Anthropogenic sinkholes' susceptibility assessment in Palermo, Italy, using a machine learning algorithm

Giuseppe Bausilio[1], Vincenzo Allocca[1], Chiara Cappadonia[2], Diego Di Martire[1], Luigi Guerriero[1], Marcello Panzica La Manna[3], Rita Tufano[1] & Domenico Calcaterra[1]

[1]Department of Earth Sciences, Environment and Resources, University of Naples Federico II, Napoli, Italy.
[2]Department of Earth and Marine Sciences, University of Palermo, Palermo, Italy.
[3]Geologist, Palermo, Italy.

GB, 0000-0003-0977-2253; VA, 0000-0001-7146-2438; CC, 0000-0002-5119-3004; DDM, 0000-0003-0046-9530; LG, 0000-0002-5837-5409; RT, 0000-0002-3188-2291; DC, 0000-0002-3480-3667.

Rend. Online Soc. Geol. It., Vol. 67 (2025), pp. 15-20, 3 figs., 1 tab. https://doi.org/10.3301/ROL.2025.26

Short Note

Corresponding author e-mail: giuseppe.bausilio@unina.it

Citation: Bausilio G., Allocca V., Cappadonia C., Di Martire D., Guerriero L., Panzica La Manna M., Tufano R. & Calcaterra D. (2025) - Anthropogenic sinkholes' susceptibility assessment in Palermo, Italy, using a machine learning algorithm. Rend. Online Soc. Geol. It., 67, 15-20, https://doi.org/10.3301/ROL.2025.26.

Guest Editor: Guido Leone

Submitted: 13 January 2025
Accepted: 20 August 2025
Published online: 16 October 2025

SOCIETÀ GEOLOGICA ITALIANA ETS
FONDATA NEL 1881 · ENTE MORALE R. D. 17 OTTOBRE 1885

ABSTRACT

Anthropogenic sinkholes are a widespread phenomenon in Italy, and Palermo, the capital city of the Sicilian Region in South Italy, is one of the urban areas most affected. Anthropogenic sinkholes refer to vertical depressions, usually circular or sub-circular in plan. They can vary from localised subsidence to actual collapses, often caused by either the presence of unstable underground man-made cavities or by voids related to aqueduct or sewer leakages. The machine learning algorithm Maximum Entropy was employed to evaluate the anthropogenic sinkhole susceptibility in the Palermo urban area. Given the outstanding results provided by this machine learning algorithm (ROC/AUC score = 0.926), it can be considered a valuable tool in urban planning and cultural heritage protection.

KEYWORDS: anthropogenic sinkhole, susceptibility, machine learning, Palermo, Italy.

INTRODUCTION

Sinkholes are a widespread phenomenon consisting of a vertical enclosed depression often related to the presence of karst environments or evaporite rocks, regardless of the existence of an internal drainage network (Waltham et al., 2005). Sinkholes' shape appears to be mainly circular or semi-circular in plan, with a cylindrical or a funnel-like three-dimensional shape. Sinkhole dimensions can vary from sub-metrical up to tens of meters both in depth and diameter (Gutiérrez, 2016). Sinkholes can affect the human sphere and the social fabric causing significant economic damage or loss of human life (Weary, 2015).

When these phenomena affect the urban fabric or man-made cavities, they are also called "anthropogenic sinkholes" (Guarino & Nisio, 2012; Tufano et al., 2022) and, among the Italian cities, Palermo (Sicily Region, South Italy) is one of the most affected (Nisio, 2018).

To comprehend the processes involved and to provide a usable tool for urban development, an anthropogenic sinkhole susceptibility map was produced. To this purpose, a Machine Learning (ML) algorithm Maximum Entropy (MaxEnt - Phillips et al., 2006), was employed. The latter is an ML algorithm largely used in literature, with good results, from landslide susceptibility assessment (Park, 2015), to forest fire (Arpaci et al., 2014), to anthropogenic sinkhole susceptibility assessment (Bausilio et al., 2022).

STUDY AREA

Palermo is located in northwestern Sicily, at the centre of the Mediterranean Sea. (Fig. 1). The city was founded by the Phoenicians during the eighth century BC. The main urban fabric of the city has grown within the so-called "*Conca d'Oro*", a coastal plain surrounded by mountains. The geomorphology reflects the geological setting of the city: the mountains and the

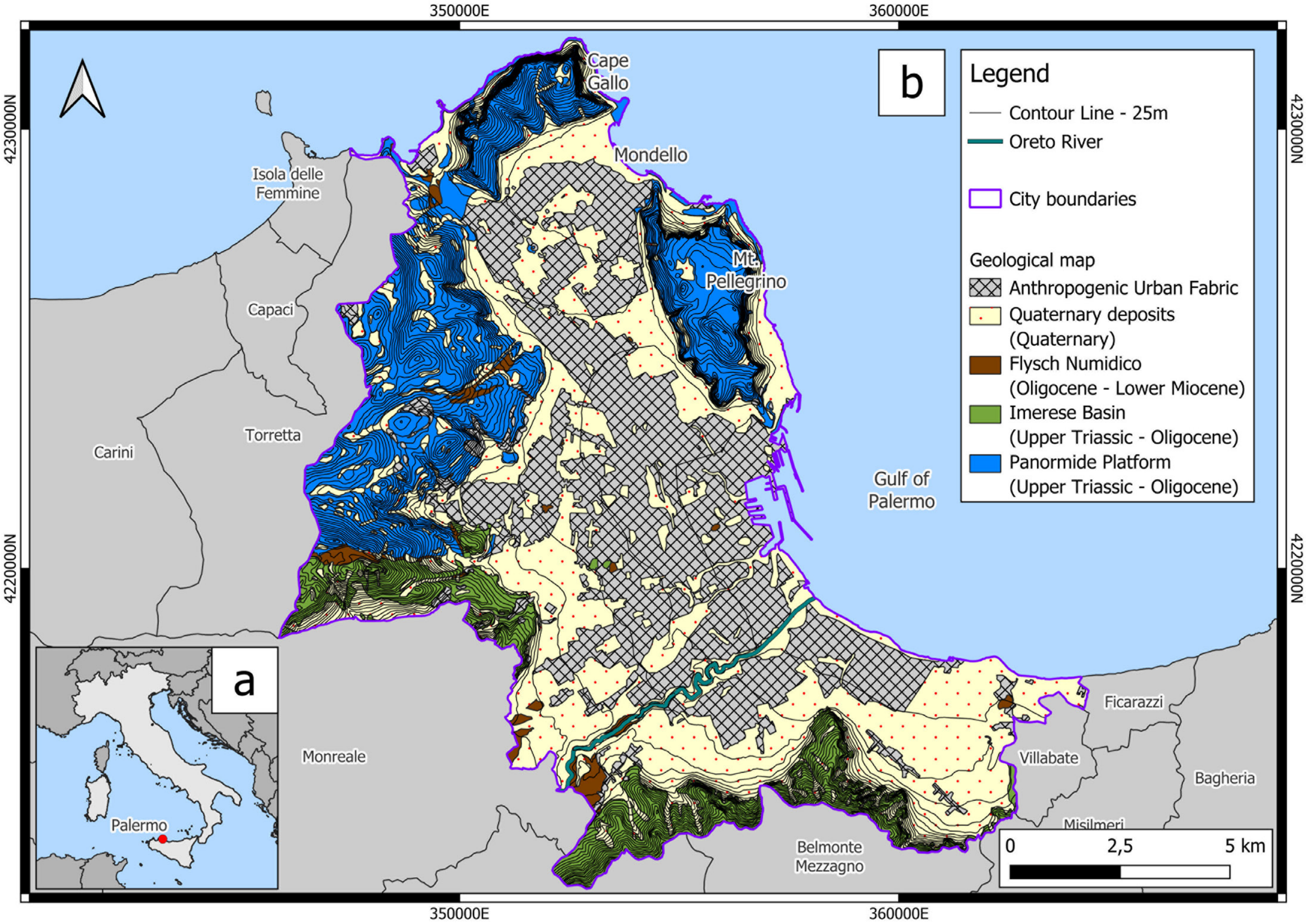

Fig. 1 - a) Location of the city of Palermo (red dot); b) Geological map of the city of Palermo (simplified from Catalano et al., 2013a, b; Servizio Geologico d'Italia, 2013a, b).

topographic highs are mainly made up of carbonate rocks, while the topographic lows and the coastal plains are mainly related to the presence of normal faults or clayey deposits, which are more easily eroded (Catalano et al., 2013a; Servizio Geologico d'Italia 2013a). As for the hydrographic network, the city was crossed by multiple rivers in the past. As of right now, only the Oreto river crosses the city, while the Kemonia, Papireto, and Maltempo were channeled and buried (Tranchina et al., 2008; Catalano et al., 2013a; Servizio Geologico d'Italia 2013a). The mountains that surround the city are part of the Sicilian-Maghrebian fold and thrust belt, formed due to the convergence of Africa and Europe and the successive Ionian lithosphere subduction roll-back (Catalano et al., 2000).

The main geological formations are related to the Panormide Platform, in the Northern sector of the Palermo Mountains, and the Imerese Basin, mostly located in the Southern sector of the Palermo Mountains (Abate et al., 1982). The Panormide Platform is mainly made up of carbonate rocks, mainly limestones, marly limestones, calcareous and dolomitic breccias, and argillite levels, while the Imerese Basin is mostly made of limestones, dolomitic limestones, and calcareous dolomites (Catalano et al., 2013a, b; Servizio Geologico d'Italia 2013a, b). The Numidian Flysch, mostly located at the top of the calcareous formations or between the Imerese Basin and the Panormide Platform, is mainly made of argillites, alternating levels of argillites and quartz arenites and brown pelites. Lastly, the quaternary deposits are related to the deposits that can be found along the slopes (debris, colluvial and fluvial deposits) or along the "Conca d'Oro" plain (alluvial, marine, and aeolian deposits).The urban area is developed mainly on the quaternary marine/coastal deposits of the Marsala synthem, which are a hundred meters thick and consist principally of bioclastic calcarenites, sands, and greyish clayey silts (Catalano et al., 2013a, b; Servizio Geologico d'Italia 2013a, b; Cappadonia et al., 2020). Palermo, just like other Italian cities such as Rome or Naples, is characterised, since ancient times, by the presence of underground cavity networks used for the extraction of a Pleistocene calcarenite largely used as building material (Todaro, 2004). On top of the quarries created for the extraction of building materials, catacombs, places of worship, cellars and hydraulic works firstly built by the Arabs for irrigation purposes (the so-called qanats) (Todaro et al., 2020) are also present within the city boundaries, forming a widespread network of underground man-made cavities that nowadays also represent an economic resource as a tourist attraction.

METHODOLOGY

The anthropogenic sinkhole susceptibility of Palermo City was assessed using the Maximum Entropy machine learning algorithm (Phillips et al., 2006); to this scope, the Maxent software was employed (http://biodiversityinformatics.amnh.org/open_source/maxent/ - Version 3.4.4). The Maximum Entropy algorithm relies on the assumption that the model that can better describe the spatial distribution of a species can be achieved by maximizing the entropy, which is the degree of resolution of the system, and it is also used to evaluate the environmental variable importance within the model (Phillips et al., 2006). Maximum Entropy algorithm gains information about the predisposing factors in correspondence of the so-called Presence points (the anthropogenic sinkhole inventory) and compares this data with the information obtained in correspondence of the Background points (random points within the study area). The data collected in correspondence of the Presence points provides the "habitat" of the geohazard, while the data collected using the Background points represent the entirety of the "habitats" available within the study area. By comparing these data, the algorithm evaluates whether an area falls within a collection of cells with high indices of "habitat" suitability (Merow et al., 2013). While the Background points are random and scattered all over the study area, the Presence data is part of a sinkhole inventory produced using literature data (Sottile, 2010, 2016), the ISPRA national inventory, and national and local newspapers. This inventory contains a total of 278 sinkholes within the Palermo city boundaries. Due to scarcity of information linked to the triggering factors, the aim of this work is focused on the predisposing factors. In fact, the data related to the triggers contained in the inventory are: i) aqueduct/sewer leaks (8.7%), ii) heavy rainfall (15.8%), while iii) for most of the entries (75.5%), the triggering factor data is not available.

Twelve predisposing factors were used in this analysis: i) presence/ absence of areas affected by extraction activities; ii) slope angle; iii) geological data; iv) land use data; v) cover layer thickness; vi) Topographic Wetness Index (TWI); vii) density of underground cavity network; viii) distance to underground cavity network; ix) hydrographic network density; x) distance to hydrographic network; xi) road density; xii) distance to road network.

The slope and TWI maps were obtained from the Digital Terrain Model (DTM) produced by the Italian Ministry of the Environment, Land and Sea (2005), with a resolution of 20 m × 20 m, while the cover layer thickness data has been gathered from professional geologic reports.

The predisposing factors importance is evaluated by the software using the permutation approach. While using this approach, the software replaces the values of a predisposing factor randomly, and if the new values lead to a strong reduction of the model performance, then the analyzed parameter is significant to the model (Phillips, 2017). If the permutation does not influence the final performance, then the analyzed predisposing factor is not important for modeling purposes. Any loss of performance is measured, normalised, and finally expressed as a percentage (Phillips, 2017).

A K-Fold Cross-validation approach has also been employed. This technique divides the Presence inventory into five subsets, four of them used for the algorithm training phase and one for performance evaluation purposes. This approach makes it possible to avoid positive bias related to performance evaluation carried out using data that has also been used during the training phase. Subsequently, one of the previous four training sets is used as a new performance evaluation set, while the remaining four are used to train a new model. This process is iterated until all subsets have been used as performance evaluation sets, obtaining five different models. An average of these five models is then evaluated and used as the final model.

The final performance is evaluated using the ROC/AUC approach (Fauwcett, 2006). The ROC (Receiver Operating Characteristics) curve is built by plotting the Sensitivity on the y-axis and the false positive rate (1-Specificity) on the x-axis:

$$Sensitivity = \frac{(areas\ affected\ by\ sinkhole\ and\ detected\ as\ susceptible)}{(total\ area\ affected\ by\ sinkholes)}$$

$$Specificity = \frac{(areas\ not\ affected\ by\ sinkholes\ and\ detected\ as\ not\ susceptible)}{(total\ area\ not\ affected\ by\ sinkholes)}$$

By calculating the area under the ROC curve (AUC = Area Under the Curve), which varies between 0.5 and 1.0, the ROC/AUC score is obtained. There is not a strictly defined interpretation of the ROC/AUC score. In general: i) a score equal to 0.5 is no different from a random guess; ii) a score between 0.5 and 0.7 is related to poor discrimination; iii) between 0.7 and 0.8 the performance is acceptable; iv) values between 0.8 and 0.9 are excellent; and for v) ROC/AUC scores > 0.9 the performance is outstanding (Hosmer et al., 2013).

RESULTS

The first product of this analysis is the permutation importance of the predisposing factors (Tab. 1). The permutation importance is obtained by randomly modifying one of the predisposing factor values and evaluating the performance decrease; the values shown are normalised to 100%. The most influential predisposing factor is the distance to the road network (64%), followed by the distance to underground cavities (8.0%), the geological data (6.7%), and the density of underground cavities (4.6%). The sinkhole susceptibility map obtained has been divided into five classes (Very Low, Low, Medium, High, and Very High) using the Natural Breaks Method (Jenks, 1967). The resulting classes' areal extensions are characterised by a negative trend (Fig. 2a). 78.8% of the city is covered by the Very Low susceptibility class, 6.6% by Low susceptibility, 6.3% by Medium susceptibility class, and 5.4% and 2.9% by, respectively, High and Very High susceptibility classes. The sinkhole frequency within every class (Fig. 2b) shows, instead, an increasing trend: 3.2% of the 278 sinkholes fall within the Very Low susceptibility class, 4.3% are related to the Low class, 17.3% of the sinkholes fall within the Medium susceptibility class, and 27.7% and 47.5% of the sinkholes fall within, respectively, the High and Very High susceptibility classes.

Table 1 - Permutation importance [%] of the predisposing factors. Values are an average obtained from the five models.

Variable	Permutation importance [%]
01_Mining_Areas	4.4
02_Slope	1.8
03_Geology	6.7
04_Land_Use	4.4
05_Cov_Thic	1.0
06_TWI	0.5
07_Cav_Den	4.6
08_Cav_Dis	8.0
09_Hydr_Den	2.1
10_Hydr_Dis	1.4
11_Road_Den	1.2
12_Road_Dis	64.0

As shown in Figure 3a, the High and Very High Sinkhole Susceptibility classes are mainly located in correspondence of the historical city centre of Palermo, around the port, and its immediate hinterland. Other sectors affected by anthropogenic sinkholes are the eastern and northern ones, which appear to be affected mostly by High and Medium Sinkhole Susceptibility. Regarding the areas related to the Palermo Mts. and Mt. Pellegrino, it appears that such areas are not affected significantly by Anthropogenic Sinkholes Susceptibility, as they fall mostly within the Very Low Susceptibility class. In total, the Very High Susceptibility class is characterised by a sinkhole density of 28.6 sinkhole/km^2, more than three times the sinkhole density related to the High Susceptibility class (8.9 sinkhole/km^2). The remaining classes are characterised by a density of 4.7 (Medium), 1.1 (Low), and 0.07 (Very Low) sinkhole/km^2.

DISCUSSION

The anthropogenic sinkhole susceptibility map performance was evaluated in terms of ROC/AUC score. The final mean score reached an outstanding value of 0.926 (Fig. 3b). The susceptibility map confirms the critical importance of the distance to the road network, which is mainly related to the presence of aqueduct and/or sewer networks. The related pipelines, when damaged and affected by leaks, can cause underground soil erosion and formation of unknown underground cavities (Dastpak et al., 2023). As a specific map of the aqueduct and sewer network was not available at the time of this work, the road network map has been used in its stead. The road density seems to be less important, probably because in Palermo high values of road density are widespread all over the study area.

Following the road distance map, the underground cavities and the geological map seem to be the other most important predisposing factors. As for the distance to underground cavities (and the density of underground cavities, the fourth most influential predisposing factor), their importance in sinkhole susceptibility assessment was expected as the presence of underground discontinuities (such as empty caves) is critical for sinkholes' formation. The Geological data (Fig. 1) also results as the third most important predisposing factors (Tab. 1). This result was expected as most sinkholes are located within the urban and the Quaternary deposit areas, where the city of Palermo is located. Similarly, the land use predisposing factor is mainly related to the urban areas and not the agricultural or natural areas. As for the "mining areas" predisposing factor, this data is related to underground building stone extraction activities within the city limits. The city centre of Palermo, like other Italian cities, has been affected by this practice since ancient times, affecting the final results. On the other hand, the hydrographic network does not seem to be impactful over sinkhole susceptibility, similarly to the TWI. As for the slope, this is probably due to the location of the sinkholes, as they are almost all located over the plain at the foot of the Palermo Mts.

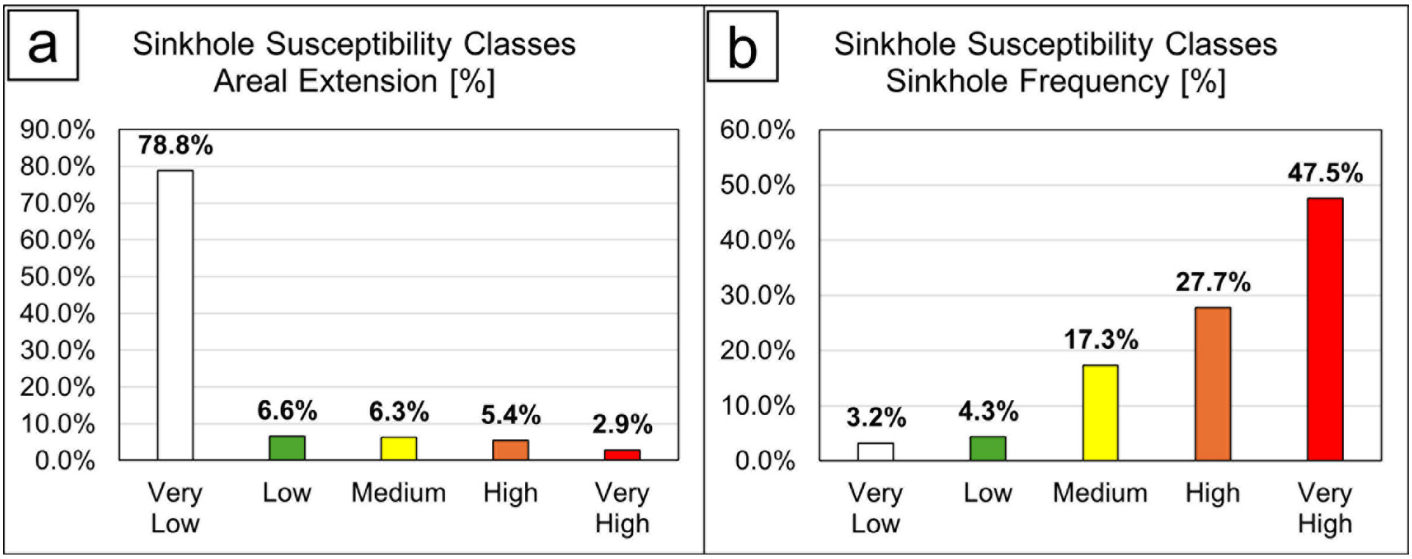

Fig. 2 - a) Areal extension of the five susceptibility classes (in percentage); b) sinkhole frequency within every sinkhole susceptibility class.

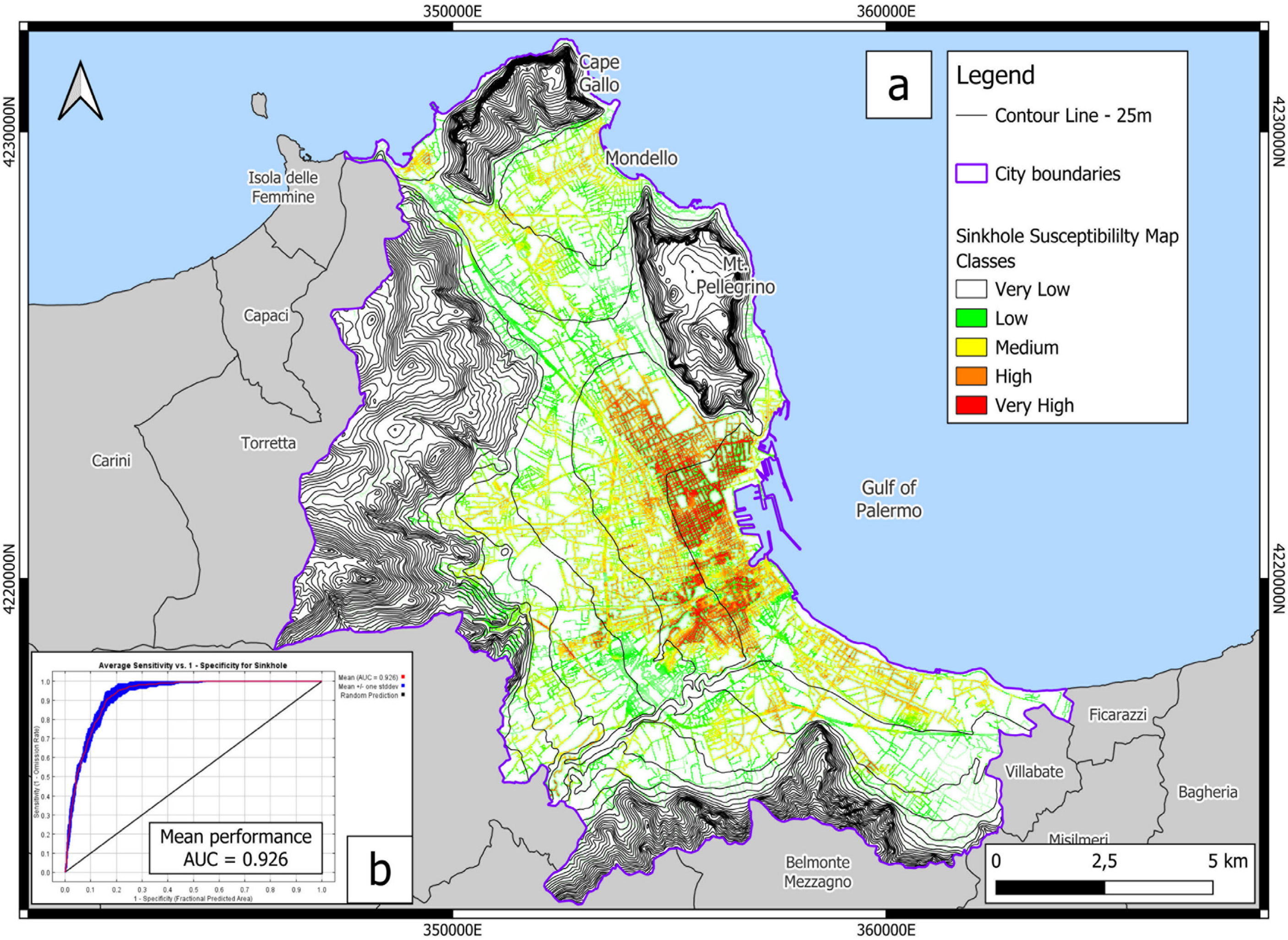

Fig. 3 - a) Sinkhole Susceptibility map for the city of Palermo; b) performance obtained for the model evaluated by MaxEnt, in red the mean AUC performance score (AUC = 0.926), in blue the standard deviation.

CONCLUSIONS

Anthropogenic sinkholes are a serious problem all over the world and, in the case of historical cities, a danger to the cultural heritage and the related economical assets. Palermo is one of the Italian cities most exposed to sinkhole-related failure risk, due to its vast underground network of man-made cavities built for different purposes, from building material extraction to irrigation. In particular, the data used to produce the Underground Cavity Density and the Distance to Underground Cavity layers are mostly related to i) qanats (43.6%), ii) drainage galleries (18.8%), iii) mobility-related galleries (underground roads and railways) (18.8%), while iv) for the remaining 18.8% of the underground cavities there is not available data. To prevent serious damage to important cultural assets or possible loss of human life, the development of tools that aim at helping the local administrations during urban planning is an important goal to achieve: the sinkhole susceptibility map here obtained is a step towards these tools. The results obtained confirm the good quality of the susceptibility maps obtained using ML algorithms, even in highly and densely urbanised areas such as the city of Palermo.

ACKNOWLEDGEMENTS

This work was carried out within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan – NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005). The authors would also like to thank the anonymous reviewers and the editors for providing valuable suggestions, which helped to improve the quality of this manuscript.

REFERENCES

Abate B., Catalano R. & Renda P. (1982) - I monti di Palermo. Mem. Soc. Geol. It., 24, 43-48.

Arpaci A., Malowerschnig B., Sass O. & Vacik H. (2014) - Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. Appl. Geogr., 53, 258-270.

Bausilio G., Annibali Corona M., Di Martire D., Di Napoli M., Guerriero L., Tufano R. & Calcaterra D. (2022) - Evaluation of anthropogenic sinkhole susceptibility in the city of Naples (Italy) using a presence only algorithm. Rend. Onlien Soc. Geol. It., 56, 45-49.

Cappadonia C., Di Maggio C., Agate M. & Agnesi V. (2020) - Geomorphology of the urban area of Palermo (Italy). J. Maps, 16(2), 274-284.

Catalano R., Franchino A., Merlini S. & Sulli A. (2000) - Central western Sicily structural setting interpreted from seismic reflection profiles. Mem. Soc. Geol. It, 55, 5-16.

Catalano R., Avellone G., Basilone L., Contino A., Agate M., Di Maggio C., Lo Iacono C., Sulli A., Gugliotta C., Gasparo Morticelli M., Caputo G., Albanese C., Di Maio D., Vattano M. & Lo Cicero G. (2013a) - Note Illustrative della Carta Geologica d'Italia alla scala 1:50.000, F. 595 Palermo. 271 pp., ISPRA, Roma.

Catalano R., Basilone L., Di Maggio C., Gasparo Morticelli M., Agate M., Avellone G., Mancuso M., Contino A., Valenti V., Vaccaro F., Lena G., Sulli A., Di Stefano E. & Caputo G. (2013b) - Note Illustrative della Carta Geologica d'Italia alla scala 1:50.000, F. 594-585 Partinico-Mondello. 271 pp., ISPRA, Roma.

Dastpak P., Sousa R. L. & Dias D. (2023) - Soil erosion due to defective pipes: a hidden hazard beneath our feet. Sustainability, 15(11), 8931.

Fawcett T. (2006) - An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

Guarino P.M. & Nisio S. (2012) - Anthropogenic sinkholes in the territory of the city of Naples (Southern Italy). Physics and Chemistry of the Earth, Parts A/B/C, 49, 92-102.

Gutiérrez F. (2016) - Sinkhole hazards. In Oxford research encyclopedia of natural hazard science.

Italian Ministry of the Environment, Land and Sea (2005) - DTM20, Digital Terrain Model of the Italian Peninsula with 20 Meters of Resolution. https://gn.mase.gov.it/portale/wcs.

Hosmer Jr D.W., Lemeshow S. & Sturdivant R.X. (2013) - Applied logistic regression. John Wiley & Sons.

Jenks G.F. (1977) - Optimal data classification for choropleth maps occasional paper no 2. University of Kansas, Department of Geography.

Merow C., Smith M.J. & Silander Jr J.A. (2013) - A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography, 36(10), 1058-1069.

Nisio S. (2018) - I Sinkholes Antropogenici nelle Città Italiane. In XIV Rapporto Qualità dell'ambiente urbano—Edizione 2018. ISPRA. In Italian.

Park N.W. (2015) - Using maximum entropy modeling for landslide susceptibility mapping with multiple geoenvironmental data sets. Environ. Earth Sci., 73, 937-949.

Phillips S.J., Anderson R.P. & Schapire R.E. (2006) - Maximum entropy modeling of species geographic distributions. Ecol. Model., 190(3-4), 231-259.

Phillips S.J. (2017) - A Brief Tutorial on Maxent. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2024-10-7.

Servizio Geologico d'Italia (2013a) - Carta Geologica d'Italia alla scala 1:50.0000, F. 595 Palermo. ISPRA, Roma.

Servizio Geologico d'Italia (2013b) - Carta Geologica d'Italia alla scala 1:50.0000, F. 594-585 Partinico-Mondello. ISPRA, Roma.

Sottile R. (2010) - Problematiche geologiche legate agli sprofondamenti in area urbana. L'esempio della città di Palermo: Rifugi antiaerei, Canali sotterranei, Cavità Antropiche. Master's degree Thesis. University of Palermo. In Italian.

Sottile R. (2016) - Census and geological study of the anthropogenic sinkholes in the urban area of Palermo (Southern Italy). PhD Thesis. University of Siena.

Todaro P. (2004) - Le cave di calcarenite del palermitano: Mucati e pirrere. Cave–Arte e Cultura. Palermo: Istituto Statale d'Arte, 29-42.

Todaro P., Barbera G., Castrorao Barba A. & Bazan G. (2020) - Qanats and historical irrigated landscapes in Palermo's suburban area (Sicily).

Tranchina L., Basile S., Brai M., Caruso A., Cosentino C. & Miccichè S. (2008) - Distribution of heavy metals in marine sediments of Palermo Gulf (Sicily, Italy). Water Air Soil Pollut., 191, 245-256.

Tufano R., Guerriero L., Annibali Corona M., Bausilio G., Di Martire D., Nisio S. & Calcaterra D. (2022) - Anthropogenic sinkholes of the city of Naples, Italy: An update. Nat. Hazards, 112(3), 2577-2608.

Waltham T., Bell F.G. & Culshaw M.G. (2005) - Sinkholes and subsidence: karst and cavernous rocks in engineering and construction. Springer Science & Business Media.

Weary D.J. (2015) - The cost of karst subsidence and sinkhole collapse in the United States compared with other natural hazards.