It is clear that here one should also attribute object  $x$  to that  $H_i$  class for which a simple ratio holds:

$$\rho_{a_1, \dots, a_i} = \psi_{a_1, \dots, a_i}^j. \tag{4}$$

Note that here  $\{0 \leq j \leq k-1\}$ ; ratio (4) represents some classification rule, and it is clear to me that the greater the  $\rho$  value the higher its effectiveness.

Since the only information that represents the partitioning of  $H_0, \dots, H_{k-1}$  representations is the initial TS, the  $H_j$  class denotes the totality of all training pairs  $(x_s, f_R(x_s))$  within TS, which satisfy the  $f_R(x_s) = j$  ratio, that is, meet the condition of belonging.

Similarly, the average effectiveness of the  $H_0, \dots, H_{k-1}$  representation set recognition, which are assigned by the relevant TS using the recognition algorithms  $a_1, \dots, a_n$ , is estimated by the following value:

$$F_S(a_1, a_2, \dots, a_n) = \sum_{a_1, \dots, a_i} \delta_{a_1, \dots, a_i} * \rho_{a_1, \dots, a_i}. \tag{5}$$

Thus, in a given case, the value of  $F_S(a_1, \dots, a_n)$  can be considered an assessment of the approximation of the initial TS using a set of independent classification algorithms  $a_1, \dots, a_n$ .

Taking into consideration the very idea of an algorithmic classification tree, the value of  $F_S(a_1, \dots, a_n)$  can be derived for the following reasons. The recognition function  $F$  would denote some representation that assigns to each  $a_1, \dots, a_n$  set some element from the set  $\{0, \dots, k-1\}$ , that is the corresponding class number.

Therefore,  $F$  as RF represents a function in the form of  $F_S(a_1, \dots, a_i)$ , where  $a_1, \dots, a_i$  accept values from set  $\{0, 1\}$ . Accordingly, the RF  $F_S(a_1, \dots, a_i)$  object  $x$ , ( $x \in G$ ) clearly refers to that class (representation) of  $H_j$ , ( $0 \leq j \leq k-1$ ), for which the following ratio holds:

$$F(a_1, \dots, a_i) = l, \quad (0 \leq l \leq k-1). \tag{6}$$

Let the initial TS be assigned in the form of (1), one then assumes that the RF  $F(a_1, \dots, a_i)$  correctly classifies the set  $(x_j, f_R(x_j))$ , ( $1 \leq j \leq m$ ) of data array (1) if  $F(a_1(x_j), \dots, a_i(x_j)) = f_R(x_j)$ ; otherwise, the RF  $F(a_1, \dots, a_i)$  incorrectly classifies the training pair  $(x_j, f_R(x_j))$ .

Next, let the value  $m_F$  be the number of all inputs of training pairs  $(x_j, f_R(x_j))$  to the initial TS, which are correctly classified by the RF  $F(a_1, \dots, a_i)$ . Introduce the following quantity:

$$\tau_F = \frac{m_F}{m}. \tag{7}$$

Note that this quantity  $\tau_F$  can be considered the overall efficiency of the RF  $F(a_1, \dots, a_i)$  for the initial TS of power  $m$  relative to some set of classification algorithms  $a_1, \dots, a_n$ .

In the next step, we express a given quantity  $\tau_F$  using the previously proposed quantities  $\delta$  and  $\psi$ ,  $i=1, \dots, n; j=0, \dots, k-1$ . To this end, calculate the number of those training pairs  $(x_S, f_R(x_S))$  that correctly classify the RF  $F(a_1, \dots, a_i)$  and for which the membership relationship  $x_S \in G_{a_1, \dots, a_i}$  holds. Let  $F(a_1, \dots, a_i)=l$ , then the number of all training pairs  $(x_S, f_R(x_S))$  that are properly classified (when meeting the basic membership condition  $x_S \in G_{a_1, \dots, a_i}$ ,  $f_R(x_S)=l$  is equal to  $S_{a_1, \dots, a_i}^l, \dots, a_i$  (according to the above expressions (3)). It then becomes so obvious that the quantity  $m_F$  introduced above is calculated from the following formula:

$$m_F = \sum_{0 \leq a_1, \dots, a_i \leq 1} S_{a_1, \dots, a_i}^{F(a_1, \dots, a_i)}. \quad (8)$$

It should be noted that one can take into consideration in formula (8) only those training sets for which the ratio  $S_{a_1, \dots, a_i} \neq 0$  is valid (the  $S_{a_1, \dots, a_i}$  quantity is similar to the quantities in (3)). Thus, given the just mentioned comment on (8), one can represent the  $m_F$  quantity in the following form:

$$m_F = \sum_{0 \leq a_1, \dots, a_i \leq 1} S_{a_1, \dots, a_i} * \psi^{F(a_1, \dots, a_i)}. \quad (9)$$

Recall that the  $\tau_F$  quantity determines the effectiveness of the RF  $F(a_1, \dots, a_i)$  for the given TS of power  $m$  relative to the fixed set of classification algorithms  $a_1, \dots, a_n$ . Note that here the value of  $\psi$  was predefined in expressions (3). In the following step, formulae (8), (9) produce the following:

$$m_F = \frac{m_F}{m} = \sum_{0 \leq a_1, \dots, a_i \leq 1} \delta_{a_1, \dots, a_i} * \psi^{F(a_1, \dots, a_i)}. \quad (10)$$

It should be noted that since  $\psi^{F(a_1, \dots, a_i)} \leq \rho_{a_1, \dots, a_i}$ , the following situation emerges:

$$\tau_F \leq F_S(a_1, \dots, a_n). \quad (11)$$

Next, the recognition function (4) would be denoted through  $F_0$ ; the  $F_0$  function would be assigned by the following ratio:

$$F_0(a_1, \dots, a_n) = l,$$

if

$$\psi_{a_1, \dots, a_i}^l = \rho_{a_1, \dots, a_i} = \max_{0 \leq j \leq k-1} \psi_{a_1, \dots, a_i}^j. \quad (12)$$

Then formulae (12) and (10) would directly yield the following:

$$\begin{aligned} \tau_{F_0} &= \frac{m_{F_0}}{m} = \sum_{0 \leq a_1, \dots, a_i \leq 1} \delta_{a_1, \dots, a_i} * \psi_{a_1, \dots, a_i}^{F(a_1, \dots, a_i)} = \\ &= \sum_{0 \leq a_1, \dots, a_i \leq 1} \delta_{a_1, \dots, a_i} * \rho_{a_1, \dots, a_i} = F(a_1, \dots, a_n). \end{aligned} \quad (13)$$

Thus, it follows from formulae (11) and (13) for all RF  $F(a_1, \dots, a_n)$ :

$$\tau_F \leq \tau_{F_0}. \quad (14)$$

Consequently, the quality assessment of the set of algorithms  $W(a_1, \dots, a_n)$  is the effectiveness of recognizing some built ACT structure. Here, the structure of the classification

tree is assigned by a set of algorithms  $a_1, \dots, a_n$ ; the  $W$  value can be determined for an arbitrary recognition algorithm in the structure of the classification tree.

Given the above, one can assume that the ACT structure would take a general form shown in Fig. 1. Here, each tier of the structure determines the stage of ACT construction in the form of the approximation by the current classification algorithm  $a_i$  of a certain part of TS. At every step of the ACT model generation (Fig. 1), there is a separate  $a_i$  classification algorithm and a separate corresponding TS or a subset of the initial TS; the initial TS in complete composition is given only in the first step. During the subsequent stages of building a classification tree, the power of the TS data array would drop due to a set of built-in generalized features (GF)  $f_j$  that would approximate a certain part of the initial TS data. It is also important to note that depending on the structure of the scheme of ACT construction and the features of the current  $a_i$  algorithm, it is possible, at every step, to generate more than one GF  $f_j$ . In the next step, for the ACT method, the basic criteria for building a classification tree model are introduced: the criterion for terminating a branching procedure  $K_{Stop}$ . This criterion regulates the complexity and accuracy of the resulting ACT model. The criteria for selecting the branching  $W(a)$  or selecting a classification algorithm in the current step for the classification tree being built.

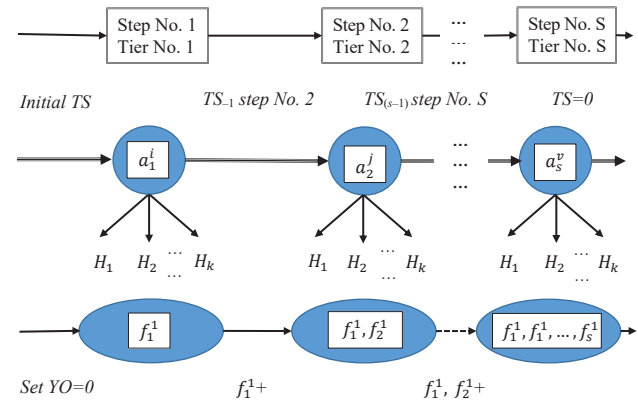


Fig. 1. The general scheme of ACT structure (type I)

Thus, based on the above, it is advisable to introduce the criterion  $K_{Stop}$  for terminating the branching process the type of (Boolean) procedure for ACT construction, which implies checking the power  $P_{pt}$  ( $HB$ ) of the training sample in the following form:

$$K_{Stop} = \begin{cases} 0, & \text{if } P_{pt}(HB) = 0, \\ 1, & \text{if } P_{pt}(HB) > 0. \end{cases} \quad (15)$$

Note that the classification tree construction procedure continues until  $K_{Stop}=1$ ; the opposite situation where  $K_{Stop}=0$  signals the completion of the ACT synthesis phase. Note that it is a fundamental issue for the methods of ACT to choose the criterion of branching, evaluation, and selection of the  $a_i$  classification algorithm for the current step in the structure of the classification tree model under construction. It is clear to me that, by analogy with the method of TS approximation using a set of ranked elementary features as a branching criterion, one can propose an initial assessment of the effectiveness of the set of algorithms  $(a_1, \dots, a_i)$  in the following form:



$$W(a_i) = \frac{1}{P_{pr}(HB)} * \sum_{j=1}^k \left( T_{Uz} + S_{Uz} + \frac{E_{Uz}}{S_{Uz}} \right) \quad (16)$$

Note that in the proposed functionality (16) the values introduced have the following interpretation:

- 1)  $k$  is the total number of classes of the current task, assigned by the data partitioning  $R$  of the initial TS;
- 2)  $T_{Uz}$  characterizes the total time spent building the current GF  $f_j$ ;
- 3)  $E_{Uz}$  is the information capacity, the structural complexity of the GF  $f_j$  built in the current step of the generation of an ACT model;
- 4)  $S_{Uz}$  represents the total number of discrete objects  $x_i$  in the initial TS that are generalized by a given GF  $f_j$ ;
- 5)  $P_{pr}(HB)$  is the power of the initial TS or its fixed part for the current step of the scheme of ACT construction.

Note that in the formula (16) the summation is carried out in all classes assigned by the data array from the initial TS, although there may be limits on the summation, which are due to the structure of the algorithm of building the classification tree. An important point in the scheme of building the ACT model (Fig. 1) is that at every step of the tree algorithm one actually builds its own fixed, one, or more – depending on the structure of the algorithm of the ACT GF  $f_j$ . The total number of GF increases with each step of the classification tree algorithm, namely the ACT with a set of classification algorithms  $(a_1, \dots, a_m)$  generates, replicates a tree-like structure – a tree of generalized features (TGF) with the corresponding set of GF  $(f_1, \dots, f_z)$ .

The ACT structure makes it possible to effectively adjust the complexity, to retrain the models of the classification tree under construction, or to build models of recognition trees with predetermined accuracy. It is crystal clear that to compare and select a particular model of ACT from a fixed set, it is necessary to highlight the most important characteristics of them – the dimensionality of feature space, the number of vertices, the number of transitions to the construction of the tree, etc. It is necessary to determine their error relative to the input array. An important point is the analysis of the quality criteria of the derived information models, which depend on the error of the model, the power of the initial array of data in TS and TeS, the number of structural parameters of the model, and so on. It is clear to all that the critical parameters of the built ACT model, which need to be minimized, are model errors on the data sets of TS and TeS for each of the classes. Note that one of the most important indicators, which characterizes the basic properties of the resulting ACT models, is the basic indicator of the generalization of data from an initial TS by a classification tree. This indicator is calculated as follows:

$$I_{Main} = \frac{m * O_{Uz}}{Fr_{All} + V_{All} + N_{All} + 2P_{All}} \quad (17)$$

A given indicator of the generalization of the model of a classification tree, the ACT structure reflects its basic characteristics and can be applied as a criterion of optimality in the procedure of evaluation of an arbitrary tree-like recognition scheme. For an arbitrary application task, it is important to maximize the  $I_{Main}$  parameter, which makes it possible to achieve the optimal structure of the classification tree. It provides for the maximum compression of the initial TS data,

makes it possible to represent an array of initial data by the tree minimal in its structural complexity. Thus, an important indicator of the quality of the constructed model in the form of a classification tree, taking into consideration the parameters of ACT structure, is a general integrated indicator of model quality in the following form:

$$Q_{Main} = \frac{Fr_{All}}{O_{Uz} * \sum_i P_i} * e^{\frac{Er_{All}}{M_{All}}} \quad (18)$$

Note that in functionalities (17), (18) the parameters are interpreted as follows:

- $Er_{All}$  is the total number of errors in the ACT model on the data sets of the initial test and training samples  $Er_{All} = Er_{tr} + Et_{tr}$ ;
- $M_{All}$  is the total power of these two data arrays  $M_{All} = m + T$ , where  $m$  and  $T$  are the TS and TeS powers, respectively;
- $Fr_{All}$  is the quantity that characterizes the total number of vertices in the resulting ACT model with resultant  $f_R$  values, that is, the classification tree leaves;
- $O_{Uz}$  is the quantity that represents the total number of all generalized features, sets of GF in the structure of the ACT model;
- $V_{All}$  is the quantity that represents the total number of all types of vertices in the structure of the ACT model;
- $N_{All}$  is the total number of different autonomous classification algorithms used in the classification tree model;
- $P_{All}$  is the total number of transitions between the vertices, tiers in the structure of the built model of the classification tree.

The  $p_i$  parameter set for the integrated assessment of the quality of the classification tree model represents the most important characteristics of the classification tree. Note that this integrated quality indicator of the ACT model would take values within zero and unity.

Thus, the proposed integrated assessment of the quality of the classification tree model, the ACT structure reflects its basic parameters, characteristics, and can be applied as a criterion of optimality in the procedure of evaluation of an arbitrary tree-like recognition scheme.

One of the possible directions to apply the concept of algorithmic classification trees is the tasks associated with forecasting and categorization of flood-related events (based on the arrays of meteorological and hydrographic data). It is known that flood-related events of the Carpathian region cause significant damage to the economic and ecological sector of the state economy and business. Therefore, a comprehensive study of the conditions that form flood-related events in terms of the classification of meteorological (hydrographic) phenomena under modern changing climatic conditions is a necessary stage in the calculation and prediction of environmental critical situations.

Note that the regime of surface water runoff of the Tisza river basin in the Transcarpathian region is characterized by significant territorial unevenness due to climatic, temperature factors, and the influence of mountain ranges of the Carpathians. A set of these factors has a decisive impact on the formation of seasonal drains. To analyze the conditions and causes of the formation of flood-related events on the Tisza river in the Transcarpathian oblast in the fall-spring period, data from one hydrological (observation) post were used, with the monitoring period from the beginning of 1995 to 2012 inclusive. Note that, according to its water regime,

the river Tisza can be attributed to rivers with a flood regime of the Carpathian subtype – where flood-related events are observed mainly in the time interval of the fall-spring period. Usually, such flood-related events are termed the floods of the cold season; they are observed between October/November and April/May. It is clear to all that modern climate change should be accounted for – the limits of warm and cold periods are not constant from year to year; the basis for determining their onset and end were mainly basic meteorological conditions.

An important feature of the Tisza river basin, unlike other river basins in the western region of Ukraine, is the formation of flood-related events of mixed origin on the basis of snow and rain drains, most often with a dominating rain component, which occur in most observations during the fall-spring period of the year. Moreover, this mode of water performance is due to the climatic features of Transcarpathia, the presence of mountain ranges, the movement of atmospheric fronts, and other factors. It should be emphasized that in the basins of the year within the Transcarpathian region rainy, snow-rain flood-related events of different intensity and duration are repeated with a frequency of up to 4–6 times per calendar year of observations. On the other hand, it should be noted that as a result of intense snowfall, the corresponding process of snow retreat, purely snowy flood-related events are observed in the Tisza river basin. Moreover, it should be noted that annual maxima only due to the melted non-waterways of the Transcarpathian region are quite common. Thus, the annual highs of cold-season floods often significantly exceed the highs of the warm period, respectively. Flood phenomena of the cold period are longer in time and occur more often; the maximum water flows of cold season floods are greater than the maximum of the warm season by almost 2.5 times.

The initial parameters for a given applied problem to categorize flood-related events in the Tisza river basin are given in Table 1.

Parameters for the problem to categorize flood-related events in the Tisza river basin

Station number	Feature space dimensionality – $N$	Initial TS data array power – $M$	TeS data array power – $S$	Total number of classes based on the TS data parsing – $l$	Ratio of objects from different TS classes – $(H_1/H_2/H_3)$
1	18	6.118	500	3	76/108/5.934

Note that the monitoring of a river basin has specific features (it may even be unique within a certain observation area), which are predetermined by that the river is one of the quite dynamic objects of nature. Moreover, it requires a largely individual approach, due to geography, conditions of formation, water regime, etc. Consequently, the monitoring of the riverbed is associated with a large space and significant time costs, as well as a change in the frequency of observations. Thus, the tasks of the observation station include a complete characteristic not only of purely hydrological objects' attributes but the general geological and geomorphological conditions of the river basin. The general model of flood-related events is described on the basis of 18 features that have different natures and are formed on the basis of many years of observations in the Tisza river basin. Among

these attributes, the following defining characteristics can be distinguished, the sets of parameters of the river basin:

1. The total density of the river canopy of the observation sector is the ratio of the length of all surface runoffs of the riverbed (kilometers) to the size of the total area of the basin (in square kilometers) in the corresponding observation sector. Note that a given parameter is calculated separately for each observation station.

2. The coefficient of riverbed curvature at the observation site is the ratio of the actual length of the river (observation area) to the length in a straight line from the source to mouth.

3. The parameter of the type of river terraces – a feature that characterizes the geological peculiarities of the riverbed structure at the observation site; it accepts values from the set {1...6}.

4. The average width of the riverbed (AWR) at the observation site is the parameter for the Tisza river basin characterized by sharp changes in values depending on the observation station where measurements are carried out.

5. Conditional water level (CWL) is the parameter that is typically the average level of water observed during the year over the period when the river is free of ice cover. This level is defined as the arithmetic average of daily level values over a limited period (from flood decline to the onset of ice phenomena). They approximately determine the period and height of the low, constant, and intermediate level, which is accepted as conditional; the next step implies the calculation of the operational level that exceeds the conditional level. In the event of exceeding, it is used as a growth value to reduce all current observations to a conditional level.

6. The parameter of the mean depth of a riverbed (MDR) at the observation site. For the Tisza river basin, it is characterized by quite large discrepancies for different areas of observation.

7. The parameter of the general mode of the river is the parameter that characterizes the features of water supply and runoff of the riverbed at the observation site; it accepts values from the set {1...10}.

Table 1

8. The parameter of a river temperature regime (RTR) characterizes the temperature regimes of the river at the observation site; it accepts values from the set {1...50}.

9. The parameter of an average river flow rate (RFR) characterizes the average rate of water flow in the middle of the riverbed at the observation site.

10. The parameter of the maximum surface river flow rate is different from the RFR parameter and is measured on a regular basis for the section of observations of the riverbed.

11. The parameter of average daily water consumption of the river at the observation site.

12. The parameter of riverbed relief type characterizes the complexity of the riverbed relief at the observation site; it accepts values from the set {1...20}.

## 5. The study results

### 5.1. A method of algorithmic classification trees

The devised ACT method makes it possible to build the models of classification trees of different types; all of them are

based on a simple idea of approximating the initial sample by a set of autonomous classification algorithms and representing the resulting model in the form of some tree-like scheme. An arbitrary ACT model, in turn, would represent a new recognition algorithm. In other words, the concept of ACT is a method of synthesis of new algorithms of classification of the predefined accuracy relative to TS based on a set of already known ones. The scheme of stepwise approximation of the array of initial data of TS by a set of selected and evaluated independent classification algorithms makes it possible to build different types of ACT models. Moreover, the function of form recognition (5) of ACT structure is the most effective of all classification rules, which cover the data of initial TS based on a set of classification algorithms  $a_1, \dots, a_n$ . Note that the  $F_S(a_1, \dots, a_n)$  quantity produces the greatest effectiveness of recognition of the initial TS data using a fixed set of independent classification algorithms  $a_1, \dots, a_n$ .

Thus, given the above, one of the possible schemes of ACT construction (type I) shown in Fig. 2 can be suggested.

**Stage 1.** At the first stage of ACT construction, one fixes in a library of the algorithms of an information system a set of classification algorithms  $(a_1, \dots, a_m)$ . Note that the classification algorithms and their number are selected depending on the conditions and aspects of the application problem.

**Stage 2.** In the second stage of ACT construction, a selected set of classification algorithms  $(a_1, \dots, a_m)$  is evaluated and ranked on the basis of functionality (16) according to TS data in the set according to their effectiveness. It should be noted that there are two options here by analogy with the LCT – depending on the algorithmic scheme of building a classification tree:

a) an option when assessing the effectiveness and ranking of a set of classification algorithms  $(a_1, \dots, a_m)$  is carried out only once at this stage. Next, at every step of ACT construction, one fixes for the approximation of the data the following algorithm  $a_i$  of the initial sequence. This approach significantly saves the hardware resources of the information system but negatively affects the complexity of the resulting model of the classification tree;

b) an option when the evaluation of the effectiveness and ranking of a set of classification algorithms  $(a_1, \dots, a_m)$  is carried out at every step of ACT construction according to the relevant data on the subsets of the initial TS. This approach makes it possible, for fewer steps, to complete the approximation of TS and get a more economical design of the ACT compared to option (a). This variant requires much more hardware information system resources for the second stage of the scheme of ACT construction and requires considerable attention to the initial selection of algorithms  $(a_1, \dots, a_m)$ .

**Stage 3.** At the third stage of the scheme of ACT construction, one fixes the initial ACT vertex as the algorithm of classification  $a_i$  of the greatest efficiency from the sorted set  $(a_1, \dots, a_m)$ ; the initial TS is submitted to the input in the form of a sequence of training pairs. The  $a_i$  algorithm provides the generation of one or more GF  $f_i$  of the first tier, which approximate a certain part of TS. The number of GF generated at each step is specified by the parameters of the algorithm of ACT construction.

**Stage 4.** At the fourth stage, one selects the classification algorithm  $a_i$  of the ranked sequence  $(a_1, \dots, a_m)$  as the vertex of the second tier, and the procedure for building the third stage of GF is repeated. One submits to the input the already limited TS without training pairs, which are approximated by the GF vertex of the first tier, and so on. The procedure

for building the ACT would be repeated for the next effective  $a_i$  algorithm from the set  $(a_1, \dots, a_m)$ . There is a stepwise approximation of TS parts and a check of the criterion for terminating the branching. The terminating criterion signals the completion of the procedure for building the ACT model, obtaining at the output of the tree of algorithms  $\alpha_i$  and the tree of generalized features  $f_i$ .

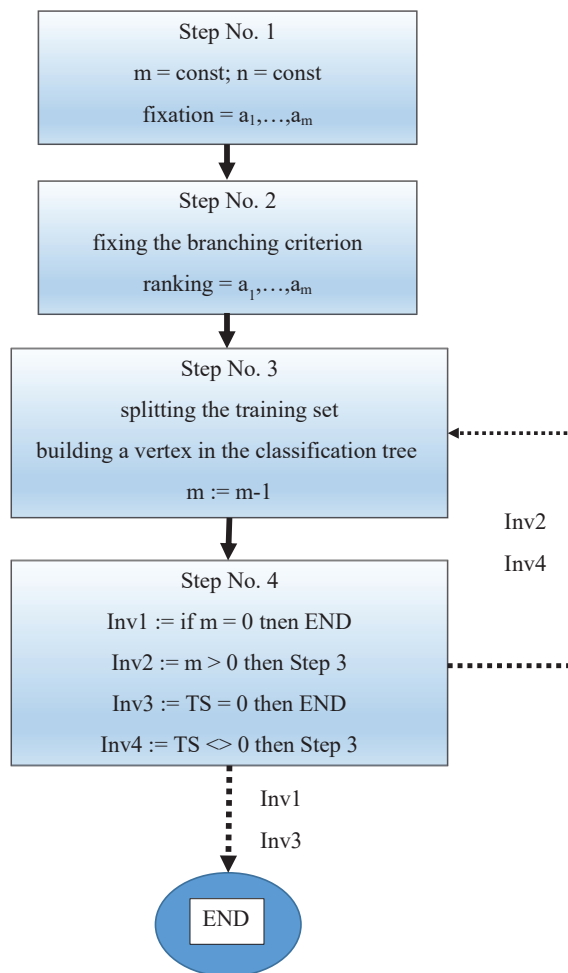


Fig. 2. ACT model construction scheme

Note that there are other implementation options for the scheme of ACT construction of type I, which differ from the proposed scheme with variations on the number of GF, which are built at every step, by the criteria, and sequence of stages of quality assessment of classification algorithms. Thus, it is possible to use a limited number of classifiers, to provide for the approximation of each TS class by a set of the selected algorithms.

**5. 2. Categorization models of flood-related events**

Based on a set of the represented hydrographic characteristics, I have built the models for the classification of flood-related events for the Tisza river over 18 years in the form of structures, models of ACT. Thus, a library of 15 algorithms, methods, and recognition schemes was used to build classification tree models. The main array of TS was composed of objects from three basic classes; at the stage of the exam, the classification system should ensure effective recognition of objects of unknown categorization relative to these three classes. General parameters of the built structures, models of ACT, are given in Table 2.

Table 2

General parameters of the built ACT models

Number of built ACT model	ACT model construction method (algorithm)	Total number of different classification algorithms involved in ACT – $N_{All}$	Total number of GF (GF sets) in the ACT structure $O_{Uz}$	Total number of all vertices (together with resultant) in the ACT structure	Total time to generate the ACT structure of current TS, s
Observation station 1					
1	Complete ACT method (type I)	5 constraint for the sequential application of one algorithm	36	75	602
2	Complete ACT method (type II)	5 constraint for the number of GF generations per one step in ACT construction	38	89	679
3	Limited ACT method ( $Z=10$ )	1 hypersphere algorithm	31	63	382
4	Limited ACT method ( $Z=5$ )	1 hypercube algorithm	50	102	801
5	Limited ACT method ( $Z=3$ )	1 hyper ellipse algorithm	33	71	413
6	Complete ACT method (type I)	2 hypersphere algorithm; hypercube algorithm;	36	79	679

At the initial stage of software system operation, the training sample was automatically checked for correctness. Note that the array of initial training information was dominated by the  $H_3$  class learning pairs of neutral zone situational objects, the green marker. The second place was occupied, with a significant margin, by the training pairs from class  $H_2$ , the objects of the situational status of the observation zone, the yellow marker. The third place was occupied by the training pairs of flood-related events, the red marker objects, class  $H_1$ . The  $H_2$  class power marginally outweighs the  $H_1$  class power. This is due to the dynamics of change in a flood-related event over time, which can return to normal, the neutral zone –  $H_3$  class phenomena, and, in most cases, transfers into crisis, the red flood-related event zone – class  $H_1$  (Fig. 3). The TS array consisted of 6.118 objects, the sets of known classification for two monitoring points at the site of the city of Uzhhorod. Moreover, the effectiveness of the constructed recognition system was estimated on a test sample with a volume of 500 objects for each observation station. The array of TeS data represented a separate part of the initial TS and consisted of discrete objects of known classification.

This volume of a test sample is not enough for a comprehensive analysis of the quality of the built models of classification trees; but even such a TeS makes it possible to evaluate and analyze the basic parameters of ACT structures. The constructed ACT structures provided the necessary level of accuracy of classification specified by the conditions of the applied problem. The models demonstrated different structural complexity in the construction of classification trees, parameters of the complexity of the ACT design – the sets of algorithms, sets of generalized features, the number of vertices in the classification tree structure. Based on the ACT method proposed within the framework of this study, the models of classification of flood-related events for the Tisza river were built. A fragment of the main results of the integrated assessment of the quality of the constructed structures, the ACT models, is given in Table 3.

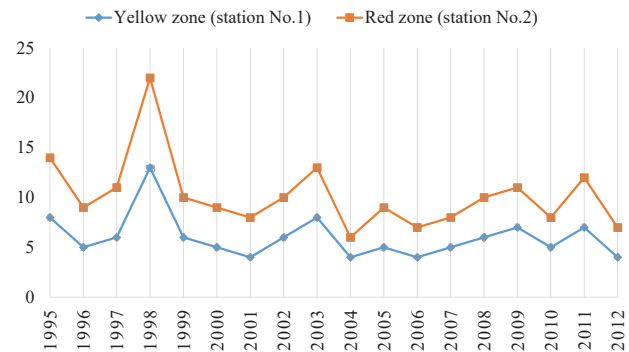


Fig. 3. Flood-related events in the yellow and red zones of the Tisza river over the period of 1995–2012

Table 3

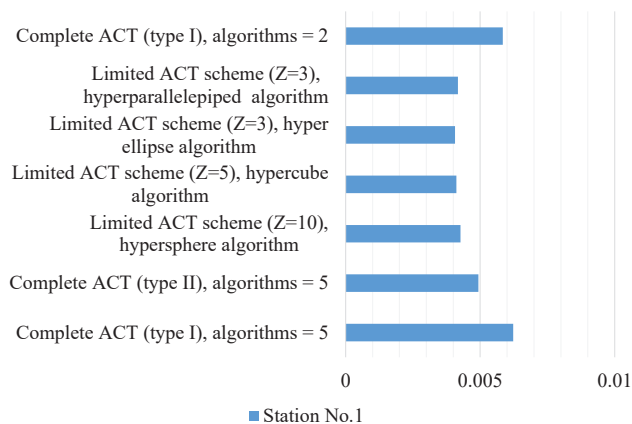
Comparative table of classification models (ACT structures) to categorize flood-related events for the Tisza river

No.	Method (scheme) to synthesize the structure (model) of a classification tree (LCT/ACT)	Integrated quality indicator of classification tree $Q_{Main}$	Total number of model errors on TS and TeS $Er_{All}$
1	Complete ACT method (type I) (the number of algorithms is 5, restrictions for the sequential use of one algorithm)	0.006229	0
2	Complete ACT method (type II) (the number of algorithms is 5, constraint for the number of GF generations per one step in ACT construction)	0.004934	0
3	Limited ACT method ( $Z=10$ ) (the number of algorithms is 1, hypersphere algorithm)	0.004271	0
4	Limited ACT method ( $Z=5$ ) (the number of algorithms is 1, hypercube algorithm)	0.004119	8
5	Limited ACT method ( $Z=3$ ) (the number of algorithms is 1, hyper parallelepiped algorithm)	0.004065	3
6	Limited ACT method ( $Z=3$ ) (the number of algorithms is 1, hyper ellipse algorithm)	0.004172	2
7	Complete ACT method (type I) (the number of algorithms is 2, hypersphere algorithm; hypercube algorithm)	0.005843	0



**6. Discussion of results of studying the built classification models**

The built classification trees, ACT models, provided for the necessary quality and speed of schemes for the classification of flood-related events for the Tisza river with a sufficiently compact structure of the tree structure itself (Fig. 4). The sets of independent classification algorithms, selected for the generation of GF groups, have also confirmed their effectiveness within a given applied task. The highest quality, in terms of the integrated evaluation, was demonstrated by the ACT model of type I ( $Q_{Main}=0.006229$ ) based on five basic geometric classifiers. Moreover, all the ACT structures of the first and second types did not produce classification errors on the data of training and test samples. The second model of complete ACT of type I, based on two geometric classifiers, showed a close result ( $Q_{Main}=0.005843$ ) due to the presence of a universal hypersphere algorithm in the scheme. Note that the ACT structure of type II showed poorer quality ( $Q_{Main}=0.004934$ ) compared to the structures of the trees of the first type of algorithms. This is due to the more complex design of the model ( $V_{All}=89$ ), as a result of the greater ACT generation time – Table 2. The ACT models based on the limited method for generating classification trees showed the worst result – Table 3. To a large extent, this is due to the peculiarities of the limited scheme where the addition of the ACT structure takes place only in the directions with the greatest number of classification errors. It is also necessary to take into consideration the features of the selected geometric classifiers (hypercube algorithm, hyper parallelepiped algorithm), which do not always allow for the effective approximation of TS data. The disadvantage of the presented ACT models, which was revealed under a given problem’s conditions, is relatively large time costs for the stage of synthesis of the models of classification trees themselves compared to the LCT structures. The difference, compared to the LCT structure, based on the branched choice of features with a stepwise assessment of informativeness for the ACT models of type I was almost 20 %.



**Fig. 4. An integrated quality indicator of the built ACT models**

The reported models of classification trees can be used to assess the general condition of the Tisza river basin, and to identify the situation of the red (flood) zone based on the current measurements by observation stations. This allows me to recommend the application of an ACT model construction method for a wide range of applied classifica-

tion and recognition tasks. The only drawback is the much higher operating memory costs and processor time of an information system compared to LCT methods. To a large extent, it depends on the peculiarities of the implementation of the classifiers, the number of algorithms in the ACT scheme, the type of ACT model. The structural complexity of the built ACT models is due to that the branching criterion (16) makes it possible to construct sufficiently compact ACT structures. On the other hand, the initial assessment of the quality of classification algorithms according to a given criterion requires significant time costs, which depend on the number of selected algorithms in the set. Moreover, a fundamental feature of the proposed ACT method, compared to LCT methods (a standard scheme by Quinlan [29–32]), is that the resulting structures are characterized by high compactness and speed of operation. In addition, the concept of ACT makes it possible to build models of classification trees on a modular principle, using, as a structural element of the vertices of the classification tree, the C5.0 algorithm itself. This method of organization of classification scheme allows the rational use of the already accumulated potential of methods and algorithms of recognition theory for problems involving data in arbitrary scales where information is set in a natural form. The only fundamental limitation of this work in the generation of ACT structure is the limitation in terms of time spent assessing the effectiveness of the initial algorithms, based on the fixed criterion of branching the ACT structure. A possible area of further research may be the expansion of the list of classification algorithms in the ACT scheme, as well as additional conditions and restrictions on the generation of GF sets for each step of the classification tree scheme, the ACT structure.

A promising direction to resolve the task of classification of the situational status of a river basin may also be the use of methods for building LCT models with the possibility of comparing them with the built ACT structures. It is important to expand the volume of a library of recognition algorithms, classifiers, optimization of software implementations of methods of ACT construction, as well as practical approbation of the built models of classification trees on a set of TS data from more observation stations in the Tisza river basin. An example of such effective software optimization could be the evolutionary transition from the C4.5 to the C5.0 algorithm.

**7. Conclusions**

1. The method and model to construct the ACT structures have been devised in order to categorize flood-related events based on the assessment and ranking of a set of autonomous recognition algorithms to generate the structure of the classification tree. Moreover, at every step of branching the classification tree, a certain part of the TS data is approximated. This makes it possible to build different classification models, ACT structures, based on the quality assessment of a set of algorithms in the scheme of ACT methods. A given approach ensures high versatility in terms of type and structure of TS, the ability to work with large data arrays. A branching criterion (16) for ACT structures can be used not only to evaluate the quality of individual classification algorithms but also to calculate the effectiveness of linked sets of algorithms. This makes it possible to achieve the optimal structure of a synthesized ACT according to the initial

TS data. The built ACT models of the first and second type showed the best result based on the integrated quality indicator and the absence of classification failures on the data of the training, test sample.

2. A set of classification models, ACT structures have been built to analyze the situational status of the Tisza river basin. Moreover, the constructed models of classification trees could be used to assess the general condition of a river basin and to identify the situation of the red (flood), yellow (dangerous) zones based on the current monitoring. This work suggests indicators of ACT models – a general

integrated indicator of quality and a basic indicator of data generalization, which make it possible to effectively represent the general characteristics of the built classification tree. These indicators register the most important parameters of ACT structure operation – the number of vertices, the number of transitions of the structure, the number of the tree's leaves, the number of generalized features, the number of classifiers, the number of errors of all types, the power of the training and test samples. It is possible to use them for the selection of the optimal ACT from the set built.

## Reference

1. Srikant, R., Agrawal, R. (1997). Mining generalized association rules. *Future Generation Computer Systems*, 13 (2-3), 161–180. doi: [https://doi.org/10.1016/s0167-739x\(97\)00019-8](https://doi.org/10.1016/s0167-739x(97)00019-8)
2. Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning*. Springer. doi: <https://doi.org/10.1007/978-0-387-84858-7>
3. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1 (1), 81–106. doi: <https://doi.org/10.1007/bf00116251>
4. Utgoff, P. E. (1989). Incremental Induction of Decision Trees. *Machine Learning*, 4, 161–186.
5. Vasilenko, Y. A., Vasilenko, E. Y., Kuhayivsky, A. I., Papp, I. O. (1999). Construction and optimization of recognizing systems. *Information technologies and systems: scientific and technical journal*, 1, 122–125.
6. Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31, 249–268.
7. Mingers, J. (1989). An empirical comparison of pruning methods for decision tree induction. *Machine Learning*, 4 (2), 227–243. doi: <https://doi.org/10.1023/A:1022604100933>
8. Alpaydin, E. (2010). *Introduction to Machine Learning*. London: The MIT Press, 584.
9. Povhan, I. F. (2020). Logical recognition tree construction on the basis of a step-to-step elementary attribute selection. *Radio Electronics, Computer Science, Control*, 2, 95–105. doi: <https://doi.org/10.15588/1607-3274-2020-2-10>
10. Povkhan, I. F. (2020). The general concept of the methods of algorithmic classification trees. *Radio Electronics, Computer Science, Control*, 3, 108–120. doi: <https://doi.org/10.15588/1607-3274-2020-3-10>
11. Povkhan, I., Lupei, M. (2020). The Algorithmic Classification Trees. 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP). doi: <https://doi.org/10.1109/dsmp47368.2020.9204198>
12. Povkhan, I., Lupei, M., Kliap, M., Laver, V. (2020). The Issue of Efficient Generation of Generalized Features in Algorithmic Classification Tree Methods. *Data Stream Mining & Processing*, 98–113. doi: [https://doi.org/10.1007/978-3-030-61656-4\\_6](https://doi.org/10.1007/978-3-030-61656-4_6)
13. Povhan, I. (2016). Designing of recognition system of discrete objects. 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP). Lviv, 226–231.
14. Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. (2017). *Classification and regression trees*. Boca Raton, 368. doi: <https://doi.org/10.1201/9781315139470>
15. Miyakawa, M. (1989). Criteria for selecting a variable in the construction of efficient decision trees. *IEEE Transactions on Computers*, 38 (1), 130–141. doi: <https://doi.org/10.1109/12.8736>
16. Povkhan, I. (2020). A constrained method of constructing the logic classification trees on the basis of elementary attribute selection. *Proceedings of The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020)*. Zaporizhzhia, 843–857.
17. Mitchell, T. (1997). *Machine Learning*. McGraw-Hill, 432.
18. Amit, Y., Geman, D., Wilder, K. (1997). Joint induction of shape features and tree classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (11), 1300–1305. doi: <https://doi.org/10.1109/34.632990>
19. Dietterich, T. G., Kong, E. B. (1995). Machine learning bias, statistical bias, and statistical variance of decision tree algorithms. Available at: <http://www.cems.uwe.ac.uk/~irjohnso/coursenotes/uqc832/tr-bias.pdf>
20. Karimi, K., Hamilton, H. J. (2011). Generation and Interpretation of Temporal Decision Rules. *International Journal of Computer Information Systems and Industrial Management Applications*, 3, 314–323.
21. Deng, H., Runger, G., Tuv, E. (2011). Bias of Importance Measures for Multi-valued Attributes and Solutions. *Artificial Neural Networks and Machine Learning – ICANN 2011*, 293–300. doi: [https://doi.org/10.1007/978-3-642-21738-8\\_38](https://doi.org/10.1007/978-3-642-21738-8_38)
22. Koskimaki, H., Juutilainen, I., Laurinen, P., Roning, J. (2008). Two-level clustering approach to training data instance selection: A case study for the steel industry. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). doi: <https://doi.org/10.1109/ijcnn.2008.4634228>
23. Subbotin, S. A. (2019). Construction of decision trees for the case of low-information features. *Radio Electronics, Computer Science, Control*, 1, 121–130.

24. Kamiński, B., Jakubczyk, M., Szufel, P. (2017). A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research*, 26 (1), 135–159. doi: <https://doi.org/10.1007/s10100-017-0479-6>
25. Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Machine learning*, 40, 139–157. doi: <https://doi.org/10.1023/A:1007607513941>
26. Subbotin, S. A. (2013). Methods of sampling based on exhaustive and evolutionary search. *Automatic Control and Computer Sciences*, 47 (3), 113–121. doi: <https://doi.org/10.3103/s0146411613030073>
27. De Mántaras, R. L. (1991). A distance-based attribute selection measure for decision tree induction. *Machine Learning*, 6 (1), 81–92. doi: <https://doi.org/10.1007/bf00153761>
28. Painsky, A., Rosset, S. (2017). Cross-Validated Variable Selection in Tree-Based Methods Improves Predictive Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (11), 2142–2153. doi: <https://doi.org/10.1109/tpami.2016.2636831>
29. Saha, S. (2018). What is the C4.5 algorithm and how does it work? Available at: <https://towardsdatascience.com/what-is-the-c4-5-algorithm-and-how-does-it-work-2b971a9e7db0>
30. C5.0 Classification Models. Available at: <https://cran.r-project.org/web/packages/C50/vignettes/C5.0.html>
31. C5.0 Decision Trees and Rule-Based Models. Available at: <https://topepo.github.io/C5.0/reference/C5.0.html>
32. C5.0: An Informal Tutorial. Available at: <https://www.rulequest.com/see5-unix.html>
33. Vasilenko, Y. A., Vashuk, F. G., Povkhan, I. F. (2010). The importance of discrete signs. XX International Conference Promising ways and directions of improving the educational system. Uzhgorod, 217–222.
34. Subbotin, S. A., Oliinyk, A. A. (2016). The Dimensionality Reduction Methods Based on Computational Intelligence in Problems of Object Classification and Diagnosis. *Advances in Intelligent Systems and Computing*, 11–19. doi: [https://doi.org/10.1007/978-3-319-48923-0\\_2](https://doi.org/10.1007/978-3-319-48923-0_2)
35. Subbotin, S. (2019). A random forest model building using a priori information for diagnosis. *Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019)*, 2353, 962–973.