

Machine Learning Techniques for the Semiautomated Recognition of Urban and Peri-Urban Configurations

Chiara Di Dato(✉), Federico Falasca, and Alessandro Marucci

Department of Civil, Building-Architecture and Environmental Engineering, University of L'Aquila, Via Giovanni Gronchi 18, 67100 L'Aquila, Italy
chiara.didato@graduate.univaq.it

Abstract. Smart technologies are becoming increasingly essential in the city of the future. Urban and peri urban dynamics description is a crucial element to manage the changes undergone in the urban settlements. Achieving the sustainability goals set at the global and local levels depends on this. At the same time, the possibility to generalize such approaches to make them universally applicable sees machine learning (ML) techniques as a powerful ally. The enhancement of these models is a topic long since addressed by the scientific community. Nevertheless, such applications focus on a limited in time and space scale.

In the present study a methodology capable of combining semiautomated recognition of urban and peri-urban spatial configurations has been set up. The spatial configurations have been chosen within three different functional urban areas (FUA), considering the broad territorial scale of the Italian central Apennines.

Results represent a first significant step toward the semiautomated urban and peri urban configuration pattern recognition, using supervised machine learning techniques, for an expert based characterization of spatial dynamics.

Keywords: Supervised machine learning · Spatial configurations · Urban planning

1 Introduction

Urban settlements are heterotrophic systems whose sustenance depends on input fluxes from elsewhere [1]. Urban areas are immersed in a territorial framework with which they have direct and indirect interactions. Both these systems are characterized by input and output [2]. The way different urban configurations and forms have different consequences on the environmental, social, and economic aspect of urban and no urban areas has been well deepened in the scientific literature [3].

The transition from urban settlements to natural areas is not even clear because various configurations can occur. The anthropized environment decreases moving away from cores cities and it creates a gradient of different forms of urban structures. The peri-urbanization process has characterized the urban expansion tendence of the 21st

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023 Workshops, LNCS 14109, pp. 16–29, 2023.
https://doi.org/10.1007/978-3-031-37120-2_2

century and it has increased the spreading of artificial urban surface and fragmentation [4, 5].

Peri-urbanization process refers to the process of rapid creation of new settlements surrounding existent cities and along the main road network. The resulting territory is multifunctional, extensive, with a scattered morphology. It has low-density settlements which requests high mobility and transport dependence longer commuting distances, and it can also imply fragmented communities and degraded landscapes [6, 7]. Generally, the causes are articulate and interconnected such as demographic changes, variation in settlement intensity and economic structures [8].

Studies are extensive because researchers dealt with the topic working on different subjects, scale, methodologies, and purpose [9, 10]. The lack of one singular definition reveals the complexity of associate one singular term to a broad concept. Moreover, as every process which involves land, peri-urbanization is a site-specific issue, and it varies depending on characteristics of geographical context. The uncertainty is given by the genesis of these areas which can be uncontrolled and scattered. Generally, it refers to areas of transition and interaction between urban and rural, which are hybrid landscapes with dispersive urban growth. Peri-urban areas are characterised by an extensive use of the space and low-density settlements [7] which are highly cost-demanding [11].

A univocal systematic approach to classify urban and peri-urban configurations is still missing [10]. Characterization and then classification of different configurations would be indeed desirable. The ability to discriminate such gradient is hence necessary to better define the dynamics occurring between urban and natural territorial systems in terms of pressures, threats, and opportunities.

To accomplish a similar task, machine learning (ML) techniques are nowadays being employed to decline different aspects of territorial dynamics [12].

Among the declinations of these technologies, land use change and urban form hold a lot of importance, together with socio economic considerations, infrastructures, and risk analysis [13, 14].

Mohajeri et al. (2023) used machine learning techniques (Random Forest algorithm) to develop density scenarios for the Oxford–Cambridge Arc region [15]. Similarly, machine learning approaches are used to generate landscape typologies, exploiting remote sensing data [16], as well as to extract, quantify and map urban built-up impervious areas [17].

Machine learning techniques are also used to describe important features of the built environment [18], as well as to model urban forms and their relations with sustainability, socio economic and land use issues [19–21].

The potentialities of ML techniques allow a further deepening of these aspects, also offering predictions and scenarios useful for planning actors and policies.

In the present study a methodology able to discriminate spatial configurations has been set up. Six different supervised machine learning (SML) algorithms have been trained. The semiautomatic recognition task works on an expert-based classification of the urban – peri urban gradient (Fig. 1).

The aim of the present work is to set up as linking point between the unsupervised recognition of the urban and no urban shapes [12] and a deepened description of urban configurations [22].

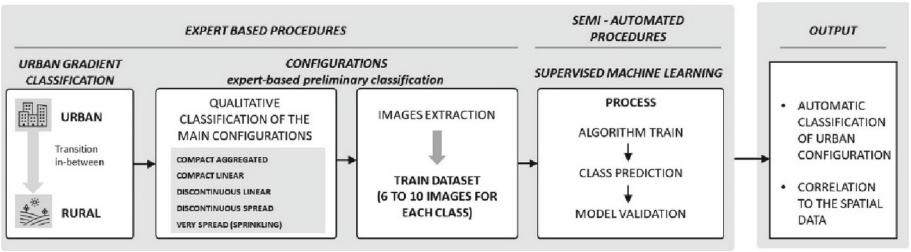

Fig. 1. The main steps of the methodology: the expert based procedure and the semi-automated procedure

2 Materials and Methods

This paragraph aims at comparing three areas around urban centres in central Italy: Perugia, Terni and L'Aquila. The geographical setting is Central Apennines: Perugia and Terni are in Umbria Region and L'Aquila is in Abruzzo Region. Respectively, city's inhabitants are 164880, 110003 and 70019 [23]. Territorial continuity and similarities among the three areas permit to consider them together in this further focus (Fig. 2). They all represent urban poles on the Central Apennines, and they have been involved in recurring seismic events. The geographic area has been subject to seismic events for the last 30 years. Earthquakes of 2009 and 2016 have stroked this geographical area causing differential grades of damage revealing high local level of fragility [24, 25]. Spatial configurations, socio-economic implications and reconstruction processes have influenced actual conditions. Probably, such characteristics will impact also on future development [26]. Phenomena such as seismic events, in fact, emphasise fragilities related to depopulation and abandonment. Such conditions are shared also around such urban centres and then they involve the surrounding minor centres [27].

The analysis is settled inside the perimeters of Functional Urban Areas (FUA) due to reliability of the European standard of Urban Atlas [28] which provides analytical identification of urban poles and their next commuting areas. The Urban Atlas provides pan-European comparable Land Use and Land Cover (LULC) data which describes the spatial distribution analyzed in the present work. Also, the context of FUA is interesting to consider in terms of urban gradient expansion. Such area implicates various processes of land take because its land is the most required for housing, transportation infrastructure, or economic development [5].

2.1 Expert Based Classification of Urban and Peri-Urban Configurations

The first part of the methodology aims at investigating the physical structure of the urban environment. The object is the urban gradient which develops from the main compact urban morphology to the peri-urban areas. Then, the methodology aims at overcoming the vagueness which arises from the multitude of definitions around the urban gradient and peri-urban areas [10]. The methodology introduces the fundamental phase of the expert-based procedure in the process of qualitative evaluation which brings to the further semi-automated procedure.

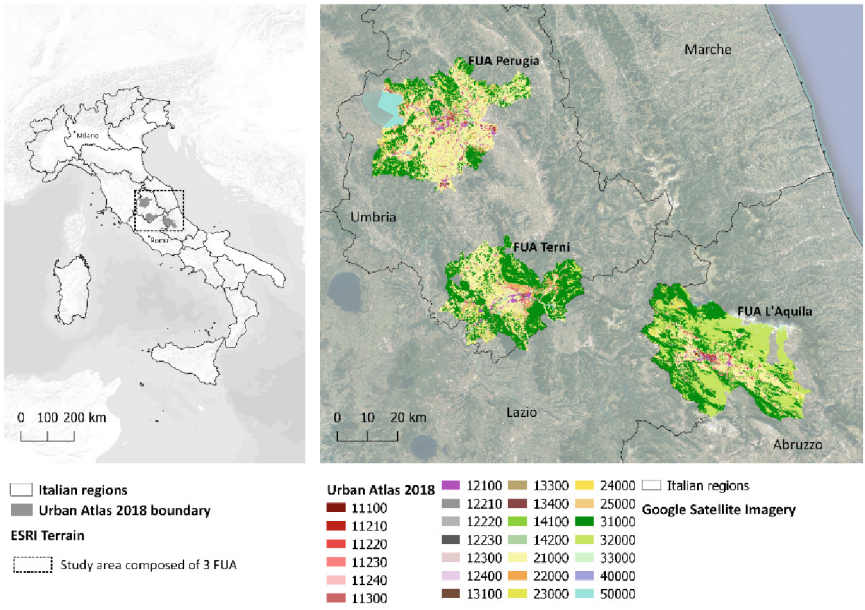

Fig. 2. The FUA of Perugia, Terni and L'Aquila in the context of Central Italy Apennines. Source: Urban Atlas. Authors's elaboration.

LULC data [28] and Google satellite imagery are the main materials of the present analysis. The CORINE Land Cover's layer of artificial surfaces (classes from 1.1 to 1.4) is considered for setting the scale of analysis [29].

Since the medium value of single polygons of artificial surfaces in the CLC is about 100 ha (86,5 ha), the scale of the window of selection is chosen as the double of this data for framing the main geometries of artificial surfaces. In other words, such dimension assures to capture and frame the medium size of the settlements or relevant parts of them. Then, the dimensions of 2 km^2 is taken as reference for capturing the recurrent configurations.

Urban and peri-urban configurations are framed in hexagonal windows of selection. The hexagonal shape is considered suitable for many analyses [12] and it is used for finding recurrent configurations in the case study gradient. The window is arbitrarily stopped where it frames one recurrent configuration.

The process of research of main recurrent configurations is carried out through two steps. Firstly, identification of real settlements approaches the local specificities of the case study. Secondly, the definition of the main classes is guided by the necessity of giving coherence to the specificities in homogeneous classification. The first step of analysis considers qualitative aspects. A punctual, visual, and qualitative analysis is necessary together with process of classification. If classification permits to allocate different examples in different categories due to a process of simplification of the reality, qualitative analysis is fundamental for describing specific characteristics that contribute to allocate examples in classes. Secondly, the classification is organised through the main structures and aggregation features. The shape and structure are taken as primary feature

used to classify the settlements. Shapes refers to main conformation of the artificial area as it appears at the chosen scale.

This work implicates the reduction of complex system (the reality) in different possibilities of settlements (the model) which correspond to parts of the real gradient. The match of identification of the urban structure with interpretation of quality of the urban environment is the step that make the classification possible.

Six main models then are classified from the most compact and aggregated tissue to the absence of buildings and urban shapes.

1. Aggregated, compact and dense. Portions of the urban area which are generally comprehensive of historical centre or they are in proximity of that.
2. Linear, compact, and dense. Settlement is aggregated along a main path which can be a provincial or regional road, or a high-speed road.
3. Linear and discontinuous. Settlement is aggregated along a main path and tissue appears discontinuous.
4. Spread and discontinuous tissue. Settlement is discontinuous and scattered with variable density. There is no one path of reference, but some main infrastructures can be present.
5. Very spread tissue in rural context. Settlement is discontinuous and scattered with very low density. In this model rural context is dominant, whereas buildings are spots which follow streets that connect residential areas.
6. Free areas in rural context. The dominant element is a natural area, such as fields, woods, orchards. Buildings are excluded from this pattern.

As the model of the gradient does, such classification comprehends the free areas. Instead, free areas are excluded from the present analysis. It considers the part of the classification which comprehends urban settlements, and the analysis has been conducted on the first 5 classes which are identified with the related acronyms COMP_A, COMP_L, DISC_L, DISC_R, AGR_D (Fig. 3).

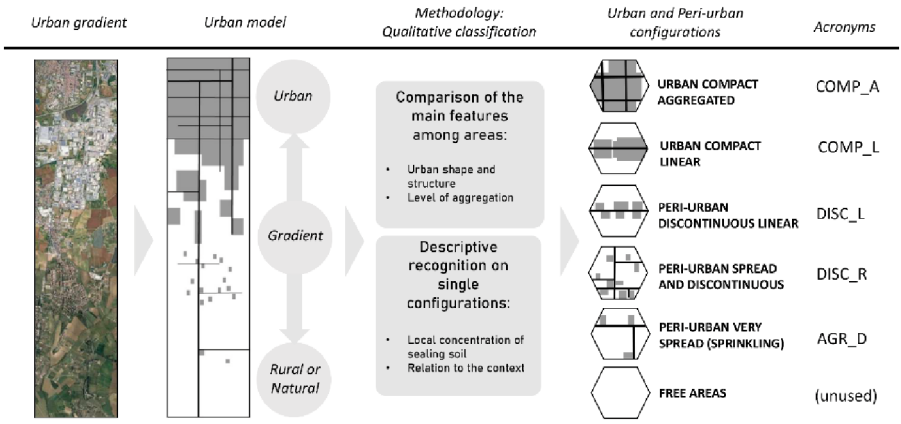

Fig. 3. The six configurations from the analysis on the urban gradient. Source: Google Imagery. Authors' elaboration

2.2 Supervised Machine Learning Algorithms

To perform the analyses open-source software Orange has been used (https://orangedatamining.com/docs/). Despite other software allow a deeper analysis (MATLAB, Rstudio and so on), Orange has been chosen thanks to its user-friendly interface and the possibility to perform machine learning analyses without entering specific procedures that could be less clear for a non-expert user.

To analyse spatial configurations, orange image analysis add-on has been used. This extension allows the user to work with a relatively simple interface, operating some important image analysis operations.

All the following steps have been realized through a train dataset, composed firstly of 30 images, and then expanded to 50 images, equally distributed among 5 classes. Each image has been converted into vector descriptors. To do this, VGG 19 Convolutional Neural Network (CNN) model has been used. This kind of model, pretrained on the ImageNet dataset, has been found to be very useful for image analysis [30].

Based on the information from vector descriptors, the algorithms learn and then formulate a series of rules. Such rules will then help in the association of the categories. In this preliminary step, six different algorithms have been performed (Fig. 4).

For each algorithm cross and stratified validation (k-fold = 5) has been realized. The evaluation of the best algorithm focused on two metrics: Area Under Curve (AUC) and F1-score. AUC is the value of the area under the Receiver Operating Characteristic Curve (ROC curve), which is related to rate of true positive (TP) and rate of false positive (FP). TP are elements assigned to the real class of belonging by the algorithm, whereas FP are assigned to the wrong class compared to the real one. UC values vary from 0 to 1. Values from 0.9 to 1 then provide a highly accurate test [31]. F1-score metrics comes from the harmonic mean of precision and recall. Precision is given by number of TP (elements assigned to the real class of belonging by the algorithm) divided by total number of elements belonging to the class considered (TP and FP) [32]:

$$Precision : \frac{TP}{TP + FP} \tag{1}$$

Recall is the number of TP divided by total number of elements belonging to the class considered (TP and FN). FN represents the number of elements belonging to the class considered but assigned to another class (False Negative):

$$Recall : \frac{TP}{TP + FN} \tag{2}$$

Values of F1-score varies from 0 (worst result) to 1 (best result). Both F1 and AUC metrics are indicative of the model's efficiency based on how it has been trained with the train set.

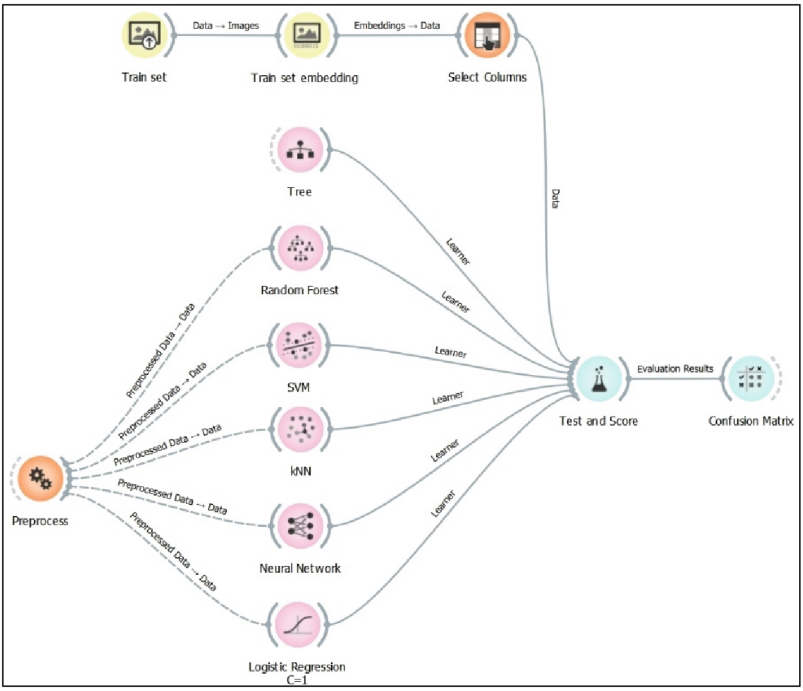

Fig. 4. Learners' performance evaluation. The metrics have been calculated for six main algorithms: Tree; Random Forest; Support Vector Machine (SVM); K-nearest neighbors; Neural Network; Logistic regression. Scheme elaborated by the authors and extracted in Orange. (Colour figure online)

Subsequently, a validation dataset has been realized. It is composed of 25 images and allows validation of the chosen best algorithm, trained with the train set.

The following validation analysis has been conducted firstly associating to each image of the validation set one class, then evaluating the misclassified and the correct ones through a confusion matrix (Fig. 5).

Also for the evaluation of this last step, AUC and F1-score metrics have been used.

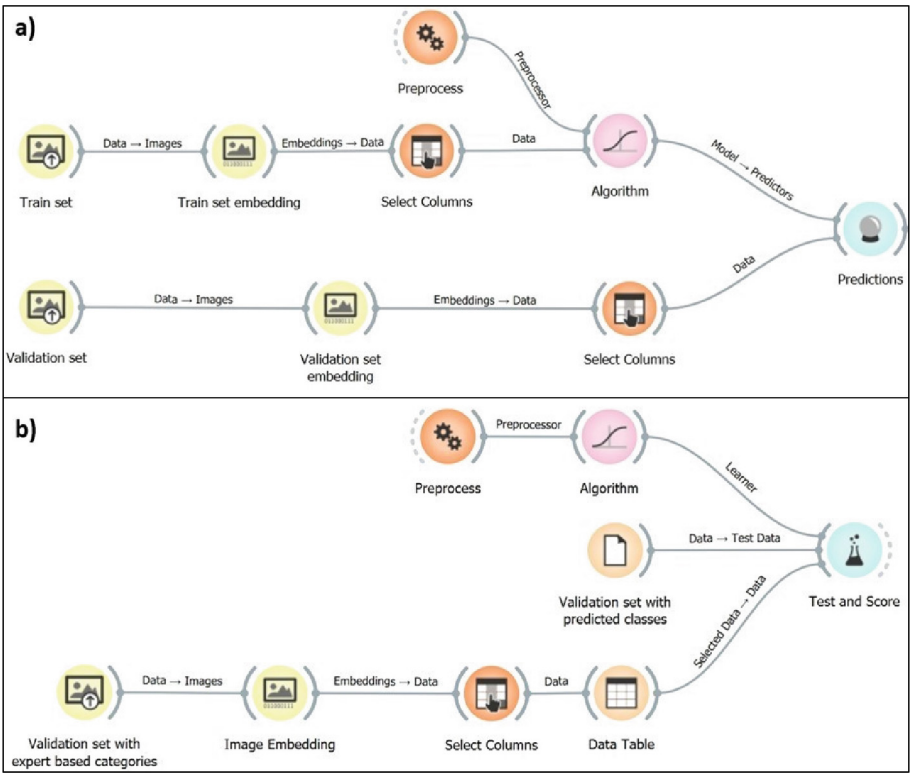

Fig. 5. 1a) First step of class assignment from the chosen best algorithm. After an image embedding, the algorithm learns through the train set and assigns a class to the validation one. 1b) performance evaluation of the validation set, through a comparison with the same images, corresponding to the correct classes (assigned by the operator).

Figure 6 shows a sum up of the followed procedures.

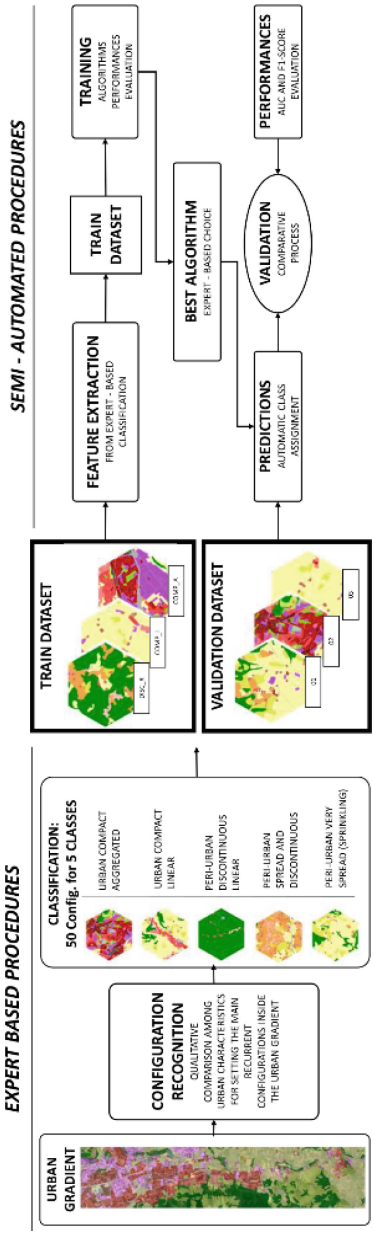

Fig. 6. Explicative workflow of the main passages of the methodological procedures.

3 Results

The first evaluation of algorithms performance returns the highest AUC values for Logistic regression (0.928) and Neural Network (0.898). Instead, highest values of F-1 score corresponds to Logistic regression (0.710), and Random Forest (0.676) algorithms (Table 1). Furthermore, AUC metric slightly decreases for kNN, Tree and SVM algorithms, while remaining quite high for the other ones. F-1 score has instead a fluctuation distributed among all the algorithms, having a clear rise for the logistic regression.

Table 1. Performance metrics algorithms for the training set (30 images). Stratified cross validation (k = 5)

MODEL	AUC	CA	F1	Precision	Recall
Logistic regression C = 1	0.928	0.700	0.710	0.734	0.700
Neural network	0.898	0.600	0.550	0.523	0.600
Random Forest	0.891	0.667	0.676	0.702	0.667
kNN	0.869	0.667	0.670	0.708	0.667
Tree	0.625	0.400	0.405	0.463	0.400
SVM	0.585	0.500	0.508	0.563	0.500

The relative confusion matrix of the train set (composed of 30 images) reports a misclassification of 9 images (about 30%) for logistic regression algorithm. Hence, for the second step of model validation, logistic regression has been used.

The latter algorithm shows performance values for the second step of model validation of 0.796 for the AUC value and 0.556 for the F1-score value, correctly classifying 14 images (56%) out of 25 (Table 2).

Table 2. Confusion matrix validation dataset (train made with 30 images equally distributed among 5 classes, logistic regression algorithm). Acronyms indicate the 5 classes considered out the 6 urban and peri-urban configurations.

		Predicted					
		AGR_D	COMP_A	COMP_L	DISC_L	DISC_R	∑
Actual	AGR_D	4	0	2	2	2	10
	COMP_A	0	2	0	0	0	2
	COMP_L	1	2	3	0	0	6
	DISC_L	0	0	0	2	0	2
	DISC_R	0	1	0	1	3	5
	∑	5	5	5	5	5	25

The same computational phases performed with a train set of 50 images, equally distributed into the 5 classes, show the highest performance values for the random forest algorithm (AUC = 0.854, F1 – score = 0.685). The latter correctly classify 34 (68%) out of 50 images (Table 3).

Table 3. Performance metrics algorithms for the training set (50 images). Stratified cross validation (k = 5)

MODEL	AUC	CA	F1	Precision	Recall
Random Forest	0.854	0.680	0.685	0.697	0.680
Neural Network	0.842	0.580	0.582	0.618	0.580
Logistic Regression C = 1	0.828	0.580	0.585	0.597	0.580
kNN	0.799	0.580	0.583	0.597	0.580
SVM	0.613	0.460	0.470	0.557	0.460
Tree	0.613	0.380	0.369	0.380	0.380

These performance values for the validation phase show an AUC value of 0.849 and an F-1 score of 0.692, correctly classifying 17 images (68%) out of 25 (Random Forest algorithm) (Table 4).

Table 4. Confusion matrix validation dataset (train made with 50 images equally distributed among 5 classes, Random Forest algorithm). Acronyms indicate the 5 classes considered out the 6 urban and peri-urban configurations.

		Predicted					
		AGR_D	COMP_A	COMP_L	DISC_L	DISC_R	∑
Actual	AGR_D	5	0	1	2	0	8
	COMP_A	0	3	0	0	1	4
	COMP_L	0	1	3	0	0	4
	DISC_L	0	0	0	2	0	2
	DISC_R	0	1	1	1	4	7
	∑	5	5	5	5	5	25

4 Discussion and Conclusion

Though AUC for the logistic regression algorithm has a value of 0.928, it isn't the same for the F1 metric, of 0.710. This means that Logistic regression discriminates quite adequately true positive and true negatives but has some difficulties in correctly classifying false positives and false negatives, having direct consequences when assigning new classes to the validation set.

Using a train set of 30 images, model can correctly classify more than half of the validation set (56%).

When considering the train set implemented with 20 more images, performance values slightly decrease compared to the 30 images dataset train and validation phases.

Despite correctly classifying the 68% of the validation set, performance values of the model trained on 50 images are lower than those of the first 30 train set images analysis (AUC drops from 0.928 to 0.854 and F-1 score drops from 0.710 to 0.685).

This means that the number of images describing each class lowers these values due to the amount of additional information that the model receives. However, the number of images correctly classified increases, potentially meaning that random forest algorithm better describes the classes and their differences. Hence, further studies will be aimed at deepening random forest's performance values, also due to its ability to avoid overfitting and to work with non – linear relationships [33].

Furthermore, being the train set firstly composed of 30 images equally distributed into 5 classes, the capacity of the trained algorithm to assign the correct class is based on the parameters extracted from only 6 figures per class from the train set. Hence, it is likely that the dataset still doesn't reach its full potential, to the point of minimizing its performances. This aspect is supported by the comparison with the same analysis performed using a 50 images train set that, despite being lower than the first train, it returns a higher percentage of correctly classified images.

The passages presented in this article belong to two different approaches, apparently distant from each other. While the classification is based on an expert-based lecture of the spatial configurations, to reach the full potential in its applicability to a wider scale it is necessary to introduce Machine Learning techniques to "teach" to a machine to do this work. Contextually, this kind of technologies can't fulfill the tasks for which they are trained without an operator driven process that connects all the steps to build a reliable, efficient system.

In this sense, integration is the key element able to make it work.

Specifically, the presented work is a systematization aimed at integrating a qualitative-descriptive and a computation approach, for producing integrated land planning tools [34].

Such as other approaches based on machine learning techniques for the automatization of complicated procedures and tasks, also this model is subject to under or overfitting risk [35]. Referring to these aspects, further studies will be aimed not only at extending the areas on which train dataset will be constructed, but also to further deepen the performances of such model, for example in the recognition of other territorial scopes. Regarding this, expert based classification will then need to be finalized to avoid all these potential criticalities, to have not only a model that performs well in the analyzed

territorial scope, but is also able to recognize the correct class, regardless of where it belongs.

Although in its embryonal phases, this research encloses a first fundamental step towards the semi-automated recognition of spatial configurations.

The next phase/step will be a finer model tuning and a deeper description of the classes that will then be used to evaluate and map territorial configurations.

References

1. Pickett, S.T.A., et al.: Urban ecological systems: linking terrestrial ecological, physical, and socioeconomic components of metropolitan areas. Annu. Rev. Ecol. Syst. **32**, 127–157 (2001)
2. Pilogallo, A., Saganeiti, L., Fiorini, L., Marucci, A.: Ecosystem services for planning impacts assessment on urban settlement development. In: Gervasi, O., Murgante, B., Misra, S., Rocha, A.M.A.C., Garau, C. (eds.) ICCSA 2022. LNCS, vol. 13380, pp. 241–253 (2022). https://doi.org/10.1007/978-3-031-10542-5_17
3. Marucci, A., Fiorini, L., Dato, C.D., Zullo, F.: Marginality assessment: computational applications on Italian municipalities. Sustainability **12**, 3250 (2020). https://doi.org/10.3390/SU12083250
4. Ravetz, J., Fertner, C., Nielsen, T.S.: The dynamics of peri-urbanization. In: Nilsson, K., Pauleit, S., Bell, S., Aalbers, C., Sick Nielsen, T. (eds.) Peri-Urban Futures: Scenarios and Models for Land use Change in Europe, pp. 13–44. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-30529-0_2
5. European Environment Agency: Land take and land degradation in functional urban areas (2021)
6. ISTAT: Forme, livelli e dinamiche dell'urbanizzazione in Italia (2017)
7. Piorr, A., Ravetz, J., Tosics, I.: Peri-urbanisation in Europe Towards European Policies to Sustain Urban-Rural Futures SyNThESIS REPORT (2011)
8. Cattivelli, V.: Methods for the identification of urban, rural and peri-urban areas in Europe: an overview. J. Urban Regeneration Renewal **14**, 240–246 (2021)
9. Cattivelli, V.: Delimiting rural areas: evidence from the application of different methods elaborated by Italian scholars. Land **11**, 1674 (2022). https://doi.org/10.3390/LAND11101674
10. Sahana, M., Ravetz, J., Patel, P.P., Dadashpoor, H., Follmann, A.: Where is the peri-urban? A systematic review of peri-urban research and approaches for its identification and demarcation worldwide. Remote Sens. **15**, 1316 (2023). https://doi.org/10.3390/RS15051316
11. Manganelli, B., Murgante, B., Saganeiti, L.: The social cost of urban sprinkling. Sustainability **12**, 2236 (2020). https://doi.org/10.3390/SU12062236
12. Fiorini, L., Falasca, F., Marucci, A., Saganeiti, L.: Discretization of the urban and non-urban shape: unsupervised machine learning techniques for territorial planning. Appl. Sci. **12**, 10439 (2022). https://doi.org/10.3390/app122010439
13. Casali, Y., Aydin, N.Y., Comes, T.: Machine learning for spatial analyses in urban areas: a scoping review. Sustain Cities Soc. **85**, 104050 (2022). https://doi.org/10.1016/J.SCS.2022.104050
14. Dobesova, Z.: Experiment in finding look-alike European cities using urban atlas data. ISPRS Int. J. Geoinf. **9**, 406 (2020)
15. Mohajeri, N., et al.: A machine learning methodology to quantify the potential of urban densification in the Oxford-Cambridge Arc, United Kingdom. Sustain Cities Soc. **92**, 104451 (2023). https://doi.org/10.1016/J.SCS.2023.104451

16. van Strien, M.J., Grêt-Regamey, A.: Unsupervised deep learning of landscape typologies from remote sensing images and other continuous spatial data. Environ. Model. Softw. **155**, 105462 (2022). https://doi.org/10.1016/J.ENVSOFT.2022.105462
17. Misra, M., Kumar, D., Shekhar, S.: Assessing machine learning based supervised classifiers for built-up impervious surface area extraction from sentinel-2 images. Urban For. Urban Green. **53**, 126714 (2020). https://doi.org/10.1016/J.UFUG.2020.126714
18. Koc, M., Acar, A.: Investigation of urban climates and built environment relations by using machine learning. Urban Clim. **37**, 100820 (2021). https://doi.org/10.1016/J.UCLIM.2021.100820
19. Koumetio Tekouabou, S.C., Diop, E.B., Azmi, R., Jaligot, R., Chenal, J.: Reviewing the application of machine learning methods to model urban form indicators in planning decision support systems: potential, issues and challenges. J. King Saud Univ. Comput. Inf. Sci. **34**, 5943–5967 (2022). https://doi.org/10.1016/J.JKSUCI.2021.08.007
20. Lin, J., Qiu, S., Tan, X., Zhuang, Y.: Measuring the relationship between morphological spatial pattern of green space and urban heat island using machine learning methods. Build Environ. **228**, 109910 (2023). https://doi.org/10.1016/J.BUILDENV.2022.109910
21. Kafy, A.A., et al.: Predicting the impacts of land use/land cover changes on seasonal urban thermal characteristics using machine learning algorithms. Build. Environ. **217**, 109066 (2022). https://doi.org/10.1016/J.BUILDENV.2022.109066
22. Marucci, A., Di Dato, C.: Urban and peri-urban shapes for sustainable governance. In: Wolski, J., Regulska, E., and Affek, A. (eds.) IALE 2022 European Landscape Ecology Congress Book of Abstracts, p. 341. IGiPZ PAN, Warsaw (2022). https://doi.org/10.7163/Konf.0004
23. ISTAT: Istituto Nazionale di Statistica. https://www.istat.it/. Accessed 14 Apr 2023
24. Forino, G.: Disaster recovery: narrating the resilience process in the reconstruction of L'Aquila (Italy). Geografisk Tidsskrift Danish J. Geogr. **115**, 1–13 (2015). https://doi.org/10.1080/00167223.2014.973056
25. Romano, B., Fiorini, L., Sette, C.: #comeradovera Quale città e quale territorio: L'Aquila a 13 anni dal sisma. EcoWebTown **25**, 28–33 (2022)
26. Compagnucci, F., Morettini, G.: Abandoning the Apennines? The anthropo-systemic value of the Italian inner areas within the 2016-17 seismic crater (2021)
27. Di Dato, C., Marucci, A.: Fragile territories around cities: analysis on small municipalities within functional urban areas. In: Gervasi, O., Murgante, B., Misra, S., Rocha, A.M.A.C., Garau, C. (eds.) ICCSA 2022. LNCS, vol. 13378, pp. 427–438 (2022). https://doi.org/10.1007/978-3-031-10562-3_30
28. Copernicus Land Monitoring Service: Urban Atlas. https://land.copernicus.eu/local/urban-atlas. Accessed 14 Apr 2023
29. Copernicus Land Monitoring Service: CORINE Land Cover. https://land.copernicus.eu/pan-european/corine-land-cover. Accessed 14 Apr 2023
30. Godec, P., et al.: Democratized image analytics by visual programming through integration of deep models and small-scale machine learning. Nat. Commun. **10**(1), 1–7 (2019). https://doi.org/10.1038/s41467-019-12397-x
31. Manning, C.D.: An Introduction to Information Retrieval. Cambridge University Press, Cambridge (2009)
32. Fawcett, T.: An introduction to ROC analysis. Pattern Recognit. Lett. **27**, 861–874 (2006)
33. Lee, H., Wang, J., Leblon, B.: Using linear regression, random forests, and support vector machine with unmanned aerial vehicle multispectral images to predict canopy nitrogen weight in corn. Remote Sens. **12**, 2071 (2020)
34. Koutra, S., Ioakimidis, C.S.: Unveiling the potential of machine learning applications in urban planning challenges. Land **12**, 83 (2023)
35. Jabbar, H., Khan, R.Z.: Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). Comput. Sci. Commun. Instrum. Devices **70**, 163–172 (2015)