

Article

Prospectivity Mapping of Heavy Mineral Ore Deposits Based upon Machine-Learning Algorithms: Columbite-Tantalite Deposits in West-Central Côte d'Ivoire

Kassi Olivier Shaw ^{*}, Kalifa Goïta  and Mickaël Germain 

Center for Applications and Research in Remote Sensing (CARTEL), Department of Applied Geomatics, University of Sherbrooke, 2500 Boulevard de l'Université, Sherbrooke, QC J1K 2R1, Canada

^{*} Correspondence: kassi.olivier.m.shaw@usherbrooke.ca

Abstract: This study aimed to model the prospectivity for placer deposits using geomorphic and landscape parameters. Within a geographic information system (GIS), spatial autocorrelation analysis of 3709 geochemical samples was used to identify prospective and non-prospective targets for columbite-tantalite (Nb-Ta) placer deposits of Hana-Lobo (H-L) Geological Complex (West-Central Côte d'Ivoire, West Africa). Based on mineralization system analysis, hydrologic, geomorphologic and landscape parameters were extracted at the locations of the identified targets. Supervised automatic classification approaches were applied, including Random Forest (RF), K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) to find a prospectivity model complex enough to capture the nature of the data. Metrics such as cross-validation accuracy (CVA), Receiver Operating Characteristic (ROC) curves, Area Under Curve (AUC) values and F-score values were used to evaluate the performance and robustness of output models. Results of applying machine-learning algorithms demonstrated that predictions provided by the final RF and KNN models were very close ($\kappa = 0.56$ and $CVA = 0.69$; $\kappa = 0.54$ and $CVA = 0.68$, respectively) and those provided by the SVM models were slightly lower with $\kappa = 0.46$ and $CVA = 0.63$. Independent validation results confirmed the slightly higher performance of both KNN and RF prospectivity models, compared to final SVM. Sensitivity analyses of both KNN and RF prospectivity models for medium and high-grade Nb-Ta deposits show a prediction rate of up to 90%.

Keywords: predictive mineral “prospectivity” mapping; GIS; machine-learning algorithms; mining exploration; columbite-tantalite; Ivory Coast



Citation: Shaw, K.O.; Goïta, K.; Germain, M. Prospectivity Mapping of Heavy Mineral Ore Deposits Based upon Machine-Learning Algorithms: Columbite-Tantalite Deposits in West-Central Côte d'Ivoire. *Minerals* **2022**, *12*, 1453. <https://doi.org/10.3390/min12111453>

Academic Editors: Ehsan Farahbakhsh, Rohitash Chandra and Mohammad Parsa

Received: 22 October 2022

Accepted: 14 November 2022

Published: 17 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Supervised or semi-supervised machine-learning algorithms are often used to model the prospectivity of new exploration target areas in the context of complex and non-linear relationships between ore deposit targets and input proof functions, or when the input datasets have different statistical distributions [1]. These algorithms include Support Vector Machines (SVM) [2–6], K-Nearest Neighbors (KNN) [7], or Random Forest (RF) [8–12]. With these machine-learning algorithms, new prediction methods are increasingly being used in Mineral Prospectivity Mapping (MPM) to predict the prospectivity of new regions at satisfactory rates, especially when the entry space is complex and there are non-linear relationships between entry criteria and mineral exploration targets. For instance, predictive models based on logistic regression functions [13,14], Bayesian networks [15], SVM [4,6], neural network systems [16,17], or decision trees, such as RF [10], can be used to model, predict, and map the prospectivity of new mineral exploration targets.

Prediction methods based on automatic machine-learning algorithms using expert system products such as restricted Boltzmann machines [18,19], wavelet neural networks [20], or group method of data-handling neural networks [21], demonstrated the high potential of artificial neural networks (ANN) to model the prospectivity of new mineral targets.

However, these methods require large quantities of input data for learning, otherwise their performance is degraded [22,23]. Moreover, according to [1], the learning processes of ANN algorithms require the estimation of many parameters that can have a significant impact on the robustness of the final model. In summary, the relationships between targeted deposits and input evidence functions are more often complex and non-linear. Furthermore, the available input datasets may present different statistical distributions [1], or be of limited quantity. In such complex situations, supervised or semi-supervised automatic machine-learning algorithms, including SVM, KNN or RF have shown real potential in predicting the mineral prospectivity of new regions (e.g., [6,7,10]).

The columbite-tantalite (Nb-Ta) mineralization system of the Hana-Lobo (H-L) Geological Complex (central-western Côte d'Ivoire) has been the subject of several geological [24–27], and metallogenic research works [28–30]. These studies revealed that the Nb-Ta placer deposits mineralization system is linked to landscape spatial factors, and the relationships between the placer deposits area and mineralization evidence are complex and non-linear. In such a context of complex and non-linear spatial relationships between evidence criteria and the mineral exploration target, the predictive modelling of prospectivity, using machine-learning algorithms, is a valid tool.

MPM is a typical classification method that can be used to solve problems in finding prospective or non-prospective areas [2]. To this end, several MPM techniques are commonly implemented in a GIS environment. MPM approaches can be used to model the prospectivity of placer deposits and assess probabilities to find new prospective and non-prospective areas in regions that are little or poorly explored [31–33], or brownfield. Additionally, with MPM, after analyzing the mineralization system, important mapping indicators (known as input criteria) can be extracted and then the exploration target areas can be identified using the data or knowledge-based methods provided.

MPM methods proposed in the literature are generally classified into three categories. The first category consists of MPM methods based on knowledge of the mineral system, which use expert judgement to assign weight to layers of evidence [34,35], to predict the mineral prospectivity of new, poorly-explored or brownfield regions. In this case, weighting the evidence criteria requires a good understanding of the mineral processes and the relationships between the evidence factors and the mineralization system being studied. This sometimes requires the acquisition of a large amount of in-situ or remotely detected exploration data.

The second category is data-based methods. They use the locations of known mineral occurrences as a training site, to model, based on evidence criteria, the prospectivity of new, poorly-explored or brownfield regions [2,9,36]. These methods are based on an estimate of spatial associations between different measurements performed within the target data (minerals occurrences) to assign a weight to each layer of informative evidence [37]. These weights are computed from the target data.

The third category is a hybrid approach that uses both the locations of known mineral occurrences and expert opinions to assign weight to the evidence criteria [38,39].

In practical terms, the application of these three categories generally falls within a spectrum between conceptual and empirical targeting. Conceptual targeting will attempt to understand the process controlling the spatial distribution of the mineral exploration target under consideration. Empirical targeting will focus on the recognition of spatial models, based on known geological controls that can be used as a prospecting guide. Where existing spatial relationships between mineral targets and evidence criteria are complex and non-linear, conceptual, and empirical targeting approaches can be complementary and used in tandem [40].

The novelty of this study lies in prospectivity modeling for Nb-Ta placer deposits using geomorphic and landscape parameters in an area with limited data. The literature together with field work were necessary to provide data for this study.

2. Geology and Mineralization

2.1. Study Area and Geology

The study area covers 645 km² in the northern part of the H-L Geological Complex within the 6°26'00" N and 6°42'30" N and 6°26'30" W and 6°23'00" W coordinates. The H-L Geological Complex is in the center west of Côte d'Ivoire (West Africa) (Figure 1).

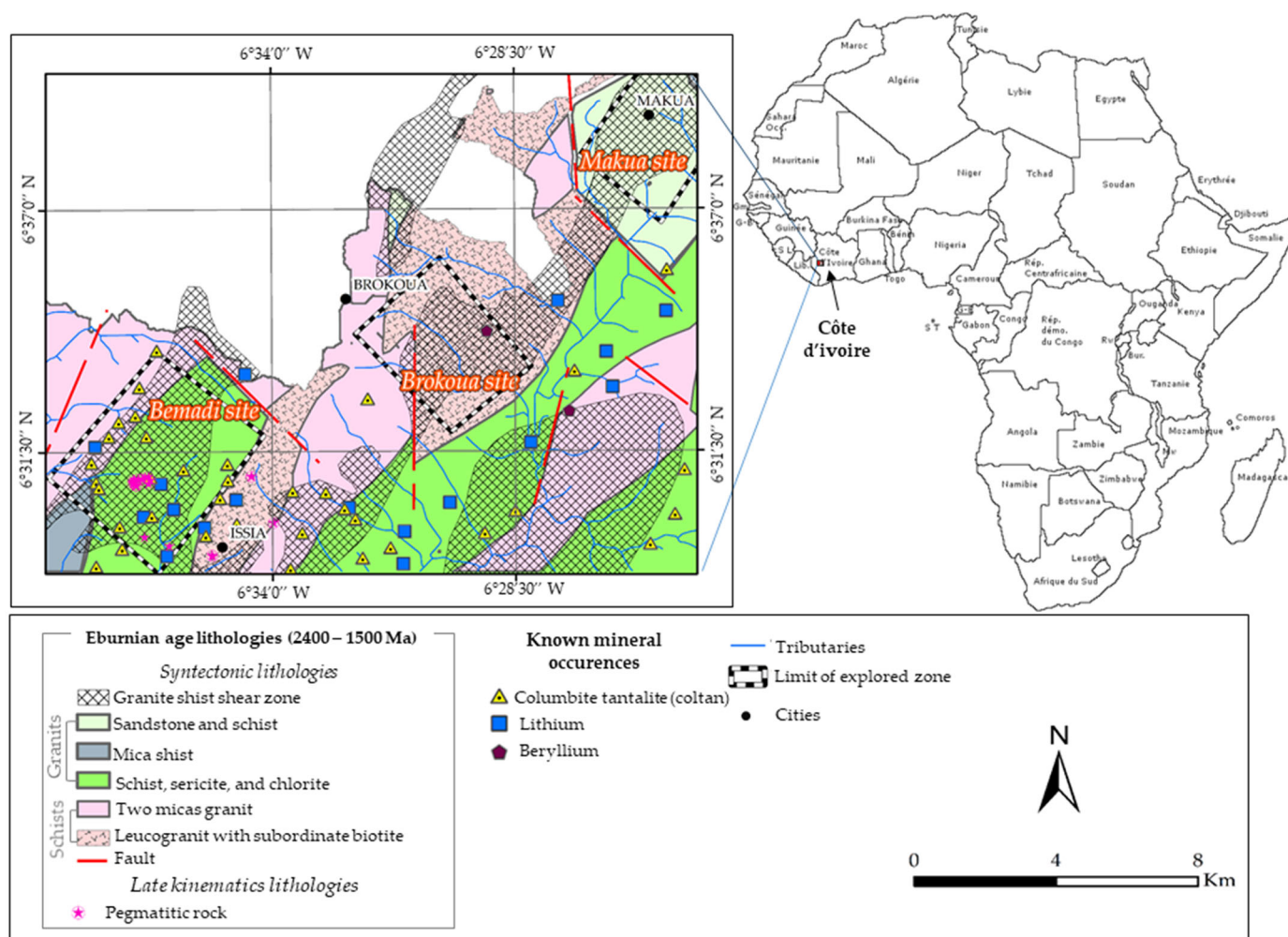


Figure 1. Location of the study area with regard to the granite schist shear zone (by Cruys, 1965) [25] (map of Africa source: © mappemonde.net, (accessed on 18 October 2022)).

Figure 1 shows the location of the study area, particularly the Bemadi site (located southwest of the study area) that covers an area of 41.81 km², the Brokoua site (at the center of the study area) covering an area of 26.56 km² and the Makua site (northeast of the study area) covering an area of 24.47 km² (Figure 1).

Only 14.4% (92.84 km²) of the study area has been sampled by soil geochemistry, especially in the Bemadi, Makua and Brokoua sites. The three sites explored (white and black square in dotted line in Figure 1) are located on the periphery (Bemadi site) or in the vicinity (Makua and Brokoua sites) of the pegmatitic-granite massifs.

The Precambrian rocks occupy 97% of the superficial area of the country and include rocks of two orogenic episodes, the Liberian (3000–2580 Ma) and the Eburnean of lower Proterozoic (2400–1550 Ma) ages.

The Precambrian base rock is a domain of meta-somatic granitic rocks (calco-alkaline and leucocratic granites containing two micas) intruded by schistose veins or rocks with sedimentary origin. These veins include mica schists, sericite schists, schistose chlorite, and micaceous and sandstone schists. Due to rather pronounced geomorphogenesis,

the region is almost entirely covered by eluvial, colluvial and alluvial deposits, which occupy almost all surfaces being studied [30]. Moving inland from the Atlantic coast, the tropical rainforest that covers the southern part of the country is eventually replaced by semi-deciduous tropical forest and, ultimately, by tropical savannah [41].

Rarely visible in outcrops, H-L volcano-sedimentary rocks are observable in altered state, through exploration pits. Lithology consists primarily of sericito-schistose and chlorito-schistose rocks, sandstone, micaceous and arkose schists and mica schists (shown as grey and green panels on Figure 1), homogeneous leucocratic granites, calco-alkaline granites, and two-micas granites showed as pink panels on Figure 1).

Granites appear in the region in the form of domes and extend northeast. Pegmatitic granite massifs of the region (Issia region ~384 km from Abidjan) were developed through several tectonic episodes with an important phase of contact metamorphism affecting the adjoining geological lithologies composed of schists and mica schists [30], that belong to a shear zone described by Cruys [25] as favorable to Nb-Ta placer deposits. The geological complex is also marked by several kilometric faults bearing NS, NNW-SSE and NS-SE (showed as red continuous or discontinuous lines in Figure 1).

2.2. Mineralization System Analysis and Data Used

Within the H-L Geological Complex, several studies have focused upon the complex and non-linear relationship between known exploration indicators and Nb-Ta placer deposit formation and emplacements [28–30]. The mineral system under study belongs to the Man Shield (West African Craton) and its emplacement is related to the Eburnean Orogeny, which occurred 2100 to 2070 Ma BP. Primary mineralization was emplaced in granites and pegmatites in a multi-phase tectonic context [30]. The current location of the Bemadi, Brokoua and Makua sampled deposits (i.e., on the periphery or in the vicinity of mineralized granite pegmatites massifs) does not cast doubt on their granitic and pegmatitic origin [30]. Moreover, a previous study [30], demonstrated that the primary mineralization system of the Nb-Ta fits into a fractional crystallization model that explains the origin of Nb-Ta mineralization (Figure 2a).

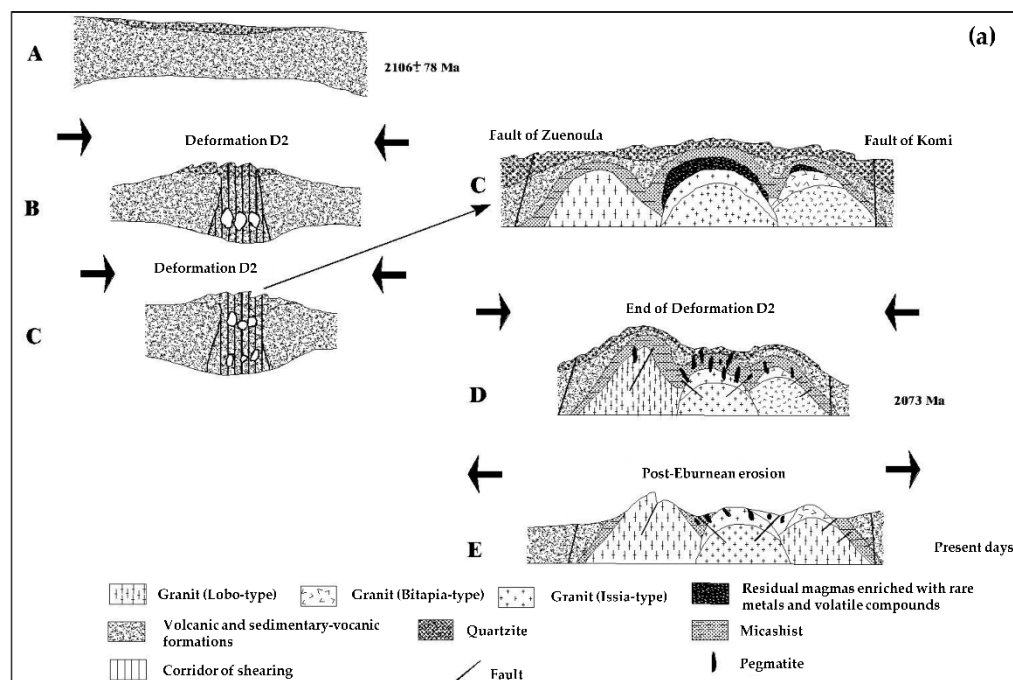


Figure 2. Cont.

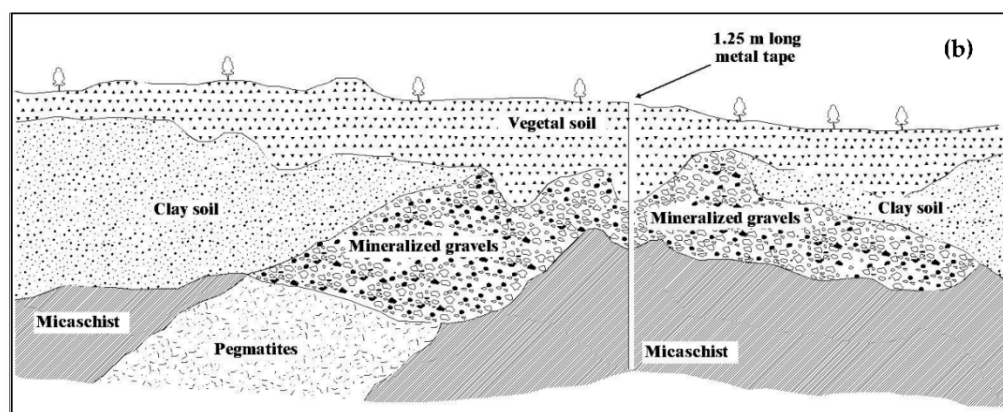


Figure 2. (a) Modified geodynamic model for the placement of columbite-tantaliferous pegmatites and associated rocks of Issia [30] (p. 273). (b) Modified general architecture of a weathering mantle within a studied Nb-Ta placer deposit [30] (p. 196). Original figure from [30] has been modified with the consent of its author.

Chronologically, the geodynamic model [30] (p. 273) established primary metalliferous concentrations of beryl, lithium, and especially Nb-Ta, with four orogenic stages (A, B, C and D) that explains the formation of granitic and pegmatitic rocks. Stage A is marked by erosion of sedimentary rocks and sedimentary volcanic rocks (2106 ± 78 Ma). Stages B, C and D are marked by a transcurrent tectonic deformation (D2) (2073 Ma) accompanied by a process of progressive granitization (towards the surface) of a mineralized magma in niobium-tantalum (Nb-Ta), lithium (Li) and, beryllium (Be).

This means that massifs of pegmatitic granites mineralized in Nb-Ta were set up in favor of the movement of the geological faults (tensile fissures, Riedel fractures). This indicates that the fracturing system of the Precambrian base rock is involved in the location of the mineralized pegmatitic granite massifs and their associated Nb-Ta placer deposit in proximity. Otherwise, there is a complex spatial relationship between the fracturing density system and emplacement of the soil geochemical pits acquired into the known Nb-Ta Bemadi, Brokoua and Makua deposits.

Stage E is marked by a period of erosion due to the action of meteoric water. Observations in field works confirmed that Nb-Ta placer deposits sampled are small amplitudes and are located near the surface, under a low vegetal cover and in a flat layer of mica schists or altered pegmatitic rocks. The deposits are usually made up of two gravelly layers of fine, medium, or coarse grain sizes (characteristic of lateritic soils) [30] (Figure 2b).

Placer deposits of Nb-Ta were described as secondary mineralization built from accumulations of valuable mineral commodities, which are formed by gravimetric separation, with specific gravities greater than that of quartz (>2.65), such as Nb-Ta minerals [25,26,30]. In addition, Nb-Ta placer deposits were developed in a humid tropical context that is favorable first, to the disintegration of mineralized primary rocks, secondly to the transportation of altered products such as gravel, and thirdly to the deposition of altered heavy minerals and sediments in natural trap zones. Placer deposits are a favorable environment for accumulation of sediments and minerals resulting from a succession of several hydro-morphologic processes. Hydro-morphologic processes are marked by major drainage phases and alluviation phases, followed by a phase of fragmentation and wear of vein rocks (hydrothermal alteration) such as pegmatites and quartz [25,26]. Otherwise, there is a complex spatial relationship between hydro-morphologic factors and emplacement of the soil geochemical pits acquired into the known Nb-Ta Bemadi, Brokoua and Makua deposits. Analysis of mineralization systems indicate that hydro-morphologic parameters (such as drainage processes), geomorphic and landscape parameters (such the unevenness of the ground and fracturing of Precambrian base rocks) are involved in the mineralization processes and the locations of the Nb-Ta placer deposits studied.

2.3. Data Used

Table 1 summarizes the main type of data that were considered.

Table 1. Summary of data that were used in this study.

Nature	Format	Source	Spatial Resolution	Spectral Resolution	Study Parameters	Usages
Mineralogy	Vector (points)	Fieldwork	25 m × 50 m		Concentrated Nb-Ta content in gravels.	Identification of mineral targets.
Hydrology	Matrix	ASTER (ASTGTM2_N06 W007_dem) of 17 October 2011.	30 m		Surface accumulation of runoff, flow amplitude (or magnitude and direction).	Identification of areas suitable for sediment transport and deposition.
Morphology	Matrix	ASTER (ASTGTM2_N06 W007_dem) of 17 October 2011.	30 m		Degree of slope inclination, profile and curvature of slope and, relief.	Identification of natural trap areas (favorable for erosion and deposition).
Structural	Matrix	Sentinel 2B MSI, level 1C, Date 24 December 2017	10, 20, 60 m	13 bands between 443 and 2190 nm	Discontinuity In pixels	Extraction of lineaments
		Sentinel 1A SAR (VV & HH), Date 6 January 2018	40 m	Band C		
		ESRI world imagery (Maxar, Geoeye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN and, the GIS USER community)	15 m		Actual appearance of objects	Validation of faults and fractures.
		SPOT-5	2.5–10 m	5 bands between 0.5 and 0.89 µm	Actual appearance of objects	
Geology	Vector	Geological map (Government of Côte d'Ivoire)	1:200,000		Lithological surfaces, Pegmatite emplacements, Faults, Fractures	Understanding the geological context

Geochemical data were acquired from field works by sampling soil. A total of 3709 geochemical samples were taken within the granite schist shear zones described by Cruys [25], in Bemadi, Makua and Brokoua sites. Sampling pits measured 1.20 m long by 0.80 m wide, with a maximum depth of 3 m. The pits were positioned on a regular grid that was 100 m wide by 50 m long. Average Nb-Ta concentrate grades (in grams per cubic meter) of the 3709 gravel samples were then calculated. The geochemical exploration survey data collected for this study, consist of the geographical coordinates of the pit, and the average Nb-Ta grade sampled in the pit.

Geological data in vector format were taken from the geological map (1: 200,000 scale) covering the studied area. Geological information used in this study consists of locations

of rocky outcrops of granitic pegmatites, together with known Nb-Ta, Be and Li mineral occurrences, and granitic and schist formations hosting the mineralization.

Structural data were extracted from different satellite images. These included a Sentinel 2B-MSI Level 1C multispectral image (acquired 24 December 2017 in a descending orbit). The image contains 13 wavelength bands, ranging from the Visible and Near Infrared (VNIR) to the Shortwave Infrared (SWIR), i.e., between 443 and 2190 nm. The second image that was used is a Sentinel-1A wide interferometric C-band radar image (HH and VV single polarization, 40 m resolution), which was acquired on 6 January 2018 in a descending orbit. These data were used to extract natural lineaments and to map the fracture system of the study area. Finally, Esri world imagery and SPOT image mosaics were used to validate the natural lineaments visually, and to analyze and validate the appearance of the objects at resolutions of 15 m to 2.5 m.

3. Methodology

Given the lack of data that would be required to investigate existing spatial relationships between exploration criteria and Nb-Ta placer deposits, the MPM selected approach used the predictive power of supervised automatic classification algorithms to model the prospectivity of new prospective and non-prospective Nb-Ta placer deposits. The overall approach is detailed in the flowchart depicted in Figure 3. The workflow is structured around six main steps, denoted S1 to S6 and described below.

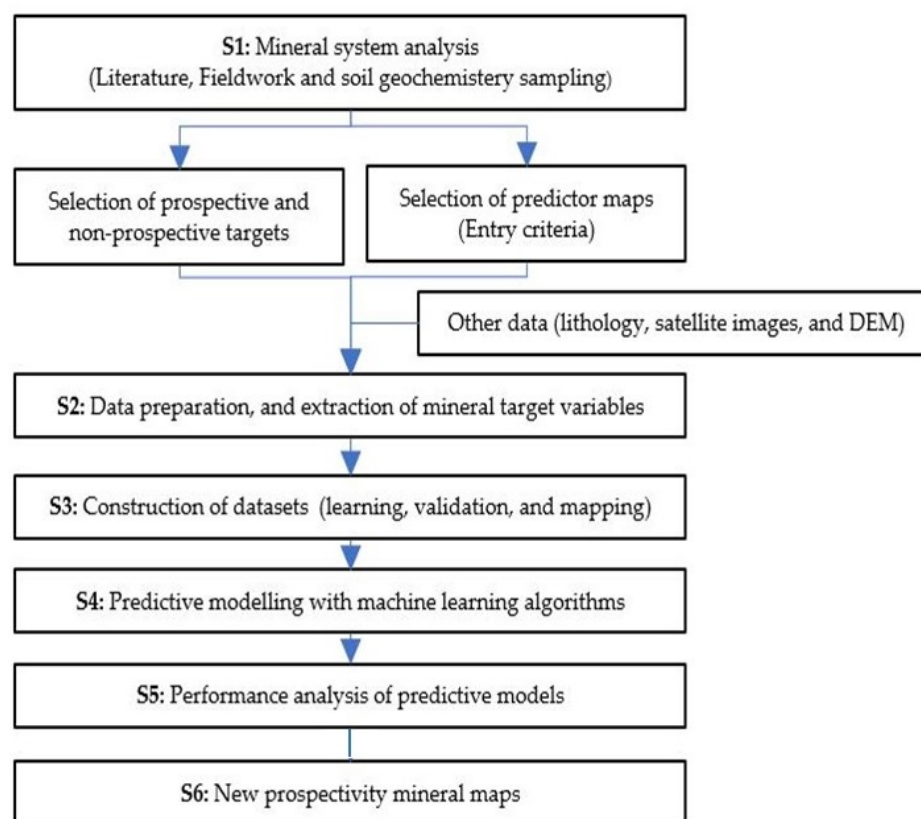


Figure 3. Flowchart summarizing the six research steps.

3.1. Classification with SVM, KNN and RF Algorithms Machine-Learning

Algorithms such as SVMs were designed for binary classification and do not natively support classification tasks with more than two classes. One approach for using binary classification algorithms for multi-classification problems is to split the multi-class classification dataset into multiple binary classification datasets and fit a binary classification model on each one. A binary classifier can be trained on each binary classification problem

and predictions are made using the model that is the most confident. Two different examples of this approach are the One-vs.-All (OvA) and One-vs.-One (OvO) strategies. The SVM model has been widely documented in the literature (e.g., [42,43]). SVM partitions a k -dimensional space into $k-1$ hyperplanes. The algorithm uses these optimal-margin hyperplanes to separate the data into different classes, based on a training vector dataset. This classification and regression model is nonparametric, given that it is insensitive to the statistical distribution of the data, which is an important advantage. In this study, the classification results were SVM models with Gaussian, cubic, and quadratic kernels, with varying levels of flexibility and interpretability.

The RF classification method is an ensemble classifier whose predictions are based on complex decision trees. In effect, RF uses a set of classification and regression trees that are created from training datasets in a replacement process, referred to as “bagging.” The broader technique of bagging, or bootstrap aggregation that is performed within RF, is an ensemble-based algorithm that fits multiple models on different bootstrapped samples of the training dataset, and then aggregates the predictions made by the decision trees that are formed by each of these models. This process increases the diversity of the trees in the RF, while avoiding the problems of correlation among the different trees, together with overfitting [44]. RF, which uses subsamples of the predictors, has been widely used in classification problems for more details regarding the method [43,45,46]. Generally, RF generates efficient classification results, especially by over-sampling classes with large misclassification costs and under-sampling those with low misclassification costs. In this study, parameters were varied during the learning process to obtain the best possible results. For example, the number of learners and the maximum number of splits were used to control the depth of the trees and to increase the flexibility of the different output predictive models.

The KNN algorithm is well known in the field of remote sensing classification; it uses a similarity measure based on a distance matrix, which was calculated in a manner that was Euclidean or weighted Euclidean to determine the nearest neighbors [47]. In this study, several variants of KNN were analyzed, i.e., fine, medium, coarse, cosine, cubic and weighted models. The training parameters were varied to exploit the flexibility and interpretability of the different output models.

3.2. Selection and Extraction of Predictor Maps

3.2.1. Extraction of Structural Evidence Map

The fractures and tensile fissures generated by the transcurrent deformation (D2) (see Figure 2a) contributed to the migration of the mineralizing fluids responsible for the placement of the mineralized pegmatitic granite. Only one structural evidence criterion was considered in this study, i.e., the fracturing density system. This system is likely to favor the formation of topographic anfractuosités (i.e., winding, or circuitous channels or passages), as well as influencing the emplacement of potentially mineralized host rocks (pegmatitic granite).

To map fracturing density, lineaments were extracted from Sentinel optical and radar images (see Table 1). The lineaments extraction technique used was based on previous studies [48–50]. The first one [48] showed—by comparing five satellite image enhancement techniques (mean value of all bands, Principal Component Analysis (PCA), band ratio, histogram equalization and pass-through filter top)—that the PCA is an effective technique for automatic identification of geological lineaments. Later, the second approach [49] used the second principal component to map geological lineaments from Radarsat-1, and, the third one [50] used the VNIR and SWIR bands of the OLI and ASTER images.

In this study, the B11 (SWIR) and B08 (NIR) bands were re-sampled at 10 m resolution. The VV bias gamma band of the Sentinel-1 SAR IW GRD image was corrected radiometrically and geometrically at 10 m resolution. Previous works [51,52] recommend a minimum length of 10 pixels as an acceptable detection threshold (LTHR); a value between 3 and 8 pixels as a filter radius (RADI) to avoid the introduction of noise into the process;

a value between 10 and 70 pixels as the minimum edge detection level (GTHR); a value between 2 and 5 as the maximum tolerance level (FTHR) to adjust a polyline curve; a value between 3 and 20 degrees recommended as the maximum angle (ATHR) not exceeded for a two-polyline link, and finally a value between 10 and 45 pixels as the maximum distance (DTHR) acceptable for linking two polylines. In this study, lineaments were extracted with the following optimal parameters: RADI: 5; GTHR: 30; LTHR: 5; FTHR: 5; ATHR: 30; and DTHR: 30. A regional bearing (angle from the geographic north direction) rosette was calculated using the Rose and Stereonet chart-plotting program (Rose.net), which can draw structural geology rose charts [53]. Figure 4 shows the major regional fracture directions (angle from the geographic north direction) that were extracted from lineaments, calculated with the Rose and Stereonet chart-plotting program.

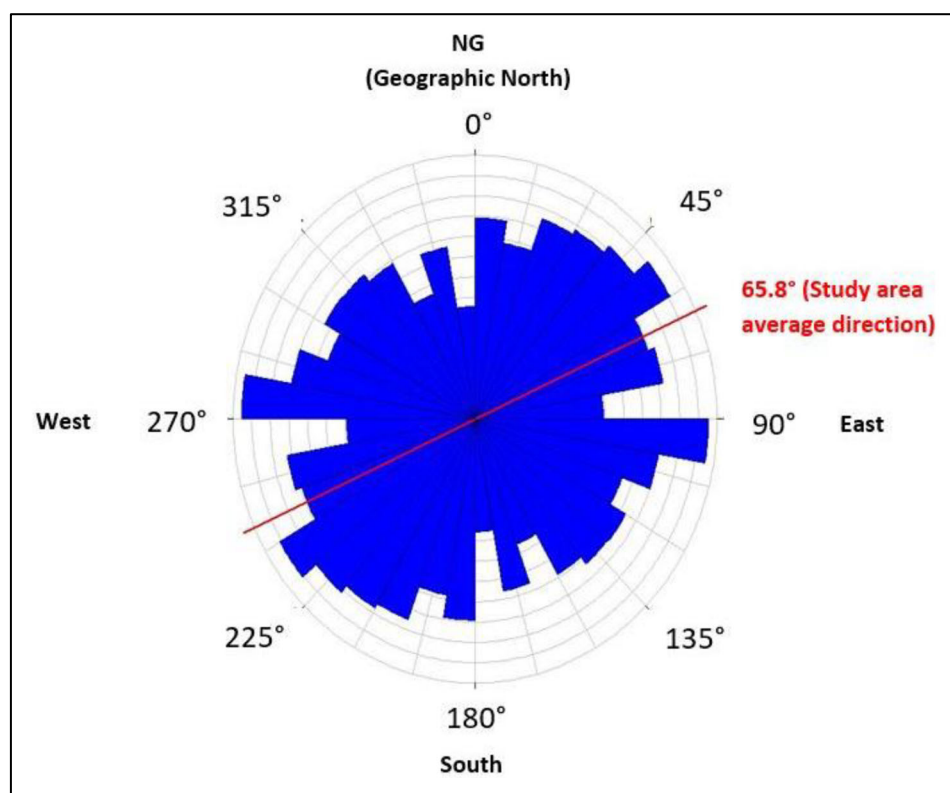


Figure 4. Rosette diagram of fractures orientations calculated from extracted lineaments.

Major fracture orientations identified in this study were north to south (0 to 10° north), northeast to southwest (20 to 50° north), east to west (90 to 110° north), and southeast to northwest (120 to 150° north). The study area average bearing was estimated as 65.8° (red line on Figure 4). The obtained geological lineaments were validated using the literature and the high-resolution satellite images available on the study area. The directions agreed well with the regional tectonic deformation system that gave rise to the straight and isoclinal folds observed in the shear zone (granite schists) that have been previously mentioned by field geologists [25,29,54]. These studies also agree with results of structural mapping works [55,56] related to the bedrock of the Lobo Basin in west-central Côte d'Ivoire. Geological lineaments were visually validated using three data sources: a false-color composite (red, green, blue) from bands 4, 3, and 2 of the Sentinel-2B image; Esri world imagery and SPOT-5 image mosaics corresponding to the study area; and the geological map. This task was necessary to separate the natural lineaments from anthropogenic lineaments; the latter included boundaries of cultivated land or forests, boundaries of built-up areas, and land or rail transport routes [52,57,58]. Therefore, a fracturing density map was constructed in a GIS from the validated geological lineaments. Figure 5a shows

the fracture density map generated and Figure 5b,c shows hydrological predictors used. Figure 5b,c, respectively, illustrate flow accumulation areas and flow amplitude (i.e., flow magnitude and flow direction), which are the two hydrological evidence criteria that were considered in this study.

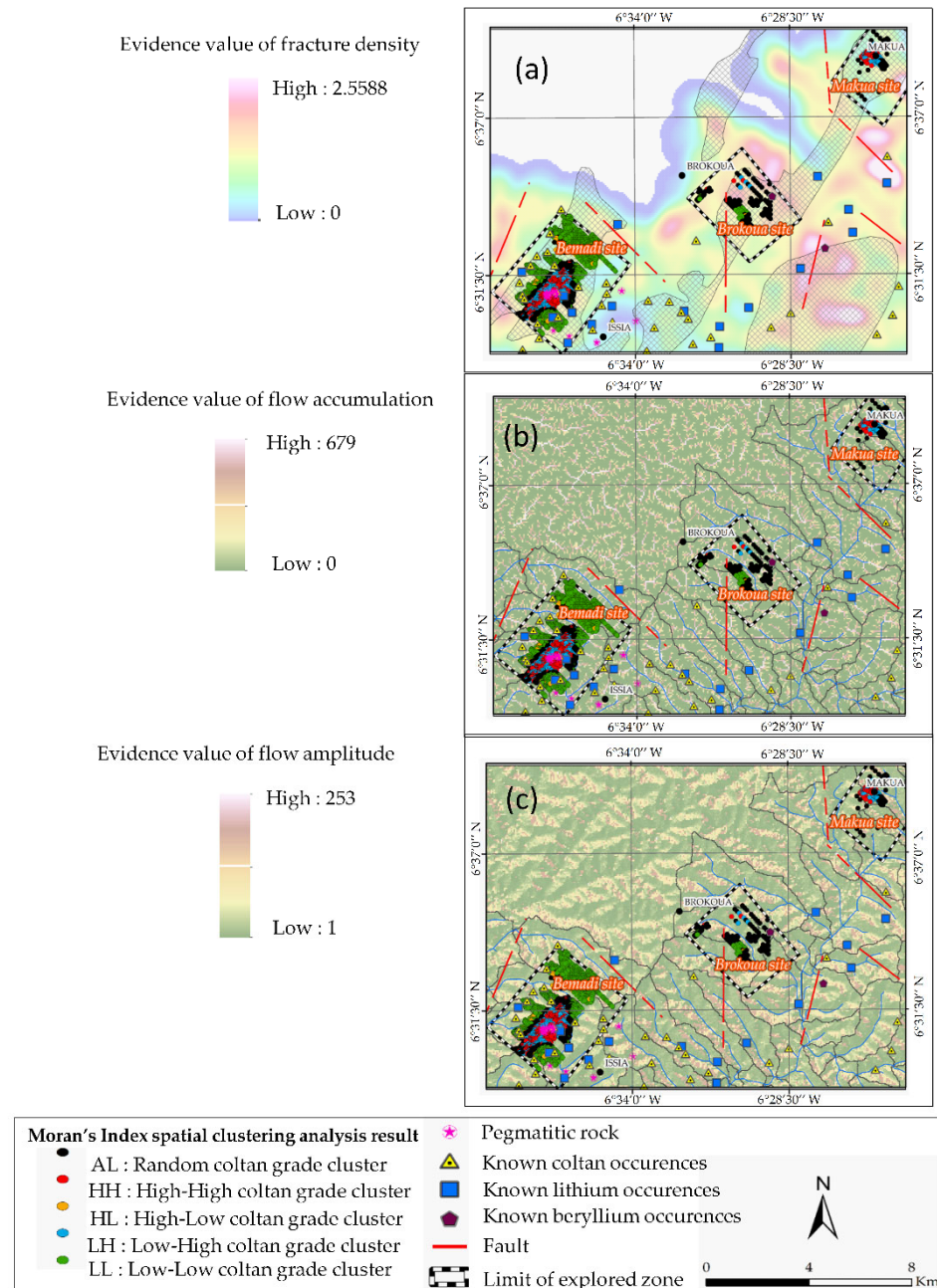


Figure 5. (a): Evidence maps of structural parameters (fracture density). (b): Evidence map of hydrological parameters (flow accumulation). (c): Evidence map of hydrological parameters (flow amplitude).

The values of the density parameter increase in the green to red color range (0 to 2.6). Yellow to yellow-orange areas represent areas of medium fracturing. The heavily fractured areas are magenta to red and the less fractured areas are green, dark blue and turquoise blue (Figure 5a). In step S2, fracture density parameter values of the landscape were extracted at the geographic coordinates of each pixel of the study area. The extracted values corresponding to each pit location were used as the structural predictor variable with machine-learning algorithms.

3.2.2. Extraction of Hydrological Evidence Criteria

Field observation of the pit profiles showed a sequence of grano-graded gravel layers covering the Precambrian age base rocks. The observation of a grano-classification phenomenon in a placer-type deposit involves control of hydrologic processes (such as drainage) and geomorphologic or landscape factors (such as relief and slope shapes) that control, first, the erosion processes of the primary mineralized rocks, and second, the transportation and accumulation of erosion products (i.e., sediments and altered minerals) into natural trap areas.

Therefore, the mapping of natural concentration axes of surface runoff highlights the ability of surface runoff to mobilize, transport, erode and deposit eroded sediments and altered Nb-Ta minerals within catchments. Hydrologic exploration criteria of the landscape considered in this study are the accumulation area and the drainage flow amplitude parameters, together with the hydrographic network. The catchment areas of runoff and the network through which water flows to an orifice are called drainage systems. The amplitude and flow accumulation surfaces represent the spatial variation of the water flow phenomenon, parameterized according to a value, extracted at each pixel (or cell) of a digital elevation model covering the extent a study area.

The accumulation capacity of catchment runoff increases towards the riverbeds, i.e., grading from yellow-orange to white pixels (see Figure 5b). Areas that are favorable for runoff accumulation occupy the riverbeds and banks of rivers or marshes. They have high values and are the source of strong river currents that favor soil erosion and sedimentation processes. Lower accumulation values (green pixels on Figure 5b) occupy large areas of valleys or alluvial plains, which are favorable for long-distance sediment transport and deposition.

Flow amplitudes or directions also highlight, on either side of the marshes and rivers, areas of asymmetric interfluvies (green pixels on Figure 5c) and topographic irregularities (yellow-orange) that occupy areas of ridges, flanks or summits of the massifs, or peneplains of the region. The direction of runoff flow is naturally from orange-yellow pixels (elevated areas) to green pixels of increasingly lower topographic relief (see Figure 5c).

In step S2, hydrologic parameters were extracted using algorithms that were available in ArcGIS v.10.2. The hydrologic parameter values of the landscape were extracted at the geographic coordinates of each pixel of the study area. The extracted values corresponding to each pit location were used as the hydrological predictor variable with machine-learning algorithms.

3.2.3. Extraction of Geomorphologic Evidence Criteria

Landscape geomorphologic parameters are likely to influence the acceleration, deceleration, convergence, or divergence of runoff at the scale of a study sub-catchments. Relevant elements that were mapped in this study are the curvature surfaces (i.e., concavity, convexity) or slope profile, the degree of slope inclination and the landform. The GIS software that was used has all functionality required to make the necessary calculations from the DEM (Bonham-Carter, 2013) [34]. Figure 6a–c illustrates the geomorphological evidence maps used.

Slope curvatures (Figure 6a) show areas with very concave slopes (dark brown) that are favorable for erosion, areas with very convex slopes (dark blue) that are favorable for runoff acceleration, and quasi-planar areas (light brown) that are unfavorable for sediment erosion or runoff.

With respect to degrees of slope inclination (see Figure 6b), steep slopes (dark brown) favor sediment flow and drainage processes, while increasingly gentle slopes contribute to deceleration of runoff and, consequently, trapping of heavy minerals over time in areas with topographic irregularities. Relief map evidence shows that all mineral targets occur at an altitude ranging from 200 to 393 m (see Figure 6c). Dark brown pixels indicate alluvial, colluvial and eluvial zones that are occupied by placer deposits.

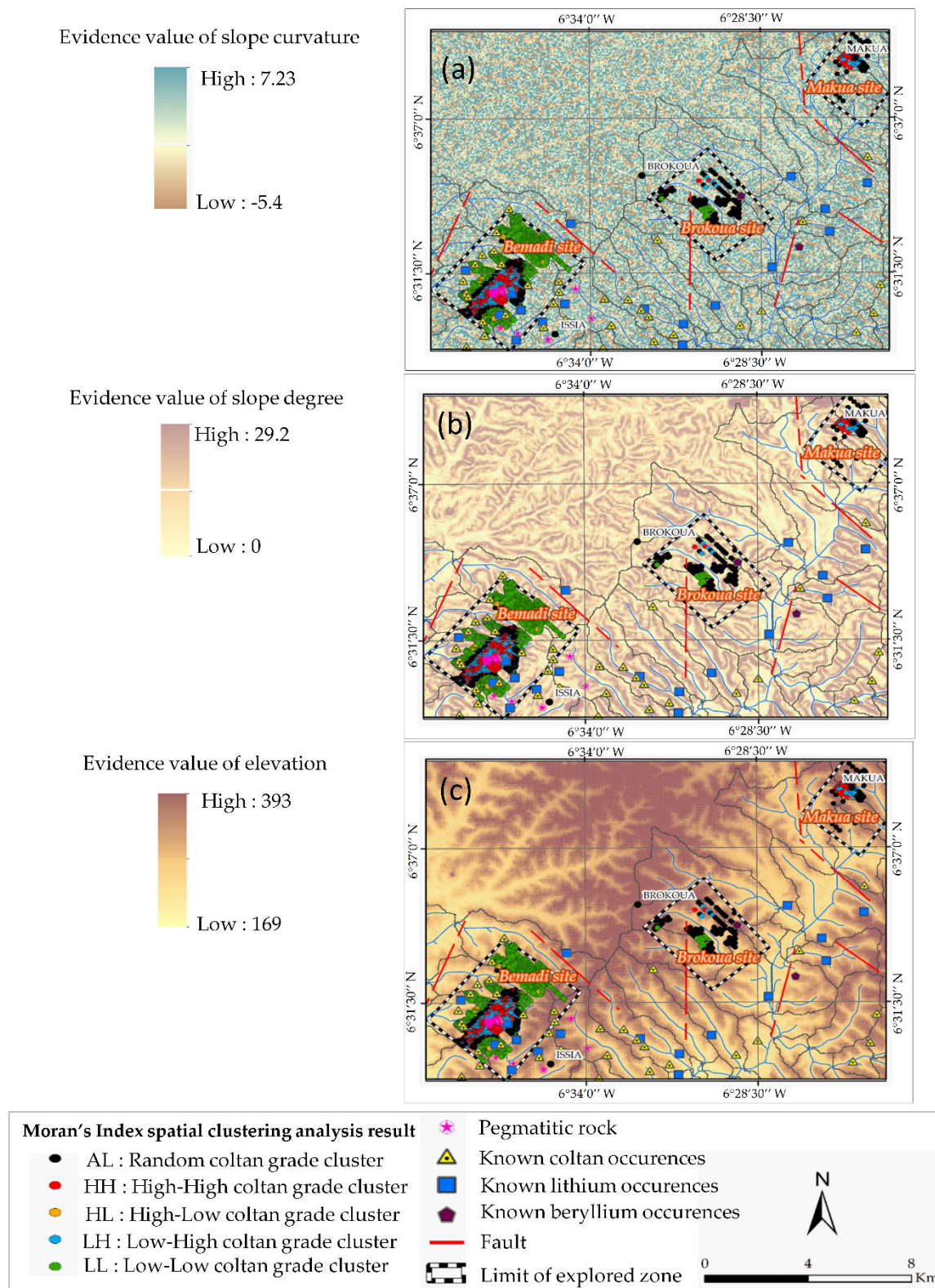


Figure 6. Evidence maps of geomorphologic parameters, (a): Slope curvature; (b): Slope inclination; and (c): Elevation (relief).

Ochre-yellow to light brown pixels are the lowest topographic levels (riverbeds and gullies), while dark-brown pixels indicate increasingly higher elevation areas ranging from shallows to hillsides, plateaus, and ridges. In step S2, the geomorphologic parameter values

of the landscape were extracted at the geographic coordinates of each cell of the study area. The scale of most evidence maps that were used is 1:30,000. This scale provides sufficient accuracy for camp-scale mineral predictive mapping [42]. Finally, all the geomorphologic, hydrologic, and structural parameter values that were extracted from the placer deposit emplacements were used as predictor variables with machine-learning algorithms.

3.3. Selection and Extraction of Mineral Targets

The first step (S1) involved a statistical analysis of autocorrelation (or spatial aggregation technique) of the geochemical dataset (called dataset 1). Dataset 1 is a table consisting of 3709 rows (each row corresponds to a sample) with four columns of values populated by the latitude coordinate, the longitude coordinate, the sample name, and the average grade value of Nb-Ta expressed in grams per cubic meter.

We considered the Nb-Ta average grade value and the geographic coordinates of each pit. We analyzed the geostatistical distribution of dataset 1. Then we distinguished between different types of mineral prospective and non-prospective targets of Nb-Ta placer deposits inside the H-L Complex. Dataset 1 was implemented in a GIS (ArcGIS Version 10.2, ESRI, Redlands, CA, USA) and we used Moran's statistical clustering analysis technique to determine (based on their respective locations) whether spatial patterns of mineral exploration targets were clustered, dispersed or randomly distributed among the placer deposits studied. Given that the pits were arranged in a regular grid work (50 m by 100 m), we calculated Moran's I, i.e., spatial autocorrelation [59], using the inverse of the distance between pits as weights. This form of the weight matrix is a rational a priori choice of hypothesis and is used to characterize the existing spatial relationships between the different mineralized gravel locations that were sampled.

The analysis yielded global values of Moran's I that determined whether the overall pattern simultaneously expressed among spatial locations and mineral grades was clumped (negative-binomial distribution), over-dispersed (binomial), or random (Poisson). Local Moran's I [60] was used to indicate both statistically significant and non-significant spatial aggregations (clusters) of mineral grades (e.g., local patterns of High-High, Low-Low, Low-High and High-Low). Clusters resulting from the statistical analysis of autocorrelation of dataset 1 may represent different types of prospective and non-prospective Nb-Ta samples.

The appropriate tool (ArcGIS Pro 2.7: Cluster and Outlier Analysis) not only calculates these local indices, which can be used to identify spatial outliers (high-low and low-high groupings), but also their associated Z-scores, pseudo-probability values (permutation-based estimates), and codes (COType) for each feature in the geochemical survey data table (dataset 1). The COType (Cluster/Outlier Type) assigned to each feature in the pit grid is used to determine whether a sample belongs to a spatial cluster (or mineral target type). Interpretation of these groupings is carried out in the context of a null hypothesis. For example, when the pseudo-*p*-values of the features composing the clusters of points (high, low, or outliers) are less than ≤ 0.05 , these points are then considered to be statistically significantly aggregated with a probability of 95%. Clusters are thus clumps or aggregates of High-High (HH) or Low-Low (LL) sample locations embedded in a matrix of low or high values, respectively. Outliers occur when a high value is surrounded by predominantly low values (HL), or conversely, when a low value is surrounded by predominantly high values (LH). The significant clustering response of HH or LL (indicated by the associated Z-score and *p*-value) could imply the existence of underlying factors that would control the spatial distribution of these mineral exploration targets at the observed population level. The spatial aggregation analysis step (S1) produces prospective and poorly prospective mineral target classes, which are then used for modeling mineral prospectivity of placer deposits, using geomorphic and landscape parameters. The map shown in Figure 7 represents clusters and outliers of the prospective and non-prospective samples selected from the Bemadi, Makua and Brokoua sites.

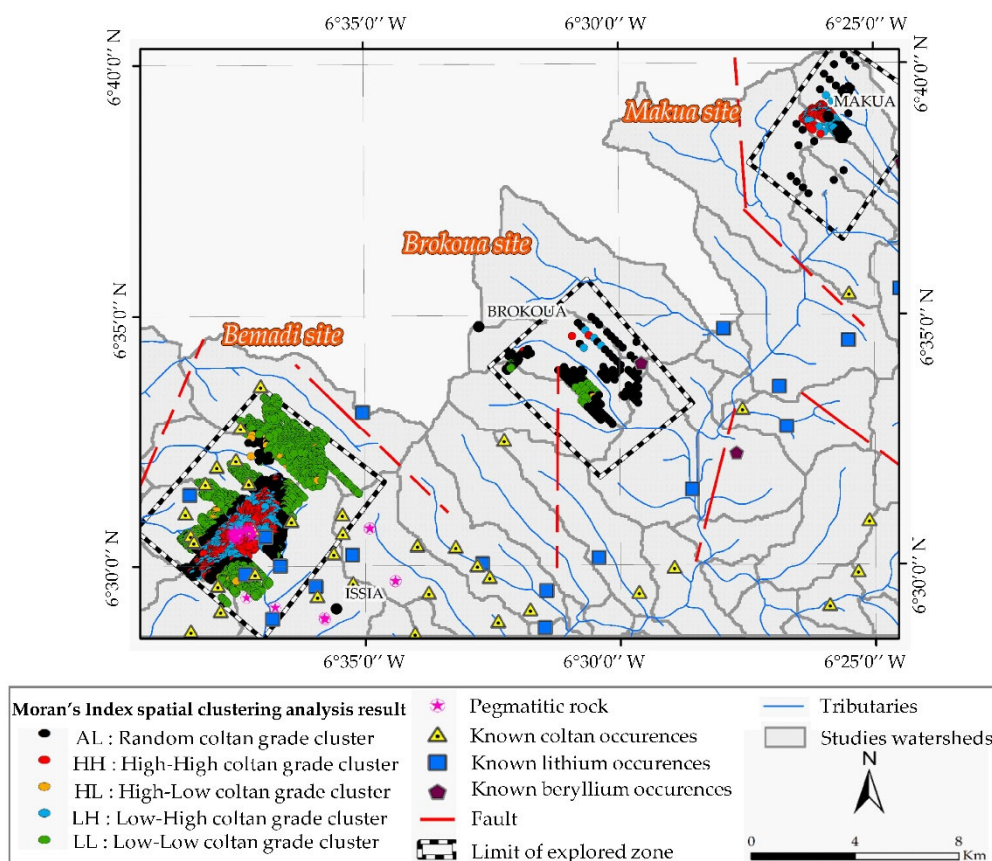


Figure 7. Map of spatial aggregates in the three different sites of the study area.

Based on the type of cluster or outlier, the threshold value of the cluster or outlier and the geoeconomic importance of the mineral target (see Section 5 for details) were used to select three classes of prospective samples. The AL cluster (referred to as Class 1 (AL)) was selected as the class of non-prospective samples. Cluster LL (referred to as Class 2 (LL)) was the mineral prospective target of low-grade placer deposits. Outlier LH (referred to as Class 3 (LH)) is the mineral prospective target of medium-grade placer deposits. The cluster HH and the outlier LH were grouped together as a single class (referred to as Class 4 (HH/HL)) of mineral prospective for targets of high- and very high-grade placer deposits.

Table 2 summarizes all the evidence layers and features used to build the machine-learning prediction models in this study.

Table 2. Summary of all the evidence layers and features used to build the machine-learning prediction models.

N°	Evidence Layer Selected as Entry Criteria	Predictive Parameters Selected to Build the Machine-Learning Prediction Model
1	Evidence map of fracture density (Figure 5a).	Value of the fracture density extracted at the sample locations into deposits (clusters HH/HL, LL and LH) and non-deposit (cluster AL) targets.
2	Evidence map of flow accumulation area (Figure 5b).	Value of the flow accumulation surface extracted at the sample locations into deposits (clusters HH/HL, LL and LH) and non-deposit (cluster AL) targets.
3	Evidence map of flow accumulation or magnitude areas (Figure 5c).	Value of the flow magnitude extracted at the sample locations into deposits (clusters HH/HL, LL and LH) and non-deposit (cluster AL) targets.

Table 2. Cont.

N°	Evidence Layer Selected as Entry Criteria	Predictive Parameters Selected to Build the Machine-Learning Prediction Model
4	Evidence map of slope curvature (Figure 6a).	Value of the slope curvature extracted at the sample locations into deposits (clusters HH/HL, LL and LH) and non-deposit (cluster AL) targets.
5	Evidence map of slope degree (Figure 6b).	Value of the slope inclination extracted at the sample locations into deposits (clusters HH/HL, LL and LH) and non-deposit (cluster AL) targets.
6	Evidence map of relief (Figure 6c).	Value of the elevation extracted at the sample locations into deposits (clusters HH/HL, LL and LH) and non-deposit (cluster AL) targets.

3.4. Datasets and Processing

In step three (S3), four datasets (datasets 1, 2 and 3) were constructed. First, dataset 1 was used for the spatial aggregation statistical analysis (see Section 3.3). Based on the spatial aggregation statistical analysis results (expressed in the CoType field), the new dataset 2 was constructed. Prospectivity of placer deposits were weighted from 1 to 4. Value 1 was assigned to the non-prospective mineral target type (AL); value 2 to the prospective target type (LL); value 3 to the prospective target-type (LH), and value 4 to the prospective target type (HH/HL). Dataset 2 was composed of 2920 prospective LL, LH, HH/HL samples and 789 non-prospective AL samples.

Second, dataset 2 was randomly partitioned on two sub-datasets (called dataset 3 and dataset 4) for the learning phases. The learning dataset used for machine-learning algorithms is generally composed of 60%–80% initial data or observations [61–63]. In this study, training dataset 3 selected a ratio of 80% of dataset 2. Dataset 3 was composed of 172 non-prospective AL, 172 prospective HH/HL, and 90 prospective LH samples. The test dataset 4 was a ratio of 20% of dataset 2. These data were scrutinized to evaluate the robustness and accuracy of the output models in Section 6.

3.5. Evaluation of Predictive Models and Validation of Results

In step S5 of the methodological approach (Figure 3), several metrics were considered for evaluating the performance and robustness of the various SVM, KNN and RF predictive models, and then to evaluate the sensitivity of prospectivity maps generated with the best performing models (see Section 6.4). These metrics are cross-validation accuracy (CVA) calculated with Equation (5), sensitivity (true positive rate calculated with Equation (1)), specificity (true negative rate calculated with Equation (2)), and AUC (Area Under Curve) values based on ROC (Receiver Operating Characteristic) curves. Positive and negative predictive values (calculated with Equations (3) and (4)) were calculated from the true positive and the true negative rates in confusion matrices, as well as F-scores (calculated with Equation (6)). The F-score metric was defined as the harmonic mean of the model's precision and recall (i.e., sensitivity). Details of the different metrics and their applications can be found in the literature (e.g., [64]).

$$\text{Sensitivity (or true positive rate)} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Specificity (or true negative rate)} = \frac{TN}{TN + FP} \quad (2)$$

$$\text{Positive predictive value} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Negative predictive value} = \frac{TN}{TN + FN} \quad (4)$$

$$\text{Crossvalidationaccuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$F - \text{score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (6)$$

where:

TP: the rate (expressed as a percentage) of true positives (refers to well-ranked samples).

FN: number of false negatives (refers to misclassified positive samples).

FP: number of false positives (refers to misclassified negative samples).

TN: the rate (expressed as a percentage) of true negatives (refers to well-ranked negative samples).

Confusion matrices were subsequently constructed and used to calculate and compare the values of the evaluation metrics during training phases. For each category of machine-learning algorithms used in this study, the most precise ones were selected. Consequently, the three best models were used for the independent validation and the robustness-testing phase with dataset 4.

3.6. Prospectivity Mapping Process

GIS-based predictive modelling of mineral prospectivity is very practical and is being used increasingly to delineate repeatable mineral exploration targets [42]. A grid of 71,730 cells with an elementary cell size of 25 m² covering all the studied area was constructed. First, we superimposed the grid with the six evidence criteria maps (see Section 3.2) and we estimated the centroid point of each cell in GIS (ArcGIS v.10.2). By considering each pit emplacement as the centroid of the grid cell, we extracted the corresponding hydrological, geomorphical, and structural predictors parameter values into a new table. This table, called dataset 5, was used with the final KNN, SVM and RF models to predict prospectivity of the new prospective and non-prospective areas in the H-L Complex. Results are the model responses or the predicted coded values estimated by the final KNN, SVM and RF models. Finally, the model responses (predicted coded value) at each centroid point were rasterized using the most-frequent-value assignment rule. The most common coded value of the mineral target among neighboring cells was assigned in priority. Finally, color coding was applied to each grid cell to represent graphically the Nb-Ta placer deposit prospectivity. The result is the prospectivity map of placer deposits in the northern region of the H-L Geological Complex (see Section 6.4).

In MPM, the locations of known deposits and non-deposit locations can be used to assess the sensitivity of mineral prospecting models [19,65,66]. The locations of exploration targets generated by a MPM approach should be as coincidental as possible with non-deposit locations where mineral deposits are least likely to be present, due to inadequate geological parameters and lack of exploration indicators [35,65,67]. Several approaches were used to assess the sensitivity of mineral prospectivity models. Bonham-Carter et al., 1994 [68] applied the weights of classes of spatial values divided by their corresponding occupied area (the area occupied by each class of evidential values) to estimate the probability of discovering mineral deposits in the classes. Yousefi and Carranza, 2015a, 2016 [69,70] developed a Prediction-Area (P-A) plot through which the percentage of known deposits anticipated by prospectivity classes (prediction rate) and the occupied areas of the corresponding prospectivity classes are used to quantify the relative importance of different prospectivity models. By developing the P-A plot, both the prediction rate and the occupied area of exploration targets contribute to the evaluation of prospectivity models [71–77].

In this study, we used several known deposit and non-deposit cells to assess the sensitivity of predicted values by KNN, SVM and RF final models. The reference deposits and non-deposit areas were composed by each cluster and outlier point (see Figure 7) that was rasterized to a 25 m² cell. The reference map comprised placer deposit and non-deposit cells with 950 AL non-prospective cells (1.4 km²) and 2972 (LL, LH and HH/HL) prospective cells (5.8 km²) of the Bemadi, Brokoua and Makua sites. Then, the generated cells of prospective and non-prospective areas of the Bemadi, Brokoua and Makua sites were

superimposed with the prospectivity maps generated by best-performing models. Finally, the known prospective and non-prospective reference classes areas (in km²) corresponding to the new prospective and non-prospective classes areas predicted by prospectivity models were extracted and calculated. Then, confusion matrices were constructed and used to quantify the relative importance of the different prospectivity models. Thus, if two different prospectivity models delineated exploration targets in different occupied areas, but with the same prospectivity score, the performance of the prospectivity model with the smaller target areas is higher than that of the model with larger target areas [78]. Results of sensitivity analysis of prospectivity models are presented in Section 6.4.

4. Machine-Learning Algorithms Selection

Indeed, automatic classification algorithms (SVM, KNN and RF) require less learning data (thereby increasing the degree of classifier automation) while making them more easily configurable [79]. Their use is more valuable than traditional statistical techniques. In addition, the SVM automatic classification algorithms integrate during their learning phase, an estimate of the complexity of the data to limit the phenomenon of over-adjustment and to find a compromise between the adequacy of the data and the complexity of the data to be generalized.

Decision trees (such as RF) use a hierarchical tree algorithm to make classification decisions. The advantage is that the tree structure is transparent, which in comparison with ANN is easier to interpret [1]. The automatic RF classification method combines the performance of several decision-tree algorithms to classify or predict the value of a target variable [1,44]. To avoid the correlation of different trees, RF increases the diversity of trees by growing them from different subsets of learning data, created through the “bagging” procedure. This process provides greater stability and robustness to the output model, with slight variations in the input data, while increasing the accuracy of the prediction [44].

KNN is a simple algorithm developed by Fix and Hodges in 1951 [80]. This machine-learning algorithm stores all available cases and classifies new cases according to a measure of similarity (weight or distance). This method is a simple but effective way to classify new items. It is one of the simplest and most intuitive techniques commonly used in the field of statistical discrimination [7].

5. Training of SVM and Other Models

At the step S4, training SVM and other models were conducted using a ‘methods and models’ learning approach. All processing of classification approaches was performed using Matlab Version R2019a software. The tool provides a variety of supervised classification techniques that can be used to generate multiple output KNN, SVM and RF models by optimizing hyperparameters and testing different classification methods. The software can tune output models by selecting different advanced options. Some of these options are internal parameters or hyperparameters of the model that can strongly affect its performance. For a given model type, the software tries different combinations of hyperparameter values by using an optimization scheme that seeks to minimize model classification error and returns a model with optimized hyperparameters.

In addition, software (Matlab Version R2019a) used in this study allows the selection of cross-validation options during the training phase. Applying cross-validation methods also helps with choosing the best performing model by calculating the error using a testing dataset which has not been used for training. The testing dataset helps to calculate the accuracy of the model and how it will generalize with future data. In this study a k-fold cross-validation method with $k = 5$ (given the small size of the learning dataset 3) was used. The k-fold cross-validation method partitions data into k randomly chosen subsets (or folds) of roughly equal size. During the learning process, the tool uses one subset to train, and the model trained is validated with the remaining subsets. For each round, « x » samples are selected to constitute the learning sample. The remaining $k-x$ samples are used to evaluate the performance of the model. The process is repeated k times, such

that each subset is used exactly once for validation. To build the following model, samples were selected differently so that learning and validation samples were never the same. The average error across all k partitions was used to evaluate the accuracy and performance of the learning model.

Determining the host parameters to be used for the training of automatic classification algorithms is a crucial step in a predictive modeling process. This step involves developing a key classification procedure to obtain reliable predictions. However, it is difficult to a priori specify appropriate configuration procedures with the desired precision, as there are fewer rules for determining optimal host parameters in the real world [42]. Applying the grid search method, which consists of testing the hyperparameters of a model by simply crossing each hypothesis in order to create a model for each combination of hyperparameters, SVM, KNN and RF were trained in this study (interactively) by optimizing and testing different classification methods. For each hyperparameter, a set of values were tested. Table 3 presents methods used to form all models tested in this study.

Table 3. Learning hyperparameters and methods used to train SVM, KNN and RF machine-learning algorithms.

Learning Hyperparameters to Train SVM Machine-Learning Algorithms					
Two multi-class classification strategies tested: One-vs.-All (OvA) and One-vs.-One (OvO)					
Value of the box constraint level parameter used: 1					
Model Category	Type of Output Model	Detection Function Parameter	Value of Detection Parameter	Kernel Scale Parameter	Value of Kernel Scale Parameter
SVM	Linear SVM	Linear	1	-	-
	Quadratic Core SVM	Quadratic	1	-	-
	Cubic Core SVM	Cubic	1	-	-
	Fine Gaussian Kernel SVM	Gaussian	0	Fine	61
	Medium Gaussian Kernel SVM	Gaussian	2	Medium	4
	Coarse Gaussian Core SVM	Gaussian	9	Coarse	8
Learning Parameters Used to Train KNN Machine-Learning Algorithms					
		Separation Parameter	Number of Neighbors	Metric Distance Parameter	Weighting Parameter
KNN	Fine KNN	Fine	1	-	-
	Median KNN	Medium	10	-	-
	Coarse KNN	Coarse	100	Coarse	-
	Cosine KNN	Medium	10	Cosine	-
	Cubic KNN	Medium	10	Cubic (Minkowski)	-
	Equal-weighted KNN		1	Weighted	Equal
	Euclidean-weighted KNN		10	Weighted	Euclidean
	Inverse distance weighted KNN		100	Weighted	Inverse distance

Table 3. Cont.

Learning Hyperparameters to Train SVM Machine-Learning Algorithms				
Two multi-class classification strategies tested: One-vs.-All (OvA) and One-vs.-One (OvO)				
Value of the box constraint level parameter used: 1				
	Inverse squared distance weighted KNN	-	100	Weighted Inverse square distance
Learning Hyperparameters Used to Train RF Machine-Learning Algorithms				
Assembly method used: Bagging trees; range of maximum number of splits tested: 10, 20 and 30; range of numbers of learners tested: 30–60–100 and 500; learning rate values used: 0.1 and 1; sub-space dimension value used: 1				
Random Forest	RF-10 splits and 30 learners	RF-20 splits and 60 learners		
	RF-10 splits and 60 learners	RF-20 splits and 100 learners		
	RF-10 splits and 500 learners	RF-20 splits and 500 learners		
	RF-30 splits and 500 learners	RF-30 splits and 500 learners		
	RF-30 splits and 100 learners	RF-30 splits and 60 learners		
	RF-10 splits and 100 learners	RF-30 splits and 30 learners		
	RF-20 splits and 30 learners	-		

Since the range of values of raw data varies widely, in some machine-learning algorithms, objective functions will not work properly without normalization. For example, many classifiers calculate the distance between two points by the Euclidean distance. If one of the features has a broad range of values, the distance will be governed by this feature. Therefore, the range of all features should be normalized so that each feature contributes approximately proportionately to the final distance. All training procedures were performed within Matlab Version R2019a, with a standardized data option set “True”. Data normalization and standardization is a good practice. It normalizes each column of an array using standard scores or the feature-scaling normalization procedure [81].

For optimizing hyperparameters, the Matlab software used provides pre-set SVM models with kernel-scale advanced options and a manual kernel scale mode. SVM performance greatly depends on the kernel functions considered. Consequently, several activation functions were tested in this study, including linear, quadratic, cubic, and Gaussian (fine, medium, and coarse) functions. Box constraint level parameter value used for all multiclass classification approaches was fixed at 1; kernel scale parameter was varied among positive values log-scaled at 4, 8 and 61; and value of detection parameter was varied among positive values log-scaled at 0, 1, 2 and 9.

To optimize RF hyperparameters in this study, the maximum number of split parameters were tested among integers log-scaled of 10, 20 and 30; the number of learner parameter was tested among integers log-scaled of 30, 60, 100 and 500, and learning rate parameters were tested with a real value log-scaled of 1.

To optimize KNN hyperparameters in this study, a number of neighbor parameters were tested among integers log-scaled of 1, 10 and 100. Distance metric parameters were tested among Euclidean, Minkowski (cubic) and cosine to determine the distance to points. Distance weight parameters was tested among equal-weighted, Euclidean-weighted, and inverse squared distance weighted.

6. Results

6.1. Discrimination of Mineral Targets

Results of the spatial autocorrelation analysis shown in Table 4 reveals that the geostatistical distribution of the 3709 Nb-Ta mineralized gravel samples was not the result of a random process (Table 4).

Table 4. Results of Moran's statistical analysis.

CO Type	Number of Points	Nb-Ta Grade Range (g/m ³)	Spatial Cluster/Outlier Nb-Ta Grade (g/m ³)	Min/Max Moran's I Index	Min/Max <i>p</i> -Value	Min/Max Number of Neighbors
HH	767	45 to 1157	136.61	0.0005/14.35	0.002/ 0.038	7/554
HL	98	45 to 821	74.08	−1.65/0	0.002/ 0.034	33/442
LH	513	0 to 44	23.5	−0.59/0	0.002/ 0.038	9/546
LL	1542	0 to 44	5.63	0/0.43	0.002/ 0.038	13/459
AL	789	0 to 1126	40.98	−0.08/2.34	0.04/0.498	1/545
Total	3709	-	-	-	-	-

Note: COType (Cluster/Outlier) type codes are described in the text, while *p*-values in boldface type are the maximum pseudo-probability estimates that do not exceed the *p* = 0.04 threshold.

Figure 8 depicts the standard normal distribution of the pseudo-probability values (*p*) against their associated Z-scores. Here, 2920 samples (i.e., 79% of the surveyed placer deposits) can be grouped into five types of spatial clusters and outliers (or mineral target classes): High-High (HH), color-coded red; High-Low (HL), color-coded orange; Low-Low (LL), color-coded green; and Low-High (LH), color-coded blue. The estimated *p*-values of the 2920 samples range from 0.002 to 0.038 and, therefore, lie below the *p* = 0.04 threshold. The relevant features exhibit statistically significant spatial aggregation at 96%. The remaining samples (789 samples i.e., 21% of the surveyed placer deposits), having rejected this hypothesis (*p* > 0.04), are grouped into their own cluster in this study, which is referred to as AL (color-coded black). This cluster does not show statistically significant spatial aggregation and is characterized by a random distribution pattern.

Analysis of the distributions of mean grade of Nb-Ta values for the output cluster types indicates that a 44 g/m³ threshold can discriminate between different types of deposits or mineral targets. In contrast, spatial variation in Nb-Ta grades for samples in Class AL (0–1126 g/m³) shows very little similarity among centroids that are in proximity. Class AL represents areas where the probability of finding an HH, HL, LL or LH placer deposits is <4% and, therefore, could be considered negligible. It is thus a class that defines areas of non-prospectivity relative to other deposit types.

The LL cluster type represents a zone of accumulation of mineralized gravels with low Nb-Ta grades (<44 g/m³: averaging 5.63 g/m³). The LL cluster therefore corresponds to prospective areas of low grade. Type LH represents an accumulation zone of medium-grade mineralized gravels (<44 g/m³: averaging 23.5 g/m³), i.e., prospective medium-grade zones. This type is located at the edge of HH deposits, which represent very high-grade mineralized gravel accumulation zones (>44 g/m³: averaging 136.61 g/m³). Finally, the HL outlier cluster type represents a zone of high-grade mineralized gravel accumulations (>44 g/m³: averaging 74.08 g/m³), which are generally surrounded by LL type deposits.

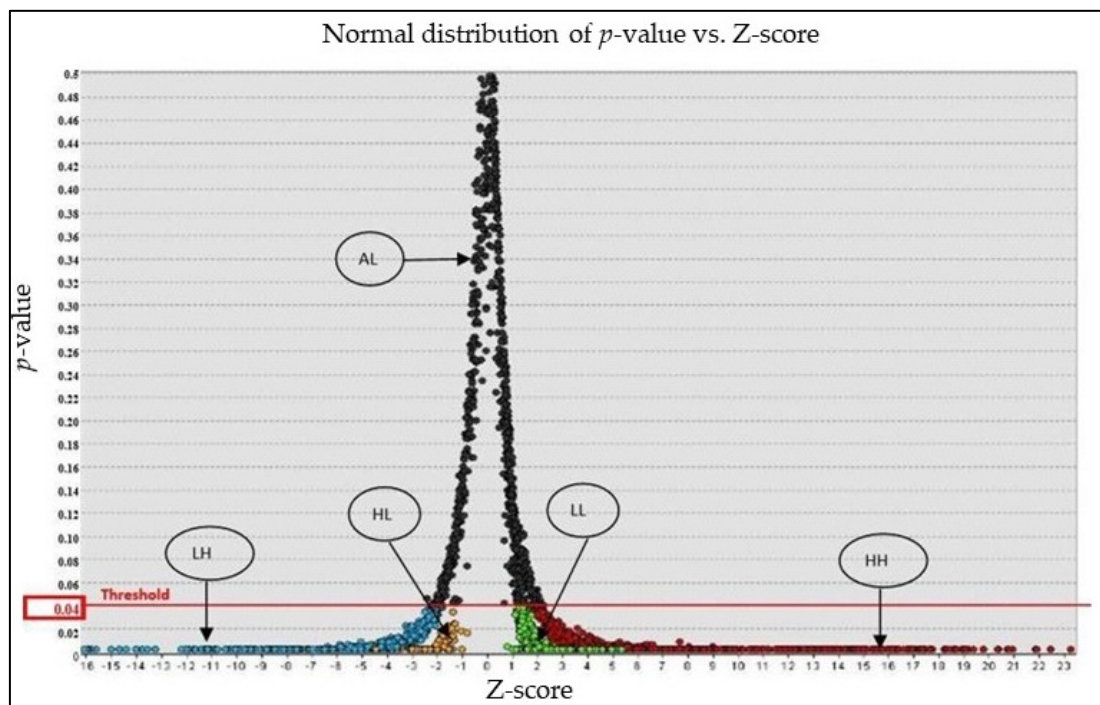


Figure 8. Spatial autocorrelation curve (p -value vs. Z -score) of Nb-Ta mineralized gravel samples (AL: black dots; LH: blue dots; HL: orange dots; LL: green dots, and HH: red dots).

6.2. Machine-Learning Training Phases Analysis

As indicated in Section 5, the SVM, KNN and RF approaches were used to develop predictive models for Nb-Ta placer deposits within the H-L Geological Complex. The results that were obtained during the training phases are presented here.

A compilation of results (Table 5) shows that regardless of the training method that was used, the linear, quadratic and cubic function SVMs, as well as the medium and coarse Gaussian kernel SVMs, produced CVAs results of 52% or less. Only the SVM algorithm with a fine Gaussian kernel produced better results among the SVM models that were tested. CVA that was obtained with the SVM model with a fine Gaussian kernel was 61.9% with the OvO strategy, and 62.4% with the OvA. Moreover, this model had the highest AUC values (Table 5) and managed to correctly predict 75% of the observations in Class 3 (LH), 78% in Class 4 (HH/HL), 79% in Class 1 (AL), and 85% in Class 2 (LL).

The results obtained with the KNN model are shown in Table 6. The three coarse KNN models are clearly the least accurate in the series, with CVAs below 46% and lowest AUC values (63% to 67%). On the one hand, the various KNN algorithms (medium, coarse, cosine, cubic, weighted-Euclidean) display poor accuracy results (CVA <51%). On the other, the fine KNN models with cubic, Euclidean or cosine metric distances show clear improvement in results, with CVAs varying between 62% and 64%. Without question, the best performance is obtained with the KNN approach when weighting is based on inverse distance (IDW) or inverse distance squared (IDW2). In these cases, CVA reaches just over 65%, and AUC values (between 75% and 89%) are the best of those obtained for each of the four classes. As shown in Table 6, the IDW- or IDW2-weighted Euclidean KNN models predict 82% of the observations in Class 1 (AL), 89% in Class 2 (LL), 75% in Class 3 (LH), and 78% in Class 4 (HH/HL).

Table 5. Results of the automatic classifications with SVMs. The best performing model that was selected for the SVM category is shown in bold type.

Name of Model	Multi-Class Classification Strategy Used	Value of Detection Parameter	Value of Kernel Scale Parameter	CVA (%)	AUC Value (%)			
					Class1 (AL)	Class 2 (LL)	Class 3 (LH)	Class 4 (HH/HL)
Linear kernel SVM	One vs. One	1	-	41.6	63	58	51	55
	One vs. All	1	-	39.8	58	63	58	54
Quadratic kernel SVM	One vs. One	1	-	45.2	67	66	62	66
	One vs. All	1	-	42	55	67	63	60
Cubic kernel SVM	One vs. One	1	-	52	71	74	70	71
	One vs. All	1	-	41.8	65	64	63	66
Fine Gaussian kernel SVM	One vs. One	0	61	61.9	79	85	75	78
	One vs. All	0	61	62.4	79	85	75	78
Medium Gaussian kernel SVM	One vs. One	2	4	50.5	70	73	65	69
	One vs. All	2	4	49.9	63	74	68	68
Coarse Gaussian kernel SVM	One vs. One	9	8	41.7	65	65	59	58
	One vs. All	9	8	42.7	59	65	59	56

Table 6. Results of automatic classifications with KNN models. The best performing model that was selected for the KNN category is shown in bold type.

Name of Model	Metric Distance Type	Distance Weighting Type	Number of Neighbors	CVA (%)	AUC Value (%)			
					Class 1 (AL)	Class 2 (LL)	Class 3 (LH)	Class 4 (HH/HL)
Fine KNN	Euclidean	Equal	1	63.7	73	81	67	73
	Cubic	Equal	1	63.1	73	80	66	73
	Cosine	Equal	1	62.9	73	80	67	72
Medium KNN	Euclidean	Equal	10	50.5	71	76	69	72
	Cubic	Equal	10	49.7	71	75	69	72
	Cosine	Equal	10	49.1	70	75	69	73
Coarse KNN	Euclidean	Equal	100	45.9	66	67	64	64
	Cubic	Equal	100	45.8	65	66	64	64
	Cosine	Equal	100	45.9	65	67	63	64
Cosine KNN	Euclidean	Equal	10	50.5	71	76	69	72
	Cosine	Equal	10	49.1	70	75	69	73
Cubic KNN	Euclidean	Equal	10	50.5	71	76	69	72
	Cubic	Equal	10	49.7	71	75	69	72
Weighted-Euclidean KNN	Euclidean	Equal	10	50.5	71	76	69	72
IDW-weighted KNN	Euclidean	Inverse distance	10	65.3	82	89	75	78
IDW2-weighted KNN	Euclidean	Inverse distance-squared	10	65.3	82	89	75	78

The last model category to be considered is the RF. Results are summarized in Table 7. CVAs that were obtained vary between 67% and 69%. There are few significant differences in the results that were obtained. For example, the RF model (30 splits (sp)/100 learners (L)) can predict 87% of Class 1 (AL) mineral targets compared to 88% for the RF model (30 sp/500 L). The two models correctly predict 91% and 81% of Class 2 (LL) and four (HH/HL) observations. The RF model (30 sp/100 L) predicts 82% of the observations in Class 3 (LH) compared to 81% for the competing model. Given the importance that is attributed to Classes 2, 3 and 4 in this study, RF model (30 sp/100 L) has been selected for further study in this category.

Table 7. Results of automatic classifications with RF. The best performing model that was selected for the RF category is shown in bold type.

Name of Model	Maximum Division	Number of Learners	Learning Rate	CVA (%)	AUC Value (%)			
					Class1 (AL)	Class 2 (LL)	Class 3 (LH)	Class 4 (HH/HL)
RF (10sp/30L)	10	30	0.1	67.3	87	91	80	81
RF (10sp/60L)	10	60	0.1	68.2	87	91	81	80
RF (10sp/100L)	10	100	0.1	68.3	87	91	81	81
RF (10sp/500L)	10	500	0.1	68.4	88	91	82	81
RF (20sp/30L)	20	30	0.1	67.2	86	90	80	80
RF (20sp/60L)	20	60	0.1	67.5	87	91	80	81
RF (20sp/100L)	20	100	0.1	68.3	87	91	81	81
RF (20sp/500L)	20	500	0.1	68.4	88	91	81	81
RF (30sp/30L)	30	30	0.1	67.3	86	90	80	80
RF (30sp/60L)	30	60	0.1	68.1	87	91	81	81
RF (30sp/100L)	30	100	0.1	69	87	91	82	81
RF (30sp/500L)	30	500	0.1	68.5	87	91	81	81

Figure 9 compares the performance of the three best performing models during the training phase: the fine Gaussian kernel SVM model, the IDW-weighted KNN model, and the RF model. It includes CVA and AUC values (Figure 9a), together with calculated F-scores (Figure 9b).

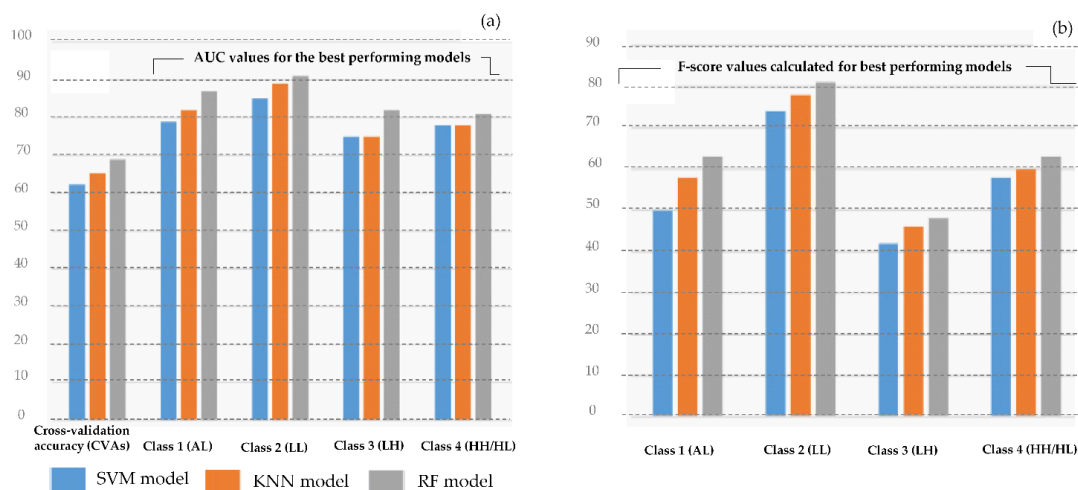


Figure 9. (a): CVA and AUC values of the selected SVM, KNN and RF models; (b): F- scores of the different models for the four classes.

Overall, predictive abilities of the three models are comparable. They are higher for Class 2, and decrease slightly for Classes 1, 3 and 4. The different indicators (CVA, AUC, and F-score values) indicate that the KNN model is slightly better than the SVM model, but the differences appear negligible. The RF model is systematically superior to SVM and KNN, particularly for Classes 1, 2 and 4. The calculated F-scores (Figure 9b) confirm the slightly higher performance of the RF model in this study, compared to SVM and KNN. Most importantly, each of the three models could be used to predict mineral targets, given the minimal differences between them.

6.3. Evaluation of Robustness of Output Models

In step S5 of the methodological approach (see Figure 3), several metrics were considered for evaluating the performance and robustness of the different SVM, KNN and RF classifiers (see Section 3.5 for details). Good performance using training dataset 3 does not guarantee that the model is sufficiently robust and that it is able to perform on new datasets against which it has not been previously trained. Therefore, the performance and robustness of the output SVM, KNN and RF, those identified as best for each category, were validated using the independent validation dataset 4 (composed with 20% of the original dataset 2). The resulting confusion matrices for each category (KNN, SVM, and RF) were analyzed considering CVA, AUC values and Kappa coefficients. Kappa values were interpreted according to the scale proposed by Landis and Koch [82]. Table 8 shows confusion matrices that were obtained during the independent validation phase.

Independent validation confirms the results that were obtained during the learning test phase. The best predictions are provided by the selected RF algorithm, with $\kappa = 0.56$ and CVA = 69%. These metrics are slightly higher than those obtained with the KNN model, which yields $\kappa = 0.54$, and similar CVA = 0.68. The SVM model provides slightly inferior results (CVA = 63%, $\kappa = 0.46$). Calculated kappa for the RF predictor model is very close to 0.6, indicating strong agreement on the scale created by Landis and Koch [82]. KNN and SVM models attain kappa values between 0.4 and 0.6, indicating moderate agreement on this same scale.

Table 8. Confusion matrices that were obtained during the independent validation phase.

Fine Gaussian Kernel SVM Confusion Matrix (Values are in %)				
	Class 1 (AL)	Class 2 (LL)	Class 3 (LH)	Class 4 (HH/HL)
Class 1 (AL)	48.84	34.88	2.91	13.37
Class 2 (LL)	6.82	86.04	1.30	5.84
Class 3 (LH)	10.00	34.44	23.33	32.22
Class 4 (HH/HL)	5.23	25.00	11.63	58.14
Number of samples for validation = 742; CVA = 63.34%; Kappa Coefficient = 0.46				
IDW-weighted KNN Confusion matrix (values are in %)				
	Class 1 (AL)	Class 2 (LL)	Class 3 (LH)	Class 4 (HH/HL)
Class 1 (AL)	60.47	20.93	9.3	9.3
Class 2 (LL)	7.14	83.12	1.62	8.12
Class 3 (LH)	12.22	16.67	38.89	32.22
Class 4 (HH/HL)	4.65	12.79	19.19	63.37

Table 8. Cont.

Fine Gaussian Kernel SVM Confusion Matrix (Values are in %)				
Number of samples for validation = 742; CVA = 67.92%; Kappa Coefficient= 0.54				
RF (30sp/100L) Confusion matrix (values are in %)				
	Class 1 (AL)	Class 2 (LL)	Class 3 (LH)	Class 4 (HH/HL)
Class 1 (AL)	63.37	23.26	5.23	8.14
Class 2 (LL)	6.82	86.04	1.30	5.84
Class 3 (LH)	10.00	20.00	36.67	33.33
Class 4 (HH/HL)	5.23	16.86	15.12	62.79
Number of samples for validation = 742; CVA = 69.41%; Kappa Coefficient = 0.56				

6.4. Evaluation of Sensitivity of the Prospectivity Maps Generated

The three predictive models that were retained following the learning and the robustness test phases provide encouraging results. However, KNN and RF are both predictive models that were used to map (following the approach described in Section 3.6) the prospectivity of new non-prospective and prospective exploration targets within the H-L Geological Complex, with a precision of 68% and 69%, respectively. Figure 10a,b illustrates the results.

The light grey areas, which are potentially non-prospective, are of little interest in terms of prospecting for new economically viable mineral targets. In contrast, two categories of potentially favorable zones for the concentration of economically valuable mineralized gravels are highlighted. The first are areas that are favorable for the concentration of medium-grade (orange) and high-grade (red) gravels. These areas correspond mainly to interfluvial and river or marshland edges, but they also include areas located near rocky outcrops of mineralized granitic pegmatites. They occupy elevated topographic positions and correspond to areas that are prone to poor drainage. These zones are marked by medium to high-fracture density. The second are zones that are potentially favorable to the concentration of low-grade mineralized gravels (in green). These are zones occupying lower topographic positions, subject to considerable drainage processes, and characterized by a lower fracturing density. They are very extensive and occupy most of the flats in the sub-watersheds under study. Table 9 shows confusion matrices used to evaluate the sensitivity of both KNN and RF prospectivity models.

As observed, the performance of both prospectivity models is very similar. RF and KNN prospectivity models are more sensitive to the wide LL areas (i.e., 3.12 km²) with low-grade deposits at 84% and 77%, respectively. RF and KNN prospectivity models are more specific to the smallest LH and HH/HL prospective areas (i.e., 0.9. to 1.81 km²) with medium-grade and high-grade deposits at 94% and 93%, respectively. Comparatively, the calculated metrics (accuracy, sensitivity, specificity, and F-score) confirm the slightly higher performance of the RF model in this study, compared to KNN.

Result of the sensitivity analysis shows that in the context of a lack of exploration data with complex and non-linear spatial relationships, between spatial evidence criteria and known mineral targets, an MPM approach—based on knowledge of the mineral system and geostatistical analysis of soil geochemistry data (acquired from the field), and using the predictive power of supervised machine-learning algorithms (e.g., SVM, KNN, RF)—is a valid approach to model at a satisfactory rate a set of exploration criteria.

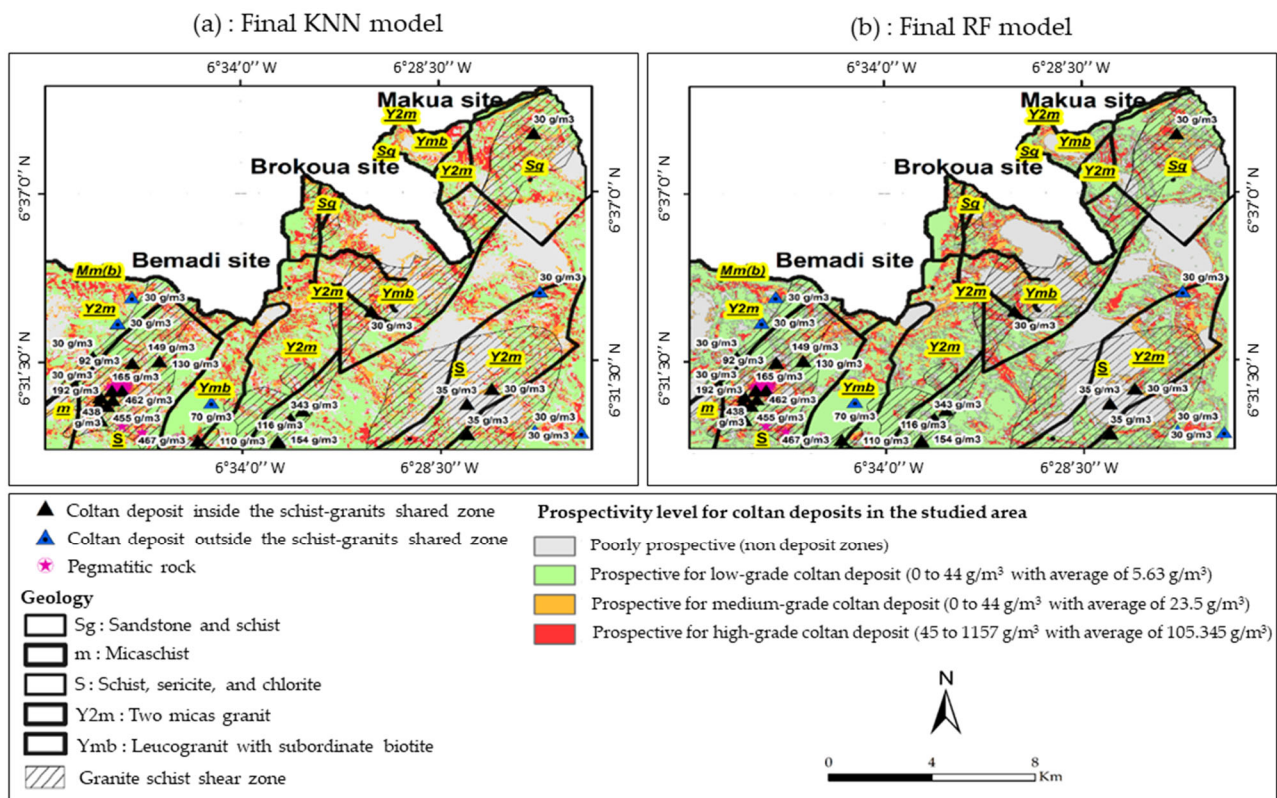


Figure 10. Geology and prospectivity map for Nb-Ta placer deposits generated with the final (a) IDW-weighted KNN model and (b) RF (30 sp/100L) model (blue pyramids are known occurrences outside the granite schist shear zone, and black pyramids are those inside the shear zone).

Table 9. Sensibility and specificity results of RF and KNN prospectivity models. The best performing metric values that were calculated per class is shown in bold type.

Confusion Matrix of the RF Model Prospectivity MAP (Occupied Area)					
Value is km ²		Predicted value from RF model prospectivity map			
		AL	LL	LH	HH/HL
Known Deposit value	AL	0.80	0.40	0.09	0.12
	LL	0.22	2.62	0.07	0.21
	LH	0.13	0.21	0.34	0.23
	HH/HL	0.16	0.48	0.25	0.91
Total		1.31	3.71	0.75	1.47
Confusion Matrix of the KNN Model Prospectivity MAP (Occupied Area)					
Value is km ²		Predicted value from KNN model prospectivity map			
		AL	LL	LH	HH/HL
Known Deposit value	AL	0.70	0.41	0.12	0.17
	LL	0.26	2.40	0.10	0.36
	LH	0.15	0.20	0.30	0.25
	HH/HL	0.20	0.45	0.24	0.92
Total		1.31	3.46	0.76	1.70

Table 9. Cont.

Metrics of the RF prospectivity map (value is in percent)				
Deposit type	Accuracy	Sensitivity	Specificity	F-score
AL	61.07	57.06	91.25	59.00
LL	70.62	83.97	73.49	76.72
LH	45.52	37.36	93.56	41.04
HH/HL	62.12	50.56	89.78	55.74
Metrics of the KNN prospectivity map (value is in percent)				
Deposit type	Accuracy	Sensitivity	Specificity	F-score
AL	53.44	50.00	89.54	51.66
LL	69.28	77.02	74.14	72.95
LH	39.68	33.19	92.79	36.14
HH/HL	54.12	50.83	85.61	52.42

7. Discussion

This study is one of the few mineral prospectivity studies of Nb-Ta placer deposits in West Africa that has been based on spatial analysis and machine-learning methods. The proposed approach lies in prospectivity modeling for placer deposits using geomorphic and landscape parameters. Moran's spatial aggregation statistical analysis technique allowed us to discern geostatistical distribution patterns of 3709 locations of Nb-Ta mineralized gravel samples. The geostatistical distribution is significant at the 96% level, which could distinguish (at a threshold of 44 g/m³) several classes of prospective and non-prospective mineral exploration targets. These classes reflect the numerical composition of the different placer deposit types which were identified in previous work in this study area [24–30].

Moran's spatial aggregation resulted in automatic definition of five clusters, some of which contained large numbers of samples (e.g., the AL class), while others were under-represented. Segmentation of the input dataset for learning process (following a random process) satisfies scientific rigor and guarantees a certain a posteriori robustness of the models, but it can generate asymmetric learning and validation data subsets. In the real world, and particularly in mineral prediction, such asymmetry in training data is not the exception.

The consequence is that it can affect the results of the classifications. In semi-supervised or supervised predictive classification approaches, standard classifiers (e.g., maximum likelihood) tend to classify majority classes better, while underestimating minority classes [14]. Over large geographical areas, mineralization may be considered a rare event, with low relative representation [14,83,84]. The results obtained by standard classifiers may be affected by uncertainties in the classification of data on rare events, since the decision limits of classifiers are largely biased towards the majority class and tend to ignore rare events [14]. Results obtained confirm that the SVM, KNN and RF models can handle data asymmetry and produce improved performance in predicting mineral prospectivity in new regions of interest [85–87]. To overcome some of the limitations and potentially improve the results, balance techniques, such as the synthetic minority oversampling technique (SMOTE) [88–90], could be investigated in forthcoming studies.

Observed differences in model results for the four mineral classes are not only from inherent properties of the models, but also from the skewed distribution of the training samples. Indeed, Class 1 comprised 23% of the training dataset, compared to 46% for Class 2, 16% for Class 3, and 26% for Class 4. Analysis of the different performance metrics (CVAs, AUCs, false positives, and false negatives, among others) shows that they are partly dependent on sampling. For example, all three models have high AUCs for Class 2, which contains 46% of the samples, and these values tend to decrease with the number of samples. Overall, the SVM and KNN models were apparently more sensitive to the structure of the training dataset than was the RF predictor model.

The novelty of this study lies in prospectivity modeling for Nb-Ta placer deposits using geomorphic and landscape parameters. Based on important mapping indicators or criteria and using a soil geochemical dataset to assess Nb-Ta sample locations and grades, this study used the predictive power of supervised automatic classification machine-learning algorithms to find predictive models complex enough to capture the nature of the exploration dataset. Results obtained were very close with the three best models (SVM, KNN and RF), especially during the learning phase. A priori, the spatial evidence criteria that were considered in this study as spatial predictors for Nb-Ta placer deposits are highly relevant (see Section 3.2). Results of independent validation confirm the performance of all three predictor models, especially for the final RF model and the final KNN model. The final RF model exhibits a CVA of 69% with Kappa equaling 0.6, and the final KNN model exhibits a CVA of 68% with Kappa equaling 0.6. According to the Landis and Koch [82] rating scale, agreement between predictions and field observations is strong. The final KNN model successfully predicts 67% of known prospective occurrences that were selected as control points, and the final RF model can predict, correctly, 79%. Based on those statistical results, MPM is indeed possible here with a good level of confidence.

The overall performance of the resulting final predictive models was assessed in both training and test datasets using a confusion matrix set of statistical measurements, AUC value, and F-score. The assessment results indicate that the three machine-learning models presented in this study achieved satisfactory performance levels characterized by high predictive accuracy. The accuracy rates that were obtained with the RF, KNN and SVM models in this study are very similar to those reported in previous MPM studies [1,10,42,86]. Most of these works attest that RF and KNN are sufficiently powerful and robust classifiers to predict mineral prospectivity at satisfactory rates. As demonstrated in this paper, performance of both prospectivity models is very similar. RF and KNN prospectivity models are more sensitive to the wide LL areas (i.e., 3.12 km²), with low-grade deposits at 84% and 77%, respectively. RF and KNN prospectivity models are more specific to the smallest LH and HH/HL prospective areas (i.e., 0.9. to 1.81 km²), with medium-grade and high-grade deposits at 94% and 93%, respectively. Comparatively, the calculated metrics (accuracy, sensitivity, specificity and F-score) confirm the slightly higher performance of the RF model in this study, compared to KNN.

The results show that the new prospective areas for medium- and high-grade Nb-Ta are located within interfluvial zones, along rivers or marshes (asymmetrically) and near rocky outcrops of mineralized pegmatitic granites. This means that the mineralized gravels have not been displaced over a long distance during successive runoff events and that they have accumulated progressively in the form of placers, preferentially in colluvium and eluvium, and within the vicinity of mineralization sources. This observation confirms results of a textural study of gravels in the Issia region [30]. Indeed, the current study showed that Nb-Ta minerals released on the crests and flanks of the interfluvial zones is generally large and angular in shape. In comparison, previous work [30] showed that at the level of the interfluvial zones, lateritic soils rich in gravels with high Nb-Ta mineral contents were developed. These high Nb-Ta grades are the consequence of the in situ or nearby degradation of a large quantity of mineralized pegmatites.

Topographic crevice zones are not very favorable for the weathering of primary host rocks or drainage of poorly eroded heavy minerals. Prospectivity maps generated shows that topographic crevice zones are narrow and elongated at the edges of rivers or marshes and a little more spread out at the level of the interfluvial zones. Topographic crevice zones are real natural traps of heavy minerals, which are fixed in space and time. Previous works [91] reported similar observations involving the formation of placer-type diamond deposits along the Orange River in South Africa. The study highlighted three types of natural traps that were formed by topographic irregularities, which are fixed in space and favorable to the accumulation (in the same place) of heavy diamond minerals of high economic value. Prospective areas with low-grade minerals also deserve further attention. Our results show that these

prospective areas occupy slightly lower topographic positions and correspond to the vast areas of lowlands and valleys that are subject to important drainage processes.

The results obtained in this study have demonstrated that, considering the landscape evidence criteria, the low potential to transport fragments and erode mineralized pegmatitic host rocks can be used not only as a spatial indicator for searching new zones favorable to concentrating these little-altered pegmatites, but also as a spatial criterion guiding the search for new deposits of medium and highly mineralized placers.

8. Conclusions

Results showed that the final RF prospectivity models generated with 30 splits and 100 learners, and the final model obtained with the KNN approach when weighting is based on inverse distance (IDW) or inverse distance-squared (IDW2) with 10 as number of neighbor parameters, emerged as the best performers, followed by the final SVM (fine Gaussian kernel). The RF model, with a Kappa coefficient of about 0.6, predicts mineral targets with rather strong agreement, while the other two models (Kappa about 0.5) do so with moderate agreement. Over a broad exploration area, the RF model and KNN approach has predicted and delimited areas of topographic crevices that are potentially favorable for medium- and high-grade Nb-Ta deposits of economic interest. Moreover, sensitivity analyses of both the final KNN and RF prospectivity maps confirm the slightly higher performance of the RF model. Final RF and KNN prospectivity models are both more sensitive for wide areas occupied by low-grade placer deposits at 84% and 77%, respectively. Conversely, they are more sensitive for the smallest prospective areas of economic interest with medium-grade and high-grade placer deposits at 94% and 93%, respectively. These results help to reduce interpretation bias by geologists in predictive MPM, thereby reducing exploration risk and costs.

Author Contributions: Conceptualization, K.O.S., K.G. and M.G.; methodology, K.O.S., K.G. and M.G.; software, K.O.S.; validation, K.O.S., K.G. and M.G.; formal analysis, K.O.S., K.G. and M.G.; investigation, K.O.S., K.G. and M.G.; resources, K.G. and M.G.; data curation, K.O.S., K.G. and M.G.; writing—original draft preparation, K.O.S.; writing—review and editing, K.G. and M.G.; visualization, K.O.S., K.G. and M.G.; supervision, K.G. and M.G.; project administration, K.G.; funding acquisition, K.G. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the financial support that was provided by the Programme Canadien de Bourses de la Francophonie (Government of Canada, Canadian Partnership Branch, Department of Foreign Affairs, Trade and Development) and Natural Sciences and Engineering Research Council of Canada (NSERC Discovery grant RGPIN-2018-06101).

Acknowledgments: We are grateful to the Société pour le Développement Minier de Côte d'Ivoire (SODEMI) for their continued support and geoscientific data that were provided for the research. We thank the many people who contributed to various aspects of the study, including data collection at the study site. W.F.J. Parsons for English translation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* **2015**, *71*, 804–818. [[CrossRef](#)]
2. Zuo, R.; et Carranza, E.J.M. Support vector machine: A tool for mapping mineral prospectivity. *Comput. Geosci. Geocomputation Miner. Explor. Targets* **2011**, *37*, 1967–1975. [[CrossRef](#)]
3. Abedi, M.; Norouzi, G.-H. Integration of various geophysical data with geological and geochemical data to determine additional drilling for copper exploration. *J. Appl. Geophys.* **2012**, *83*, 35–45. [[CrossRef](#)]
4. Yu, L.; Porwal, A.; Holden, E.-J.; Dentith, M.C. Towards automatic lithological classification from remote sensing data using support vector machines. *Comput. Geosci.* **2012**, *45*, 229–239. [[CrossRef](#)]
5. Et Granek, J.; Haber, E. Data mining for real mining: A robust algorithm for prospectivity mapping with uncertainties. In Proceedings of the SIAM International Conference on Data Mining 2015, SDM 2015, Vancouver, BC, Canada, 30 April–2 May 2015; pp. 145–153. [[CrossRef](#)]

6. Shabankareh, M.; Hezarkhani, A. Application of support vector machines for copper potential mapping in Kerman region, Iran. *J. Afr. Earth Sci.* **2017**, *128*, 116–126. [\[CrossRef\]](#)
7. Nguyen, T.H.; Jones, S.; Soto-Berelov, M.; Haywood, A.; Hislop, S. A Comparison of Imputation Approaches for Estimating Forest Biomass Using Landsat Time-Series and Inventory Data. *Remote Sens.* **2018**, *10*, 1825. [\[CrossRef\]](#)
8. Carranza, E.J.M.; Laborte, A.G. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Comput. Geosci.* **2015**, *74*, 60–70. [\[CrossRef\]](#)
9. Carranza, E.J.M.; Laborte, A.G. Data-Driven Predictive Modeling of Mineral Prospectivity Using Random Forests: A Case Study in Catanduanes Island (Philippines). *Nonrenewable Resour.* **2015**, *25*, 35–50. [\[CrossRef\]](#)
10. Hariharan, S.; Tirodkar, S.; Porwal, A.; Bhattacharya, A.; Joly, A. Random Forest-Based Prospectivity Modelling of Greenfield Terrains Using Sparse Deposit Data: An Example from the Tanami Region, Western Australia. *Nonrenewable Resour.* **2017**, *26*, 489–507. [\[CrossRef\]](#)
11. Harris, J.R.; Grunsky, E.; Behnia, P.; Corrigan, D. Data-and knowledge-driven mineral prospectivity maps for Canada's North. *Ore Geol. Rev.* **2015**, *71*, 788–803. [\[CrossRef\]](#)
12. Parsa, M. A data augmentation approach to XGboost-based mineral potential mapping: An example of carbonate-hosted ZnPb mineral systems of Western Iran. *J. Geochem. Explor.* **2021**, *228*, 106811. [\[CrossRef\]](#)
13. King, G.; Zeng, L.; Tomz, M. Logistic Regression in Rare Events Data. *J. Stat. Softw.* **2003**, *8*, 137–163. [\[CrossRef\]](#)
14. Xiong, Y.; Zuo, R. GIS-based rare events logistic regression for mineral prospectivity mapping. *Comput. Geosci.* **2018**, *111*, 18–25. [\[CrossRef\]](#)
15. Bolouki, S.M.; Ramazi, H.R.; Maghsoudi, A.; Beiranvand Pour, A.; Sohrabi, G. A Remote Sensing-Based Application of Bayesian Networks for Epithermal Gold Potential Mapping in Ahar-Arasbaran Area, NW Iran. *Remote Sens.* **2020**, *12*, 105. [\[CrossRef\]](#)
16. Noack, S.; Knobloch, A.; Etzold, S.H.; Barth, A.; Kallmeier, E. Spatial predictive mapping using artificial neural networks. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *XL-2*, 79–86. [\[CrossRef\]](#)
17. Zhao, J.; Chen, S.; et Zuo, R. Identifying geochemical anomalies associated with Au-Cu mineralization using multifractal and artificial neural network models in the Ningqiang district, Shaanxi, China. *J. Geochem. Explor.* **2016**, *164*, 54–64. [\[CrossRef\]](#)
18. Chen, Y.; Lu, L.; Li, X. Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. *J. Geochem. Explor.* **2014**, *140*, 56–63. [\[CrossRef\]](#)
19. Chen, Y. Mineral potential mapping with a restricted Boltzmann machine. *Ore Geol. Rev.* **2015**, *71*, 749–760. [\[CrossRef\]](#)
20. Chen, G.; Huang, N.; Wu, G.; Luo, L.; Wang, D.; Cheng, Q. Mineral prospectivity mapping based on wavelet neural network and Monte Carlo simulations in the Nanling W-Sn metallogenic province. *Ore Geol. Rev.* **2022**, *143*, 104–765. [\[CrossRef\]](#)
21. Parsa, M.; Carranza, E.J.M.; Ahmadi, B. Deep GMDH Neural Networks for Predictive Mapping of Mineral Prospectivity in Terrains Hosting Few but Large Mineral Deposits. *Nonrenewable Resour.* **2021**, *31*, 37–50. [\[CrossRef\]](#)
22. Abedi, M.; Norouzi, G.-H.; Bahroudi, A. Support vector machine for multi-classification of mineral prospectivity areas. *Comput. Geosci.* **2012**, *46*, 272–283. [\[CrossRef\]](#)
23. Asadi, H.H.; Porwal, A.; Fatehi, M.; Kianpouryan, S.; Lu, Y.-J. Exploration feature selection applied to hybrid data integration modeling: Targeting copper-gold potential in central Iran. *Ore Geol. Rev.* **2015**, *71*, 819–838. [\[CrossRef\]](#)
24. Rombach, M. *Rapport de Synthèse de la Minéralisation Alluvionnaire de la Région de SAIOUA (Issia)*; SODEMI: Abidjan, Côte d'Ivoire, 1960.
25. Cruys, H. *Prospection pour Columbo-Tantalite dans la Région d'Issia. Campagne Juillet 1963 Avril 1965*; SODEMI: Abidjan, Côte d'Ivoire, 1965.
26. Adam, H. *Les Pegmatites de la Région d'Issia*; SODEMI: Abidjan, Côte d'Ivoire, 1968.
27. Papon, A.; Lemarchand, R. *Geologie et Mineralisations du Sud-Ouest de la Cote-d'Ivoire; Synthèse des Travaux de l'Operation SASCA, 1962–1968*; B.R.G.M.: Paris, France, 1973.
28. Feybesse, J.-L.; Milési, J.P.; Ouédraogo, M.F.; Prost, A.E. La ceinture Protérozoïque inférieure de Boromo-Goren (Burkina Faso): Un exemple d'inférence entre deux phases transcurrentes éburnéennes. *C.R. Acad. Sci.* **1990**, *310*, 1353–1360.
29. Vidal, M.; Prost, A.E.; Aric, G.; Lemoine, S. Présence d'un socle antérieur à une suture océanique du Birimien inférieur en Côte d'Ivoire (afrique de l'Ouest). *C.R. Acad. Sci.* **1992**, *2*, 1085–1090.
30. Allou, A.B. *Facteurs, Paramètres, Dynamique de Distribution et Genèse des dépôts de Columbo-Tantalite d'Issia, Centre-Ouest de la Côte d'Ivoire*; Université du Québec à Chicoutimi: Montréal, QC, Canada, 2005.
31. Porwal, A.; González-Álvarez, I.; Markwitz, V.; McCuaig, T.; Mamuse, A. Weights-of-evidence and logistic regression modeling of magmatic nickel sulfide prospectivity in the Yilgarn Craton, Western Australia. *Ore Geol. Rev.* **2010**, *38*, 184–196. [\[CrossRef\]](#)
32. Joly, A.; Porwal, A.; McCuaig, T.C. Exploration targeting for orogenic gold deposits in the Granites-Tanami Orogen: Mineral system analysis, targeting model and prospectivity analysis. *Ore Geol. Rev.* **2012**, *48*, 349–383. [\[CrossRef\]](#)
33. Joly, A.; Porwal, A.; McCuaig, T.C.; Chudasama, B.; Dentith, M.C.; Aitken, A.R. Mineral systems approach applied to GIS-based 2D-prospectivity modelling of geological regions: Insights from Western Australia. *Ore Geol. Rev.* **2015**, *71*, 673–702. [\[CrossRef\]](#)
34. Bonham-Carter, G.F.; Agterberg, F.P. Application of a Microcomputer-Based Geographic Information System to Mineral-Potential Mapping. Geological Survey of Canada Contribution Number 47488. In *Microcomputer Applications in Geology 2*; Computers and Geology; Hanley, J.T., Merriam, D.F., Eds.; Pergamon: Amsterdam, The Netherlands, 1990; pp. 49–74. [\[CrossRef\]](#)
35. Carranza, E.; van Ruitenbeek, F.; Hecker, C.; van der Meijde, M.; van der Meer, F. Knowledge-guided data-driven evidential belief modeling of mineral prospectivity in Cabo de Gata, SE Spain. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 374–387. [\[CrossRef\]](#)

36. Parsa, M.; Lentz, D.R.; Walker, J.A. Predictive Modeling of Prospectivity for VHMS Mineral Deposits, Northeastern Bathurst Mining Camp, NB, Canada, Using an Ensemble Regularization Technique. *Nat. Resour. Res.* **2022**, 1–18. [\[CrossRef\]](#)
37. Carranza, E.J.M. Geocomputation of mineral exploration targets. *Comput. Geosci.* **2011**, *37*, 1907–1916. [\[CrossRef\]](#)
38. Porwal, A.; Carranza, E.J.M.; Hale, M. A Hybrid Neuro-Fuzzy Model for Mineral Potential Mapping. *Math. Geol.* **2004**, *36*, 803–826. [\[CrossRef\]](#)
39. Yousefi, M.; Carranza, E.J.M. Fuzzification of continuous-value spatial evidence for mineral prospectivity mapping. *Comput. Geosci.* **2015**, *74*, 97–109. [\[CrossRef\]](#)
40. Mccuaig, T.C. Sherlock, R.L. Exploration Targeting. *Proc. Explor.* **2017**, *17*, 75–82.
41. Kottek, M.; Grieser, J.; Beck, C.; Rudolf, B.; Rubel, F. World map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* **2006**, *15*, 259–263. [\[CrossRef\]](#)
42. Sun, T.; Chen, F.; Zhong, L.; Liu, W.; Wang, Y. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geol. Rev.* **2019**, *109*, 26–49. [\[CrossRef\]](#)
43. Sheykhou, M.; Mahdianpari, M.; Ghanbari, H.; Mohammadimanesh, F.; Ghamisi, P.; Homayouni, S. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6308–6325. [\[CrossRef\]](#)
44. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)
45. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [\[CrossRef\]](#)
46. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
47. Balabin, R.M.; Safieva, R.Z.; Lomakina, E.I. Near-infrared (NIR) spectroscopy for motor oil classification: From discriminant analysis to support vector machines. *Microchem. J.* **2011**, *98*, 121–128. [\[CrossRef\]](#)
48. Walsh, S.J.; Mynar, F. Landsat digital enhancements for lineament detection. *Environ. Earth Sci.* **1986**, *8*, 123–128. [\[CrossRef\]](#)
49. Paganelli, F.; Grunsky, E.C.; Richards, J.P.; Pryde, R. Use of RADARSAT-1 principal component imagery for structural mapping: A case study in the Buffalo Head Hills area, northern central Alberta, Canada. *Can. J. Remote Sens.* **2003**, *29*, 111–140. [\[CrossRef\]](#)
50. Adiri, Z.; El Harti, A.; Jellouli, A.; Lhissou, R.; Maacha, L.; Azmi, M.; Zouhair, M.; Bachaoui, E.M. Comparison of Landsat-8, ASTER and Sentinel 1 satellite remote sensing data in automatic lineaments extraction: A case study of Sidi Flah-Bouskour inlier, Moroccan Anti Atlas. *Adv. Space Res.* **2017**, *60*, 2355–2367. [\[CrossRef\]](#)
51. Corgne, S.; Magagi, R.; Yergeau, M.; Sylla, D. An integrated approach to hydro-geological lineament mapping of a semi-arid region of West Africa using Radarsat-1 and GIS. *Remote Sens. Environ.* **2010**, *114*, 1863–1875. [\[CrossRef\]](#)
52. Hashim, M.; Ahmad, S.; Md Johari, M.A.; Pour, A.B. Automatic lineament extraction in a heavily vegetated region using Landsat Enhanced Thematic Mapper (ETM+) imagery. *Adv. Space Res.* **2013**, *51*, 874–890. [\[CrossRef\]](#)
53. Robson, R. A multi-component rose diagram. *J. Struct. Geol.* **1994**, *16*, 1039–1040. [\[CrossRef\]](#)
54. Triboulet, C.; et Feybesse, J.-L. Les metabasites birimiennes et archéennes de la région de Toulepleu-Ity (Côte-d'Ivoire): Des roches portées à 8 kbar (≈ 24 km) et 14 kbar (≈ 42 km) au Paléoproterozoïque. *Comptes Rendus De L'Académie Des Sci.-Ser. IIA-Earth Planet. Sci.* **1998**, *327*, 61–66. [\[CrossRef\]](#)
55. Koudou, A.; Kouamé, F.K.; Youan Ta, M.; Saley, M.B.; Jourda, J.P. Contribution des données ETM+ de Landsat, de l'analyse multicritère et d'un SIG à l'identification de secteurs à potentialité aquifère en zone de socle du bassin versant du N'zi (Côte d'Ivoire). *GéoProdig, portail d'information géographique*. Available online: <http://geoprodig.cnrs.fr/items/show/48174> (accessed on 16 February 2022).
56. Koudou, A.; Assoma, T.V.; Adiaffi, B.; Ta, M.Y.; Kouame, K.F.; et Lasm, T. Analyses Statistique et Geostatistique de la Fracturation extraite de l'imagerie Asar Envisat du Sud-est de la Côte D'Ivoire. *Larhyss J.* **2014**, *11*, 147–166.
57. Nkono, C.; Féménias, O.; Lesne, A.; Mercier, J.-C.; Demaiffe, D. Fractal Analysis of Lineaments in Equatorial Africa: Insights on Lithospheric Structure. *Open J. Geol.* **2013**, *3*, 157–168. [\[CrossRef\]](#)
58. Saadi, N.; Zaher, M.A.; El-Baz, F.; Watanabe, K. Integrated remote sensing data utilization for investigating structural and tectonic history of the Ghadames Basin, Libya. *Int. J. Appl. Earth Obs. Geoinformation* **2011**, *13*, 778–791. [\[CrossRef\]](#)
59. Moran, P.A.P. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17–23. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Anselin, L. The Local Indicators of Spatial Association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [\[CrossRef\]](#)
61. Peters, J.; De Baets, B.; Verhoest, N.E.C.; Samson, R.; Degroove, S.; De Becker, P.; Huybrechts, W. Random forests as a tool for ecohydrological distribution modelling. *Ecol. Model.* **2007**, *207*, 304–318. [\[CrossRef\]](#)
62. Toth, R.; Ribault, J.; Gentile, J.; Sperling, D.; Madabhushi, A. Simultaneous segmentation of prostatic zones using Active Appearance Models with multiple coupled levelsets. *Comput. Vis. Image Underst.* **2013**, *117*, 1051–1060. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Chen, Y.; Wu, W. Mapping mineral prospectivity using an extreme learning machine regression. *Ore Geol. Rev.* **2017**, *80*, 200–213. [\[CrossRef\]](#)
64. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. 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* **2015**, *13*, 361–378. [\[CrossRef\]](#)
65. Nykänen, V.; Lahti, I.; Niiranen, T.; Korhonen, K. Receiver operating characteristics (ROC) as validation tool for prospectivity models—A magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland. *Ore Geol. Rev.* **2015**, *71*, 853–860. [\[CrossRef\]](#)

66. Zuo, R. Selection of an elemental association related to mineralization using spatial analysis. *J. Geochem. Explor.* **2018**, *184*, 150–157. [CrossRef]
67. Chen, Y.; An, A. Application of ant colony algorithm to geochemical anomaly detection. *J. Geochem. Explor.* **2016**, *164*, 75–85. [CrossRef]
68. Bonham-Carter, G.F.; Bonham-Carter, G. *Geographic Information Systems for Geoscientists: Modelling with GIS (No. 13)*; Elsevier: Amsterdam, The Netherlands, 1994.
69. Yousefi, M.; Carranza, E.J.M. Prediction–area (P–A) plot and C–A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. *Comput. Geosci.* **2015**, *79*, 69–81. [CrossRef]
70. Yousefi, M.; Carranza, E.J.M. Data-Driven Index Overlay and Boolean Logic Mineral Prospectivity Modeling in Greenfields Exploration. *Nonrenewable Resour.* **2015**, *25*, 3–18. [CrossRef]
71. Yousefi, M.; Carranza, E.J.M. Geometric average of spatial evidence data layers: A GIS-based multi-criteria decision-making approach to mineral prospectivity mapping. *Comput. Geosci.* **2015**, *83*, 72–79. [CrossRef]
72. Du, X.; Zhou, K.; Cui, Y.; Wang, J.; Zhang, N.; Sun, W. Application of fuzzy analytical hierarchy process (AHP) and prediction-area (P–A) plot for mineral prospectivity mapping: A case study from the Dananhu metallogenic belt, Xinjiang, NW China. *Arab. J. Geosci.* **2016**, *9*, 298. [CrossRef]
73. Gao, Y.; Zhang, Z.; Xiong, Y.; Zuo, R. Mapping mineral prospectivity for Cu polymetallic mineralization in southwest Fujian Province, China. *Ore Geol. Rev.* **2016**, *75*, 16–28. [CrossRef]
74. Nezhad, S.G.; Mokhtari, A.R.; Rodsari, P.R. The true sample catchment basin approach in the analysis of stream sediment geochemical data. *Ore Geol. Rev.* **2017**, *83*, 127–134. [CrossRef]
75. Zhang, N.; Zhou, K.; Du, X. Application of fuzzy logic and fuzzy AHP to mineral prospectivity mapping of porphyry and hydrothermal vein copper deposits in the Dananhu-Tousuquan island arc, Xinjiang, NW China. *J. Afr. Earth Sci.* **2017**, *128*, 84–96. [CrossRef]
76. Almasi, A.; Yousefi, M.; Carranza, E.J.M. Prospectivity analysis of orogenic gold deposits in Sasez-Sardasht Goldfield, Zagros Orogen, Iran. *Ore Geol. Rev.* **2017**, *91*, 1066–1080. [CrossRef]
77. Roshanravan, B.; Aghajani, H.; Yousefi, M.; Kreuzer, O. Particle Swarm Optimization Algorithm for Neuro-Fuzzy Prospectivity Analysis Using Continuously Weighted Spatial Exploration Data. *Nonrenewable Resour.* **2018**, *28*, 309–325. [CrossRef]
78. Roshanravan, B.; Aghajani, H.; Yousefi, M.; Kreuzer, O. An Improved Prediction-Area Plot for Prospectivity Analysis of Mineral Deposits. *Nonrenewable Resour.* **2018**, *28*, 1089–1105. [CrossRef]
79. Bater, C.W.; et Coops, N.C. Evaluating error associated with lidar-derived DEM interpolation. *Comput. Geosci.* **2009**, *35*, 289–300. [CrossRef]
80. Silverman, B.W.; Jones, M.C.; Fix, E.; Hodges, J.L. An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951). *Int. Stat. Rev.* **1989**, *57*, 233. [CrossRef]
81. Joseph Santarcangelo (2022). Data Normalization and Standardization. MATLAB Central File Exchange. Available online: <https://www.mathworks.com/matlabcentral/fileexchange/48677-data-normalization-and-standardization> (accessed on 9 November 2022).
82. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [CrossRef] [PubMed]
83. Japkowicz, N.; Stephen, S. The class imbalance problem: A systematic study1. *Intell. Data Anal.* **2002**, *6*, 429–449. [CrossRef]
84. Chawla, N.V.; Lazarevic, A.; Hall, L.O.; Bowyer, K.W. SMOTE Boost: Improving Prediction of the Minority Class in Boosting. In Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery, Cavtat-Dubrovnik, Croatia, 22–26 September 2003; pp. 107–119. [CrossRef]
85. Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2012**, *42*, 463–484. [CrossRef]
86. Li, T.; Xia, Q.; Zhao, M.; Gui, Z.; Leng, S. Prospectivity Mapping for Tungsten Polymetallic Mineral Resources, Nanling Metallogenic Belt, South China: Use of Random Forest Algorithm from a Perspective of Data Imbalance. *Nonrenewable Resour.* **2019**, *29*, 203–227. [CrossRef]
87. Xiong, Y.; Zuo, R. Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine. *Comput. Geosci.* **2020**, *140*, 104484. [CrossRef]
88. Alyasin, E.I.; Ata, O.; Mohammedqasim, H. Novel hybrid classification model for multi-class imbalanced lithology dataset. *Optik* **2022**, *270*, 170047. [CrossRef]
89. Lui, T.C.; Gregory, D.D.; Anderson, M.; Lee, W.-S.; Cowling, S.A. Applying machine learning methods to predict geology using soil sample geochemistry. *Appl. Comput. Geosci.* **2022**, *16*, 100094. [CrossRef]
90. da Silva, G.F.; Silva, A.M.; Toledo, C.L.B.; Junior, F.C.; Klein, E.L. Predicting mineralization and targeting exploration criteria based on machine-learning in the Serra de Jacobina quartz-pebble-metaconglomerate Au–(U) deposits, São Francisco Craton, Brazil. *J. South Am. Earth Sci.* **2022**, *116*, 103815. [CrossRef]
91. Jacob, R.J.; Bluck, B.J.; Ward, J.D. Tertiary-age diamondiferous fluvial deposits of the lower orange river valley, southwestern Africa. *Econ. Geol.* **1999**, *94*, 749–755. [CrossRef]