Intelligent Decision Support System for Industrial Site Classification: A GIS-Based Hierarchical Neuro-Fuzzy Approach

Aleksandar Rikalovic, Ilija Cosic, Ruggero Donida Labati, and Vincenzo Piuri, *Fellow, IEEE*

***Abstract*—Industrial site selection involves a large number of criteria and location alternatives; consequently, the selection process leads to extended decision-making periods and requires complex knowledge management, classification, and analysis skills. The selection criteria are generally described by a number of different features expressed as both quantitative and qualitative measures that can involve some uncertainty. Moreover, the goals considered in the selection process are frequently nonlinearly related to the criteria; thus, they give rise to an optimization problem that is nonlinear with respect to each goal. Consequently, decision making requires appropriate support to enable efficient data optimization and classification under uncertainty before the final selection of an industrial site is made. This paper presents a novel intelligent decision support system for classifying industrial sites according to quality criteria estimated by exploiting a geographic information system, expert knowledge, and machine learning techniques. The proposed system is based on a geographic information system for generating location alternatives and a hierarchical neuro-fuzzy approach for site classification. The neuro-fuzzy method is based on a knowledge base designed by experts in the field and uses a neural approach to tune the parameters of the membership functions. Experimental results on real-world problems show that the proposed system provides accurate results for industrial site classification at the local level (micro locations).**

***Index Terms*—Adaptive neuro-fuzzy inference systems (ANFISs), artificial neural networks (ANNs), fuzzy inference systems (FISs), geographic information systems (GISs), industrial site classification (ISC), intelligent decision support system (IDSS).**

I. Introduction

INDUSTRIAL site selection is a key issue in corporate strategies [1] in which decision makers identify and evaluate location alternatives on the basis of technical, economic, social, environmental, and political criteria, possibly resulting in conflicting objectives [2], The goals considered during the selection process may be nonlinearly related to the criteria; thus, they can lead to a nonlinear optimization problem with respect to each goal [3]. Consequently, industrial site selection is a complex, multiobjective [4] and multicriteria [5] optimization problem.

Industrial location analysts increasingly strive to optimize various decision criteria that may conflict and to present a number of possible sites, each with specific advantages and limitations [6]. This decision-making process is overloaded with information and occurs under highly uncertain conditions in which strategic decisions on industrial location have an extremely complex and imprecise nature [7]. In fact, decision makers must make difficult and important decisions concerning industrial locations on weak bases that involve imprecise information and incomplete knowledge [8]. Therefore, the need for fast decision making has increased and there is a growing demand for an appropriate support method for these decision makers.

A number of decision support tools have been used to support industrial site selection including geographic information systems (GISs) [9], multicriteria decision analysis (MCDA) [10], fuzzy expert systems (FISs) [11], genetic algorithms (GAs) [12], artificial neural networks (ANNs) [13], and various combinations of these approaches [14], [15]. GISs are mainly used to collect spatial (geographic) data and support their spatial analyses. MCDA has been widely used for structuring and solving problems involving multiple (possibly conflicting) criteria that must be evaluated when making decisions. FIS are used to solve nonlinear optimization problems by expressing the available knowledge in a fuzzy manner. GA are often used for solving p-median problems, thus generating optimal solutions. Finally, ANNs are mainly used to recognize patterns; these help in finding relevant sites at new locations. These tools have played important roles in solving the site selection problem, but each has its limits in dealing with all the relevant criteria and reaching the most suitable solutions.

The key to making smarter long-term decisions for tasks such as industrial site selection is to optimize the large number of possible location alternatives. To simplify the choices that a human must make, it is therefore necessary to initially classify industrial sites and intelligently reduce the number of location alternatives. Typically, what decision makers truly need is a limited amount of accurate information about a small, carefully selected set of location alternatives. The initial industrial site selection task consists of acquiring discriminative data for a set of sites in a region of interest that meet the basic requirements defined by given industrial site selection criteria. The second

Manuscript received December 12, 2016; revised February 28, 2017; accepted April 13, 2017. Date of publication May 5, 2017; date of current version August 23, 2018. This work was supported in part by the Urban Institute of Vojvodina, Serbia. *(Corresponding author: Aleksandar Rikalovic.)*

A. Rikalovic and I. Cosic are with the Department of Industrial Engineering and Management, University of Novi Sad, Novi Sad 21000, Serbia (e-mail: a.rikalovic@uns.ac.rs; ilijac@uns.ac.rs).

R. D. Labati and V. Piuri are with the Department of Computer Science, Università degli Studi di Milano, Crema 26013, Italy (e-mail: ruggero.donida@unimi.it; vincenzo.piuri@unimi.it).

Digital Object Identifier 10.1109/JSYST.2017.2697043

step consists of classifying these candidate industrial sites, thus providing a good set of location alternatives for decision making. The final step is the human decision.

To address all the above aspects in an integrated way, we propose a novel human–machine collaborative system for industrial site classification, which performs spatial data mining, provides efficient data optimization, and mitigates uncertainties by combining various complex decision support systems synergistically. The proposed system consists of an innovative intelligent decision support system for industrial site classification (IDSS-ISC) based on a GIS for generating location alternatives and a hierarchical neuro-fuzzy system for site classification. The neuro-fuzzy system is based on a knowledge base designed by experts in the field and uses a neural approach to tune the parameters of the membership functions. We use fuzzy logic to cope with the contradictory nature of the data and the presence of uncertainty in the data. Different characteristics of a candidate industrial site can, in fact, be contradictory with regard to the decision process (e.g., a site may offer relatively little space but a high availability of infrastructures), and the features used for decision making are affected by uncertainty (which is higher for qualitative features).

This paper is organized as follows. Section II summarizes the available literature concerning industrial site selection and classification, emphasizing the role of MCDA, GISs, and intelligent systems. This section also provides some basic concepts that underlie computational intelligence approaches for classification and neuro-fuzzy techniques. Section III presents the proposed innovative IDSS-ISC. Section IV presents an experimental validation of the methodology applied to the Vojvodina region (Serbia) as a study case. Section V presents a detailed discussion and comparison between the proposed neuro-fuzzy system and FIS for industrial site selection. This section also provides a discussion and analysis of the application of the proposed approach by experts in the field. Section VI derives some conclusions and provides some directions for future research.

II. Background

This section reviews the related studies on industrial site selection, presents an overview of hierarchical classification strategies, and provides a brief description of neuro-fuzzy methods.

A. Industrial Site Selection

Determining locations for industrial facilities is one of the most important and far-reaching strategic decisions faced by decision makers [16]. Due to the complexity of the task, the industrial site selection process is usually divided into two phases: the selection of a macro location and, then, within this area, the selection of a micro location [17], [18]. The macro location is a geographical area (typically the administrative region) that meets the basic requirements for building and developing the facility with minimal operating costs. The micro location is the specific place within the macro location that meets specific technical, infrastructural, and working process requirements for the planned facility. However, most previous works that addressed the industrial (facility) location problem did not discuss micro location selection: they focused solely on macro location selection.

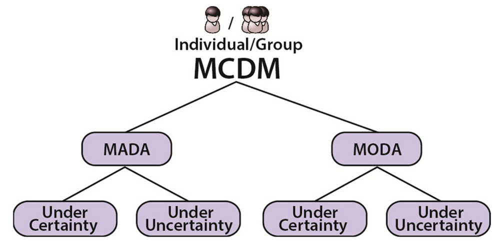

Fig. 1. Multicriteria decision making (MCDM).

Over the years, the challenge of locating industrial facilities has attracted the attention of researchers and practitioners in numerous disciplines. Traditionally, industrial location analysis was considered in the framework of operational research [19]; however, GISs [20] and intelligent systems [21] have increasingly been employed for industrial location analysis.

Most previous works have mainly focused on decision support models based on MCDA [19]. Multicriteria decision making (MCDM) problems (see Fig. 1) can be classified based on the major components of MCDA: multiobjective decision analysis versus multiattribute decision analysis, individual versus group decision-maker problems, and decision under certainty versus decision under uncertainty. Multiattribute techniques are discrete methods that assume that the number of alternatives is explicit, while multiobjective methods are mathematical programming, model-oriented techniques in which the alternatives—identified by solving a multiobjective mathematical programming problem—must be generated [19]. According to Keeney [22], the two major approaches are the alternative-focus approach, which aims at generating the decision alternatives, and the value-focus approach, which uses the values (attributes) as fundamental elements in the decision analysis.

It has been estimated that 80% of the data used by managers and decision makers for industrial site selection are geographical (spatial) in nature [23]. This fact makes GISs central in industrial location problems. GISs use powerful tools designed for spatial analysis that provide the functionality to capture, store, query, analyze, display, and output geographic information [20]. In industrial location, science GISs are irreplaceable in spatial analysis, generating location alternatives and evaluation [24]. GISs have been used to identify suitable areas for industry by using a multicriteria evaluation method based on Boolean logic to produce suitability maps, which are map-based graphic representations showing the suitability of each location for the envisioned industry [25].

B. Industrial Site Selection Using Computational Intelligence Techniques

The industrial location problem requires complex knowledge management and comprehensive analysis. A comprehensive method for industrial site selection [15], based on an IDSS for industrial location criteria analysis, a GIS for generating location alternatives, and a spatial decision support system for evaluating location alternatives is a good example of combining various complex interacting decision support systems synergistically.

Unfortunately, the approaches described above rarely adopt efficient data optimization strategies and fail to consider the uncertainty that is always present in the information acquired for the analysis of geographical sites. When there are a large number of feasible location alternatives—that is, when many decision variables and many possible values for each of them exist—it may no longer be practical to identify and simulate all feasible combinations of decision-variable values or even a small percentage of them.

ISC must be performed efficiently, addressing all the above aspects in an intelligent and coordinated way. To design an IDSS-ISC, we should first understand how a person would solve the problem and, then, understand how to translate this reasoning into something that a computer can execute. Finally, we need to develop software applications able to mimic the human reasoning for the considered application case. For this purpose, we need to create structured, quantitative data about complex technical issues by using approximations and classifications, by tolerating ranges and uncertainties in numeric measurements, by making subjective assessments, and by looking for patterns and clusters across different categories of data.

Studies exist in the literature that describe the use of machine learning techniques to make quality assessments of industrial sites. For example, ANNs are used in [26]. However, that study did not consider the use of *a priori* human knowledge to simplify the learning process. In [27], an approach based on neural networks uses fuzzy logic to create a simplified representation of the features that enable this approach to increase the classification accuracy. However, this method requires the design of a specific fuzzy system for each new application scenario. In contrast, our approach can be easily adopted in new operational scenarios by training the proposed classification method for a novel application case.

C. Classification Methods

Machine learning techniques support the implementation of classification methods able to learn from examples and adapt their parameters according to the envisioned problem [28]–[30]. It is possible to distinguish classifiers that use statistical approaches [e.g., the linear, quadratic, *k*-nearest neighbor (KNN) classifiers] from those that use computational intelligence approaches. The latter are able to learn more complex nonlinear models [e.g., feed-forward neural networks (FFNN), support vector machines, and neuro-fuzzy methods].

A single classifier may prove unsatisfactory for complex problems when the number of classes involved is greater than two. Hierarchical strategies that employ a pool of classifiers are widely used in the literature to improve classification accuracy [31]. In this context, it is possible to distinguish different strategy categories, including the flat classification approach, local classifier approaches, and the global classifier approach. In this paper, we consider basic hierarchical approaches pertaining to the first two categories because they are the methods used most often in the literature. These hierarchical neuro-fuzzy classifiers rely on expert knowledge for the case in which sufficient training data are not available to achieve automatic classification with satisfactory accuracy. Further details on these techniques are provided in the next section.

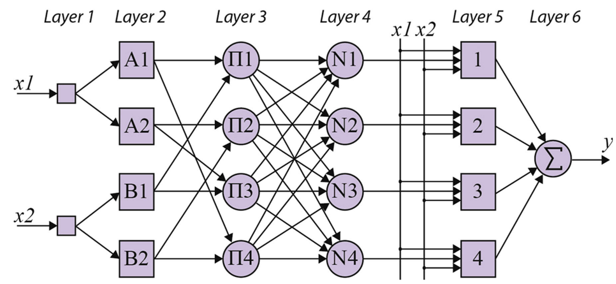

Fig. 2. Adaptive neuro-fuzzy inference system (ANFIS) [41].

D. Neuro-Fuzzy Method

As shown in [18], the fuzzy inference system (FIS) [32] can simplify the task of industrial site selection for human experts. An FIS maps a given input to an output using fuzzy logic. An FIS can be divided into three main components: the fuzzifier, the knowledge base, and the defuzzifier [32]. The fuzzifier and the defuzzifier map external information into fuzzy sets and fuzzy sets into crisp values, respectively. The knowledge base consists of the set of rules that simulates the reasoning of human experts.

In many scenarios, designing effective FISs can be particularly difficult for human experts because the design process requires tuning a large number of parameters. Therefore, adaptive techniques for tuning FIS implementations from examples have been widely used in previous studies [33], [34]. These techniques include methods that use fuzzy logic and ANNs [35] and methods that use fuzzy logic and GA [36], [37]. One of the most commonly used techniques is called the adaptive neuro-fuzzy inference system (ANFIS) [38], which adopts ANNs to tune the parameters of the FIS fuzzifier and defuzzifier.

The ANFIS architecture is shown in Fig. 2. It consists of the following six layers.

- *Layer 1 (input layer)* takes in input a set of m inputs.
- *Layer 2 (fuzzification layer)* fuzzifies the inputs according to a set of membership functions.
- *Layer 3 (rule layer)* is the knowledge base composed of a set of if–then rules. The rules are expressed using the Takagi–Sugeno model [39], which supports solving problems with multiple inputs and outputs [40]. The Takagi–Sugeno model expresses fuzzy rules using the following schema: IF x IS A and y IS B THEN $z = f(x,y)$, where A and B are fuzzy sets in the antecedent, while $z = f(x,y)$ is a crisp function in the consequent. Usually $f(x,y)$ is a polynomial of the input variables x and y, but it can be any function as long as it can appropriately describe the output of the model within the fuzzy region specified by the antecedent of the rule.
- *Layer 4 (normalized firing strengths)* evaluates every rule. Each node in the layer receives inputs from Layer 3 and computes the ratio of the firing strength of a given rule.
- *Layer 5 (defuzzification layer)* evaluates the consequent parameters of the rules.

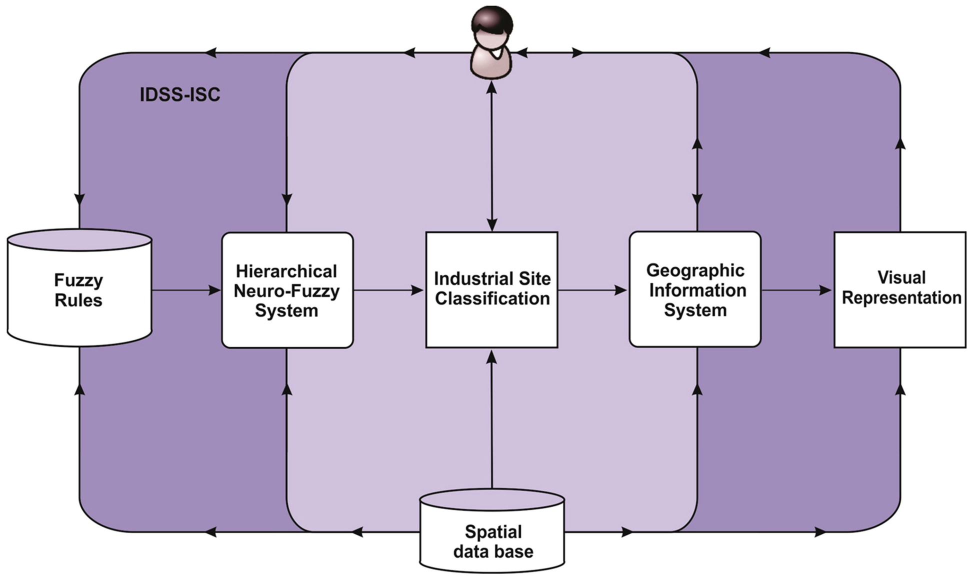

Fig. 3. Architecture of the intelligent decision support system for industrial site classification.

- *Layer 6* is the output layer, which computes the final output of the ANFIS.

III. IDSS-ISC

In this section, we introduce our novel intelligent human–machine collaborative system for industrial site classification, called the IDSS-ISC, which is a GIS-based hierarchical neuro-fuzzy approach.

The proposed IDSS-ISC is based on a set of interacting decision support systems: a GIS for generating location alternatives and a hierarchical neuro-fuzzy system for site classification (see Fig. 3). The developed GIS is a multilayer framework, based on a spatial database created from an environmental analysis that generates location alternatives. The hierarchical neuro-fuzzy system for ISC is based on a set of ANFIS [38] classifiers, which rely on expert knowledge for the case in which sufficient training data are not available to achieve automatic classification with satisfactory accuracy.

The proposed classification system can be divided into four sequential steps: *Problem definition, Generating location alternatives, Expert analysis*, and *Classification*. In the following sections, we will analyze each of these steps.

A. Problem Definition

ISC is a general process related to categorization in which features associated with a specific location must be recognized, defined, and evaluated.

The first phase of ISC involves recognizing and defining the relevant features of the classification problem (i.e., the most important location characteristics that influence the location quality for the envisioned industry).

As an example, to test our IDSS-ISC, we adopted the features for industrial location assessment identified in [41] for the case of Serbia by the Urban Institute of Vojvodina, Serbia (see Table I). For our study, we used ten features (i.e., the most influential features for assessing the suitability of an industrial site). Features i_{11} and i_{12} are not included in our study: i_{11} does not present sufficient discriminant ability because it has almost the same value in all the locations considered in our experiments, whereas i_{12} presents a strong correlation with i_{10} and is therefore unnecessary. These classification features are expressed in both quantitative and qualitative ways, thereby resulting in a nonlinear optimization problem. Some data (i_1, i_5, i_6, i_7, and i_8) are collected and treated as directly measurable quantities, even though they may be affected by uncertainty due to the measurement process. However, several characteristics of the classification problem (i_2, i_3, i_4, i_9, and i_{10}) can be described only in a qualitative manner by using natural language, which is an intrinsically approximate and imprecise representation of information.

TABLE I
INDUSTRIAL LOCATION FEATURES (URBAN INSTITUTE OF VOJVODINA STUDY)

No.	Feature
i_1	*Site size*
i_2	*Property*
i_3	*Limitations*
i_4	*Land use*
i_5	*Occupancy index*
i_6	*Construction index*
i_7	*Minimum site size for construction*
i_8	*Minimum site width for construction*
i_9	*Infrastructure*
i_{10}	*Conditions for obtaining location permit*
i_{11}	*Maximum number of floors*
i_{12}	*Urban plan*

B. Generating Location Alternatives

To provide a useful set of location alternatives, we developed a GIS to mine data and generate location alternatives.

Our GIS is based on a spatial database that contains data for all the selected features in the region of interest. Spatial data collection includes GIS maps, satellite and aircraft images, and descriptive data related to the observed location. Geographic data are obtained by remote sensing, representations of existing data (infrastructure), collecting geolocated data with a GPS, or scanning and digitizing maps. Collected data are first analyzed to ensure consistency and then stored [42]. This step is the most demanding in terms of time and costs.

The spatial data mining process focuses on the process of discovering interesting and potentially useful patterns for generating industrial locations alternatives from a spatial dataset by using a multilayer framework [i.e., an infrastructure where the spatial database is organized in multiple layers (thematic maps) that represent the selected features layer by layer (map by map)].

To mine the data and generate industrial sites, we developed a GIS using ArcGIS [42] and spatial database. Using our GIS, we first separately analyzed each layer of the spatial data in the region of interest. Then, we overlapped specific layers to obtain useful information for the subsequent expert analysis.

C. Expert Analysis

To obtain ISC and provide necessary data for computational intelligence techniques (the FIS and neural training), we asked leading experts from the field to evaluate the set of candidate industrial sites from the region of interest. Working together, experts from the field analyzed the spatial data acquired using the developed GIS and evaluated the technical suitability of industrial sites.

D. Classification

To evaluate the quality of industrial sites, we propose a classification system that assigns ranking scores to each site using a hierarchical classification strategy. In our system, each discrete score value corresponds to a distinct class.

The proposed system is based on a set of ANFIS classifiers. ANFIS classifiers are more suitable than FIS in cases in which a sufficient number of training examples are available and the design of the system requires tuning a large number of parameters.

With respect to other supervised learning techniques in the literature, ANFIS provides the advantage of being able to mix prior knowledge with machine learning algorithms. While future automatic learning methods may be better able to learn the relationships in the input compared to fuzzy rules expressed by human experts for large sets of data, the fuzzy knowledge base used by ANFIS can increase the classification performance in cases in which it is not possible to collect significantly large datasets [38]. For this reason, classification methods based on ANFIS are particularly suitable for the quality evaluation of industrial sites, in which collecting data is a complex, expensive, and time-consuming task.

The proposed neuro-fuzzy method consists of a two-class classification technique based on an FIS designed by a human expert.

The adopted hierarchical classification strategy is designed to take advantage of classifiers that estimate rank scores expressed in increasing order.

Considering *n* classes, this strategy uses *n* two-class classifiers that return a discrete value o_i, which could be equal to 0 or 1. The approach evaluates the values o_i in a cascade fashion, starting from $i = 1$, and computes the output class as follows:

$$\text{class} = \begin{cases} 0, & \text{if } o_1 = 0 \\ & \dots \\ n-1, & \text{if } o_n = 0 \\ n, & \text{othewise.} \end{cases}$$

Each classifier used by the proposed approach consists of an ANFIS classifier that computes a rank value representing a class starting from a set of features describing an industrial site.

ANFIS classifiers can be realized by starting from the knowledge expressed in an FIS or by using techniques for inferring a knowledge base from a set of training data [41].

We designed the proposed ANFIS classifiers starting from an expert system that we implemented from a knowledge base created in collaboration with other experts in the field. Similar to [15], the expert system consists of an FIS that assigns weights to each industrial site. The main difference between this approach and [15] lies in the fact that the output value of the FIS is thresholded to obtain a two-class classifier in which each rank level represents a class. We implemented this FIS using the Takagi–Sugeno model. The ANFIS is based on the same knowledge as the FIS and uses a neural approach to tune the parameters of the membership functions.

We then created an ANFIS by integrating the realized FIS with neural networks to obtain an adaptive classification method able to tune its parameters from examples. In the following, we provide a description of the main characteristics of the designed ANFIS at Layer 1, Layer 2, Layer 3, and Layer 6, which are the layers that exhibit the most important differences in ANFIS implementations. Layers 4 and 5 are implemented as described in [38].

- At Layer 1, the proposed ANFIS takes a generic set of features as input. Our implementation uses a set of ten features (features i_1 to i_{10}, described in Table II).
- Layer 2 is based on Gaussian membership functions. Table II reports the values of the membership functions designed for each of the input feature. Fig. 4 shows an example of the Gaussian membership functions designed for an input feature.
- At Layer 3, the ANFIS uses the fuzzy rules of the knowledge base for approximate reasoning. To create the knowledge base, a group of experts cooperated in defining a unique set of 36 rules applicable to different cases of study. Table III presents fuzzy rules from the knowledge base designed by human experts based on their experience in the field. The AND operator is implemented by using the product method, while the OR operator is implemented by using the probabilistic OR function (algebraic sum).
- Layer 5 uses the weighted average defuzzification method.

TABLE II
FUZZIFICATION

ID	Feature	Range	MFs
i_1	Site size	[0 100]	Very Low (VL), Low (L), Medium (M), High (H), and Very High (VH)
i_2	Property	[3 5]	Private (P) and Government (G)
i_3	Limitations	[1 5]	Forbidden Construction (FC), Limited Construction (LC), Partly Forbidden-Limited Construction (PLC), Small Part Forbidden-Limited Construction (SPLC), No Restrictions (NR)
i_4	Land use	[1 5]	Logistics-Other (LO), Services-Warehouses (SW), Technology Park (TP), Manufacturing-Free Zone(M)
i_5	Occupancy index	[40 80]	Low (L), Medium (M), and High (H)
i_6	Construction index	[0.8 2.1]	Low (L), Medium (M), and High (H)
i_7	Minimum site size for construction	[100 16 000]	Very Low (VL), Low (L), Medium (M), High (H), and Very High (VH)
i_8	Minimum site width for construction	[10 60]	Very Low (VL), Low (L), Medium (M), High (H), and Very High (VH)
i_9	Infrastructure	[1 5]	Very Low (VL), Low (L), Medium (M), High (H), and Very High (VH)
i_{10}	Conditions for obtaining location permit	[1 5]	Urban-Planning-Preparation-Needed (UPN), Based-On-Urban-Plan (BUP), Directly-From-Urban-Plan (DUP)

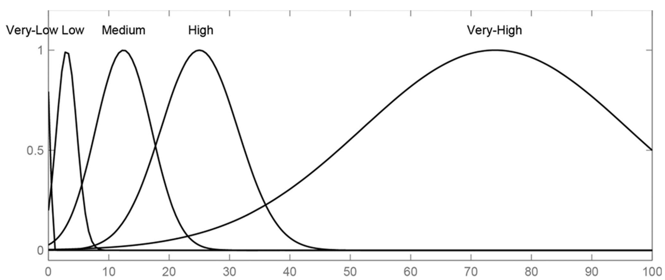

Fig. 4. Example of the Gaussian membership functions used in the proposed system.

- Layer 6 returns a floating-point output in a range from 1 to n.

In the proposed approach, the ANFIS output is normalized to integer numbers that represent classes using a rounding function.

The ANFIS can be trained in a supervised manner. To train the proposed ANFIS classifier, we used the back-propagation method as well as a hybrid learning approach combining least-squares estimation and back-propagation [38]. To provide a visual representation of the classified industrial sites in the observed geographical area, we used ArcGIS [42] and the spatial database developed in Section III-B.

IV. Experimental Results

This section presents the application scenario, the testing protocol, the performed experiments, and the obtained results to show the use of the proposed system and to evaluate its effectiveness and reliability. In particular, we evaluate the performance of the proposed system on different datasets representing application scenarios. The quality level of industrial sites is expressed by two or five classes. For each application scenario, the proposed system is compared with other techniques available from the literature. The proposed system achieved the best accuracy for each performed test and proved to be suitable for real application scenarios.

A. Application Scenario

In this section, we describe an experimental application of the proposed IDSS-ISC system for ISC described in Section III. To demonstrate the efficiency and the effectiveness of the proposed approach, we show its use in a micro location analysis in the Vojvodina region, Serbia. In this example, the ISC was performed for the manufacturing industry.

In this case, the starting point of the classification process consists of defining the features shown in Table I [41]. In the spatial data mining process using our GIS, we analyzed spatial data in the region of interest separately for each map as well as by overlapping layers of thematic maps to discover useful information. Fig. 5 shows the spatial data mining process in the GIS environment using five maps: industrial zones, land use, property, infrastructure, and limitations. The last map [see Fig. 5(f)] shows the industrial sites generated during the spatial data mining process that were selected for classification. In total, 450 industrial sites were obtained in the Vojvodina region, Serbia.

To provide target data for the FIS and the training classifiers, we analyzed each site in cooperation with a team of other experts in industrial site selection. For each site, we estimated the technical suitability for locating industry from one to five, where one denotes very low suitability, two denotes low suitability, three denotes average suitability, four denotes high suitability, and five denotes very high suitability. During the evaluation process, the experts agreed on a single group opinion for each site according to the available information. Table IV shows the results of the industrial site classifications provided by these human experts for the 450 sites.

B. Evaluation of the Proposed Classifier

We performed tests to discover the best configuration for the proposed neuro-fuzzy classifier and compared its performance with other classification methods available from the literature, considering application scenarios in which the quality level of industrial sites is expressed by two or five classes.

The features and the corresponding labels pertaining to the training set are used to train the classifier. The features pertaining to the validation set are then used as input to the classifier and the corresponding labels are compared with the expert classification results to evaluate the accuracy of the learning method. We used all the data labeled by experts in industrial site selection to create the following two datasets.

- DB_5C: This dataset includes the features extracted for all 450 of the labeled industrial sites. The industrial sites

TABLE III
KNOWLEDGE BASE

#	IF–THEN Rules
1	IF $(i_1 \rightarrow VL) \vee (i_3 \rightarrow FC) \vee (i_4 \rightarrow LO) \vee (i_9 \rightarrow VL) \vee (i_{10} \rightarrow UPN)$ THEN $(o_1 \rightarrow l_1)$
2	IF $(i_1 \rightarrow VL) \vee (i_3 \rightarrow LC) \vee (i_4 \rightarrow SW) \vee (i_9 \rightarrow VL) \vee (i_{10} \rightarrow UPN)$ THEN $(o_2 \rightarrow l_2)$
3	IF $(i_9 \rightarrow L) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_3 \rightarrow l_3)$
4	IF $(i_9 \rightarrow M) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_4 \rightarrow l_4)$
5	IF $(i_1 \rightarrow\neg VL) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow M) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_5 \rightarrow h_5)$
6	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_6 \rightarrow h_6)$
7	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_7 \rightarrow h_7)$
8	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_8 \rightarrow h_8)$
9	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_9 \rightarrow h_9)$
10	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{10} \rightarrow h_{10})$
11	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{11} \rightarrow h_{11})$
12	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{12} \rightarrow h_{12})$
13	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{13} \rightarrow h_{13})$
14	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{14} \rightarrow h_{14})$
15	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{15} \rightarrow h_{15})$
16	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{16} \rightarrow h_{16})$
17	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{17} \rightarrow h_{17})$
18	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{18} \rightarrow h_{18})$
19	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{19} \rightarrow h_{19})$
20	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{20} \rightarrow h_{20})$
21	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{21} \rightarrow h_{21})$
22	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{22} \rightarrow h_{22})$
23	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{23} \rightarrow h_{23})$
24	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{24} \rightarrow h_{24})$
25	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg LC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{25} \rightarrow h_{25})$
26	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg SW) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow H) \wedge (i_{10} \rightarrow BUP)$ THEN $(o_{26} \rightarrow h_{26})$
27	IF $(i_9 \rightarrow M) \wedge (i_{10} \rightarrow UPN)$ THEN $(o_{27} \rightarrow l_{27})$
28	IF $(i_9 \rightarrow H) \wedge (i_{10} \rightarrow UPN)$ THEN $(o_{28} \rightarrow l_{28})$
29	IF $(i_3 \rightarrow FC) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{29} \rightarrow l_{29})$
30	IF $(i_3 \rightarrow LC) \wedge (i_9 \rightarrow VH) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{30} \rightarrow l_{30})$
31	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow L) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{31} \rightarrow h_{31})$
32	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow L) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{32} \rightarrow h_{32})$
33	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow P) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow M) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{33} \rightarrow h_{33})$
34	IF $(i_1 \rightarrow\neg VL) \wedge (i_2 \rightarrow G) \wedge (i_3 \rightarrow\neg FC) \wedge (i_4 \rightarrow\neg LO) \wedge (i_5 \rightarrow\neg L) \wedge (i_6 \rightarrow\neg L) \wedge (i_7 \rightarrow\neg VL) \wedge (i_8 \rightarrow\neg VL) \wedge (i_9 \rightarrow M) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{34} \rightarrow h_{34})$
35	IF $(i_9 \rightarrow H) \wedge (i_{10} \rightarrow UPN)$ THEN $(o_{35} \rightarrow l_{35})$
36	IF $(i_9 \rightarrow L) \wedge (i_{10} \rightarrow DUP)$ THEN $(o_{36} \rightarrow h_{36})$

Note: This knowledge base was designed by human experts to be used by ANFIS. Neural networks are used to tune the parameters of the membership functions and do not modify the if–then rules.
$\rightarrow$ = IS; $\wedge$ = AND; $\vee$ = OR; $\neg$ = NOT.

are divided into five classes corresponding to the quality scores expressed by the experts.

- DB_2C: This dataset includes the features extracted for all 450 of the labeled industrial sites. The industrial sites are divided into two classes, where class 1 corresponds to scores ≤ 3 and class 2 corresponds to scores > 3.

To train and validate all the considered classification methods, we used an N-fold cross-validation scheme, where $N = 10$ [28]. In particular, for each fold, the parameters of the membership functions of the proposed ANFIS were tuned by exploiting the learning capability of the neural component of the method, which was trained using 90% of the samples of the considered dataset and validated using the remaining 10% of the samples. Training and validation were iterated ten times to ensure that all samples were used for training and to validate the classifiers. Each training set was created by performing random permutations of the data. The results reported in this paper represent the accuracy achieved for the validation sets.

The number of parameters tuned in the training step is 162, which corresponds to the number of values describing the Gaussian membership functions multiplied by the number of membership functions and summed with the number of if–then rules ($3 \times 42 + 36$).

We evaluated the accuracy of every method in terms of total classification error, standard deviation of the classification error, and confusion matrix [29].

We used all the features available to perform our tests because feature reduction strategies yielded unsatisfactory results for the considered dataset. For example, forward feature selection [29] with a linear classifier and five features yielded a classification error of greater than 35% for the five-class classification problem.

1) Two-Class Problem: To validate the effectiveness of the proposed system in application scenarios in which the quality level of industrial sites is expressed by two classes, we evaluated the performance of the proposed neuro-fuzzy classification approach in different configurations and compared its performance with other classification methods available from the literature. The proposed approach achieved a better classification performance than the compared methods.

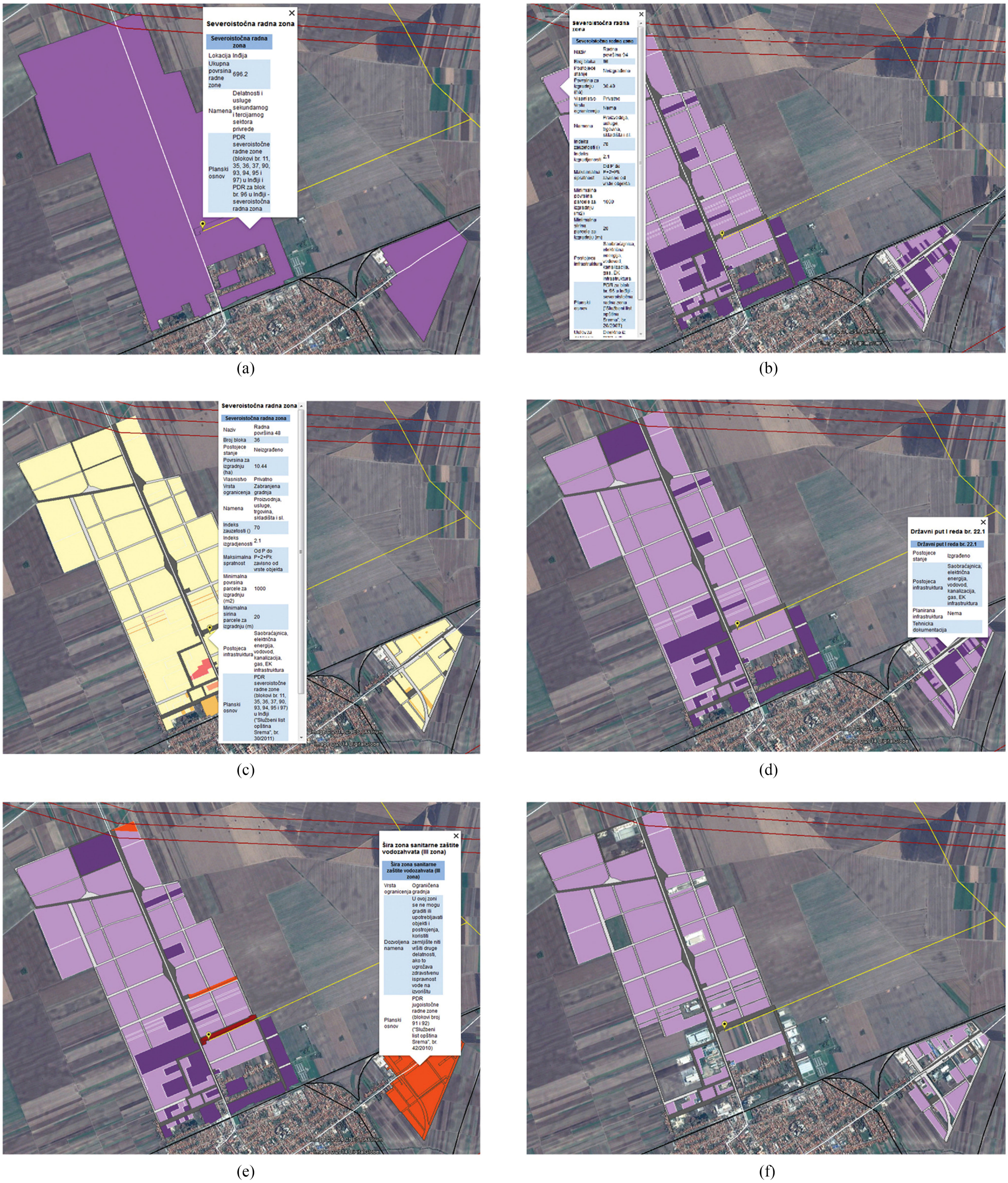

Fig. 5. Spatial data mining process: (a) industrial zones, (b) land use, (c) property, (d) infrastructure, (e) limitations, and (f) generated industrial locations.

We also evaluated different configurations of the proposed neuro-fuzzy classifier for these datasets. In particular, we tested different types of training methods (the back-propagation method and the hybrid learning [37]) and different numbers of training epochs (100, 500, and 1000). For each test, we obtained the best results by using the hybrid learning method with 500 learning epochs. The results reported below refer to this configuration.

To validate the accuracy of our system, we compared its classification accuracy with those of other well-known techniques

TABLE IV
INDUSTRIAL SITE CLASSIFICATION: EXPERT RATING

Rate Description	Rate	Number of industrial sites
Very low suitability	1	44
Low suitability	2	65
Average suitability	3	222
High suitability	4	82
Very high suitability	5	37

TABLE V
CLASSIFICATION ACCURACY OF DIFFERENT CLASSIFIERS FOR DB_2C

Classifier	Total	Std
LDC	0.113	0.317
QDC	0.080	0.272
kNN-1	0.064	0.246
FFNN-5	0.020	0.140
FIS	0.017	0.132
ANFIS-Clustering	0.047	0.211
Proposed ANFIS	0.013	0.114

Note: Total = total classification error; Std = standard deviation of the classification error; LDC = Linear Bayes Normal Classifier; QDC = Quadratic Bayes Normal Classifier; kNN-1 = k-Nearest Neighbor with $k = 1$; FFNN-5 = feed forward neural networks with five nodes in the hidden layer; FIS = the fuzzy inference system used by the proposed neuro-fuzzy approach but tuned by a human expert; ANFIS-Clustering = a neuro-fuzzy approach based on a fuzzy inference system created by using the subtractive clustering method; Proposed ANFIS = the proposed neuro-fuzzy approach.

in the literature. The compared methods are statistical and computational intelligence techniques that do not exploit any knowledge of human experts in the field. Comparing the performance of the proposed approach with those of the selected classification methods permits an evaluation of the contribution of a knowledge base designed by field experts for estimating the suitability of possible industrial sites. Specifically, we compared the accuracies of the following techniques:

- linear Bayes normal classifier [43];
- quadratic Bayes normal classifier [43];
- kNN classifier [32] with odd values of the parameter k (1, 3, 5, 7, and 10);
- FFNN [44], with different numbers of log-sigmoidal nodes in the hidden layer (1, 3, 5, 7, 10, 15, 20, 25, and 30) and a linear output node, trained with the back-propagation algorithm;
- ANFIS based on a fuzzy inference system created by using the subtractive clustering method [35].

Table V summarizes the best results achieved by the above methods in their best configuration for DB_2C. The table shows that the proposed neuro-fuzzy system achieved the best classification accuracy for DB_2C, with a total classification error of 1.3%. Table V also shows that the use of a neuro-fuzzy method increased the classification accuracy compared to the FIS tuned by a human expert. This result confirms the advantages and feasibility of the proposed approach for selecting the highest quality industrial sites from a region.

Table VI reports the confusion matrix [28], [29] obtained by the proposed approach, showing that its classification errors are similar for each class and are all close to 0. Therefore, the proposed classification system neither underestimates nor overestimates the quality of industrial sites when the quality level of the sites is expressed by two classes.

TABLE VI
CONFUSION MATRIX OBTAINED BY THE PROPOSED NEURO-FUZZY APPROACH FOR DB_2C

		Classification output	
		Class 1	Class 2
Label	Class 1	0.731	0.004
	Class 2	0.008	0.258

TABLE VII
CLASSIFICATION ACCURACY OF DIFFERENT NONHIERARCHICAL CLASSIFIERS FOR DB_5C

Classifier	Total	Std
LDC	0.284	0.452
QDC	0.336	0.473
kNN-1	0.216	0.412
FFNN-5	0.227	0.419
FIS	–	–
ANFIS-Clustering	0.302	0,459
Proposed ANFIS	0.166	0.373

Note: Total = total classification error; Std = standard deviation of the classification error; LDC = Linear Bayes Normal Classifier; QDC = Quadratic Bayes Normal Classifier; kNN-1 = k-Nearest Neighbor with $k = 1$; FFNN-10 = feed forward neural networks with ten nodes in the hidden layer; FIS = the fuzzy inference system used by the proposed neuro-fuzzy approach but tuned by a human expert; ANFIS-Clustering = a neuro-fuzzy approach based on a fuzzy inference system created by using the subtractive clustering method; Proposed ANFIS = the proposed neuro-fuzzy approach.

2) Five-Class Problem: To validate the effectiveness of the proposed approach in application scenarios in which the quality level of industrial sites is expressed by five classes, we evaluated the performance of nonhierarchical and hierarchical approaches in different configurations. We performed these tests using the DB_5C dataset and obtained better classification results compared to the compared methods available from the literature.

To evaluate the performance of the nonhierarchical configuration, we first tuned the proposed neuro-fuzzy system in the same manner as the tests performed using the DB_2C dataset (see Section IV-D). In this experiment, we also obtained the best results when using the hybrid learning method with 500 learning epochs.

Table VII summarizes the results achieved by the nonhierarchical classification methods described in Section IV-C for DB_5C. The row values for FIS are empty because human experts were not able to design an FIS for a five-class problem due to the prohibitively high complexity. The table shows that the proposed neuro-fuzzy system achieved the best classification accuracy for DB_5C, with a total classification error of 16.6%. It is worth nothing that this result would not be satisfactory in real application scenarios and justify the use of more complex hierarchical approaches for the creation of maps representing the quality of industrial sites in a geographical region.

We then evaluated the performance of the proposed hierarchical system for DB_5C and compared it to different other methods.

TABLE VIII
CLASSIFICATION ACCURACY OF DIFFERENT HIERARCHICAL CLASSIFIERS FOR DB_5C

Hierarchical strategy	Classifier	Total	Std
HA	FFNN-25	0.193	0.395
HB	FFNN-20	0.227	0.419
HA	ANFIS-Clustering	0.231	0.422
HB	ANFIS-Clustering	0.238	0.426
HA	Proposed ANFIS	0.142	0.349
HB	Proposed ANFIS	0.102	0.303

Note: Total = total classification error; Std = standard deviation of the classification error; HA = the flat hierarchical approach; HB = the proposed hierarchical approach; ANFIS-Clustering = a neuro-fuzzy approach based on a fuzzy inference system created by using the subtractive clustering method; Proposed ANFIS = the proposed neuro-fuzzy approach.

In particular, we compared two hierarchical classification strategies to the nonhierarchical classifiers that achieved the best accuracy in the previous tests (see Tables V and VII). In this test, we compared two widely used approaches in the literature, which are as follows.

- HA: The proposed local classifier approach (see Section III-D).
- HB: The flat classifier approach (see Section II-B). Considering n classes, this strategy uses n two-class classifiers returning a continuous value $o_i \in [0,1]$. The approach computes the final classification result as class $=$ argma$x_{i\,=\,1\ldots n}(o_i)$.

Table VIII summarizes the results achieved by the hierarchical classification methods for DB_5C. The table shows that hierarchical classification methods based on the proposed neuro-fuzzy approach clearly outperform the other considered classification methods by about 100%. Moreover, the hierarchical strategy (HA) applied to the proposed neuro-fuzzy approach obtained the best performance, with a total classification error of 10.0%, which can be considered as satisfactory for most of the real application scenarios.

The accuracy increase achieved by the proposed hierarchical neuro-fuzzy system compared with the nonhierarchical neuro-fuzzy approach may have occurred because the FIS that we used as the basis for implementing the proposed ANFIS was designed as a two-class classifier. It presents two membership functions in the output layer, representing class 1 and class 2, respectively. Moreover, the proposed hierarchical strategy HB obtained the best performance for the proposed neuro-fuzzy method because it better exploits the characteristics of the FIS, which was designed to classify industrial zones into two contiguous quality scores (low and high). In comparison, the HA strategy does not exploit the information that is intrinsic in the rank of the output classes. Table IX reports the confusion matrix obtained by the proposed classification system. It shows that the most erroneously classified elements are close to the diagonal of the confusion matrix and therefore represent errors of minor importance in the analysis of the quality of industrial zones in a geographical region because the estimated score classes can differ up to ± 1 from the real ones.

The classification results (five classes) were reported in the GIS environment. We produced the suitability maps by using a color gradient from red to green (see Fig. 6), where red represents a low suitability and green represents a high suitability of the alternative sites.

TABLE IX
CONFUSION MATRIX OBTAINED BY THE PROPOSED HIERARCHICAL NEURO-FUZZY APPROACH FOR DB_5C

		Classification output				
		Class 1	Class 2	Class 3	Class 4	Class 5
Label	Class 1	0,084	0,004	0,000	0,000	0,000
	Class 2	0,008	0,102	0,024	0,000	0,000
	Class 3	0,004	0,035	0,467	0,006	0,002
	Class 4	0,000	0,002	0,002	0,167	0,002
	Class 5	0,000	0,000	0,000	0,008	0,007

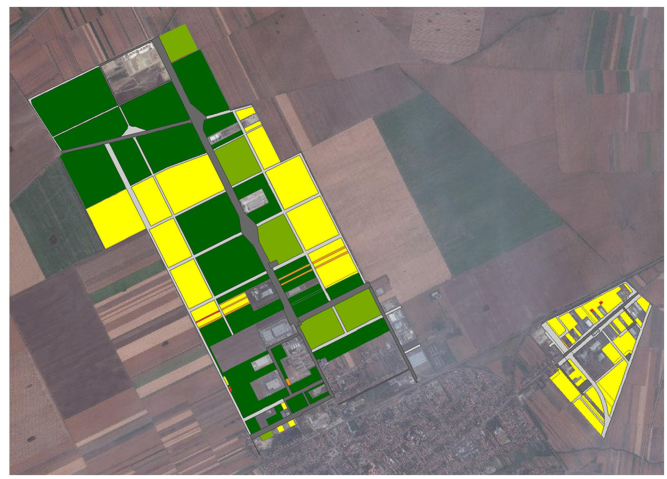

Fig. 6. Industrial site classification: a visual analytics approach.

V. DISCUSSION AND ANALYSIS

In this section, we present a detailed discussion and comparison between the proposed neuro-fuzzy system and FIS for industrial site selection [15].

This section also presents results obtained by experts in ISC by applying the proposed methods and discusses the advantages that human experts possess in performing such analyses.

In terms of accuracy, the proposed ANFIS achieved better performance compared to the corresponding FIS tuned by a human expert for the considered two-class classification problem (see Table V), obtaining a classification error of 1.3%.

Moreover, the proposed ANFIS also obtained better performance than the compared classifiers for the considered five-class classification problem (see Table VIII). It is worth nothing that human experts were not able to design and tune an FIS for this problem because it requires a wide set of if–then rules and complex settings for the membership functions.

Another important advantage of the proposed neuro-fuzzy system is the simplicity of the tuning procedure. Machine learning, in fact, can tune ANFIS classifiers in various application scenarios more easily than FIS. Moreover, the tuned FIS obtained by training ANFIS can be further finetuned by human experts, enabling a higher accuracy to be achieved and saving effort and time in designing an FIS from scratch.

To verify the actual usability of our system, we asked local experts to use it for an industrial site suitability assessment in a real scenario. Foreign investors in Serbia need an average of 13.1 months to decide where to locate an industry and an additional 7.9 months to finalize the investment decision [45]. During this long process, the experts spend time searching for specific data in maps to provide a detailed data analysis and make decisions. Fulfilling the investment decision is also time consuming because experts sometimes miss important limitations of the specific locations.

Our system can help reduce the time required for decision making and realizing the investment decision. Using our system, experts can efficiently analyze large datasets in the spatial environment presented by the GIS, perform spatial data mining, and generate feasible locations without hidden limitations (e.g., construction limitations, absence of adequate plans, unclear ownership structure).

Classifying industrial sites can take weeks without appropriate support. In our study, the experts needed only 3 days to classify 450 candidate industrial sites with the help of our GIS. If the task had involved larger numbers of industrial sites, without the appropriate tools, this process would require far more time. Our hierarchical neuro-fuzzy system can substantially reduce the classification time; for example, in our experiment, we trained and evaluated the accuracy of the proposed classification method in just a few minutes.

Specifically, we implemented the proposed classification system in MATLAB (R2015b 64 bit) using the available software libraries. The tests were performed on a computer with an Intel i7 2.70 GHz processor running the Windows 7 Professional 64-bit operating system. The training time for each of the ten iterations used to validate the ANFIS referenced in Tables VIII and IX was 11.6 s. The time required by the trained ANFIS to classify a novel site was 0.38 ms.

Using our human–machine collaborative system it is easy to understand how the system works; therefore, it is easy to modify the system if necessary. Membership functions in FIS can be tuned optionally by other experts. For example, if the variable range in another region is different from the one previously used to train the classifiers or if the membership functions have a different importance, our FIS can be easily adjusted or changed. Similarly, if some variables need to be changed or more rules need to be added for certain applications, our FIS can be easily adjusted or used as a starting point for further human analysis.

After the industrial site classification, experts can efficiently represent classes using our visual approach with different color ramps in the spatial environment. Such visual representations of classified industrial sites enable humans to understand the content more quickly in the data mining process.

VI. Conclusion

This paper proposed an innovative IDSS-ISC that adopts a GIS-based hierarchical neuro-fuzzy approach.

The proposed system uses a coordinated pair of interacting decision support systems: a GIS to generate location alternatives and a hierarchical neuro-fuzzy system for site classification. We used a GIS for data collection, spatial analysis, generating alternatives, and to produce suitability maps. We presented an innovative human–machine system for this application field to manage classifications under uncertainty and incomplete information by using a set of ANFIS classifiers and a hierarchical information fusion strategy. The proposed neuro-fuzzy classifiers consist of two-class classification techniques based on an FIS designed by human experts. The expert system is based on a Takagi–Sugeno FIS that provides a consistent framework for industrial site suitability assessments. The visual-spatial representation is based on the developed spatial base and reports the obtained results in our GIS.

The results showed that our system for ISC constitutes an efficient and highly accurate tool for decision support. In future research, the IDSS-ISC will be tested on different cases with larger datasets. Moreover, we plan to study strategies such as the use of adaptive genetic fuzzy systems to improve the robustness of the decision support system.

Acknowledgment

The authors would like to thank the Urban Institute of Vojvodina, Novi Sad, Serbia, which provided highly valuable data of industrial locations in Vojvodina. The authors would also like to thank J. Lovric, J. Ignjatic, and D. Moraca for their expert advice on performing industrial site evaluations.

References

[1] L. A. Sabater, A. A. Tur, and J. M. Navarro, "Industrial location, spatial discrete choice models and the need to account for neighbourhood effects," in *Proc. 3rd World Conf. Spatial Econometrics Assoc.*, Barcelona, Spain, Jul. 2008, pp. 393–418.

[2] M. A. Badri, "Dimensions of industrial location factors: Review and exploration," *J. Bus. Public Affairs*, vol. 1, no. 1, pp. 1–26, 2007.

[3] Z. M. Shen, C. Coullard, M. S. Daskin, and A. Joint, "A joint location-inventory model," *Transp. Sci.*, vol. 37, no. 1, pp. 40–55, 2003.

[4] M. Makowski, *Multi-Objective Decision Support Including Sensitivity Analysis, Encyclopedia of Life Support*. Paris, France: EOLSS Publishers, 2002, p. 24.

[5] W. Atthirawong and B. MacCarthy, "An application of analytical hierarchy process to International location decision making," in *Proc. 7th Annu. Cambridge Int. Manuf. Symp.*, 2002, pp. 1–18.

[6] J. R. Eastman, H. Jiang, and J. Toledano, "Multi-criteria and multi-objective decision making for land allocation using GIS multi attributes analysis for land-use management," *Environ. Manage.*, vol. 9, pp. 227–251, 1998.

[7] S. Ghazinoory, Zadeh, and A. Memariani, "Fuzzy SWOT analysis," *J. Intell. Fuzzy Syst.*, vol. 18, pp. 99–108, 2007.

[8] I. Rawabdeha, A. Al-Refaie, and H. Arabiyatd, "Developing a fuzzy logic decision support system for strategic planning in industrial organizations," *Adv. Technol. Sci.,* vol. 2, pp. 14–23, 2013.

[9] R. L. Church, "Geographical information systems and location science," *Comput. Oper. Res.*, vol. 29, no. 6, pp. 541–562, 2002.

[10] P. Y. Chang and H. Y. Lin, "Manufacturing plant location selection in logistics network using analytic hierarchy process," *J. Ind. Eng. Manage.*, vol. 8, no 5, pp. 1547–1575, 2015.

[11] A. Rikalovic and I. Cosic, "A fuzzy expert system for industrial location factor analysis," *Acta Polytech. Hung.*, vol. 12, no. 2, pp. 33–51, 2015.

[12] O. Alp, E. Erkut, and Z. V. I. Drezner, "An efficient genetic algorithm for the p-median," *Ann. Oper. Res.*, vol. 122, pp. 21–42, 2003.

[13] T. Yanar and Z. Akyüre, "Artificial neural networks as a tool for site selection within GIS," in *Proc. Int. Symp. Spatial Data Quality*, Jun. 2007.

[14] E. Fataei and A. Mohammadian, "Industrial state site selection using MCDM method and GIS in Germi, Ardabil, Iran," *J. Ind. Intell. Inf.*, vol. 3, no. 4, pp. 324–329, 2015.

[15] A. Rikalovic, I. Cosic, R. Donida Labati, and V. Piuri, "A comprehensive method for industrial site selection: The macro-location analysis" *IEEE Syst. J.*, vol. 11, no. 4, pp. 2971–2980, Dec. 2017.

[16] R. Bhatnagar and A. S. Sohal, "Supply chain competitiveness: Measuring the impact of location factors, uncertainty and manufacturing practices," *Technovation*, vol. 25, no. 5, pp. 443–456, 2005.

[17] U. Bankhofer, *Industrial Location Management*. Heidelberg, Germany: German Univ. Verlag, 2001.

[18] D. Zelenovic, "Location of production systems," in *The Design of Production Systems*, 2nd ed. Novi Sad, Serbia: Faculty Tech. Sci., 2003, pp. 373–394.

[19] W. Atthirawong and B. MacCarthy, "An application of analytical hierarchy process to International location decision making," in *Proc. 7th Annu. Cambridge Int. Manuf. Symp.*, 2002, pp. 1–18.

[20] J. Malczewski, "GIS-based land-use suitability analysis: A critical overview," *Prog. Plan.*, vol. 62, no. 1, pp. 3–65, 2004.

[21] R. Wyat and H. Hossain, "A GIS-based, genetic algorithm for addressing the facilities-location problem," in *Proc. 8th Int. Conf. Comput. Urban Plan. Urban Manage.*, 2005, pp. 697–709.

[22] R. L. Keeney, *Value-Focused Thinking: A Path to Creative Decision Making*. Cambridge, MA, USA: Harvard Univ. Press, 1992.

[23] J. R. Eastman, "Multi-criteria evaluation and GIS," *Geographic Inf. Syst.*, vol. 1, pp. 493–502, 1999.

[24] A. Rikalovic, I. Cosic, and D. j. Lazarevic, "The role of GIS in industrial location analysis," in *Proc. 24th DAAAM Int. Symp. Intell. Manuf. Autom.* 2013, pp. 1054–1063.

[25] M. Reisi, L. Aye, and A. Soffianian, "Industrial site selection by GIS in Isfahan," in *Proc. 19th Iran Int. Conf. Geoinformat.*, 2011, pp. 1–4.

[26] T. A. Yanar and Z. Akyurek, "Aritificial neural networks as a tool for site selection within GIS," in *Proc. 5th Int. Symp. Spatial Data Quality*, Jun. 2007, pp. 1151–1156.

[27] C. O. Benjamin, S. Chi, T. Gaber, and C. A. Riordan, "Comparing BP and ART II neural network classifiers for facility location" *Comput. Ind. Eng. Archive*, vol. 28, no. 1, pp. 43–50, 1995.

[28] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley-Interscience, 2000.

[29] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer-Verlag, 2006.

[30] L. Rutkowski, *Computational Intelligence: Methods and Techniques*, 1st ed. Berlin, Germany: Springer, 2008.

[31] C. N. Silla, Jr. and A. A. Freitas, "A survey of hierarchical classification across different application domains" *Data Mining Knowl. Discovery*, vol. 22, pp. 31–72, Jan. 2011.

[32] E. Czogala and J. Leski, *Fuzzy and Neuro-Fuzzy Intelligent Systems*. Berlin, Germany: Springer, 2000.

[33] S. Kar, S. Das, and P. K. Ghosh, "Applications of neuro fuzzy systems: A brief review and future outline," *Appl. Soft Comput.*, vol. 15, pp. 243–259, 2014.

[34] H. Ishibuchi, R. Fujioka, and H. Tanaka, "Neural networks that learn from fuzzy if–then rules," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 85–97, May 1993.

[35] N.Walia, H. Singh, and A. Sharma, "ANFIS: Adaptive neuro-fuzzy inference system—A survey," *Int. J. Comput. Appl.*, vol. 123, no. 13, pp. 32–38, 2015.

[36] D. P. Pitambare and P. M. Kamde, "Literature survey on genetic algorithm approach for fuzzy rule-based system," *Int. J. Eng. Res.*, vol. 2, no. 2, pp. 29–32, 2013.

[37] H. Ishibuchi, T. Nakashima, and T. Murata, "Performance evaluation of fuzzy classifier systems for multi-dimensional pattern classification problems," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 29, no. 5, pp. 601–618, Oct. 1999.

[38] J. -R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst. Man Cybern.*, vol. 23, no. 3, pp. 665–685, May 1993.

[39] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-15, no. 1, pp. 116–132, Jan./Feb. 1985.

[40] L. S. Riza *et. al.*, "Fuzzy rule-based systems for Classifcation and regression in R," *J. Statist. Softw.*, vol. 65, no. 6, pp. 1–30, 2015.

[41] J. Lovric, J. Ignjatic, D. Moraca, and A. Rikalovic, "Study of deployment of industrial zones on the territory of AP Vojvodina," Urban Inst. Vojvodina, Novi Sad, Serbia, 2014.

[42] J. McCoy and K. Johnston, *Using ArcGIS Spatial Analyst*. Redlands, CA, USA: Environ. Syst. Res. Inst., 2001.

[43] A. R. Webb and K. D. Copsey, *Statistical Pattern Recognition*, 3rd ed. New York, NY, USA: Wiley, 2011.

[44] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.

[45] "Investment climate in Serbia-investors' perspective," Strategic Marketing Res., USAID, Municipal Econ. Growth Activity, 2008.

Aleksandar Rikalovic is an Assistant Professor with the Department of Industrial Engineering and Management, University of Novi Sad, Novi Sad, Serbia. He has been a Visiting Professor with the Department of Computer Science, Università degli Studi di Milano, Milano, Italy. His main research interests include production systems, intelligent systems, decision support systems, geographic information systems, and multimedia. He has authored or co-authored more than 20 papers in international journals, proceedings of international conferences, books, and book chapters. He has also participated in many international research projects and projects in cooperation with industry.

Ilija Cosic is a Professor Emeritus with the Department of Industrial Engineering and Management, University of Novi Sad, Novi Sad, Serbia. He has been a Visiting Professor with Stuttgart University, Stuttgart, Germany. His main research interests include work study, industrial systems, effective production systems, production systems design, operations management, and production management. He has authored or co-authored 24 books and more than 350 scientific and research papers. With his colleagues he has completed 22 scientific projects and 32 projects in cooperation with industry.

Ruggero Donida Labati received the Ph.D. degree in computer science from the Università degli Studi di Milano, Milano, Italy, in 2013.

He is an Assistant Professor with the Università degli Studi di Milano. His research interests include computational intelligence, signal and image processing, and biometric systems. He has authored or co-authored more than 40 papers in journals, conference proceedings, and books. He co-authored *Touchless Fingerprint Biometrics* (CRC Press, 2015).

Vincenzo Piuri (S'84–M'86–SM'96–F'01) has been a Full Professor of computer engineering with the Università degli Studi di Milano, Milano, Italy, since 2000. He has been a Visiting Professor with The University of Texas at Austin, Austin, TX, USA, and George Mason University, Fairfax, VA, USA. His main research interests include intelligent systems, computational intelligence, machine learning, pattern analysis and recognition. He has authored or co-authored more than 400 papers in international journals, proceedings of international conferences, books, and book chapters.