Optimization of Object-Based Image Analysis with Genetic Programming to Generate Explicit Knowledge from WorldView-2 Data for Urban Mapping

Azmi Rida, Amar Hicham, and Norelyaqine Abderrahim

Abstract

Object-based image analysis techniques give accurate results when a good knowledge base is extracted from remote sensing imagery. Data mining algorithms, especially the evolutionary process, can extract useful knowledge that can be used in different fields. In this paper, object-oriented classification was used, more particularly, the object-based image analysis approach (OBIA) is used to classify a large feature space composed of a very high spatial resolution (VHR) satellite image. The genetic programming (GP) concept was applied to extract classification rules with an induction form. This study aims to examine how data mining techniques based on the GP method can help to discover knowledge and extract classification rules automatically to illustrate well this knowledge. These rules are expected to enrich an anthology in the urban remote sensing domain. A comparison of the performance of three GP algorithms (Bojarczuk_GP, Falco_GP, and Tan_GP) was made using the JCLEC framework. Results showed two main conclusions. The first showed that generated rules can classify and extract useful knowledge from VHR satellite data using GP algorithms. The second demonstrates that the Bojarczuk model is efficient on accuracy classification than the Falco and Tan models.

1 Introduction

Knowledge-based systems (KBS) are becoming more and more important in various domains, especially in high-dimensional feature space where information is variable, and knowledge in this context is still complex to produce [1]. Indeed, acquiring and representing knowledge is a tedious process and the multiple steps involved in their creation can be very different according to the studied domain. This heterogeneity led to multiple questions and propositions, and the expert is often lost when the time comes to choose a solution. However, the advantages of representing and storing domain knowledge are undeniable. Indeed, it is then possible to produce intelligent systems based on the use of the acquired knowledge and to better explain and understand the domain under consideration.

In remote sensing, domain knowledge extraction is a tedious task. This is due to the complexity of the feature space, which is generally a satellite image with multiple spectral bands. The 1980s saw the emergence of satellites capable of producing high-resolution (HR) images between 30 and 10 m (Landsat-4, 1982; SPOT3, 1993). However, the 2000s appeared very high spatial resolution (VHR) satellite images whose spatial resolution is less than 5 m (QuickBird, 2001; PLEADEES, 2011). VHR satellites currently make it possible to obtain images with a resolution up to 0.5 m per pixel on the panchromatic band. Therefore, these images offer a much higher level of detail than HR images.

A new era has come to advance the semi-automatic extraction of objects from digital images. In the remote sensing field, multispectral imagery (MSI) captures reflected radiation over a series of adjoining bands, covering a very

A. Rida (✉)
Center of Urban Systems, Mohammed VI Polytechnic University (UM6P), 43150 Ben Guerir, Morocco
e-mail: rida.azmi@um6p.ma

A. Hicham
Mining Environment & Circular Economy, Mohammed VI Polytechnic University, 43150 Ben Guerir, Morocco

N. Abderrahim
Laboratoire de géophysique appliquée, de géotechnique, de géologie de l'ingénieur et de l'environnement (L3GIE), Mohammed V University – Mohammadia School of Engineering, Rabat, Morocco

large range of the electromagnetic spectrum for every pixel in the image. In the last decade, a new series of high spatial and spectral resolution imagery has become accessible and used more in different fields. Such images with sub-metric spatial resolution can provide features pertinent to the classification task, by enhancing accuracy and reducing spectral confusion in some cases.

However, the classification methods used with high spatial and spectral resolution data apply a new analysis technique called object-oriented image analysis approach or object-based image analysis approach (OBIA). These techniques are usually based on the use of domain knowledge [2]. The key issue in this approach is the obtainment of this knowledge, which is usually implicit and not formalized. Analysis methods must reduce the dimensionality of this very high-dimensional feature space to make any classification analysis more accurate [3, 4]. HR and VHR images have been increasingly used for the classification of land use land cover (LULC), but the spectral variation within the same class, the spectral confusion between the different land covers, and the shadow problem make per-pixel classifiers less efficient. The object-oriented classification approach is designed to deal with the problem of heterogeneity of the environment; it no longer treats the pixel in isolation but a group of pixels (objects) in their context [5].

The key parameter of the OBIA approach is the extraction of primitive objects from raw images, where each object corresponds to a group of homogeneous pixels. To recognize objects (or using methods able to detect objects), several techniques are generally based on the use of knowledge related to spectral, spatial, and contextual properties (e.g., spectral and textural values of an object, shape, length, area, form factor, etc.) [6].

About a decade ago, came the launch of the first software package specializing in OBIA: a revolutionary development in the remote sensing world that led to improvements within a wide field of applications. Over the past few years more packages have been developed, both specialized, and modules of existing image-analysis software.

A brief literature search reveals that publications in the early period of OBIA (2000 to 2003/04) were dominated by conference proceedings and "grey literature" but increasing numbers of empirical studies published in peer-reviewed journals have subsequently provided sufficient proof of the improvements that OBIA offers over per-pixel analyses. Figure 1 shows the increasing number of peer-reviewed articles published, and the number was doubled between 2006–2008.

The dimensions of the features extracted from image objects are much larger than pixels, which mainly contain spectral-based information (e.g., mean, ratio, and standard deviation). In object-based classification, hundreds of features involving the spectral, geometry, and texture features can be obtained from the image objects. However, large amounts of features participating in classification always give rise to the "complexity of dimensionality", which decreases the classification accuracy. As some features make contributions to the classification and others have less influence on the result, features are commonly divided into relevant features, redundant features, and irrelevant features [2].

To yield better classification results, the irrelevant information should be removed, as much as possible, and the utilization of relevant information should be maximized. Therefore, feature selection before the object-based classification of high-resolution remote sensing images is a prerequisite. After the redundant and irrelevant features are removed, the training time is reduced, and the classification efficiency can be improved [7].

In the literature, only a few works focus on the development of a knowledge base to identify objects from remote sensing data. However, building a knowledge base in this context is not an easy task since the information required is generally variant and not formalized. This paper is organized into three sections as follows. In Sect. 2, the principles of knowledge extraction from remote sensing data and its relationship with GP were presented, as well as the algorithms used to realize this study. The methodology and the experiments were detailed in Sect. 3. Finally, Sect. 4 discussed the results and presented the concluding remarks.

2 Genetic Programming

Data mining technologies, e.g., fuzzy classifications [8], object-oriented classification (based on multiresolution segmented data) [7], per-pixel maximum likelihood [9], or artificial neural networks [10], have been used in several studies as a supervised or unsupervised remote sensing classification technique [11, 12]. However, using data characterized by huge volumes, high dimensionality, and having spatial attributes will be a tedious task capable of giving a result attended to be a suite. And of highly complex, high-dimensional, diversified, and variant datasets that present significant analysis challenges solving a problem automatically has always been the main interest. It was an idea that began in the late 1940s [13]. The domain of intelligent systems has always aimed at producing systems with supposedly intelligent behavior. GP is inspired by the design of natural evolution and seeks to solve problems automatically. An approach that requires intelligence if the same task is accomplished by a human being, is none other than the definition given by Arthur Samuel [13] on the purpose of automatic learning and intelligent systems. GP is a method inspired by the theory of evolution as it has been defined by Darwin [14], in particular its biological mechanisms. It aims to find programs that best meet a specified task. However,

Fig. 1 Cumulative number of publications using OBIA approach between 1990 and 2020 —scopus® database
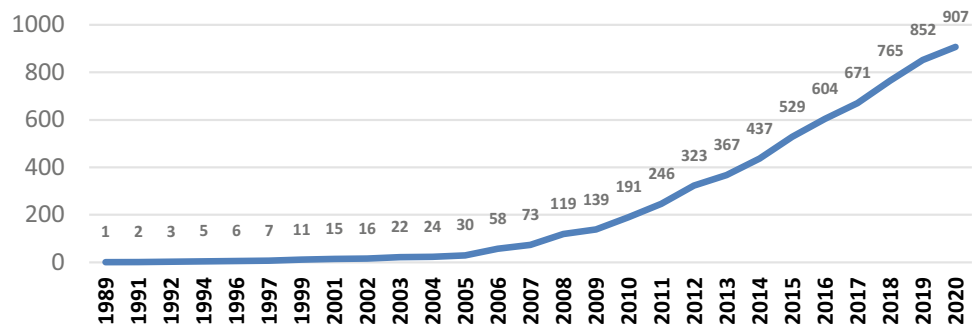
the GP concept allows the machine to learn, using an evolutionary approach, to optimize the programs' population.

Within the framework of GP from the first population of stochastically generated programs and using operators inspired by Darwinism, the GP evolves this population in a stochastic way. By reiterating this process, it is hoped to make the population converge toward solutions (programs) that respond to the problem to be solved. The flowchart showed in Fig.2 gives an idea of the general functioning of GP.

This diagram represents the operating cycle of a genetic program. First, a base is implemented to be able to start generating programs (initialization phase). Then, several individuals that generate future generations are obtained. At this time, a check is made to see if one of the solutions offered by these individuals is satisfactory (Evaluation Block). If no solution is suitable, a selection of the best must be made to generate descendants using different techniques like selection phase and crossover/mutation. Finally, these descendants will come to replace the previous generation by being, in turn, the parents, and the cycle then begins again with the evaluation block. Roughly, in biology, the

information carried by a gene is called a genotype, and the character expressed by this gene is called a phenotype [15]. By transposition in GP, a program can be seen from two angles: genotypic, the form on which the genetic operators apply, and phenotypic, the form in which the objective function or fitness function will be evaluated. The most common genotypic form in GP is the tree form, where each program is encoded as a tree. Reference [16] used this form to implement programs; it is the direct transposition of the prefixed form, used for example by the Lisp language [16].

This paper is motivated by the works of [17, 18] in analyzing and presenting the data structure of remote sensing data as a knowledge base to extract useful classification rules.

Finding a solid technique to extract knowledge from a feature space (VHR images) has two advantages: (i) being intuitively comprehensible to the user and (ii) being easily interpretable by problem-domain experts.

The induction form is one of the powerful techniques used in data mining techniques. Applying a rule-based system using the statement IF (conditions) THEN (predicts
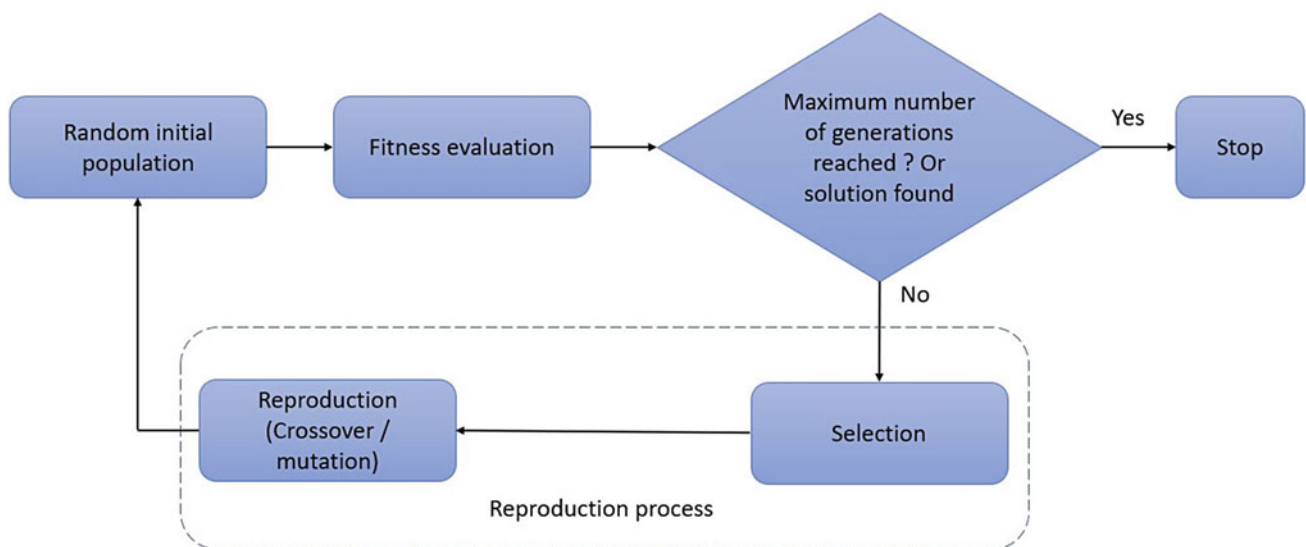

Fig. 2 General flowchart of genetic programming concept

class) is the challenge in this work. In the literature, there are several rule induction algorithms to discover such classification rules [19, 20]. A particularly famous strategy in computer science consists of the sequential covering approach, where in the essence the algorithm discovers one rule at a time until (almost) all examples are covered by the discovered rules (i.e., match the conditions of at least one rule). In contrast, sequential covering rule induction algorithms are mostly greedy, and they can perform a local search in the rule space.

An alternative approach to discover classification rules consists of using an evolutionary algorithm (EA), which performs a more global search in the rule space. Indeed, there are many EAs for discovering a set of classification rules from a given dataset [212223–24].

3 Methodology and Experiments

All data mining tasks involve at least three steps: (1) data preparation, (2) data analysis, and (3) decision-making. This work consists of the three fundamental steps listed as follows: first, the input data were preprocessed and prepared to extract knowledge from an imagery of WorldView-2 satellite sensor taken in 2011 [25]. The new knowledge base was analyzed to identify the relation between attributes and reduce spectral confusion in the dataset, and finally, GP was integrated as an optimization technique capable to find new and innovative classification rules. Figure 3 shows the complete processing chain for the proposed classification approach.

3.1 Study Area

The study area is Rabat city, the political capital of Morocco, located in the north-west of Morocco. Administratively, its territory has an area of 118.5 km^2, composed of the urban municipality of Rabat, divided into five districts. At the last census conducted in 2014, its population was 5,77,827, making Rabat the seventh-largest city in the kingdom. With its suburbs, it forms the second-largest agglomeration of the country after Casablanca. Since June 2012, a group of city sites is inscribed on the UNESCO World Heritage List as cultural property. The heart of Rabat city is made up of the old city, to the west, and along the seafront, there is a succession of modern neighborhoods, and to the east, along the Bouregreg river. Between these two axes, going from north to south, there are three main neighborhoods: The first one is Agdal, which is a very lively neighborhood of buildings mixing residential and commercial functions, mostly intended for the middle classes. The second is Hay Riad, the neighborhood with high-class areas that have experienced a

surge of dynamism since the 2000s, tending to become the new business center of Rabat. The last one is the Souissi neighborhood, consisting mainly of residential areas.

Hay Riad neighborhood made up of high standing houses with modern architecture was the study area of this work, where the roads are very clear, and the streets are also well visible. Rooftops have a unified geometry as well as their density allows a good segmentation of an input image.

3.2 Preprocessing of Input Data

The input data is generated mainly through a WorldView-2 satellite image which has eight multispectral bands: four (4) standard colors (red, green, blue, and near-infrared 1) and four (4) new bands (coastal, yellow, red edge, and near-infrared 2) [26]. WorldView-2 products are available as part of the DigitalGlobe Standard Satellite Imagery products from the QuickBird, WorldView-1/-2/-3, and GeoEye-1 satellites [27]. With the additional four spectral bands, WorldView-2 offers unique opportunities for remote sensing analysis of vegetation, coastal environments, agriculture, geology, and many other fields. This satellite image is characterized by high spatial resolution with 4 m for multispectral (MS) bands, and 0.5 m for the panchromatic (PAN) one. With its enhanced agility, WorldView-2 is capable of acting like a paintbrush, sweeping back and forth to collect very large areas of multispectral imagery in a single pass. The sensor can collect nearly 1 million km^2 every day; its high altitude allows it to typically revisit any place on earth in 1.1 days.

Radiometric calibration

As a preprocessing step, a radiometric correction was used to prepare the data for segmentation and extraction of the knowledge base. Radiometric correction of MS and PAN data was used to calibrate aberrations in data values due to specific distortions from atmosphere effects (such as haze) or instrumentation errors (such as striping) [21].

DigitalGlobe sensor products (image pixels) are radiometrically corrected image pixels. Their values are a function of how much spectral radiance enters the telescope aperture and the instrument conversion of that radiation into a digital signal [28]. Therefore, image pixel data are unique to each sensor.

A calibration step has been performed (at provider level) and these data are provided in the *.IMD* metadata file that is delivered with the imagery. Since its launch, DigitalGlobe performs an extensive vicarious calibration campaign to provide an adjustment to the prelaunch values. The top of atmosphere radiance (L) in units of [Wμm^{-1} m^{-2} sr^{-1}] is
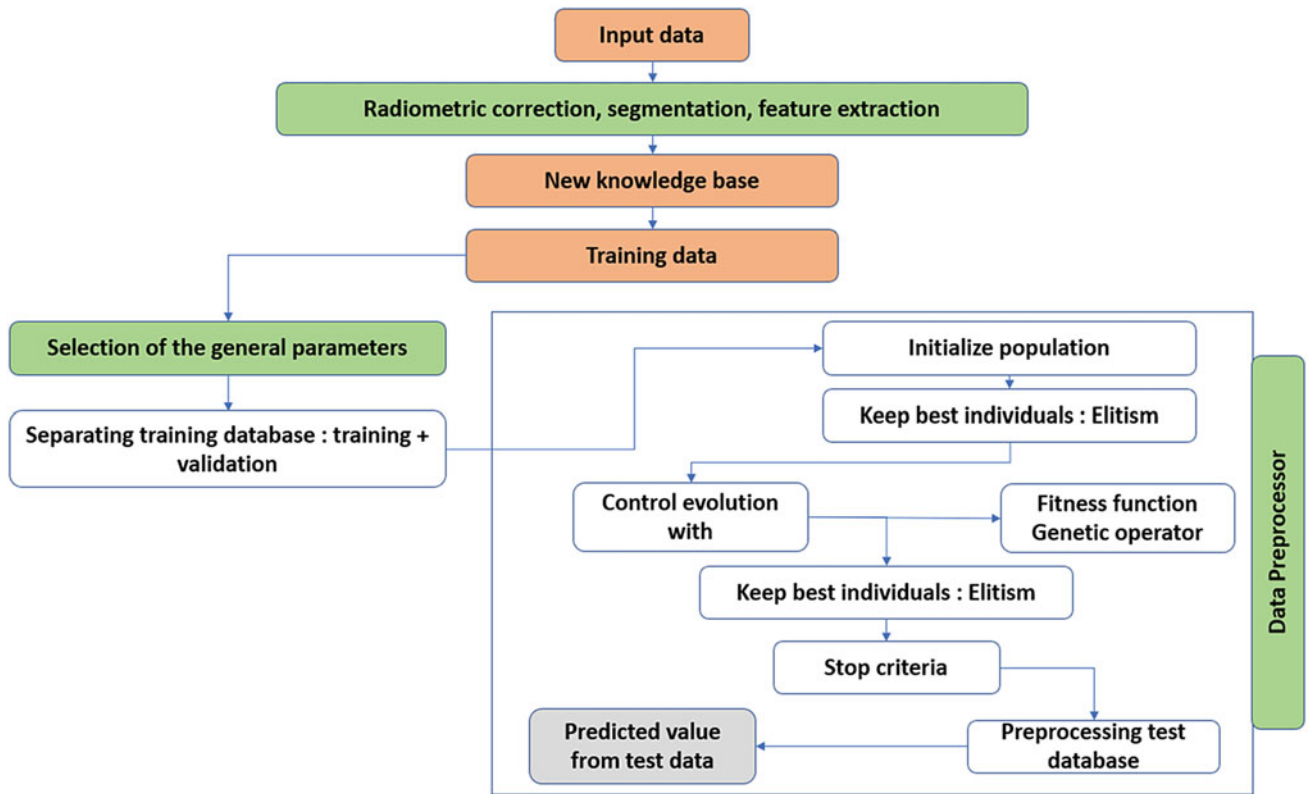
Fig. 3 Processing chain from data preparation to rule-based classification

then calculated for each band by converting from digital numbers (DN).

Equation (1) is used to convert the at-sensor radiance to top of atmosphere reflectance where calculations are performed independently for each band and pixel:

$$\rho_{\lambda_{pixelBand}} = \frac{L_{\lambda_{pixelBand}} * d_{ES}2 * \pi}{E_{sun\lambda_{Band}} * \cos(\Theta)_s} \qquad (1)$$

- "**L**" is at-sensor radiance calculated from data provided in .IMD file;
- "**d$_{ES}$**" is the Earth–Sun distance in astronomic unit;
- **E$_{sun}$** is the band-averaged solar exoatmospheric irradiance;
- **Θ** is the solar zenith angle (90-meanSunEl from IMD file).

Pan-sharpening

The second preprocessing step is the pan-sharpening of the data, where this technique was used to enhance spatial resolution. Recently, several applications, such as land-cover classification, feature extraction, image segmentation, and change detection, require both spatial and spectral images for fine features detection in suburban or urban scenes. The literature shows a large collection of pan-sharpening

methods developed and used to enhance spatial resolution and preserve spectral information. In this study, *NNDeffuse* algorithm developed by [29] was used to fusion MSI and PAN data.

Segmentation

Segmentation is a main preprocessing step that allows the user to identify the object that has similar spectral characteristics pixels. It is the process of completely portioning a scene (in this case remote sensing image) into non-overlapping regions (segments) in scene space. In the segmentation process, all objects are outlined without any class label. Usually, the outlined objects should have one specific object, to generate appropriate segments capable to distinguish semantic objects (Fig. 4).

Many powerful algorithms have been developed within pattern recognition and computer vision since the 1980s, where research led to successful applications in disciplines like medicines or telecommunication engineering. However, their application in the fields of remote sensing and photogrammetry was limited to special purpose implementations only. Nevertheless, this limitation is due to the complexity of the underlying object models and the heterogeneity of sensor data in use. With the appearance of high spatial resolution
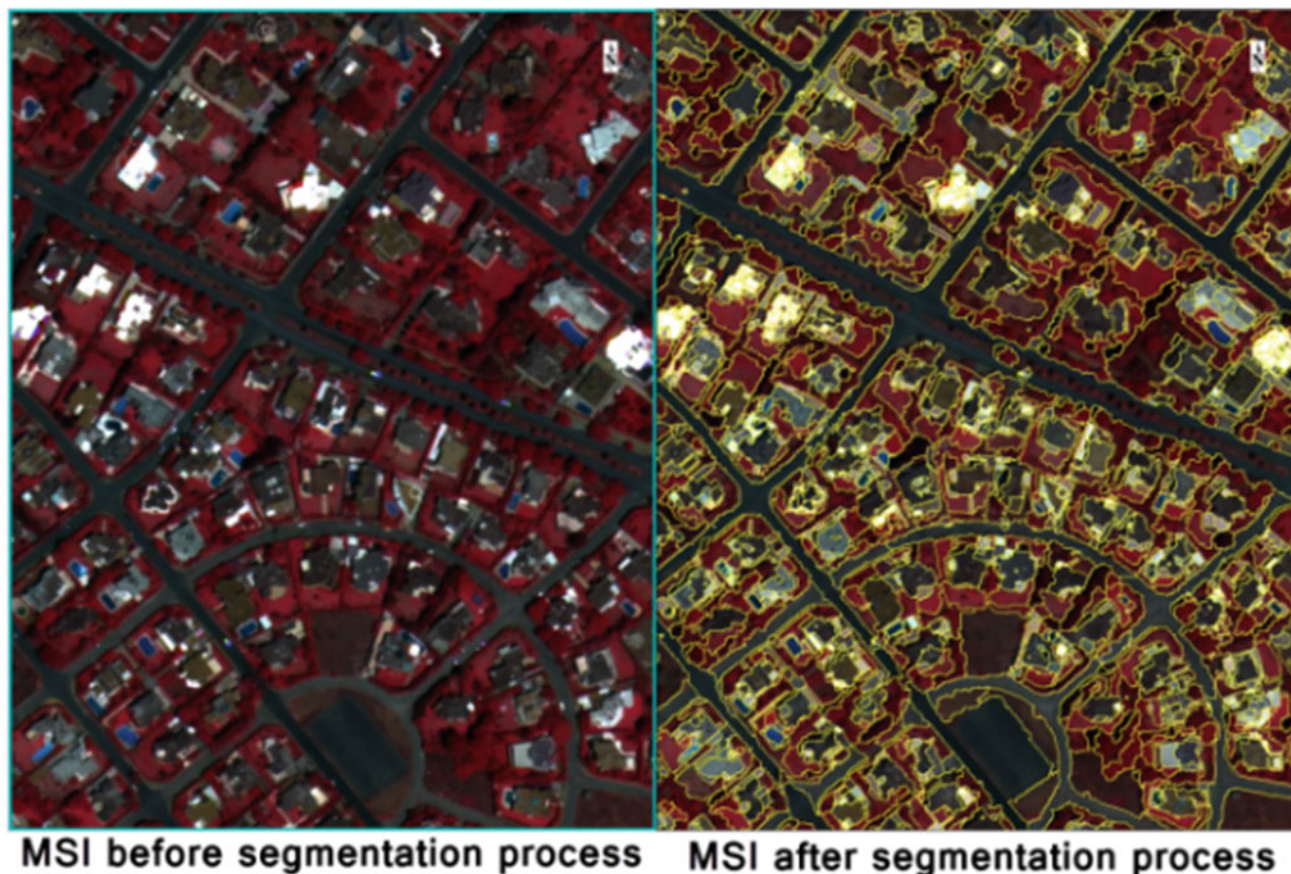
Fig. 4 Example of Hay Ryad district (in Rabat city—Morocco) showing WorldView-2 image before and after segmentation process

satellite data as well as multisource data sources, the segmentation methods have become evident again, and significant progress has been made with the introduction of the first commercial and operational software product (eCognition by Definiens-Imaging) in 2000 [30].

Segmentation methods follow two strongly correlated principles of neighborhood and value similarity. A watershed algorithm (WA) for segmentation is used. This method integrates duplicate neighboring areas based on a combination of spectral and spatial information. The WA transform is based on the concept of hydrologic watersheds, where basins fill up with water starting at the lowest points, and dams are built and water coming from different basins would meet. The process stops when the water level has reached the highest peak in the landscape [31]. A similar process is applied in digital imagery using the luminosity of the pixel; the darker the pixel, the lower the elevation. A watershed algorithm sorts pixels by increasing the grayscale value and then begins with the minimum pixels and "floods" the image, partitioning the image into regions with similar pixel intensities based on the computed watersheds. The result is a segmented image, where each region is assigned to the mean spectral values of all the pixels that belong to that region.

In this study, the Full Lambda-Schedule algorithm developed by [32] is used to merge segments. The algorithm iteratively merges adjacent segments based on a combination of spectral and spatial information as mentioned above. Figure 4 shows results after segmentation over Hay Riad district where individual buildings are surrounded as well as the green spaces (of grass and trees) and road networks.

3.3 Feature Extraction

All segmented objects from VHR images have spectral, spatial, and textural features, to have an accurate classification process. More than one attribute characterizing an object must be found to explain this accurate classification; for instance, shadow class has a high spectral value in near-infrared bands, and grass has a high rectangularity index in urban areas with a coarse texture and mean NDVI values. Combining several distinctive attributes for each class will facilitate the extraction of useful classification rules.

After the segmentation process, attributes of each segment were calculated and extracted using a feature extraction module implemented in ENVI 5.0 software [33]. Attributes were

categorized as spatial, spectral, and textural attributes. Additional data was calculated using a normalized band ratio (infrared and red) and calculation of hue, saturation, and intensity (HSI) attributes. The database extracted from ENVI's module was a *dbase file composed of 111 attributes. Details about calculated attributes can be found in [34]. A number of 590 segments as a training set was used, where the proposed classes are shown as follows (Shadow, Built up_Roofs, Built Up_Roads, Vegetation_Grass, Vegetation_Trees, Bare soil, and Water).

In Fig. 4, there are three levels of object classes. The first level contains the main component of the urban ecosystem (Built up, Water, and vegetation). In the case of VHR image, the shadow class was added due to the high buildings and trees. The second level of image object contains derived information from the three components of the first-class object. In the last level (3), there are more details about one specific class or sub-classes (Fig. 5).

Table 1 resumes all calculated and extracted attributes from a segmented image. All attributes that were in the rule-based system are implemented in ENVI software to classify the input image. Attributes were divided into two types of bands (spectral and derived bands), respectively, spectral indices and calculated attributes such as spectral, geometric, and textural attributes.

3.4 Feature Selection

In modern machine learning algorithms, there are methods used to reduce dimensionality [35, 36]. In general, these tasks are rarely performed in isolation. Instead, they are often preprocessing steps to support other tasks. In literature, there are two main strategies of dimension reduction: (i) Feature selection techniques that are typically grouped into three approaches, namely filter, embedded, and wrapper methods that extract subsets from existing features, and (ii) feature extraction (e.g., principal component analysis— PCA) [37]. The key difference between feature selection and extraction is that feature selection keeps a subset of the original features while feature extraction creates brand new ones.

In this paper, the selection of attributes was made for the supervised classification. In this context, the objective of selection is finding an optimal subset of attributes that can be composed of relevant attributes and must seek to avoid redundant ones. In addition, this set must make it possible to best meet the objective set, namely the accuracy of learning, the speed of learning, or even the applicability of the proposed classifier.

ReliefF-based feature selection method was used in this paper, where it takes a filter method approach [38]. The proposed method was used to calculate a feature score for each feature. This score can be applied to rank and select top-scoring features. Many researchers adopted the ReliefF algorithm to preliminary filter high-dimensional features in the feature database [38, 39].

By applying the ReliefF method on the input dataset, results made it possible to select the 20 best attributes of which it proposed bands 6, 7, and 8, which are, respectively, red edge, near-infrared—NIR-1, and near-infrared—NIR-2. Also, the hue, saturation, and intensity (HSI) transformation from RGB bands were highly ranked and used in the new filtered dataset.
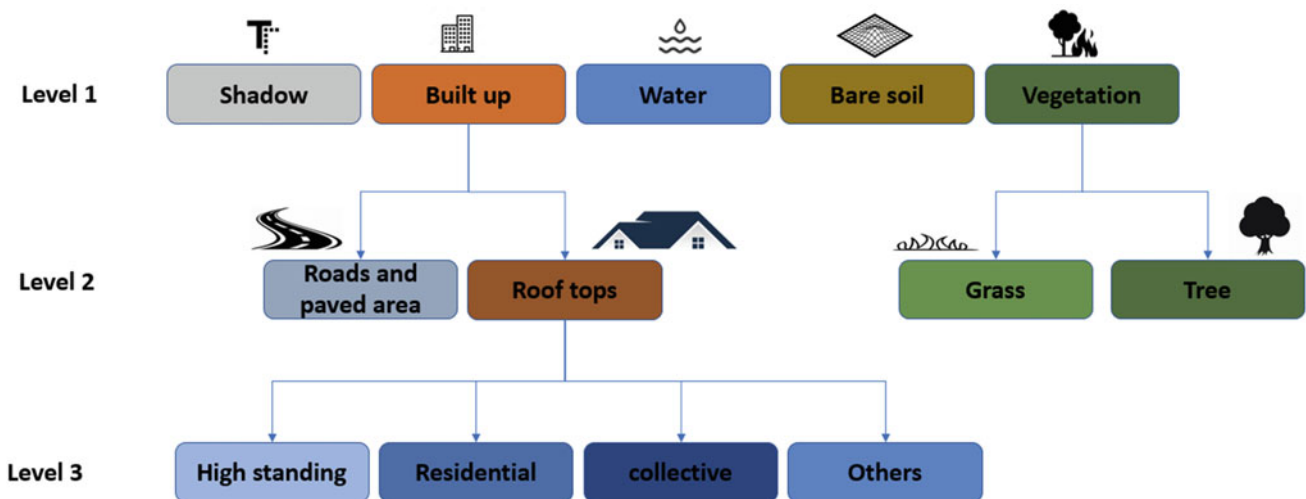

Fig. 5 Object classes hierarchy used in this study

Table 1 Extracted parameters from preprocessed satellite image

Attributes	Class attribute
Spectral bands	Spectral information was calculated using the 8 input bands where (Min, Max, Mean, and Standard Deviation was calculated)
Derived bands	Hue—Saturation—Intensity
Spectral Indices	Band ratio using Red and NIR bands
Geometric calculation	12 variables are calculated
Textural calculation	4 variables are calculated

3.5 Generating Classification Rules

In this paper, knowledge is presented as multiple IF–THEN rules in the decision rules list. Such rules state that the presence of one or more conditions (antecedents) implies or predicts the presence of other conditions (consequents). A typical rule has the form of: If X1 and X2 and … Xn as conditions THEN Y, where Xi ∈ (1, 2,..,n) is the antecedent that leads to the prediction of consequent Y. The reason why classification rules were used instead of the decision tree is because each rule can be seen as an independent piece of knowledge. Thus, newly generated rules can be added to an existing ruleset without disturbing existing ones. Multiple rules can be concatenated to form a set of decision rules. This last is usually listed according to their accuracy, where the best rule was listed first.

The Java Class Library for Evolutionary Computation—JCLEC Framework developed in Java Environment was used. JCLEC is a representative example of an evolutionary optimization framework designed for one main objective, to maximize its reusability and adaptability to new paradigms with a minimum of programming effort [40, 41]. The implemented classification module is an intuitive, usable, and extensible open-source module for GP classification algorithms [42]. This module is a part of open-source software for researchers and end-users to develop and use classification algorithms based on GP and grammar-guided genetic programming (3GP) models [43], an extension of GP which makes the knowledge extracted more expressive and flexible using a context-free grammar.

JCLEC classification module houses three 3GP classification algorithms listed as follows: (Bojarczuk_*GP* [44], Falco_GP [21] and Tan_GP [45]). JCLEC extends a class called PopulationAlgorithm. This parent class defines the main steps of the evolutionary process. To initialize the population, a component is triggered with the number of solutions to be created as a parameter; in this case the number of solutions is equal to the n class in the dataset. Each solution individual should contain a fitness object representing its quality.

Bojarczuc Model

The author used GP standard operators to evolve decision trees using a defined syntax. Bojarczuk used a GP-based approach, where a set of functions applicable to different types of attributes is defined to represent the rules as a disjunctive normal form. Several constraints are placed on the tree structure to express a valid rule. This type of GP is also referred to as constrained syntax GP [44, 46]. The fitness function used in the Bojarczuk model evaluates the quality of each individual (a rule set where all rules predict the same class) according to two basic criteria, namely its predictive accuracy and its simplicity [47]. Implementation of this fitness function in the JCLEC module is a subclass called BojarczukEvaluator. The fitness function in this case evaluates the confusion matrix for each of the data classes [44].

Falco Model

The author used GP to evolve comprehensible simple rules by combining the parallel searching ability of genetic programming. Falco used a classifier tree that is constructed using logical functions and attribute values. A grammar has been designed that can represent such rules. The author has shown that the evolved rules are comprehensible, emphasize discriminating variables, and achieve compatible performance as compared to other classification algorithms on benchmark datasets [21]. The fitness function used in this case evaluates the number of prediction errors for the class of the current algorithm's execution [21].

Tan Model

Tan model is based upon a modified version of steady-state GP in [48]. The fitness function evaluates the quality of each rule or individual, which is based on the evaluation function defined in Eq. (2). In other words, the fitness function evaluates the confusion matrix for the data class of the current algorithm's execution.

$$Fitness = \frac{Tp}{Tp + W1 * Fn} * \frac{Tn}{Tn + W2 * Fp} \qquad (2)$$

where Tp, Fp, Tn, and Fn stand for true positive, false positive, true negative, and false negative, respectively. $W1$ and $W2$ are the weights; they enable the dependency of fitness function on different concepts.

Nevertheless, the performance of three modules has been tested and validated in the following section with statistical tests like R-squared, p-value, and ROC area.

4   Results and Discussion

GP models interpretation

The models used in this study generated explicit rules to simplify knowledge from high-dimensional feature space. Rule induction technique that creates the "If—Then—Else" type was used; it generates rules from a set of input variables, and it can work with both numerical and categorical values. In this case, an inductive prediction that concludes a future instance from a past sample is as follows:

```
IF (antecedents)₁ THEN class₁
ELSE IF (antecedents)₂ THEN class₂
ELSE default class
```

All the models proposed the same inductive structure with a variety of attributes that were candidates to generate an optimal solution and help the user in the choice of the most representative attributes in the search space. The interpretation of these choices is based on three main classes: shadow, roofs, and trees.

The rules below show an example for the result of Bojarczuk's model, presented in the inductive form. For shadow class, the algorithm proposed 2 bands, respectively, the blue and near-infrared band 2—(NIR-2). Dark objects, which confound many shadow detection algorithms, often have much higher reflectance in the NIR band. The blue band is also considered to be an excellent choice because several studies have shown that shadow pixels are illuminated by the predominantly blue, diffuse sky radiation. Bojarczuk model suggested a combination between NIR-2 and blue band where the real interpretation of the rule is *IF average value of blue band < = 19,188 AND the average value of NIR-2 band < = 37,71,600 THEN objects belong to shadow class.* On the other hand, Falco's model proposed the texture of band 6 (red-edge band) to represent the shadow class. The red-edge band is a division of the red spectrum between 700 and 750 μm. However, the red band

reflects a small part of the dark pixels which is considered a poor choice for this class. The Tan model suggested a complicated rule composed of four conditions and two logical operators ("AND" & "AND NOT"); the NIR-1 and NIR2 bands, the ratio between the red band and NIR (which is the ratio of the Normalized Difference Vegetation Index—NDVI). NDVI band should not be taken into account according to the Tan model using the logical operator (AND NOT), and finally the minimum value and the NIR-1 band. This complexity can cause interference between conditions.

```
IF (AND < = AVG_B2 191,880,007 < = AVG_B8 377,160,068)
THEN (Class = Shadow.
ELSE    IF    (AND < = AVG_B9    -0,608,389 > TXAVG_B7
430,725,437)
THEN (Class = vegetation_2)
ELSE    IF    (AND > AVG_B2    191,880,007 < = AVG_B8
377,160,068)
THEN (Class = water)
ELSE IF (AND < = MAX_B9 -0,317,148 AND < = TXAVG_B7
430,725,437 > MIN_B10 107,685,121)
THEN (Class = Vegetation_1)
ELSE IF (AND < = TXRAN_B11 0,083,049 AND < = TXAVG_B7
430,725,437 > MAX_B9 -0,317,148)
THEN (Class = Built_up_1)
ELSE IF (AND > AVG_B9 -0,608,389 > AVG_B8 377,160,068)
THEN (Class = Built_up)
ELSE IF (AND < = TXRAN_B11 0,083,049 AND < = MIN_B10
107,685,121 < = TXAVG_B7 430,725,437)
THEN (Class = Bare_Soil)
ELSE (Class = Built_up)
```

For buildings and rooftops, Bojarczuk's model proposed the average value of NDVI band and NIR-2 band. Buildings and rooftops have particular characteristics relative to other features. For example, the shape of rooftops approximates a rectangle, the area of rooftops of residential buildings is within a certain range, compared to industrial or other types of buildings. In our case, rooftops of interest are relatively dark, so they should have a low average pixel value (<0.4). However, NDVI would be a good criterion to start with in this example, where the buildings have smaller NDVI values than vegetation.

The Falco model suggested the average value of the blue band and the minimum of the NIR-2 band. The near-infrared bands may contain low reflectance for dark pixels, which may meet our needs but not with great certainty. Dark pixels can also exist in the roads and bare ground classes. The Tan model once again proposed a rule with three conditions, but very interesting attributes. He suggested the minimum values for the NDVI band (which is a very good choice), the

average texture values for NIR-1 band, and texture range for band 11 (the HSI transformation of RGB bands).

In the case of trees class, the Bojarczuk model proposed the maximum value of the NDVI, which was predicted, and the mean texture of the NIR-1 band and the minimum value of the band 10 (which is the derived HIS transformation from RGB bands). In the literature, it is known that trees are more textured than grass, so the choice of the mean texture within the infrared band is well done for the Bojarczuk model.

In the case of the Falco model, the algorithm proposed NDVI combined with the coastal blue band (band 1). This combination can lead to inaccuracy of the generated rule due to the coastal blue band reflectance values. Thus, the Tan model has proposed the minimum value of band 10.

4.1 Validation Metrics

As a validation method, the confusion matrix was used to evaluate classification accuracy, which is a common way of presenting true positive (TP), true negative (TN), false positive (FP), and false-negative (FN) predictions. Those values are presented in the form of a matrix where the Y-axis shows the true classes while the X-axis shows the predicted classes.

Table 2 shows that the Bojarczuk model's classification gives a diagonal matrix, except for the confusion between (roads and roofs) classes, a large part of the roofs has been classified as roads. This is generally due to the spectral properties of the infrared band where both classes contain dark pixels. Also, a spectral confusion between roofs and water areas has been provoked. This is due generally to the low reflectance of water pixels in some areas. A second confusion has been shown in Table 3 with the trees and grass classes. This confusion is due to the spectral rapprochement between trees and grass classes in the NDVI index.

The Falco and Tan models show the same confusion (roads and roofs) with a slight difference between the number of misclassified lines.

4.2 Statistical Metrics

The confusion matrix was presented in Table 3 to evaluate the behavior of the three models in terms of classification accuracy. On the other hand, it can be more flexible to predict the probabilities of an observation belonging to each class in a classification problem rather than predicting classes directly. This flexibility comes from the way that probabilities may be interpreted using different thresholds that allow the model to trade off concerns in the errors made by the model, such as the number of false positives compared to the number of false negatives. This is required using models where the cost of one error outweighs the cost of other types of errors.

ROC areas and precision-recall curves (PRC) were used to explain the probabilistic forecast for binary (two-class) classification predictive modeling problems [42]. The metrics are used to evaluate the performance of the three models. Precision can be understood as a measure of accuracy or quality, while recall is a measure of completeness or quantity [49].

A measure that combines precision and recall in their harmonic mean, called F-measure or F-score, is used to estimate model performance. However, one rule is used: the higher the score, the better the model. This parameter combines precision and recall into one metric. Table 3 shows that Bojarczuk has the best F-score for all classes, followed by Falco and Tan. Another measurement parameter that distinguishes the performance of several models is the ROC area. In general, ROC curves are based on the rate of true positives (TP Rates) and the rate of false positives (FP Rates). These are relationships that do not depend on the distribution of classes. This robust method eliminates the need to know the costs of classification and the distribution of classes. To calculate the points of a ROC curve, several evaluations of a logistic regression model are performed by varying the classification thresholds, but this would be ineffective. In other words, the AUC provides an aggregated measure of performance for all possible classification thresholds. AUC can be interpreted as a measure of probability for the model to classify a random positive example above a random negative example. Table 3 shows the AUC values for the three models for all classes. Bojarczuk's model showed again a great score compared to other models.

In general, a value of ROC-AUC greater than 0.7 is a good representative value for a model. However, Table 3 shows that the three models were able to classify the seven classes with a score beyond 0.7, except for the road and trees classes in the Falco model. It seemed, however, that the Falco model is unable to distinguish between these two classes precisely. A second anomaly is noticed between (bare soil) and (roads), where the Tan model found difficulty in classifying these two instances correctly.

It is highly recommended to use precision-recall curves as a supplement to the routinely used ROC curves to get the full picture when evaluating and comparing tests. It is used less frequently than ROC curves but as we shall see PRC may be a better choice since the current dataset contains imbalanced data. Since precision-recall curves do not consider true negatives, they should only be used when specificity is of no concern for the classifier. In other words, the PRC area represents a different trade-off which is between the true positive rate and the positive predictive value.

PRC is simply a graph with precision values on the y-axis and recall values on the x-axis. In other words, the PRC

Table 2 Confusion matrix for the Bojarczuk, Falco, and Tan models

	Confusion matrix	Bare soil	Grass	Road	Roofs	Shadow	Trees	Water
Bojaczuk model	Bare soil	42	2	9	10	0	3	0
	Grass	0	110	0	0	0	4	0
	Road	1	0	30	2	1	0	0
	Roofs	8	1	27	72	6	0	15
	Shadow	3	0	2	1	51	5	0
	Trees	11	18	3	4	8	61	0
	Water	0	1	0	1	0	0	70
Falco model	Bare soil	57	0	0	8	0	1	0
	Grass	16	98	0	0	0	0	0
	Road	25	0	1	4	0	0	4
	Roofs	46	4	0	76	1	0	2
	Shadow	32	0	0	0	26	4	0
	Trees	65	7	0	1	0	32	0
	Water	8	0	0	1	1	0	62
Tan model	Bare soil	19	1	15	17	7	6	1
	Grass	3	105	1	3	0	2	0
	Road	1	0	11	11	10	0	1
	Roofs	2	3	11	102	3	1	7
	Shadow	0	0	2	4	40	12	4
	Trees	4	13	1	1	14	72	0
	Water	1	1	1	0	0	0	69

**Table 3** Evaluation parameters of the three models

–	Bojarczuk			Faclo			Tan		
	ROC Area	PRC Area	F-Score	ROC Area	PRC Area	F-Score	ROC Area	PRC Area	F-Score
Bare soil	0.80	0.45	0.64	0.75	0.21	0.36	0.64	0.24	0.40
Grass	0.95	0.81	0.90	0.91	0.80	0.88	0.95	0.80	0.88
Roads	0.90	0.38	0.57	0.52	0.08	0.057	0.64	0.12	0.28
Roofs	0.75	0.54	0.65	0.77	0.59	0.70	0.85	0.63	0.76
Shadow	0.90	0.65	0.80	0.70	0.45	0.57	0.80	0.39	0.58
Trees	0.78	0.56	0.68	0.64	0.39	0.45	0.82	0.59	0.72
Water	0.97	0.80	0.90	0.93	0.80	0.88	0.96	0.81	0.90

contains TP/(TP + FN) on the y-axis and TP/(TP + FP) on the x-axis. Both precision and recall are important metrics to evaluate the performance of the binary classification model. The corresponding PRC values in Table 3 show the loss of precision, even the ROC-AUC area of bare soil class that was 0.80 in the Bojarczuk model, barely touches on 0.45 precision. This deficit allows concluding that even higher ROC-AUC can hide a lot of imprecision in some cases. Falco and Tan models show much lower values in the PRC area.

Finally, a weighted average value of all metrics shows that the Bojarczuk model had high accuracy followed by Tan and *Falco*. The performance evaluation of the three models with the application of the AUC-ROC curve (specificity vs sensitivity) and PRC demonstrates that the Bojarczuk model is efficient than the Falco and Tan models.

There is a correlation between statistical metrics and attributes generated in the proposed rules. The model that performed well in terms of statistical attributes is the same model that proposed good rules. Based on developed

Fig. 6 Graph showing the accuracy chart for the three models based on statistical metrics
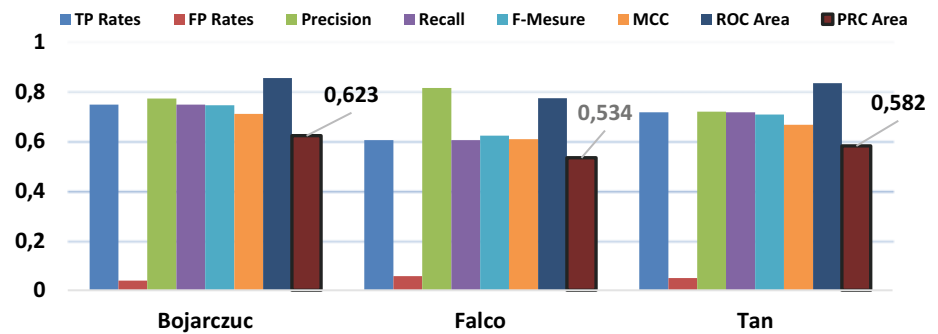
expertise in choosing the right attributes (spatial, spectral, textural, or even derived products such as NDVI), the interpretation of the rules generated by Bojarczuk's model shows a good choice of these attributes (Fig. 6).

In this paper, the evaluation of the models was limited at the level of their statistical metrics. Execution of the rule classification has shown that the models can extract each class separately from the other classes; in other words, the classification rules generated by the three models can make a good extraction of one class at a time. Tests performed in the ENVI software, using its feature extraction module, have shown that applying the rule to a single class is capable to improve accurately extracting the class.

5 Conclusion

The performance of the evolutionary approach was tested; particularly, genetic programming is used to extract explicit knowledge from VHR satellite images. Genetic programming algorithms have shown their performance in explicit knowledge extraction, especially in a complex feature space.

Genetic programming has shown its ability to simulate human expertise in the choice of the most representative variables to apply a rule-based supervised classification. Despite advances in the development of various proposed algorithmic models, the evolutionary approach is still unable to detect a precise threshold value for a given class.

However, a perspective can be retained from this work, focusing on strengthening the algorithmic model so that it can detect more accurate threshold values. This is feasible if a large amount of training data is given, as well as eliminating the preprocessing part that allows filtering of variables that have more influence on the feature space.

References

1. M.L. Brodie, J. Mylopoulos, *on Knowledge Base Management Systems: Integrating Artificial Intelligence and Database Technologies*. (Springer Science & Business Media, 2012)

2. L. Zhang et al., A knowledge-based procedure for remote sensing image classification, in *2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. (IEEE, 2014)

3. R. Azmi, O.B. Alami, A.E. Saadane, I. Kacimi, T. Chafiq, A modified and enhanced normalized built-up index usingmultispectral and thermal bands. Ind. J. Sci. Technol. **9**(28), art. no. 87405, 1-10 (2016). https://doi.org/10.17485/ijst/2016/v9i28/87405

4. L. Gu, Q. Cao, and R. Ren, Building extraction method based on the spectral index for high-resolution remotesensing images over urban areas. J. Appl. Remote Sens. **12**(4), 045501 (2018).https://doi.org/10.1117/1.JRS.12.045501

5. T. Blaschke, Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. **65**(1), 2–16 (2010)

6. L.J. Quackenbush, A review of techniques for extracting linear features from imagery. Photogram. Eng. Remote Sens. **70**(12), 1383–1392 (2004)

7. M.R. Rahman, S.J. Saha, Multi-resolution segmentation for object-based classification and accuracy assessment of land use/land cover classification using remotely sensed data. J. Indian Soc. Remote. Sens. **36**(2), 189–201 (2008)

8. A. Amo et al., Fuzzy classification systems. **156**(2), 495–507 (2004)

9. P. Bolstad, T. Lillesand, R. Rapid maximum likelihood classification. Photogram. Eng. Remote Sens. **57**(1), 67–74 (1991)

10. Zurada, J.M., *Introduction to Artificial Neural Systems*, vol. 8. (West Publishing St. Paul., 1992)

11. J. Han, J. Pei, M. Kamber, *Data Mining: Concepts and Techniques*. (Elsevier, 2011)

12. R. Azmi, A. Saadane, I. Kacimi, and M. Hakdaoui, Hybrid classification approach combining objectoriented method and expert system for extracting land cover map from very high spatial resolution image – case study city ofRabat – Morocco. Int. J. Innovation Appl. Stud. **10**(2), 594–603 (2015)

13. C.J.M. Machinery, Computing machinery and intelligence-AM Turing. **59**(236), 433 (1950)

14. S.J.J.S. Gould, Darwinism and the expansion of evolutionary theory. **216**(4544), 380–387 (1982)

15. C.H. Waddington, *An Introduction to Modern Genetics*. (1939)

16. J.R. Koza, J.R. Koza, *Genetic Programming: on the Programming of Computers by Means of Natural Selection*, vol. 1. (MIT Press, 1992)

17. K. Mertens et al., Using genetic algorithms in sub-pixel mapping. **24**(21), 4241–4247 (2003)

18. D Sheeren et al., Discovering rules with genetic algorithms to classify urban remotely sensed data, in *Proceedings IEEE International Geoscience and Remote Sensing Symposium (IGARSS'2006)* (2006)

19. P. Clark, R. Boswell. Rule induction with CN2: some recent improvements, in *European Working Session on Learning*. (Springer, 1991)

20. J.W. Grzymala-Busse, *A new version of the rule induction system LERS*. Fandom. Inform. **31**(1), 27–39 (1997)

21. I.D. Falco, A.D. Cioppa, E. Tarantino, *Discovering interesting classification rules with genetic programming.* Appl. Soft Comput. **1**(4), 257–269 (2002)

22. M.V. Fidelis, H.S. Lopes, A.A. Freitas, *Discovering comprehensible classification rules with a genetic algorithm*, in *Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No. 00TH8512).* (IEEE, 2000)

23. A.A. Freitas, A survey of evolutionary algorithms for data mining and knowledge discovery, in *Advances in evolutionary computing.* (Springer, 2003), pp. 819–845

24. P.J. Sharma, Discovery of classification rules using distributed genetic algorithm. Procedia Comput. Sci. **46**, 276–284 (2015)

25. M.A.C. Tarantino, G. Pasquariello, F. Lovergine, P.B. Valeria, Tomaselli, 8-Band Image Data Processing of the Worldview-2 Satellite in a Wide Area of Applications. Earth Observation. (2012)

26. R.P.S. Landry, A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. Remote Sens. Environ. **124**, 516–533 (2012)

27. T. Novack et al., Machine learning comparison between WorldView-2 and QuickBird-2-simulated imagery regarding object-based urban land cover classification. Remote Sens. **3**(10), 2263–2282 (2011)

28. C. Tarantino et al., 8-band image data processing of the worldview-2 satellite in a wide area of applications. *Earth Observation*. (2012) p. 137–152

29. W. Sun, B. Chen, D.W. Messinger, Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images. Opt. Eng. **53**(1), 013107–013107 (2014)

30. J. Tian, D.M. Chen, Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition. Int. J. Remote Sens. **28**(20), 4625–4644 (2007)

31. J. Roerdink, A. Meijster, The watershed transform: definitions, algorithms and parallelization strategies, Fundam. Inform. **41**(1,2) 187–228 (2001)

32. D.J. Robinson, N.J. Redding, D.J. Crisp, *Implementation of a Fast Algorithm for Segmenting SAR Imagery*. (Electronics Research Lab Salisbury, Australia, 2002)

33. ITT, *ENVI EX User's Guide.* (ENVI, 2009) p. 52–75

34. ENVI, *Feature Extraction with Example-Based Classification Tutoria.* (2010) http://www.harrisgeospatial.com/portals/0/pdfs/envi/FXExampleBasedTutorial.pdf

35. P.S. Bradley et al., *Method of Reducing Dimensionality of a Set of Attributes Used to Characterize a Sparse Data Set*. (Google Patents, 2004)

36. G. Chandrashekar, F.J.C. Sahin, E. Engineering, A survey on feature selection methods. **40**(1), 16–28 (2014)

37. R. Bro, A.K. Smilde, Principal component analysis. Anal. Methods **6**(9), 2812–2831 (2014)

38. K. Kira, L.A. Rendell, A practical approach to feature selection, in *Machine Learning Proceedings 1992*. (Elsevier, 1992), pp. 249–256

39. Z. Wang et al., Application of ReliefF algorithm to selecting feature sets for classification of high resolution remote sensing image. in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. (IEEE, 2016)

40. A. Cano et al., A classification module for genetic programming algorithms in JCLEC. **16**(1), 491–494 (2015)

41. JCLEC, *JCLEC Classification Module/User Manual and Developer Documentation.* (2014), http://jclec.sourceforge.net/data/JCLEC-classification.pdf

42. R. Azmi, H. Amar and A. Norelyaqine, Generate knowledge base from very high spatial resolution satellite image usingrobust classification rules and genetic programming. 2020 IEEE International conference of Moroccan Geomatics (Morgeo), 1–6 (2020). https://doi.org/10.1109/Morgeo49228.2020.9121914

43. R.I. Mckay et al., Grammar-based genetic programming: a survey. **11**(3–4), 365–396 (2010)

44. C.C. Bojarczuk et al., A constrained-syntax genetic programming system for discovering classification rules: application to medical data sets. **30**(1), 27–48 (2004)

45. K.C. Tan et al., Mining multiple comprehensible classification rules using genetic programming. in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*. (IEEE, 2002)

46. H. Jabeen, A.R. Baig, Review of classification using genetic programming. Int. J. I. Eng. Sci. Technol. **2**(2), 94–103 (2010)

47. C.C. Bojarczuka, H.S. Lopesb, A.A. Freitasc, Data mining with constrained-syntax genetic programming: applications in medical data set. **6**, 7–0 (2001)

48. W. Banzhaf et al., *Genetic Programming*. (Springer, 1998)

49. T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. Plos One **10**(3), e0118432 (2015)