Journal of Hydrology: Regional Studies 44 (2022) 101245

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Multiclass spatial predictions of borehole yield in southern Mali by means of machine learning classifiers

Gómez-Escalonilla V [a,*], O. Diancoumba [b], D.Y. Traoré [c], E. Montero [a], M. Martín-Loeches [d], P. Martínez-Santos [a]

[a] UNESCO/UNITWIN Chair Appropriate Technologies for Human Development. Department of Geodynamic, Stratigraphy and Paleontology, Faculty of Geology, Complutense University of Madrid, C/José Antonio Novais 12, 28040 Madrid, Spain
[b] Department of Life and Earth Sciences (DER-SVT), Normal Superior School of Bamako (ENSup), B.P. 241 Bamako, Mali
[c] Department of Geology, Faculty of Sciences and Techniques (FST), University of Sciences, Techniques and Technology of Bamako (USTTB), B.P.E 3206 Bamako, Mali
[d] Department of Geology, Geography and Environment. Geology UD, University of Alcalá, Alcalá de Henares, Madrid, Spain

ARTICLE INFO

Keywords:
Machine learning
Groundwater exploration
Yield prediction
GIS
Mali

ABSTRACT

Study region: Regions of Bamako, Kati and Kangaba, southwestern Mali
Study focus: Machine learning-based mapping of borehole yield. Three algorithms were trained on an imbalanced multiclass database of boreholes, while twenty variables were used as predictors for borehole yield. All models returned balanced and geometric scores in the order of 0.80, with area under the receiver operating characteristic curve up to 0.87. Three main methodological conclusions are drawn: (a) The evaluation of different machine learning classifiers and various resampling strategies and the subsequent selection of the best performing ones is shown to be a good strategy in this type of studies; (b) ad hoc calibration tools, such as data on borehole success rates, provide an apt complement to standard machine learning metrics; and (c) a multiclass approach with an unbalanced database represents a greater challenge than predicting a bivariate outcome, but potentially results in a finer depiction of field conditions.
New hydrological insights for the region: Alluvial sediments were found to be the most productive areas, while the Mandingue Plateau has the lowest groundwater potential. The piedmont areas showcase an intermediate groundwater prospect. Elevation, basement depth, slope and geology rank among the most important variables. Lower values of clay content, slopes and elevations, and higher values of basement depth and saturated thickness were linked to the most productive class.

1. Introduction

Groundwater plays a vital role in drinking water supplies, food security and ecosystem services. An estimated 2.5 billion people worldwide rely exclusively on groundwater to meet their daily needs, while hundreds of millions of farmers depend on groundwater to sustain their livelihoods (UNESCO, 2015). Adelana and MacDonald (2008) explain that the increasing prominence of groundwater as a drinking water source in Africa is motivated by three main reasons, namely, the large natural storage capacity of aquifers, the often

* Corresponding author.
E-mail address: vigome01@ucm.es (Gómez-Escalonilla).

https://doi.org/10.1016/j.ejrh.2022.101245
Received 6 June 2022; Received in revised form 5 October 2022; Accepted 15 October 2022
Available online 20 October 2022

good water quality, and that groundwater-based infrastructures are generally more affordable for disadvantaged communities. While critical for the survival of human beings and ecosystems, groundwater also suffers from the "hidden treasure" syndrome: because groundwater is out of sight, aquifers are seldom well known, even in industrialized countries. This leads to widespread contamination, problems of unregistered use, and dropping water tables worldwide, which all hamper attempts to manage the resource sustainably. Improving hydrogeological knowledge is thus essential, particularly given the current context of climate change in which groundwater resources are expected to be increasingly relied upon.

Groundwater potential mapping (GPM) can be used to underpin groundwater resources planning and exploration (Elbeih, 2015). Díaz-Alcaide and Martínez-Santos (2019) carried out a review of GPM approaches, analyzing the definitions of GPM, the variables that are typically involved in GPM studies and the data integration procedures available in the literature. These authors found that most of the existing studies rely on expert-decision approaches like analytical hierarchy processes (Muavhi et al., 2021), weight-of-evidence approaches (Ahmed et al., 2021), fuzzy analytical hierarchy processes (Boughariou et al., 2021), multi-criteria decision-making

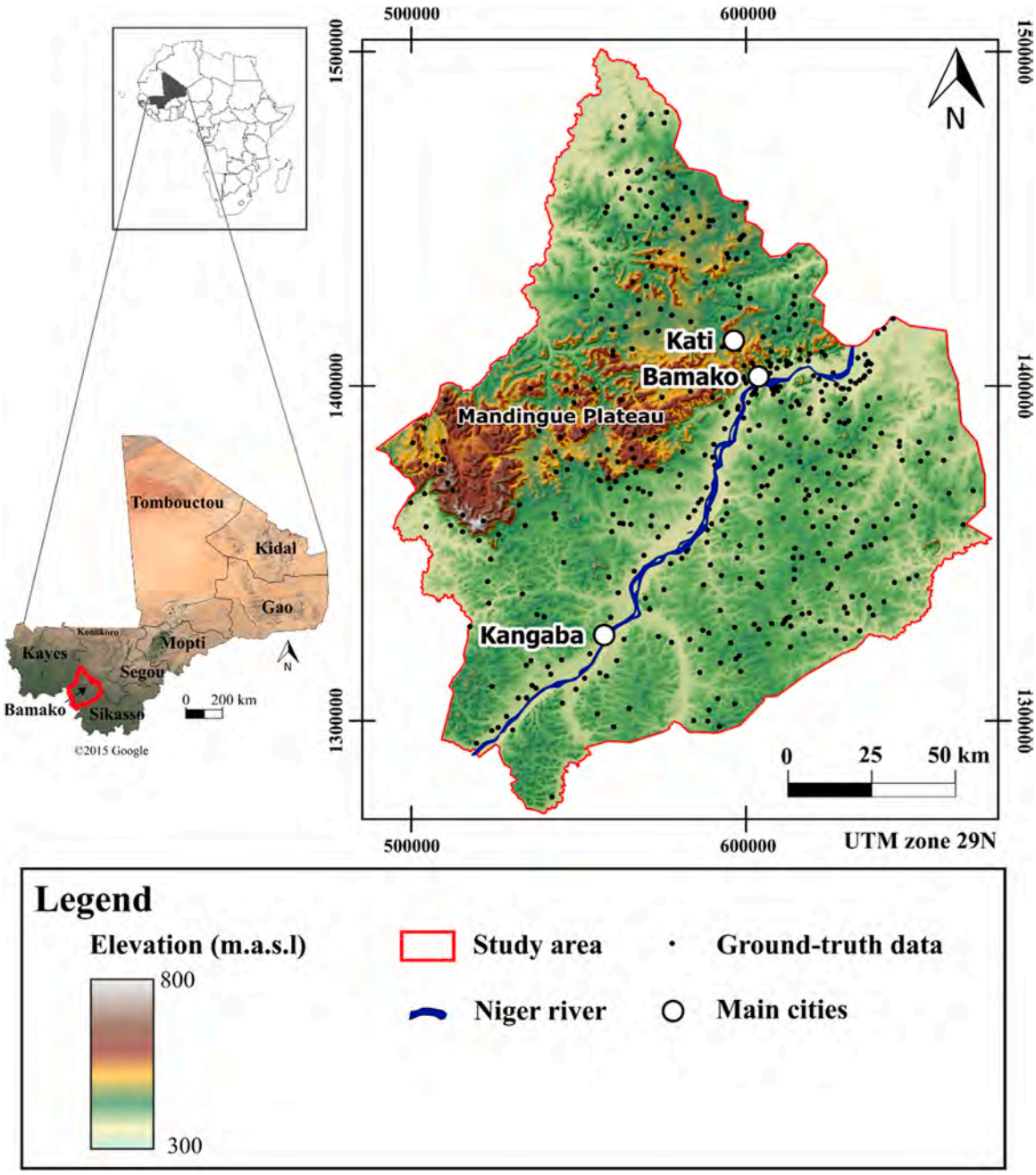

Fig. 1. Study area. Bamako, Kati and Kangaba regions located in southwestern Mali.

(Forootan and Seyedi, 2021), frequency ratio (Trabelsi et al., 2018) and Shannon's entropy methods (Khoshtinat et al., 2019). Gómez-Escalonilla et al. (2022) show that these expert-based techniques often require the reclassification of explanatory variables in intervals, which is likely to originate bias from the outset because the intervals are based almost exclusively on the criteria of the operator. The advent of machine learning approaches (ML) opens up a whole new methodological dimension to GPM, as ML approaches can overcome this disadvantage to a large extent. The recent literature showcases several examples of ML algorithms in groundwater potential mapping studies. These include support vector machines (Al-Fugara et al., 2022; Panahi et al., 2020), decision trees (Al-Abadi et al., 2021; Arabameri et al., 2021; Braham et al., 2022), artificial neural networks (Chen et al., 2021; Nguyen et al., 2020; Hakim et al., 2022), and ensemble methods like boosting, random forests and extra trees classification, among others (Bai et al., 2022; Choudhary et al., 2022; Gómez-Escalonilla et al., 2021, 2022; Martinsen et al., 2022; Sachdeva and Kumar, 2021).

ML-based predictions of groundwater potential are typically carried out within a geographic information database. The underlying logic is that the presence of groundwater, that is, the target variable, can be inferred from a series of explanatory variables. The latter typically include landforms, rainfall, lineaments, lithology, slope, weathering, and drainage density, among others (Díaz-Alcaide and Martínez-Santos, 2019). The ground-truth sample that is used to train the algorithms consists of an existing borehole database with information on positive and negative boreholes and, in some cases, characteristics such as flow rate or water table depth, among others. If the borehole database presents a sufficiently large and diverse number of records, machine learning classifiers can find those patterns of explanatory variables that lead to a positive or a negative borehole. The validated algorithm can then be extrapolated spatially to obtain a groundwater potential score for every pixel in the geographic database.

Most ML-GPM studies focus on predicting a positive or negative outcome, that is, in identifying zones of high or low groundwater prospect. In this context, our paper presents two major innovations. The first one contributes to clarify the manifold concept of "groundwater potential" (Díaz-Alcaide and Martínez-Santos, 2019) by providing groundwater potential predictions in terms of borehole yield. The second one consists of going one step beyond conventional predictive capabilities by replacing bivariate outcomes with multiclass outcomes. To the authors' knowledge, there is only one methodological precedent in this regard, in which Kumar et al. (2021) evaluated machine learning and Fuzzy AHP methods for mapping groundwater potential in data-scarce areas. This study used a data augmentation technique to increase the database available to train the algorithms. Because our database is roughly twice as large, augmentation is not needed, and thus, the present work represents a unique example in the literature based on an entirely independent set of field records. We evaluated the performance of four different resampling methods (Synthetic Minority Oversampling Technique

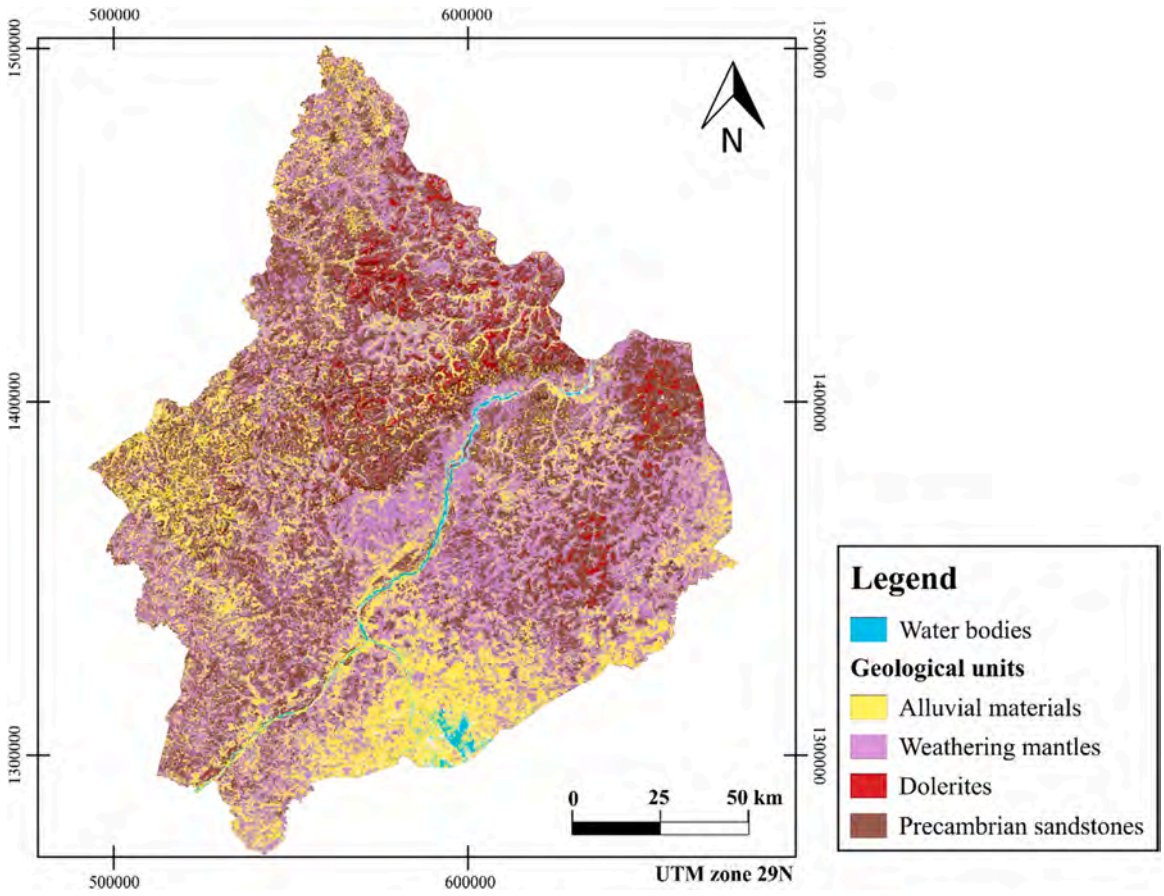

Fig. 2. Main geological units of the study area.

(SMOTE), Adaptive Synthetic Sampling approach (ADASYN), Random Oversampling (ROS) and Random Undersampling (RUS). Resampling was performed after the training/testing split and was only applied to the training dataset, which contributes to avoid the problems of data leakage and overfitting.

From a geographical standpoint, ML-GPM studies have been carried out mainly in Asia, while there are very few in the African continent (Braham et al., 2022; Gómez-Escalonilla et al., 2021, 2022; Martínez-Santos and Renard, 2020; Namous et al., 2021). Moreover, ours is one of the first studies in the literature addressing groundwater potential based on a multiclass approach. A crucial component in ensuing analyses is the need to approach the problem of predicting yield based on an unbalanced ground-truth dataset, where there is a much larger number of positive than negative points, since in many cases drilling companies do not report negative drillings to the competent authorities.

In this context, two main objectives have been established: (1) To develop a multiclass predictive borehole yield map of the study region; and (2) To evaluate different machine learning models and resampling techniques to address the imbalanced dataset approach in order to establish a strategy for potential future studies of the same nature.

2. Materials and methods

2.1. Study area

The study area spans 21,000 km^2, including the administrative region of Bamako and the municipalities of Kati and Kangaba of the Koulikoro region, southern Mali (Fig. 1). The maximum altitude in the study area is 800 m.a.s.l. in the Mandingue Plateau, while the minimum elevation is 300 m.a.s.l. in the northernmost end of the Niger River. The region is characterized by a tropical savanna climate (Traore et al., 2018), with average temperatures of 27 °C and an average rainfall around 1000 mm/year. The dry season lasts from October to May, and accounts for just 25% of the total rainfall, while the West African Monsoon months provide the rest. Diancoumba et al. (2020) estimated aquifer recharge rates to be around 3–26% of annual precipitation.

From a geological standpoint, the area consists of four geological units (Fig. 2), namely Precambrian materials (including sandstones), intrusive igneous rocks, recent alluvial sediments and weathering mantles developed over the Precambrian materials. The alluvial sediments take place mostly along River Niger and its tributaries. The map of the main geological units was obtained from

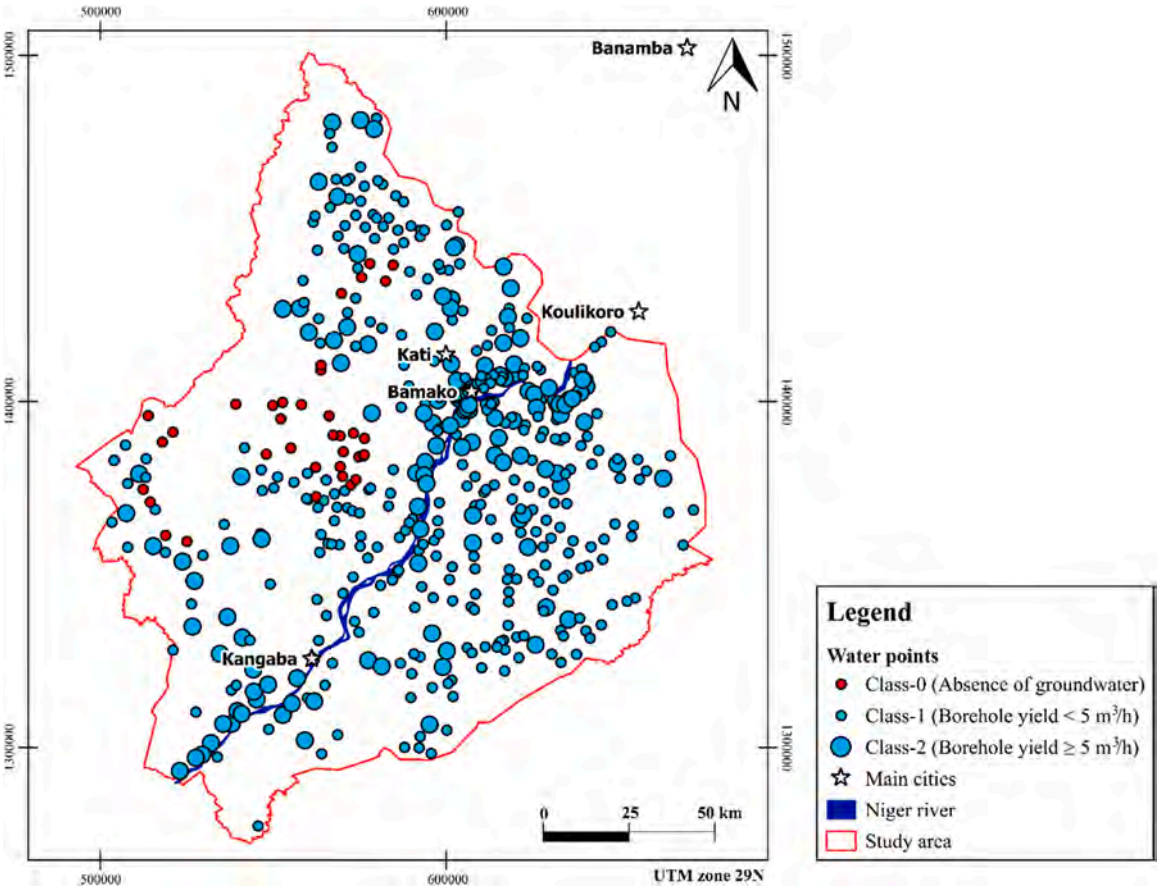

Fig. 3. Spatial distribution of the water points. Symbol size represents borehole yield.

Landsat 8 bands and was constructed with QGIS 3′s semiautomatic classification plugin (Congedo, 2021). Training points were obtained from the 1/200,000 geological map of Mali (DNGM and LGSJM, 1988).

Precambrian rocks exhibit fissured porosity and fractures are often interconnected. This provides relatively high permeability on a regional scale and favors precipitation infiltration and groundwater flow to lower elevations. Traore (1985) shows that the transmissivity of these units ranges around 130 m^2/day. This study also indicates that the thickness of fractured horizons varies between 30 and 50 m, although deeper fractures can increase the thickness of the aquifer up to 80–100 m.

Intrusive igneous rocks, more specifically dolerites, may behave either as a barrier or as preferential pathways for groundwater flow due to fracturing associated with the intrusive process. Under certain conditions, these units may constitute aquifers of local importance, however, the average transmissivity of the system fluctuates around 4 m^2/day (Traore, 1985).

Weathered mantles extend across most of the study region (PNUD, 1982). These materials present an average transmissivity of 7 m^2/day with a maximum of 350 m^2/day and a minimum of < 0.1 m^2/day, depending on the lithology (Traore et al., 2018). These authors indicate that the thickness of the productive zone varies from 10 to 50 m and the depth of the static water levels ranges between 8 and 20 m below the surface. These aquifers are drained by the Niger River system.

Recent alluvial sediments of the Niger river show relatively good hydrogeology properties, with an estimated transmissivity that varies from 0.34 to 21.6 m^2/d (Alpha et al., 1991). Thickness is highly variable but generally ranges between 2 and 20 m (MIHE, 1990). Static levels typically vary between 2 and 8 m below the surface.

2.2. Target variable

The database provided by the National Hydraulic Direction of Mali (DNH, 2010) contains 483 points with borehole yield data. These are widely distributed in space (Fig. 3). The dataset is however imbalanced in terms of yield levels: 36 points (7.4%) were classified as negative, 323 points (66.9%) present borehole yields of between 0.5 and 5 m^3/h, and 124 points (25.4%) exceed 5 m^3/h. In the experience of the authors, this is common in regional borehole databases (Gómez-Escalonilla et al., 2021), and is likely attributable to the fact that negative boreholes are reported to the authorities less frequently than positive boreholes. Because data imbalances cause difficulties to the predictive potential of the algorithms, a series of trial runs were carried out to determine the optimal threshold between borehole yield classes. These will be described later on.

Joint interpretation of Figs. 1 and 3 suggests that most of the high yield boreholes are located along the alluvial materials of River Niger, whereas the Mandingue Plateau accounts for most of the negative ones. The remainder of the study area presents a combination of low and high yield boreholes.

2.3. Conditioning factors

Díaz-Alcaide and Martínez-Santos (2019) show several variables that are commonly used in GPM studies. These include lithology (geology), geomorphology, soil, land use/land cover, topography, lineaments, drainage- and slope-related variables, rainfall and groundwater recharge. In turn, Kumar (1997) and Jyrkama et al. (2002) contend that crucial elements of groundwater potential, such as aquifer recharge, are influenced by several factors, namely, climatic conditions, soil types and characteristics, land cover, geomorphological factors, and hydrological features. Besides, the available data allows for the development of other spatially-distributed layers of information, including the expected saturated thickness of the basement depth. Thus, in this case we took into consideration twenty explanatory variables, or “conditioning factors”, as partial predictors of borehole yield (Table 1).

Table 1
Conditioning factors used to predict the borehole yield potential. Scale/resolution was derived from the original resolution of the data source.

Conditioning factor	Scale/Resolution	Source of data
Alteration Band Ratio	30 m	Elaborated from Landsat 8
Clay content	250 m	SoilGrids250m 2.0
Curvature	30.53 m	Elaborated from SRTM DEM
Saturated thickness	30.53 m	Elaborated from SRTM DEM and borehole database
Water table depth	30 m	Elaborated from Borehole database
Distance from channels	30.53 m	Elaborated from SRTM DEM
Geology	1:5 million	British Geological Survey
Geomorphology	30.53 m	Elaborated from SRTM DEM
Land use	300 m	ESA Climate Change Initiative
Soil	1:3 million	European Soil Data Centre
Rainfall	0.5°	CRU TS 3.21 dataset (Climatic Research Unit at the University of East Anglia)
Drainage density	30.53 m	Elaborated from SRTM DEM
Elevation (DEM)	30.53 m	Shuttle Radar Topography Mission (SRTM)
NDVI	30 m	Elaborated from Landsat 8
NDWI	30 m	Elaborated from Landsat 8
Slope	30.53 m	Elaborated from SRTM DEM
SPI	30.53 m	Elaborated from SRTM DEM
TWI	30.53 m	Elaborated from SRTM DEM
Surface lithology	30 m	Elaborated from Landsat 8
Basement Depth	30 m	Elaborated from Borehole database

The borehole database contains information about borehole depth, as well as static level measurements. Boreholes throughout the study area are frequently drilled to the impervious basement. Consequently, borehole depth may be interpreted as a proxy for the thickness of the overlying materials. Water table depth (Fig. 4**A)** and expected saturated thickness layer (Fig. 4**B)** may also be inferred from this information.

Soil types of the study area were obtained from the European Soil Data Centre (Dewitte et al., 2013) (Fig. 4C). Petric Plinthosols take up almost 50% of the study area, followed the Haplic Lixisols and Lithic Leptosols, which cover about 24% and 18%, respectively. Petric Plinthosols are defined as a type of soil arranged in continuous or fractured sheets rich in Fe or formed by concretions or nodules also rich in Fe connected and strongly cemented. Haplic Lixisols exhibit a higher clay content in the deeper part than in the shallower one, due to pedogenetic processes. Lithic Leptosols are very fine soils developed on continuous rock with an extremely rich content of coarse fragments. The parent rock appears less than 10 cm from the soil surface. (IUSS Working Group WRB, 2015).

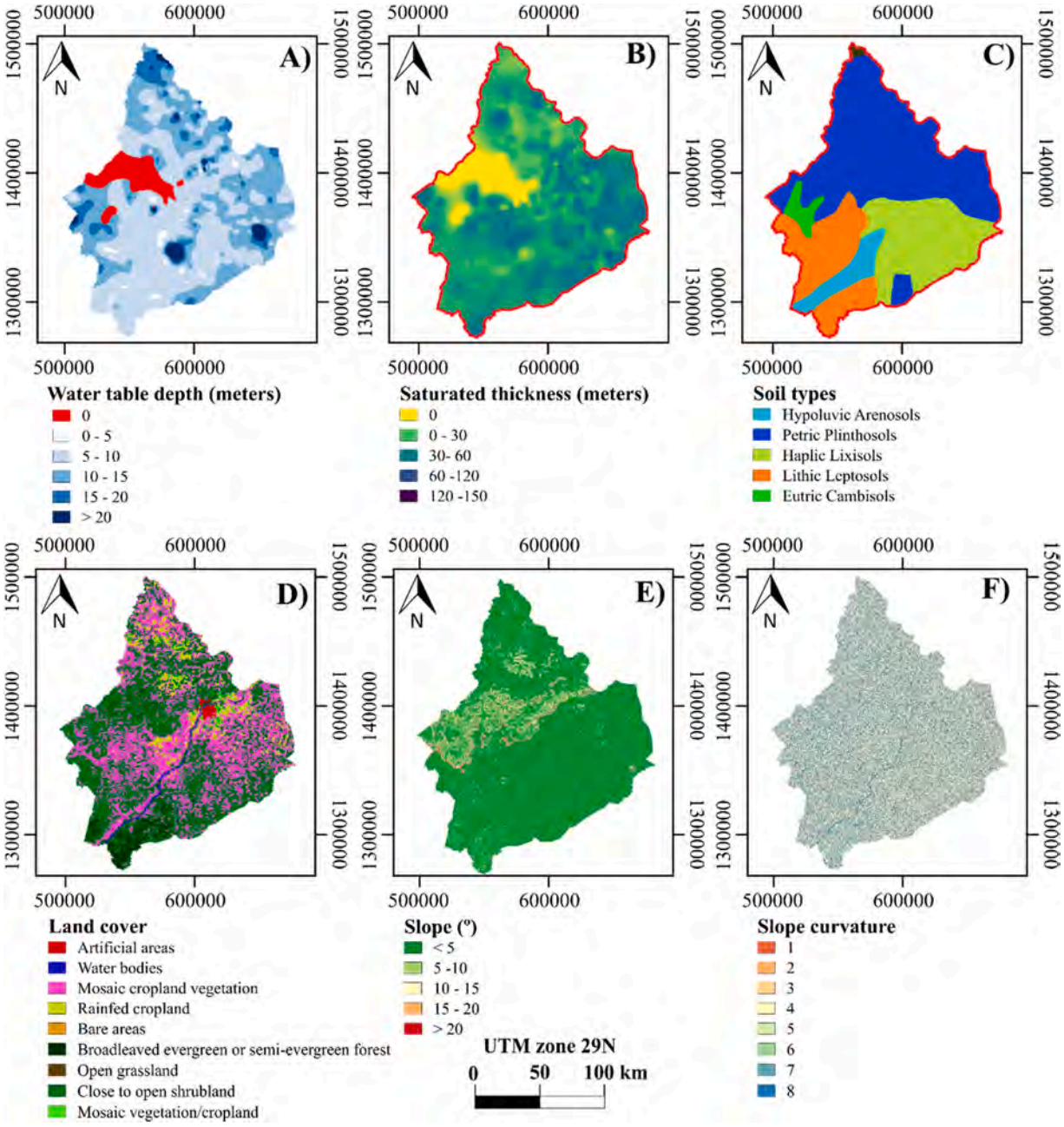

Fig. 4. Conditioning factors used in predicting borehole yield potential. (A) Water table depth; (B) Saturated thickness; (C) Soil types; (D) Land cover; (E) Slope; and (F) Slope curvature.

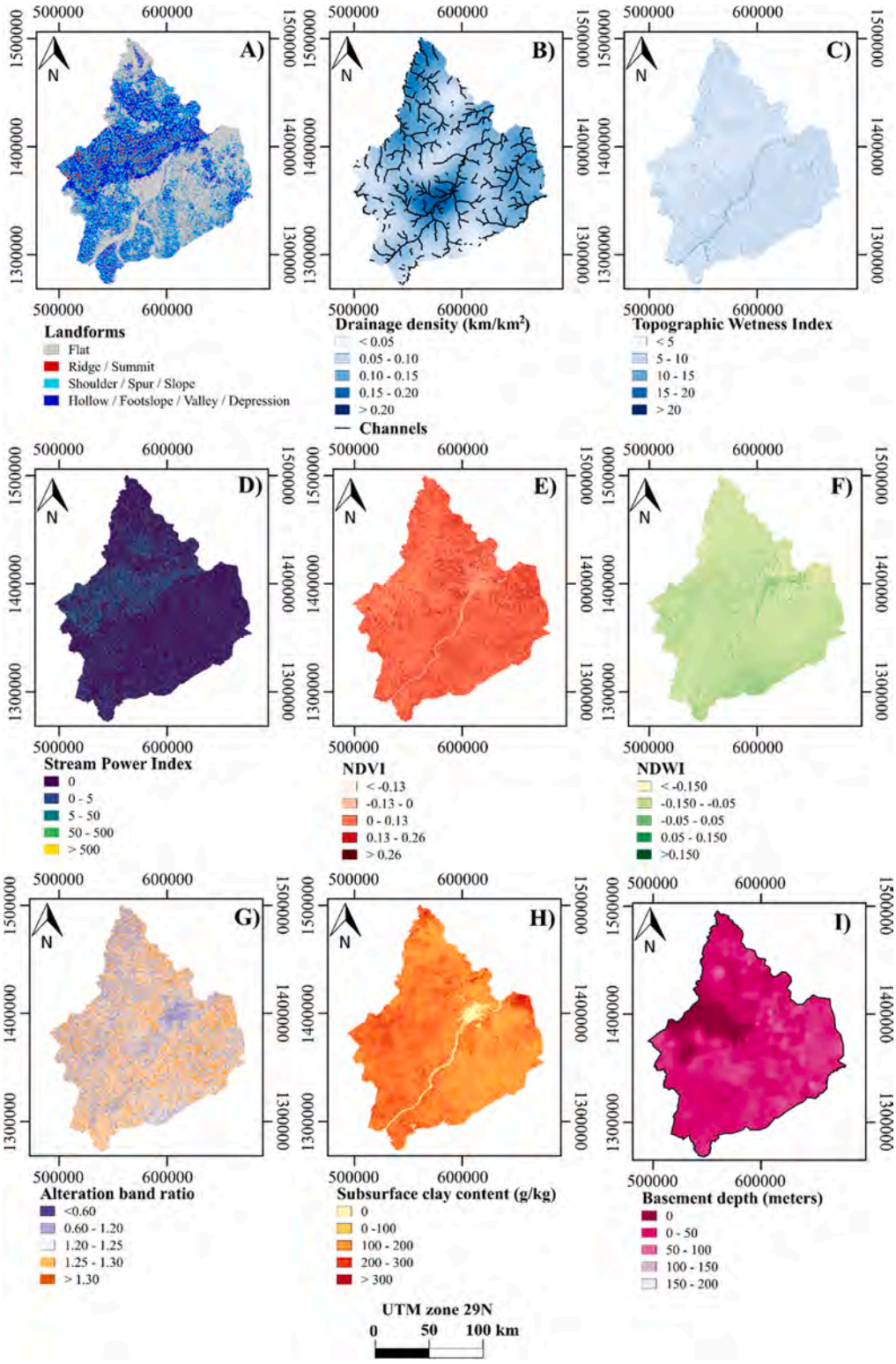

Fig. 5. Conditioning factors used in predicting borehole yield potential. (A) Landforms; (B) Drainage density; (C) Topographic Wetness Index; (D) Stream Power Index; (E) NDVI; (F) NDWI; (G) Alteration band ratio; (H) Subsurface clay content; (I) Basement depth.

Land cover information stems from the European Space Agency Climate Change Initiative (ESA, 2010). Close and open shrublands account for more than 52% of the study area, while mosaic vegetation represents about 36% (Fig. 4D). A combination of rainfed croplands and broadleaved evergreen or semi-evergreen forest add up to the remaining 9%.

The Digital Elevation Model (DEM) was obtained from the radar-based Shuttle Radar Topography Mission (available at https://earthexplorer.usgs.gov/). The DEM was used to produce several layers, including slope (Fig. 4E) and slope curvature (Fig. 4F), while landforms (Fig. 5A) were derived with QGIS 3.0′s Geomorphon plugin (Jasiewicz and Stepinski, 2013). The surface water channels extracted from the DEM analysis were used to develop the drainage density (Fig. 5B) and distance from channels layers. Topographic Wetness Index (TWI) and Stream Power Index (SPI) maps were also derived from the DEM. TWI (Fig. 5C) represents the facility for water to accumulate on the surface (Beven and Kirkby, 1979). In turn, SPI (Fig. 5D) provides a proxy of the erosive potential of runoff water (Moore et al., 1991).

Satellite imagery can provide valuable information, despite not penetrating deep into the ground, in those areas associated with

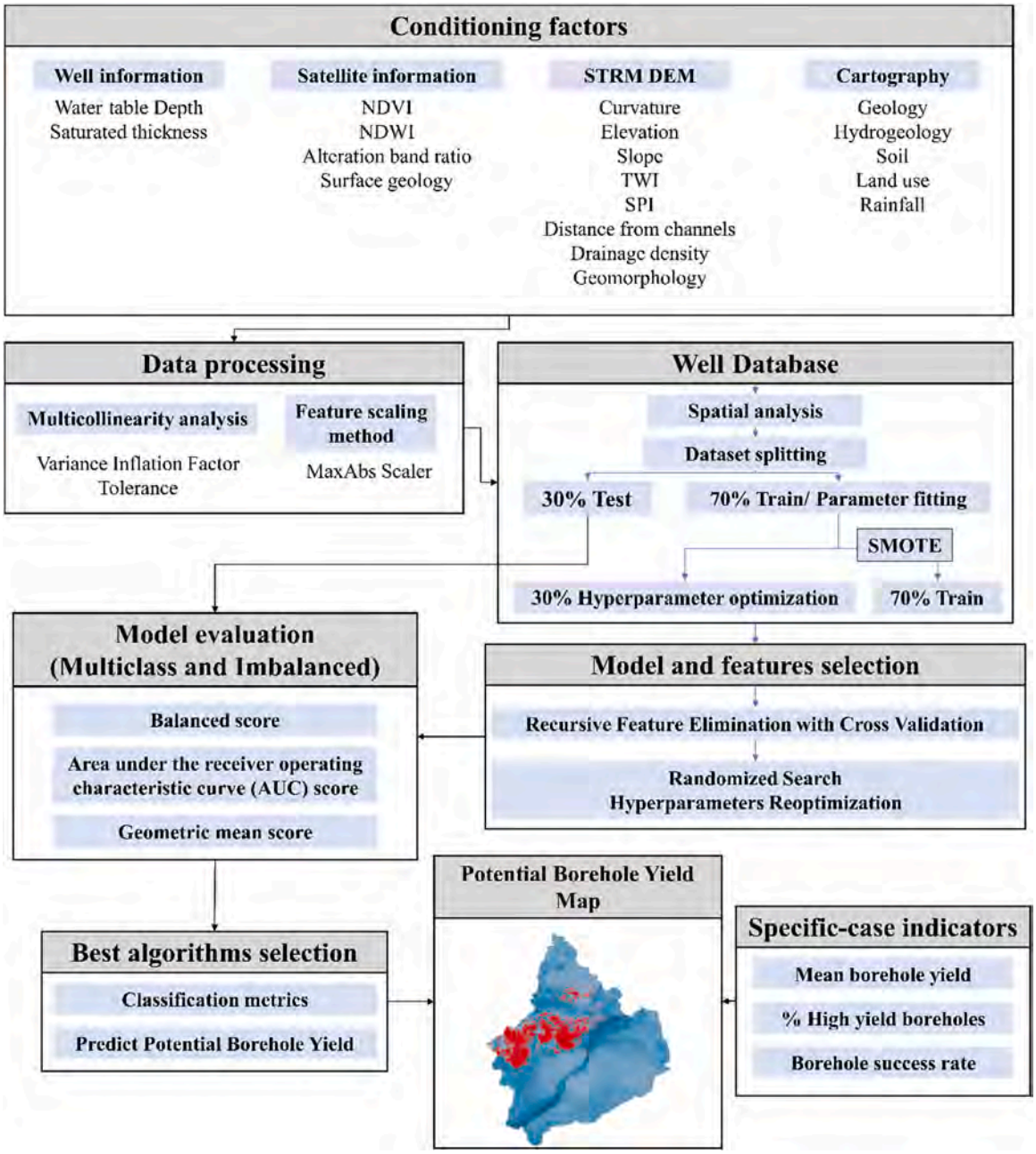

Fig. 6. MLMapper's methodological flowchart.

shallow groundwater levels (Díaz-Alcaide and Martínez-Santos, 2019). In this research, Landsat 8 bands were used to elaborate four different layers. These include a surface lithological map constructed by means Semi-Automatic Classification Plugin (Congedo, 2021) (Fig. 2); a Normalized Difference Vegetation Index (NDVI) layer (Fig. 5E), which is an estimation of the vigor of vegetation and is calculated from vegetation response to red and visible infrared wavelengths (Xie et al., 2008); the Normalized Difference Water Index (NDWI), a measure of the quantity of water in the plants or soil humidity (Xu, 2006) Fig. 5F; and the alteration band ratio or surface clay content (Fig. 5G). The latter was developed by combining Landsat 8 bands 6 (shortwave infrared 1) and 7 (shortwave infrared 2) (Ourhzif et al., 2019). This layer provides information on the clay content in the surface which, can be expected to control infiltration potential to some extent. To complement the surface clay information, a clay content layer (g/kg) in the first two meters of the subsoil was also used (Fig. 5H). Water infiltration into the underlying materials is not only controlled by the surface clay content, the first few meters of the subsoil may allow (less clay content) or impede (more clay content) infiltration down to the water table. Therefore, this layer provides additional information to that obtained with Landsat images. This conditioning factor obtained from SoilGrids 250 m 2.0 was elaborated using state-of-the-art machine learning methods by Poggio and de Sousa (2020).

2.4. Machine learning models

ML-based spatial predictions need three main elements for being conducted: (1) a spatially distributed set of explanatory variables (described in the previous section - Section 2.3 Conditioning Factors); (2) a database of points where the outcome is known (Section 2.2 Target Variable); and finally, (3) a machine learning model or models capable of learning the patterns of explanatory variables that lead to a given outcome.

In reference to the last element, ML models, the MLMapper 2.0 code (Gómez-Escalonilla et al., 2021, 2022) has been used in this research to develop spatial predictions of borehole yield. The code is an evolution of the original MLMapper tool developed by Martinez-Santos and Renard (2020). This new version, developed in Python 3.7, routinely implements multicollinearity checks, random search hyperparameter optimization, cross-validation, and recursive feature elimination (Fig. 6). In addition, it allows users to choose the target score metric to be optimized in the aforementioned processes. Finally, it provides an agreement map of the predictions of the best-performing algorithms to estimate uncertainty.

The complexity of the architecture and operation of machine learning algorithms is combined with the associations, also complex, that may exist between the explanatory variables leading to a given outcome. All this leads to the unfeasibility to forecast, at the starting point, which of the machine learning models will perform better than the others on a particular dataset. To tackle this challenge, a series of runs is performed with all the algorithms included in MLMapper. MLMapper incorporates 19 supervised classification algorithms from the SciKit-Learn toolbox (Pedregosa et al., 2011). Subsequently, the best performing algorithms based on user-defined performance thresholds are selected and the rest are discarded. The results and performance of the models were appraised on the basis of standard machine learning metrics, namely, the geometric mean, the area under the receiver operating characteristic curve (AUC score) and the balanced accuracy score (Ling and Li, 1998; Sun et al., 2006; Pedregosa et al., 2011). MLMapper works with different types of Support Vector Machines and statistical learners such as Linear Discriminant Analysis, Gaussian Naïve Bayes classification, Ridge Classifier or Logistic Regression. It also includes ensemble methods (Random Forest Classifier, Gradient Boosting Classifier, Ada-Boost Classifier, and Extra Trees Classifier), instance learners (K-Neighbor Classification), neural networks (Multilayer Perceptron Neural Network) and a simple Decision Tree Classifier, among others.

Following the methodology described above, a series of tests were performed to identify the algorithms that consistently outperformed the others on this particular dataset. This led to choosing Extra Trees (ETC), Logistic Regression (LRG) and the Gradient Boosting Classifier (GBC) as the most suitable. More information on discarded algorithms can be found in previous studies (Gómez-Escalonilla et al., 2021, 2022; Martínez-Santos and Renard, 2020; Pedregosa et al., 2011).

ETC uses an ensemble application of the simple decision tree algorithm. It builds an ensemble of unpruned decision trees according to a classical top-down process (Geurts et al., 2006). ETC uses the entire training sample to grow the trees as opposed to the widely used Random Forest which uses the bootstrapping technique. Furthermore, during the generation of a tree, ETC splits the nodes by choosing the cut points completely randomly.

GBC is based on gradient boosting technique that initially uses a decision tree classifier as a weak model. GBC builds an additive model in a forward stage-wise fashion and allows the optimization of arbitrary differentiable loss functions (Friedman, 2001; Pedregosa et al., 2011). The basic function of this method is to predict a new classification membership after each iteration in an additive way. Predictions are created from weak learners that continuously evolve on the errors of the previous learners. Misclassified samples receive higher weights in the next step, which forces the classifier to focus on their performance in the following iterations (Georganos et al., 2018).

LRG is a widely used linear model. The probabilities of each potential outcome are modeled by means of a logistic function. The multi-class setting is similar to the binary case, except the label y is now an integer in $\{1,\ldots,C\}$ where C is the number of classes.

2.5. Supervised classification procedure: an imbalanced multiclass dataset approach

Before applying all the models explained above, a multicollinearity analysis of the conditioning factors was performed. Multicollinearity arises when some of the conditioning factors used for predictions are highly correlated with each other. The presence of multicollinearity among the conditioning factors can be a potential problem and affect the performance of the algorithms. Issues can arise from attributing extra weight to an input factor or from incorporating noise into the final results. Multicollinearity among all conditioning factors was identified using the variance inflation factor and tolerance indices, as commonly used to estimate

multicollinearity of predictors in geospatial modeling (Bui et al., 2016).

The 70% of the water point database was used to train the models and the remaining 30% was employed to validate the predictive ability of the classifiers. Machine learning studies are sometimes performed on ground-truth datasets where one of the outcome classes is under-represented in relation to the others (Lemaître et al., 2017). This may lead to the problem of class imbalance (Prati et al., 2009), which essentially means that the potential to predict the under-represented class is limited. In such circumstances, performance metrics will present high accuracy in the majority classes and poor performance in the minority classes (Singh and Purohit, 2015). This will typically result in discarding some valuable information from the latter.

As mentioned earlier, dataset imbalance was perceived as a potential issue from the outset. Hence, a strategy based on comparing SkLearn's imbalanced-learn tools was developed. SkLearn provides two main different strategies to tackle imbalance (Lemaître et al., 2017), namely, undersampling and oversampling. Based on an imbalanced dataset in which *X min* and *X maj* are the subset of samples belonging to the minority and majority class, respectively, undersampling refers to the process of reducing the number of samples of the *X maj*. On the contrary, oversampling performs the data balancing by generating new samples in *X min*.

A series of tests were performed to identify the strategy to address the unbalanced dataset problem that performed best. These strategies include Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic Sampling approach (ADASYN), Random Oversampling (ROS) and Random Undersampling (RUS). Importantly, these oversampling or undersampling techniques were applied only on the training dataset points. Otherwise, the use of these resampling strategies on the entire data set could lead to overoptimism or data leakage problems (Lemaître et al., 2017; Santos et al., 2018).

Chawla et al. (2002) developed SMOTE as an oversampling approach in which the minority class is over-sampled by creating "synthetic" instances rather than oversampling with replacement. In this technique, synthetic examples are generated by operating on the "feature space". The minority class is oversampled by taking each sample of the minority class and introducing synthetic examples along the line segments linking any/all nearest neighbors of the minority class. ADASYN is an extension of SMOTE. ADASYN can decide the number of synthetic instances that need to be produced for the minority class. He et al. (2008) note that this technique not only provides a balanced data distribution, but also forces the learning algorithm to focus on complex samples in the dataset. Finally, ROS increases the number of minority class points in the training set by randomly replicating existing minority class instances, while RUS decreases the number of majority class data points by randomly eliminating majority class data points currently in the training set (Chawla et al., 2002). Liu (2004) reflects that one of the main problems with RUS is that it is not possible to control what information about the majority class is removed, so very important information about the decision boundary between the minority and majority class may be eliminated.

3. Results

3.1. Multicollinearity analysis

Table 2 presents the results of the multicollinearity analysis. Variance Inflation Factor (VIF) and Tolerance (TOL) values were calculated for all conditioning factors. VIF values range from 1 to 6.2, while TOL fluctuates between 0.16 and 0.93. This suggests that there are no multicollinearity issues among the explanatory variables (Dormann et al., 2013).

Table 2
Estimated VIF for each groundwater conditioning factor. VIFs greater than 10 or tolerance less than 0.1 indicate multicollinearity.

Conditioning factors	VIF	Tolerance
Alteration Band Ratio	1.310	0.763
Clay content	1.690	0.592
Water table Depth	1.080	0.926
Distance from channels	1.176	0.850
Geology	1.498	0.668
Geomorphology	1.284	0.779
Land use	1.206	0.829
Soil	1.506	0.664
Rainfall	1.924	0.520
SPI	1.151	0.869
Slope	1.278	0.782
TWI	1.391	0.719
Curvature	1.183	0.845
Drainage density	1.490	0.671
Saturated thickness	5.652	0.177
Elevation	2.144	0.466
NDVI	1.536	0.651
NDWI	1.384	0.722
Surface geology	1.203	0.831
Basement depth	6.212	0.161

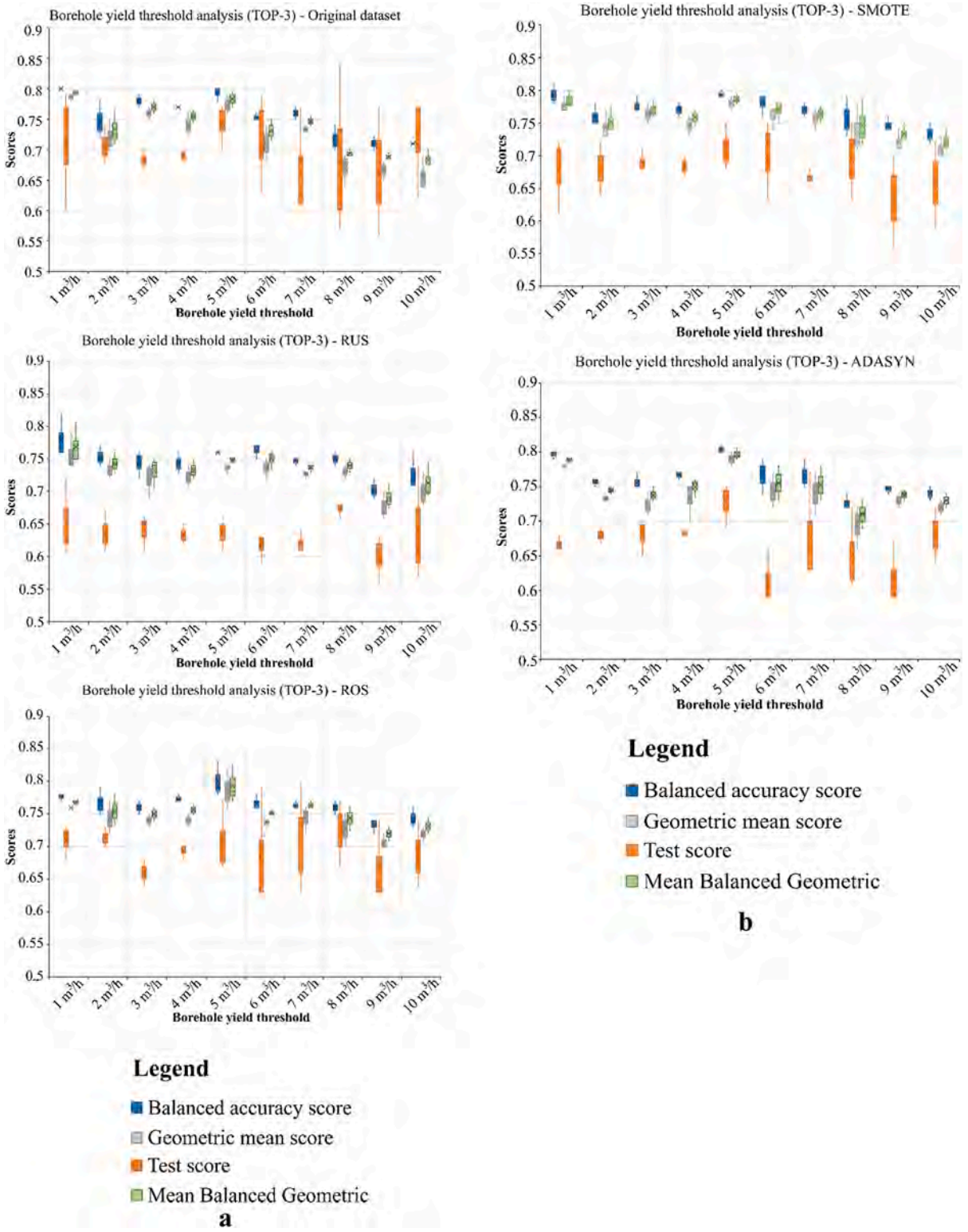

Fig. 7. **a.** Borehole yield threshold analysis outcomes of original dataset, RUS and ROS resampling methods. **b.** Borehole yield threshold analysis outcomes of SMOTE and ADASYN resampling methods.

3.2. Borehole yield threshold selection

Fig. 7a and Fig. 7b shows the results of the borehole yield threshold analysis. MLMapper was run once for each unit threshold (from 1 m^3/h to 10 m^3/h) for the original dataset without resampling and for each of the different resampling strategies (SMOTE, ADASYN, ROS and RUS). The optimal thresholds in terms of the metrics of choice, were 1 m^3/h and 5 m^3/h when both the original dataset without resampling and SMOTE, ADASYN and ROS resampling techniques were used. The most discriminating thresholds when the RUS technique was used were 1 m^3/h and 6 m^3/h.

The borehole database was subsequently split in three classes. Class-0 refers to negative boreholes, while the 5 m^3/h threshold was used to separate class-1 (less than 5 m^3/h) and class-2 (over 5 m^3/h). The 5 m^3/h threshold was picked over the 1 m^3/h for three main reasons: (1) the data would not allow for a multiclass representation if the 1 m^3/h were picked; (2) the dataset imbalance would be accentuated if the 1 m^3/h threshold were used, with 13.7% and 78.9% corresponding to class-1 and class-2, respectively, instead of 66.9% and 25.7%; and (3) the 5 m^3/h threshold conveys more valuable information for prospective uses with greater water needs.

Table 3 shows the results of the three best performing algorithms for each scenario. In addition, the means of each of the four metrics were calculated to analyze the most accurate resampling or no resampling technique. The least effective resampling technique in terms of scores was RUS, with a mean for the balanced accuracy score (BAS), test score (TS), geometric mean score (GMS) and AUC score of 0.760, 0.636, 0.736 and 0.813, respectively. The no resampling scenario obtains the best results in terms of TS (0.743), but the scores on the different ML metrics focused on imbalanced datasets were worse than those obtained with two of the resampling techniques (ROS and ADASYN). The ADASYN resampling technique obtained the best results in three of the four metrics (with the exception of TS, where it obtained the second best result). The mean of the metrics obtained with the three best performing algorithms were 0.803, 0.790 and 0.856 for BAS, MGB and RAS, respectively. This prior analysis allows the selection of ADASYN as the most efficient approach in this case. Therefore, all ensuing analyses, including the development of the borehole yield prediction maps, refer to the results obtained with this technique.

3.3. Validation of the ML models

Table 4 shows the results of the three best performing algorithms with the ADASYN resampling technique. The results of the discarded models can be shown in the Table S1 of the Supporting Information. All models returned a balanced score greater than 0.80, a geometric mean score above 0.78 and an AUC score exceeding 0.85. This implies that the predictions are sufficiently accurate for all classifiers, and thus, that the problems of data imbalance were successfully tackled. The Extra Trees classifier obtained the best results for two of the imbalance metrics, with 0.81, 0.80 and 0.85 for the balanced score, geometric mean score and AUC score, respectively. Logistic Regression and Gradient Boosting obtained similar results for the three metrics used.

Fig. 8 shows the confusion matrix for all three algorithms. F1-score for Class-0 were 0.90, 0.95 and 0.95 for ETC, GBC and LRG, respectively. This implies that all algorithms predict the class labelled as “groundwater absence” highly accurately. Class 1 (borehole yield < 5 m^3/h) rendered F1-scores of 0.79, 0.81 and 0.72 for ETC, GBC and LRG, respectively; while the F1-score for Class-2 (borehole yield ≥ 5 m^3/h) were 0.61, 0.58 and 0.59 for ETC, GBC and LRG, respectively. This all suggests that the algorithms generally found it more difficult to distinguish between Class-1 and Class-2.

3.4. Conditioning factors selection and importance

The recursive feature elimination procedure identified a different optimal number of explanatory variables depending on the

Table 3
Results of the borehole yield threshold analysis and the evaluation of the different resampling strategies. BAS = Balanced accuracy score; Mean BAS = Mean Balanced accuracy score of the technique; TS = Test score; Mean TS = Mean Test score of the technique; GMS = Geometric mean score; Mean GMS = Mean Geometric mean score of the technique; MBG = Mean of the Balanced accuracy score and the Geometric mean score of each algorithm; Mean MBG = Mean of MBG of the technique; AUC = AUC score; Mean AUC = Mean of AUC of the technique.

Technique	Algorithm	BAS	Mean BAS	TS	Mean TS	GMS	Mean GMS	MGB	Mean MBG	AUC	Mean AUC
No resampling	LRG	0.80	0.793	0.70	0.743	0.79	0.773	0.795	0.783	0.78	0.785
	GBC	0.80		0.77		0.77		0.785		0.79	
	NBA	0.78		0.76		0.76		0.770		"-"	
SMOTE	RFC	0.80	0.793	0.75	0.710	0.79	0.780	0.795	0.786	0.86	0.845
	SGD	0.79		0.68		0.78		0.785		"-"	
	ETC	0.79		0.70		0.77		0.780		0.83	
ROS	LRG	0.79	0.800	0.68	0.706	0.78	0.786	0.785	0.793	0.84	0.850
	LDA	0.78		0.67		0.76		0.770		0.84	
	ETC	0.83		0.77		0.82		0.825		0.87	
ADASYN	LRG	0.80	0.803	0.69	0.726	0.79	0.790	0.795	0.796	0.85	0.856
	GBC	0.80		0.75		0.78		0.790		0.87	
	ETC	0.81		0.74		0.8		0.805		0.85	
RUS	RFC	0.76	0.760	0.64	0.636	0.74	0.736	0.750	0.748	0.8	0.813
	LDA	0.76		0.66		0.74		0.750		0.82	
	KNN	0.76		0.61		0.73		0.745		0.82	

Table 4
Results of the top three algorithms with the ADASYN resampling technique.

Algorithm	Balanced score	Geometric mean score	ROC AUC score
Logistic Regression	0.80	0.79	0.85
Gradient Boosting	0.80	0.78	0.87
Extra Trees	0.81	0.80	0.85

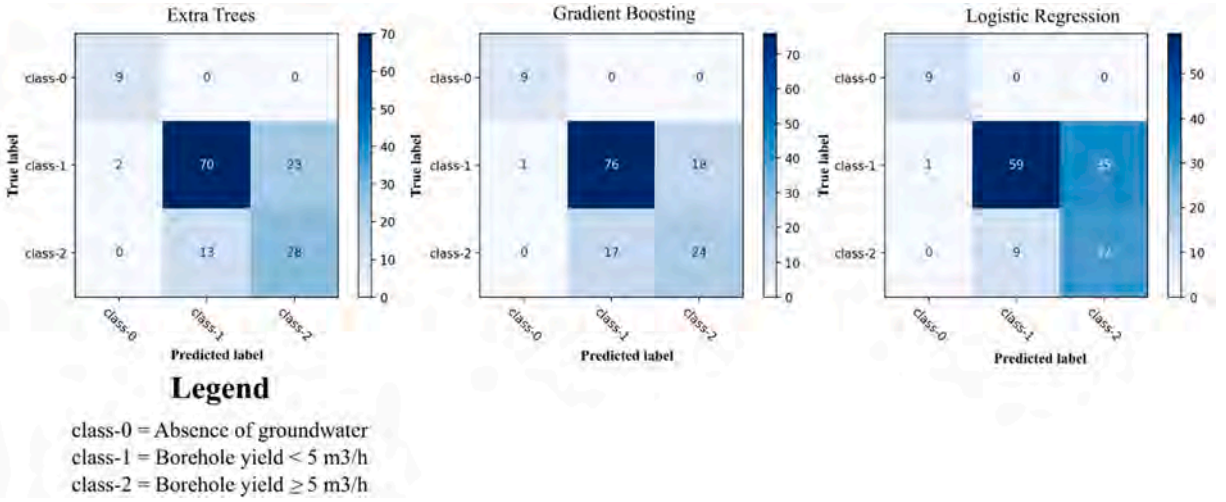

Fig. 8. Confusion matrix of the top three algorithms.

algorithm of choice. Table 5 shows the conditioning factors used by each of the three classifiers. Gradient Boosting and Logistic Regression relied on several conditioning factors with 18 and 19, respectively, while Extra Trees used the fewest (12 conditioning factors).

Clay content, geology, geomorphology, land use, rainfall, slope, curvature, drainage density, saturated thickness, elevation and basement depth were used by all three algorithms. Other explanatory variables that were used by at least two classifiers were: alteration band ratio, water table depth, distance from channels, soil, SPI, TWI, NDVI and NDWI.

Fig. 9 shows the results of the feature importance calculated for the best performing tree-based algorithms. ETC and GBC both agreed in attributing importance to elevation (0.30 and 0.25, respectively). ETC also attributed a high degree of importance to saturated thickness and land cover. This algorithm selected slope, basement depth, geology, drainage density and clay content as the next most important variables. In turn, GBC relied on basement depth, saturated thickness, land cover, clay content, slope, geology and rain among the most important factors. Both algorithms agreed in assigning the least importance to slope curvature.

Partial dependence plots (PDP) for some of the most important conditioning factors used by Extra Trees were generated in order to analyze the relationship between the values of the explanatory variables and the target variable (Fig. 10). Clay content PDP shows that higher values of clay content were associated with larger partial dependencies for class-0 and, to a lesser extent, for class-1. Class-2 exhibits mainly higher partial dependence at lower values of clay content, however, it shows a slight increase in partial dependence at higher values of clay content. Slope factor PDP shows that lower values of slope were associated with higher class-2 partial dependencies and larger values of slope were linked to higher class-0 partial dependencies. Class-1 exhibits a high variability for this variable.

Elevation PDP shows that greater values of elevation were associated with class-0. In contrast, class-1 and class-2 present greater

Table 5
Conditioning factors used by each algorithm.

Algorithm	Number of conditioning factors	Conditioning factors
Extra Trees	12	Clay content, Geology, Geomorphology, Land use, Soil, Rainfall, Slope, Curvature, Drainage density, Saturated thickness, Elevation and Basement depth
Logistic Regression	19	Alteration Band Ratio, Clay content, Water table Depth, Distance from channels, Geology, Geomorphology, Land use, Soil, Rainfall, SPI, Slope, TWI, Curvature, Drainage density, Saturated thickness, Elevation, NDVI, NDWI and Basement depth
Gradient Boosting	18	Alteration Band Ratio, Clay content, Water table Depth, Distance from channels, Geology, Geomorphology, Land use, Rainfall, SPI, Slope, TWI, Curvature, Drainage density, Saturated thickness, Elevation, NDVI, NDWI and Basement depth

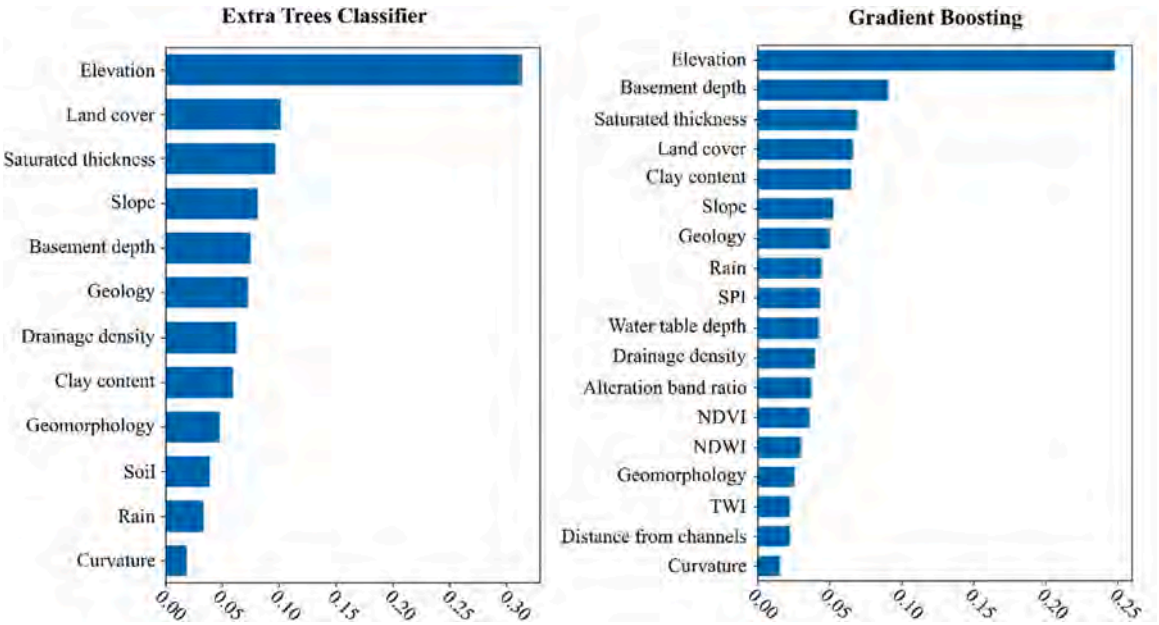

Fig. 9. Feature importance calculated for the best performing tree-based algorithms: Extra Trees Classifier (left) and Gradient Boosting Classifier (right). The sum of all variable weights equals one.

values of partial dependence for lower altitudes. This is consistent with the fact that most negative boreholes in the database are located on the heights of the Mandingue Plateau, while the productive ones tend to be located on the plain. The PDP of the basement depth shows that the lowest values of this conditioning factor are mainly associated with the highest class-0 partial dependencies. The partial dependence of class-1 presents two peaks with low-medium and high values of the basement depth factor. Class-2, on the other hand, presents a high partial dependence on the intermediate values. The saturated thickness PDP also renders intuitive results when appraised from a hydrogeological perspective. Lower saturated thickness values are associated with a higher partial dependence on class-0. On the other hand, although there is no linear relation, the trend presents higher values of this conditioning factor associated with higher partial dependencies for class-1 and class-2. The high partial dependence of the lower values of saturated thickness could be associated with the areas near the Mandingue Plateau where there are no high saturated thicknesses but percolation from the mountains could increase the groundwater potential.

3.5. Borehole yield potential mapping

Fig. 11 shows the predictive borehole yield maps obtained with the three algorithms. Green areas are those where the algorithms have found a combination of explanatory variables leading to class-2 (borehole yield ≥ 5 m^3/h), and yellow zones are those in which the algorithms have found a combination of conditioning factors that lead to class-1 (borehole yield < 5 m^3/h). Red pixels represent those areas where the algorithms found a combination of explanatory variables leading to class-0, i.e. absence of groundwater.

The three maps resemble each other, especially in two regions. The first one is the western area of the Mandingue Plateau, where the algorithms generally agreed in predicting class-0. The second major area of agreement is the south sector of the Niger riverbanks. All algorithms predicted class-2 (borehole yield ≥ 5 m^3/h) for this sector. The shores of River Niger along the Bamako sector were also predicted as class-2, but to different extents depending on the classifier. The remainder of the study area shows different degrees of discrepancy, depending on the algorithms of choice.

An ensemble map was specifically developed to analyze these discrepancies. Colors in Fig. 12 represent the arithmetic mean of all three algorithms for each pixel. Thus, the ensemble in Fig. 12 is best described as an agreement map, where all three algorithms have equal weight. The darker shades of green represent the points that were labelled as “highly productive” (class-2) by all algorithms, and red areas represent those that were labelled as “no groundwater” (class-0) by all algorithms. Intermediate colors represent different levels of agreement among the three algorithms. The results are largely consistent with those observed in Fig. 11. In addition, there were at least two algorithms that predicted the highest yield class in some areas both south and north of the Mandingue Plateau. This was reflected in the less intense green colors of the agreement map. A total of 9.6%, 57.7%, and 2.4% of the study area were predicted as class-0, class-1, and class-2, respectively. In addition, 4.0% of the total surface was predicted to be between class-0 and class-1. The remaining 26.3% was predicted as between class-1 and class-2.

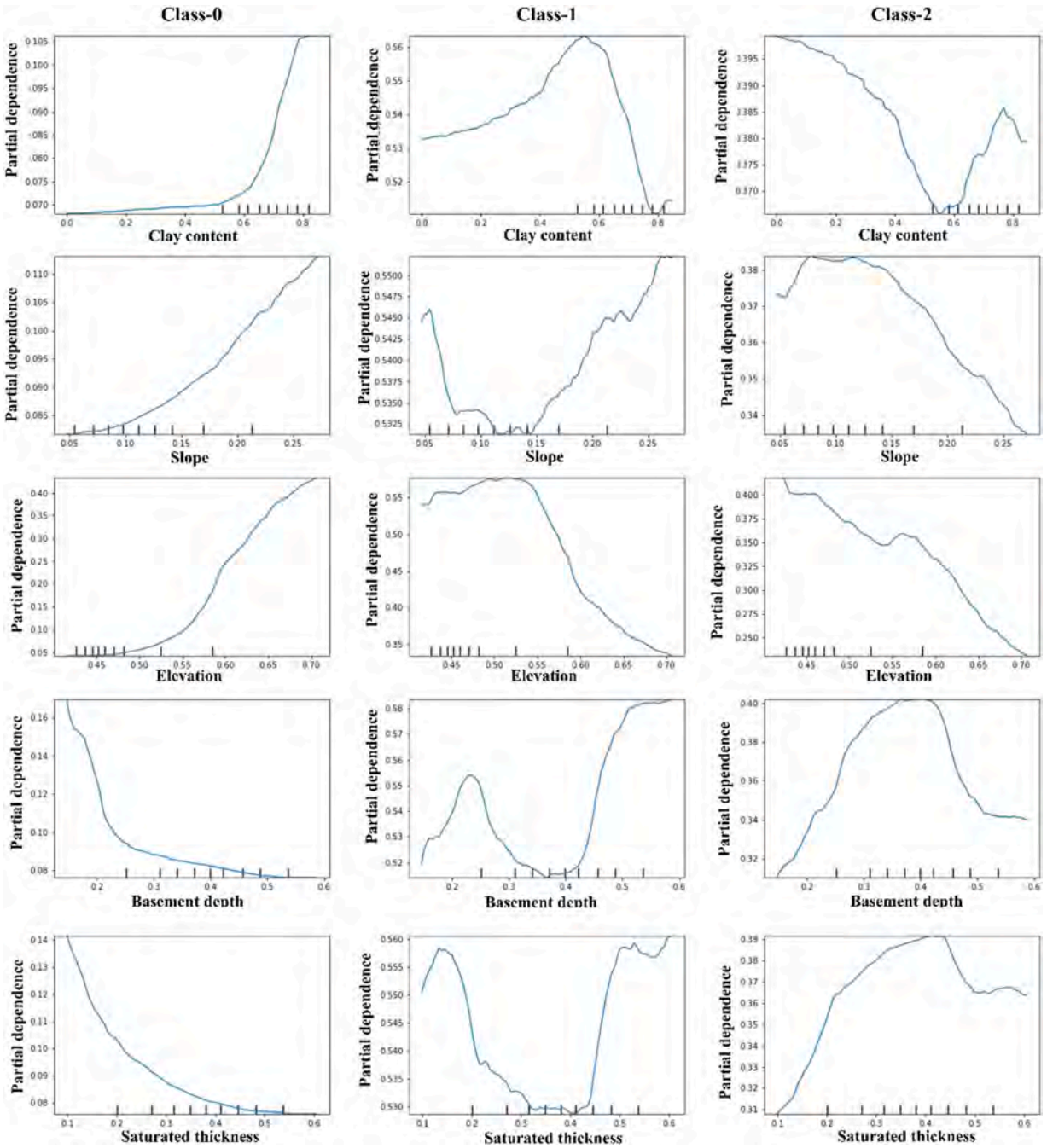

Fig. 10. Partial dependence plot for some of the most important conditioning factors of the Extra Trees Classifier. From top to bottom: Clay content, Slope, Elevation, Basement depth and Saturated thickness. From left to right: Class-0, Class-1 and Class-2.

4. Discussion

4.1. Model selection and importance of conditioning factors

The literature shows that different classifiers rely on different subsets of explanatory variables, and that no subset of conditioning factors can be readily extrapolated from one hydrogeological context to another (Martínez-Santos and Renard, 2020). This is partially due to the internal logic of each one of the classifiers and partially due to the fact that the factors governing groundwater infiltration, flow, and discharge, may differ from case to case. Considering that the internal complexity of the algorithms and the size of the databases make it impossible to predict which algorithm will perform best on a given dataset, a first conclusion that springs to mind is

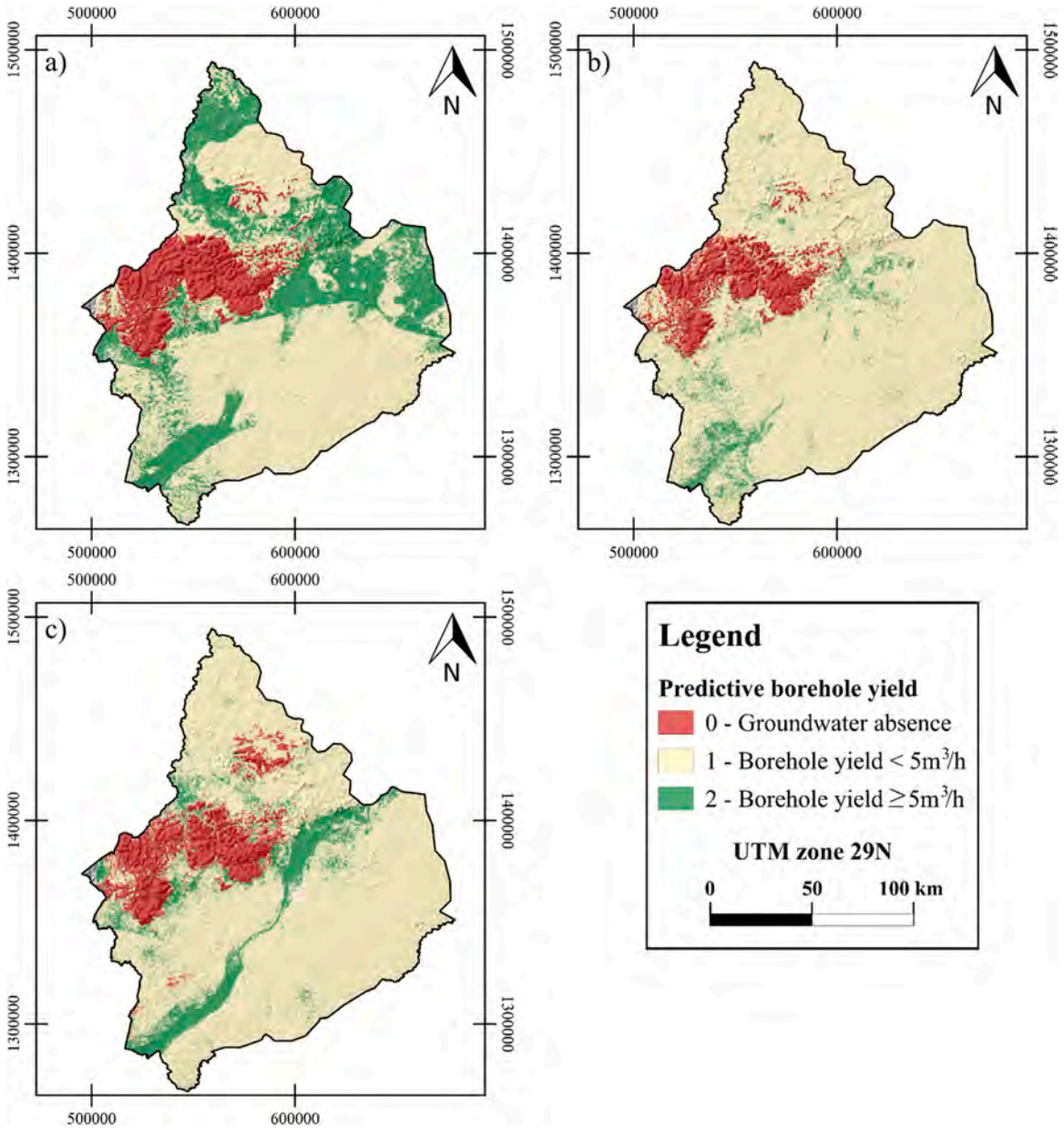

Fig. 11. Predictive borehole yield maps obtained with different machine learning algorithms. a) Extra Trees; b) Gradient Boosting; c) Logistic Regression.

that using a large number of algorithms and then picking the best performing ones is a suitable course of action in machine learning studies dealing with groundwater potential (Gómez-Escalonilla et al., 2021).

From a site-specific perspective, the feature importance analysis for the tree-based algorithms showed that elevation is the variable that correlates best with borehole yield (Fig. 9). This is consistent with the analysis of the partial dependence plots (Fig. 10), as well as with the spatial analysis of the predictive yield maps. The fact that three classifiers used the clay content, geology, geomorphology, land use, rainfall, slope, curvature, drainage density, saturated thickness, elevation and basement depth layers, suggests that these conditioning factors are all strongly correlated with borehole yield. Basement depth, saturated thickness, geology and, —to some extent—, elevation, could be expected to be solid predictors of borehole yield. However, the fact that some of the most important variables are non-hydrogeological deserves additional comment. This is the case of land use, which seems to correlate well with borehole yield. The relationship between the explanatory variable and the target variable seems to be one of consequence, rather than of cause in some cases. For example, the prevalence of irrigated croplands along River Niger obeys to the fact that the underlying aquifer is productive, not all the way around; much like there are no croplands throughout the Mandingue Plateau because there is no

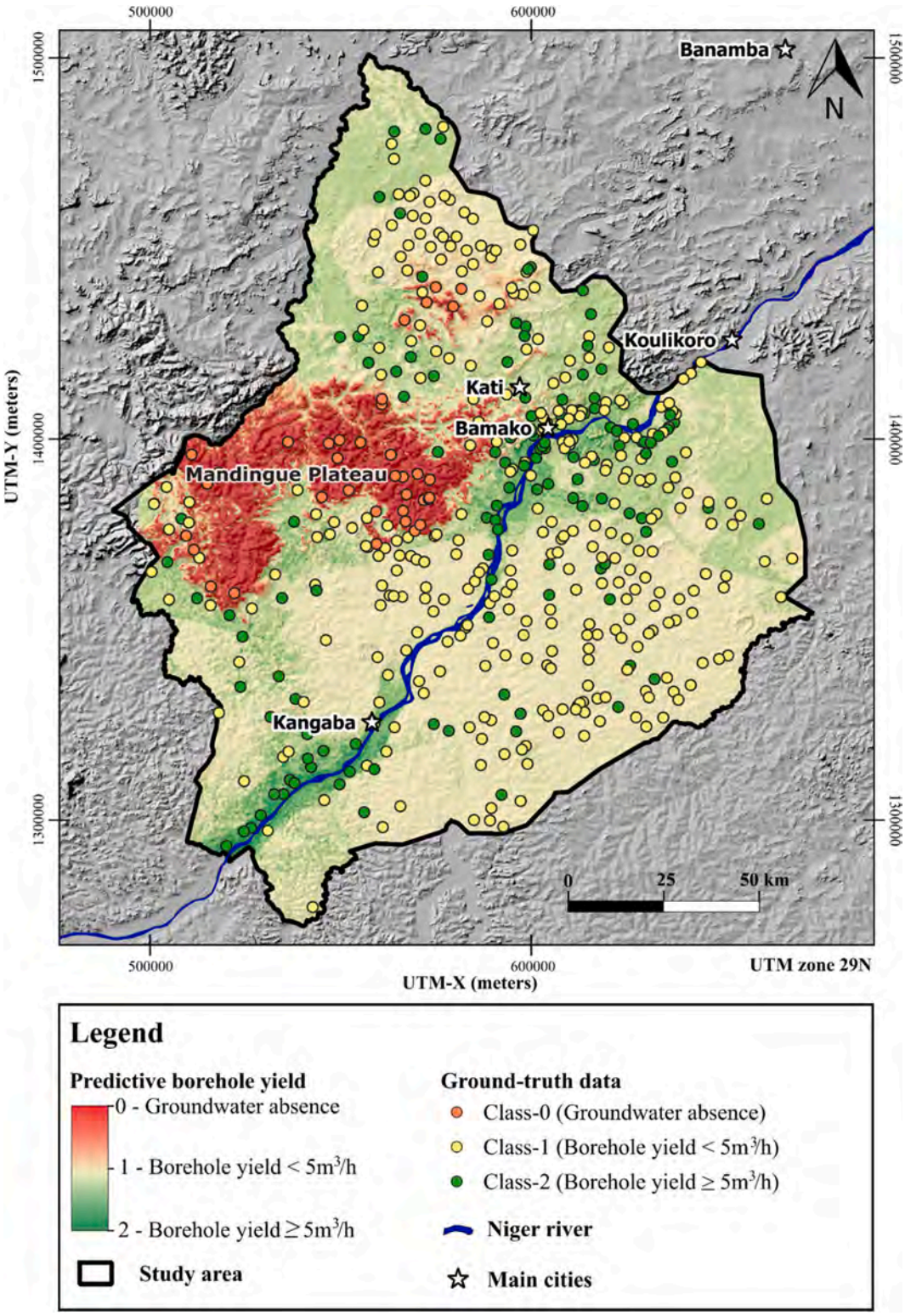

Fig. 12. Arithmetic ensemble of all three predictive borehole yield maps. Colored dots represent ground-truth borehole yield information.

groundwater to irrigate them. Despite this, the land use has been used as an explanatory variable in many studies that suggest that the different land use characteristics will modify the groundwater potential or groundwater resources, i.e., more cause than consequence (Ifediegwu, 2022; Lerner and Harris, 2009; Pham et al., 2019; Selvam et al., 2016). Among even the most favorable land uses, there are large differences related to the magnitude of groundwater recharge. Scanlon et al. (2007) shows that recharge is two orders of magnitude greater through crops than through the original native forests. It is very important to understand that the outcomes of machine learning studies require careful interpretation to avoid spurious conclusions.

Important information can be inferred from the PDP analysis. Take for instance saturated thickness, which was also used to some extent by all three algorithms. The saturated thickness PDP shows that the lowest values of saturated thickness (Mandingue Plateau) relate closely to class 0, which was seen as a likely outcome from the outset. However, the PDPs for classes 1 and 2 require a finer analysis. Intuitively, one would expect a greater saturated thickness to be better correlated with class-2 (higher borehole yields), but this is not the case. Class-2 relates to an intermediate saturated thickness, which is more commonly found along the alluvial sediments of river Niger, whereas the peak in the class-1 PDP occurs where the highest values of saturated thickness take place (i.e. weathering mantles). A plausible explanation for all this is that borehole yield in this case could be more related to hydrodynamic parameters such as hydraulic conductivity and porosity than to saturated thickness, but further data is needed to confirm this hypothesis.

Other PDPs also convey valuable information. For instance, the clay content PDPs suggest that a higher clay content leads to lower recharge of the underlying layers and, consequently, to less productive boreholes. Conversely, zones with lower clay content have a higher partial dependence on the most productive class. Similarly, the slope PDP shows that the steeper slopes favoring surface runoff are associated with the less productive classes, while flat areas and gentle slopes correlate spatially with the more productive boreholes.

Another important methodological issue is that of imbalanced classes. Imbalanced multiclass problems present added difficulties when compared to the prediction of bivariate outcomes based on balanced datasets. This is because increasing the number of classes to be predicted naturally reduces the number of examples for the algorithms to learn from. Furthermore, the more limited the number of ground-truth points is, the more difficult it will be to fine-tune the predictive ability of the classifiers. The aforementioned study by Kumar et al. (2021) used a data augmentation technique consisting of assigning the same yield class to each pixel (30 $\times$30 m) within a given radius from the actual borehole. Other mapping studies based on machine learning techniques have applied different strategies to balance the number of samples in each class. In a study about soil mapping, Taghizadeh-Mehrjardi et al. (2020) found that predictive maps obtained from imbalanced datasets tend to render results that are biased towards the most represented class. These authors evaluated the use of eight different resampling strategies including random under- and over-sampling, synthetic minority over-sampling and adaptive synthetic sampling, among others. Outcomes show that classifiers trained on the SMOTE achieved the best overall performance followed by ADASYN. This study, as well as the one conducted in our paper, reveals that the best resampling technique will vary depending on the initial dataset. In other words, the use of several strategies and the subsequent selection of the one that works best is revealed as a suitable approach when working with imbalanced multiclass datasets.

4.2. Borehole yield potential maps and limitations

Fig. 12 presents the ensemble borehole yield map. The results have been discussed to some extent in the previous section, where the feature importance and partial dependence plots of some of the conditioning factors have been shown. Aside from the Niger riverbanks (especially in the southern and northern sectors), which are clearly associated with the most productive class, the western sector presented a higher potential borehole yield than the eastern sector. As mentioned in Section 2.1, groundwater percolation from the highly-fractured Mandingue Plateau into the plains —where the weathered mantle is thickest— explains the predictions observed in the western sector of the agreement map. Furthermore, high yield predictions along the Niger riverbanks were associated with the fact that the river carries high flows throughout the year. This in turn implies a continuous supply and a hydraulic connection with the alluvial sediments. The remaining area, corresponding essentially to the weathered mantles, was predicted as the medium potential class (class-1) which is consistent with the results outlined in Table 6.

Martínez-Santos et al. (2021) argue that the results obtained by machine learning methods should be verified beyond standard machine learning indicators whenever possible. In this case, the results of the agreement map were cross-checked with variables such as the number of boreholes, the average yield of the boreholes, the percentage of wells exceeding 10 m^3/h, and the average success rate of boreholes located in the different yield domains. Table 6 shows a comparison between the agreement map and actual yield data. Boreholes located in areas predicted as class-1 (n = 260) present an average yield of 2.83 m^3/h. This accounts for 20.6% of the high-yield boreholes (>10 m^3/h), and villages in these sectors present an average borehole success rate of 80.7%. Water points located in areas labelled with intermediate values between class 1 and 2 (n = 152), present an average yield of 5.85 m^3/h, account for 57.7% of the high yield boreholes and present an average drilling success rate of 90.2%. Finally, boreholes located within class-2 areas had an average yield of 8.50 m^3/h, comprised 21.6% of high yielding boreholes (despite accounting for only 7.5% of the villages) and rendered an average borehole success rate of 99.6%. This all suggests that, despite the difficulties encountered by the algorithms in terms of telling apart classes 1 and 2, the map in Fig. 12 represents a sufficiently accurate depiction of field conditions.

While this all seems coherent with field observations, it is important to note that there are certain limitations associated with our ground-truth dataset. A recurrent problem with regional borehole databases in West Africa pertains to the widespread use of hand pumps for rural water supply. Indeed, many boreholes in the region have traditionally been fitted with hand pumps regardless their productivity. Thus, determining whether yields in the order of 1 m^3/h represent the productivity or the aquifer or that of the pump is impossible without extensive fieldwork. Furthermore, the national borehole database contains the information at a village-scale resolution. This implies that all boreholes within a given village have the same coordinates, which can be problematic if the village

Table 6
Results of the analysis based on the agreement map and borehole yield classes.

Borehole yield potential class	Description in ground-truth database	Number of points	Mean borehole yield (m^3/h)	% High yield boreholes ($>10\ m^3/h$)	Borehole success rate at village scale (%)
Class-1	$< 5\ m^3/h$	260	2.83	20.6%	80.7
Class-1/class-2	Mixed	152	5.85	57.7%	90.2
Class-2	$\geq 5\ m^3/h$	33	8.50	21.6%	99.6

overlies a heterogeneous hydrogeological environment. While this is perceived as a minor shortcoming in view of the scale of work, a more detailed knowledge of the borehole coordinates would likely render a finer depiction of field conditions. Finally, it must be acknowledged that borehole productivity is not only a function of aquifer productivity, but also of borehole construction. This means that deficient borehole implementation and/or insufficient drilling depth may result in poor production in areas where the groundwater potential is high. Enhancing the ground-truth database with pumping test data would be a welcome addition for future studies.

5. Conclusions

Groundwater is a strategic resource in arid and semiarid regions. Machine learning techniques present enormous upside in terms of interpreting large amounts of spatially-distributed groundwater data, which may be beneficial to improve water access in low income countries. While groundwater potential mapping is becoming widespread in the academic literature, predicting specific variables such as borehole yield is perceived as a major challenge ahead. Our paper represents one of the first attempts to deal with this particular issue. A key finding is that machine learning-based maps may actually depict field conditions better than standard machine learning metrics, and therefore, that combining standard metrics with ad hoc indicators is a suitable course of action in this type of study.

From a site-specific perspective the Mandingue Plateau was delineated as a low borehole yield area, while the banks of the Niger River have been found to be the most productive zones in the region. Piedmonts located north and south of the mountains represent intermediate productivity. These finding are all consistent with the available borehole yield data and could be used to inform water supply policy at the regional scale.

This research has also explored and demonstrated the possibility of dealing with imbalanced datasets in groundwater potential mapping studies. The ADASYN strategy proved to be the most appropriate in preventing the difficulties associated with class imbalance in our ground-truth sample. Furthermore, this research shows that the use of a large number of classifiers, different forms of pre-processing and, in the case of unbalanced databases, different strategies such as oversampling and undersampling, provide adequate courses of action in machine learning applications to the field of groundwater. From a methodological viewpoint, the updated version of MLMapper presents the added capability of dealing with multi-class targets, which attests to its suitability to map groundwater-related variables beyond a mere positive-negative outcome.

CRediT authorship contribution statement

Víctor Gómez-Escalonilla: Conceptualization, Formal analysis, Methodology, Software and Writing – original draft; **Oumou Diancoumba**: Validation and writing – review & editing; **Dasso Yollande Traoré**: Validation and Writing – review & editing; **Esperanza Montero**: Validation and Writing – review & editing; **Miguel Martín-Loeches**: Conceptualization, Validation and Writing – review & editing; **Pedro Martínez-Santos**: Conceptualization, Methodology, Software, Validation, Writing – review & editing and Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work has been funded under research grant PID2021-124018OB-I00 of Spain's Ministry of Science and Innovation. This work is also related to the research grant RTI2018-099394-B-I00 of Spain's Ministry of Science, Innovation and Universities.. The first author received an FPI grant from the Ministry of Science and Innovation to develop his PhD within this project (PRE2019–090026). The last author received a Salvador de Madariaga grant (PRX18/00235) from Spain's Ministry of Education, Culture and Sport to carry out a 3-month research stay at the Université de Neuchâtel, Switzerland, where the original version of the software used in this paper was developed. The authors thank the Direction Generale de l'Hydraulique of Mali for making its borehole database available. This work has also been funded and is part of the research grant 101059372 within the framework of Horizon Europe 2021 of the European

Commission.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2022.101245.

References

ESA, 2010.

Adelana, S., MacDonald, A., 2008. Groundwater research issues in Africa. In: Adelana, S., MacDonald, A. (Eds.), Applied Groundwater Studies in Africa. IAH Selected Papers on Hydrogeology, Volume 13. CRC Press/Balkema, Leiden, The Netherlands.

Ahmed, N., Hoque, M.A.A., Pradhan, B., Arabameri, A., 2021. Spatio-temporal assessment of groundwater potential zone in the drought-prone area of Bangladesh using gis-based bivariate models. Nat. Resour. Res. 1–23.

Al-Abadi, A.M., Fryar, A.E., Rasheed, A.A., Pradhan, B., 2021. Assessment of groundwater potential in terms of the availability and quality of the resource: a case study from Iraq. Environ. Earth Sci. 80, 426. https://doi.org/10.1007/s12665-021-09725-0.

Al-Fugara, A.K., Ahmadlou, M., Shatnawi, R., AlAyyash, S., Al-Adamat, R., Al-Shabeeb, A.A.R., Soni, S., 2022. Novel hybrid models combining meta-heuristic algorithms with support vector regression (SVR) for groundwater potential mapping. Geocarto Int. 37 (9), 2627–2646. https://doi.org/10.1080/10106049.2020.1831622.

Alpha, A., Traoré, A.Z., Mariko, A., Banton, O., Villeneuve, J.P., Ait-Ssi, L. 1991. Hydrogeologie et contamination de la nappe phréatique alimentant la ville de Bamako (Mali) Institut National de la Recherche Scientifique, Quebec, Canada.

Arabameri, A., Pal, S.C., Rezaie, F., Nalivan, O.A., Chowdhuri, I., Saha, A., Lee, S., Moayedi, H., 2021. Modeling groundwater potential using novel GIS-based machine-learning ensemble techniques. J. Hydrol. Reg. Stud. 36, 100848.

Bai, Z., Liu, Q., Liu, Y., 2022. Groundwater potential mapping in Hubei region of China using machine learning, ensemble learning, deep learning and AutoML methods. Nat. Resour. Res. https://doi.org/10.1007/s11053-022-10100-4.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. J. 24 (1), 43–69.

Bougharriou, E., Allouche, N., Brahim, F.B., Nasri, G., Bouri, S., 2021. Delineation of groundwater potentials of Sfax region, Tunisia, using fuzzy analytical hierarchy process, frequency ratio, and weights of evidence models. Environ. Dev. Sustain 1–26.

Braham, M., Boufekane, A., Bourenane, H., Nait Amara, B., Bensalem, R., Oubaiche, E.H., Bouhadad, Y., 2022. Identification of groundwater potential zones using remote sensing, GIS, machine learning and electrical resistivity tomography techniques in Guelma basin, northeastern Algeria. Geocarto Int. 0 (0), 1–31. https://doi.org/10.1080/10106049.2022.2063408.

Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13 (2), 361–378.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Chen, Y., Chen, W., Chandra Pal, S., Saha, A., Chowdhuri, I., Adeli, B., Janizadeh, S., Dineva, A.A., Wang, X., Mosavi, A., 2021. Evaluation efficiency of hybrid deep learning algorithms with neural network decision tree and boosting methods for predicting groundwater potential. Geocarto Int. 1–21. https://doi.org/10.1080/10106049.2021.1920635.

Choudhary, S., Pingale, S.M., Khare, D., 2022. Delineation of groundwater potential zones of upper Godavari sub-basin of India using bi-variate, MCDM and advanced machine learning algorithms. Geocarto Int. 1–31. https://doi.org/10.1080/10106049.2022.2093992.

Congedo, L., 2021. Semi-automatic classification plugin: a python tool for the download and processing of remote sensing images in QGIS. J. Open Source Softw. 6 (64), 3172.

Dewitte, O., Jones, A., Spaargaren, O., Breuning-Madsen, H., Brossard, M., Dampha, A., Deckers, J., Gallali, T., Hallett, S., Jones, R., Kilasara, M., Le Roux, P., Michéli, E., Montanarella, L., Thiombiano, L., Van Ranst, E., Yemefack, M., Zougmore, R., 2013. Harmonisation of the soil map of Africa at the continental scale. Geoderma 211–212, 138–153. https://doi.org/10.1016/j.geoderma.2013.07.007.

Diancoumba, O., Bokar, H., Toure, A., Kelome, N.C., Preko, K., 2020. Characterization of groundwater recharge using the water table fluctuation method in the koda catchment. Mali. Int J. Adv. Sci. Eng. Inf. Technol. 8 (1), 665–681.

Díaz-Alcaide, S., Martínez-Santos, P., 2019. Advances in groundwater potential mapping. Hydrogeol. J. 27 (7), 2307–2324.

Direction Nationale de l'Hydraulique, 2010. Données hydrogeologiques et des forages. Dir. Natl. De. l'Hydraulique.

DNGM) and Laboratoire de Géologie dynamique St-Jérôme Marseille (LGSJM). 1988. Carte géologique de la République du Mali à 1/200000, Feuille n° ND-29-IV, BAMAKO OUEST (MALI). Ministère des Mines, de l'énergie et de l'Eau.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C.J., Osborne, P.E., Reneking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 027–046.

Elbeih, S.F., 2015. An overview of integrated remote sensing and GIS for groundwater mapping in Egypt. Ain Shams Eng. J. 6 (1), 1–15.

Forootan, E., Seyedi, F., 2021. GIS-based multi-criteria decision making and entropy approaches for groundwater potential zones delineation. Earth Sci. Inform. 14 (1), 333–347.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Wolff, E., 2018. Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. IEEE Geosci. Remote Sens. Lett. 15 (4), 607–611.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63 (1), 3–42.

Gómez-Escalonilla, V., Vogt, M.L., Destro, E., Isseini, M., Origgi, G., Djoret, D., Martínez-Santos, P., Holecz, F., 2021. Delineation of groundwater potential zones by means of ensemble tree supervised classification methods in the Eastern Lake Chad basin. Geocarto Int. https://doi.org/10.1080/10106049.2021.2007298.

Gómez-Escalonilla, V., Martínez-Santos, P., Martín-Loeches, M., 2022. Preprocessing approaches in machine-learning-based groundwater potential mapping: an application to the Koulikoro and Bamako regions. Mali. Hydrol. Earth Syst. Sci. 26 (2), 221–243.

Hakim, W.L., Nur, A.S., Rezaie, F., Panahi, M., Lee, C.-W., Lee, S., 2022. Convolutional neural network and long short-term memory algorithms for groundwater potential mapping in Anseong, South Korea. J. Hydrol.: Reg. Stud. 39, 100990 https://doi.org/10.1016/j.ejrh.2022.100990.

He, H., Bai, Y., Garcia, E.A., Li, S., 2008. ADASYN: adaptive synthetic sampling approach for imbalanced learning (June). 2008 IEEE Int. Jt. Conf. Neural Netw. (IEEE World Congr. Comput. Intell.) 2008, 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969.

Ifediegwu, S.I., 2022. Assessment of groundwater potential zones using GIS and AHP techniques: a case study of the Lafia district, Nasarawa State, Nigeria. Appl. Water Sci. 12, 10. https://doi.org/10.1007/s13201-021-01556-5.

IUSS Working Group WRB, 2015. World Reference Base for Soil Resources 2014. FAO, Rome.

Jasiewicz, J., Stepinski, T.F., 2013. Geomorphons — a pattern recognition approach to classification and mapping of landforms. Geomorphology 182, 147–156. https://doi.org/10.1016/j.geomorph.2012.11.005.

Jyrkama, M.I., Sykes, J.F., Normani, S.D., 2002. Recharge estimation for transient ground water modeling. Groundwater 40 (6), 638–648.

Khoshtinat, S., Aminnejad, B., Hassanzadeh, Y., Ahmadi, H., 2019. Groundwater potential assessment of the Sero plain using bivariate models of the frequency ratio, Shannon entropy and evidential belief function. J. Earth Syst. Sci. 128 (6), 1–16.
Kumar, C.P., 1997. Estimation of natural ground water recharge. ISH J. Hydraul 3 (1), 61–74.
Kumar, R., Dwivedi, S.B., Gaur, S., 2021. A comparative study of machine learning and Fuzzy-AHP technique to groundwater potential mapping in the data-scarce region. Comput. Geosci., 104855
Lemaître, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. J. Mach. Learn. Res 18 (1), 559–563.
Lerner, D.N., Harris, B., 2009. The relationship between land use and groundwater resources and quality. Land Use Policy 26, S265–S273.
Ling, C.X., Li, C., 1998. Data mining for direct marketing: Problems and solutions. In: Proceedings of the Fourth Int. Conf. Knowl. Discov. Data Min. (KDD-98), Vol. 98. AAAI Press,, New York, NY, pp. 73–79.
Liu, A.Y. C. 2004. The effect of oversampling and undersampling on classifying imbalanced text datasets (Doctoral dissertation, University of Texas at Austin).
Martínez-Santos, P., Renard, P., 2020. Mapping groundwater potential through an ensemble of big data methods. Ground Water 58 (4), 583–597. https://doi.org/10.1111/gwat.12939.
Martínez-Santos, P., Díaz-Alcaide, S., De la Hera-Portillo, A., Gómez-Escalonilla, V., 2021. Mapping groundwater-dependent ecosystems by means of multi-layer supervised classification. J. Hydrol. 603, 126873.
Martinsen, G., Bessiere, H., Caballero, Y., Koch, J., Collados-Lara, A.J., Mansour, M., Sallasmaa, O., Pulido-Velazquez, D., Williams, N.H., Zaadnoordijk, W.J., Stisen, S., 2022. Developing a pan-European high-resolution groundwater recharge map – Combining satellite data and national survey data using machine learning. Sci. Total Environ. 822 https://doi.org/10.1016/j.scitotenv.2022.153464.
MIHE. 1990. Synthese hydrogeologique du Mali. Technical Report. Ministère de l'Industrie de l'Hydraulique et de l'Energie. Programme des Nations Unies pour le Développement. Bamako. 356p.
Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol. Process. 5 (1), 3–30.
Muavhi, N., Thamaga, K.H., Mutoti, M.I., 2021. Mapping groundwater potential zones using Relative Frequency Ratio, Analytic Hierarchy Process and their Hybrid Models: Case of Nzhelele-Makhado Area in South Africa. Geocarto Int 1–18. https://doi.org/10.1080/10106049.2021.1936212.
Namous, M., Hssaisoune, M., Pradhan, B., Lee, C.-W., Alamri, A., Elaloui, A., Edahbi, M., Krimissa, S., Eloudi, H., Ouayah, M., Elhimer, H., Tagma, T., 2021. Spatial prediction of groundwater potentiality in large semi-arid and karstic mountainous region using machine learning models. Water (Switz.) 13 (16). https://doi.org/10.3390/w13162273.
Nguyen, P.T., Ha, D.H., Jaafari, A., Nguyen, H.D., Van Phong, T., Al-Ansari, N., Prakash, I., Le, H.V., Pham, B.T., 2020. Groundwater potential mapping combining artificial neural network and real AdaBoost ensemble technique: the DakNong province case-study, Vietnam. Int. J. Environ. Res. Public Health 17 (7), 2473.
Ourhzif, Z., Algouti, A., Hadach, F., 2019. Lithological mapping using Landsat 8 OLI and ASTER multispectral data in Imini-Ounilla district south High Atlas of Marrakech. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. Volume XLII-2/W13, 1255–1262.
Panahi, M., Sadhasivam, N., Pourghasemi, H.R., Rezaie, F., Lee, S., 2020. Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). J. Hydrol. 588, 125033.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in python. Mach. Learn Python 12, 2825–2830.
Pham, B.T., Jaafari, A., Prakash, I., Singh, S.K., Quoc, N.K., Bui, D.T., 2019. Hybrid computational intelligence models for groundwater potential mapping. Catena 182, 104101.
PNUD. 1982. Recherche et mise en valeur des eaux souterraines. Technical Report. Programme des Nations Unies pour le Développement. Bamako. 98p.
Poggio, L. and de Sousa, L. 2020. SoilGrids250m 2.0 - Clay content, Access date: 21/11/2021.
Prati, R.C., Batista, G.E., Monard, M.C., 2009. Data mining with imbalanced class distributions: concepts and methods (December). Indian Int. Conf. Artif. Intell. 359–376.
Sachdeva, S., Kumar, B., 2021. Comparison of gradient boosted decision trees and random forest for groundwater potential mapping in Dholpur (Rajasthan), India. Stoch. Environ. Res. Risk Assess. 35 (2), 287–306.
Santos, M.S., Soares, J.P., Abreu, P.H., Araujo, H., Santos, J., 2018. Cross-validation for imbalanced datasets: avoiding overoptimistic and overfitting approaches [Research Frontier]. IEEE Comput. Intell. Mag. 13 (4), 59–76.
Selvam, S., Dar, F.A., Magesh, N.S., Singaraja, C., Venkatramanan, S., Chung, S.Y., 2016. Application of remote sensing and GIS for delineating groundwater recharge potential zones of Kovilpatti Municipality, Tamil Nadu using IF technique. Earth Sci. Inf. 9, 137–150. https://doi.org/10.1007/s12145-015-0242-2.
Singh, A., Purohit, A., 2015. A survey on methods for solving data imbalance problem for classification. Int. J. Comput. Appl. 127 (15), 37–41.
Sun, Y., Kamel, M.S., Wang, Y., 2006. Boosting for learning multiple classes with imbalanced class distribution (December). Sixth Int. Conf. data Min. (ICDM'06) 592–602.
Taghizadeh-Mehrjardi, R., Schmidt, K., Eftekhari, K., Behrens, T., Jamshidi, M., Davatgar, N., Scholten, T., 2020. Synthetic resampling strategies and machine learning for digital soil mapping in Iran. Eur. J. Soil Sci. 71 (3), 352–368.
Trabelsi, F., Lee, S., Khlifi, S., Arfaoui, A., 2018. Frequency ratio model for mapping groundwater potential zones using GIS and remote sensing; Medjerda Watershed Tunisia (November). Conference of the Arab. J. Geosci. Springer,, Cham, pp. 341–345 (November).
Traore, A.Z. 1985. Géologie et hydrogéologie des plateaux Mandingues (Mali): région de Koula-Nossombougou (Doctoral dissertation, Université Scientifique et Médicale de Grenoble).
Traore, A.Z., Bokar, H., Sidibe, A., Upton, K., Ó Dochartaigh, B., and Bellwood-Howard, I. 2018. Africa Groundwater Atlas: Hydrogeology of Mali, ⟨http://earthwise.bgs.ac.uk/index.php/Hydrogeology_of_Mali⟩, 2018.
UNESCO. 2015. Water for a sustainable world. Facts and figures. The United Nations World Water Development Report 2015. United Nations World Water Assessment Programme Programme Office for Global Water Assessment. Division of Water Sciences. Perugia, Italy. 12p.
Xie, Y., Sha, Z., Yu, M., 2008. Remote sensing imagery in vegetation mapping: a review. Plant Ecol. 1 (1), 9–23.
Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27 (14), 3025–3033.