Petrographic, mineralogical and structural characterization of Nkout center iron deposit (south Cameroon) and geological mapping using k-NN machine learning algorithm

André William BOROH
University of Ngaoundere
Amadou DIGUIM KEPMANOU
kepnamoudiguim@yahoo.fr
University of Garoua
Alpha Baster KENFACK FOKEM
University of Ngaoundere
Adama AMAYA
University of Ngaoundere
Sylvain KOUAYEP LAWOU
University of Bertoua
Ismaïla NGOUNOUNO
University of Ngaoundere

Research Article

Keywords: Geological mapping, machine learning, nearest neighbor algorithm, Nkout iron deposit

Posted Date: October 27th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7632063/v1

Additional Declarations: No competing interests reported.

Petrographic, mineralogical and structural characterization of Nkout center iron deposit (south Cameroon) and geological mapping using k-NN machine learning algorithm

Authors

André William BOROH[a], Amadou DIGUIM KEPMANOU[a,e*], Alpha Baster KENFACK FOKEM[a], Adama AMAYA[b], Sylvain KOUAYEP LAWOU[c], Ismaïla NGOUNOUNO[d]

[a] Department of Mining Geology, School of Geology and Mining Engineering, University of Ngaoundere, P.O. Box: 115, Meiganga, Cameroon

[b] Department of Geomatic and Geological Mapping, School of Geology and Mining Engineering, University of Ngaoundere, P.O. Box: 115, Meiganga, Cameroon

[c] Department of earth science and environment, High teacher training college, University of Bertoua, P.O. Box: 652, Bertoua, Cameroon

[d] Faculty of Sciences, University of Ngaoundéré, Ngaoundéré, Cameroon

[e] Faculty of Sciences, University of Garoua, Garoua, Cameroon

[*] Corresponding author, email: kepnamoudiguim@yahoo.fr

Abstract

In the process of exploring potential resources, estimation methods are crucial instruments. The identification of lithological layers in the field of geological mapping heavily depends on how well the prediction techniques work. In this work, we use artificial intelligence specifically, machine learning and its k nearest neighbour algorithm to establish a novel paradigm of geological mapping. A petrographic and mineralogical characterization of the non-mineralized deposits was done as the initial step in this investigation. The presence of plutonic rocks such amphiboles, amphibole and biotite granites, amphibole and biotite gneisses, and amphibolite's was disclosed by these formations. Gneisses and amphibolite, which are all rich in magnetite, are the mineralized formations. Banded iron formations, such as coarse banded BIF (CMB), hematite magnetite BIF (HMB), and magnetite hematite BIF (MHB), are subsequently formed. It was possible to ascertain the type and traits of the various structures in the research region thanks to the structural analysis, which was performed for ductile and brittle deformation. The superficial layers produced by the logging of the boreholes served as the basis for the geological mapping by k-NN. It was feasible to

acquire the hyperparameters, in particular on the number of k=3 with the Euclidian metric distance, thanks to model optimization. This model, which was used to create the final geological map, demonstrated performance with an average confusion matrix performance of 86%, a correlation coefficient R^2 of 91%, and an AUC for the ROC curve of 0.99.

Keywords: Geological mapping, machine learning, nearest neighbor algorithm, Nkout iron deposit

1- Introduction

The lithological layers of a region can be identified on geological maps, which makes it easier to look for minerals. Since the invention of geological mapping, there have been two main approaches used by geomaticians: the direct approach, which involves conducting field research, and the indirect approach, which uses remote sensing. However, the size of the research region and the dearth of data make it difficult to pinpoint precisely where the geological layers indicated on the map are. The use of artificial neural networks (ANN) and other machine learning techniques in the field of cartography started at the end of the 1990s. In a first study, [1] used ANN to map four lithological types from airborne and total magnetic intensity (TMI), Landsat and SPOT data. They found out that GRS and Landsat data were the most useful variables to discriminate between classes representing lithologic units with the hardest-to-classify samples located near lithologic boundaries [1]. In a similar study, [2] used airborne geophysical, Landsat TM, SAR and SPOT data to classify four distinct lithologic classes (carbonate rocks, metavolcanic gneiss rocks, clastic rocks and granodiorite) in high latitude region of North America with help of ANN. Over the past two decades, machine learning algorithms have continued to develop, contributing to increasing precision in the field of geomatics, particularly through remote sensing [3, 4, 5, 6] , photogrammetry [7, 8] , geophysical data [9, 10] , and geochemical data [11, 12]. The main motivation for using machine learning (ML) algorithms for data predictions is that structured informations (patterns) in high-dimensional data can be identified, exploited, and then interpreted. This includes the ability to integrate a large number of disparate variables in a meaningful and effective way. ML therefore provides the opportunity to formulate interpretations that might not have been obvious using traditional data modeling methods [13].

Mining exploration companies collect large amounts of data that are used in the search for mineral products or for environmental studies [14]. These data are mostly used for estimating mineral resources in the area of investigation, mapping is no longer a priority. The objective of this article is to use the method of the nearest neighbors (k-NN) of data resulting from a drilling campaign in order to make an exact local cartography. Its application will be made on the Nkout Center iron deposit located in southern Cameroon. This study will be done first of all by the geological, structural and mineralogical characterization of the deposit.

2. Natural setting

The Nkout sector (Fig. 1) is located in the Dja and Lobo department in southern Cameroon. Nkout is located about 7.5 km from Djoum. Access to Nkout Center is via the national road N9, the main road which leads to the border between Cameroon and Congo Brazzaville. The Nkout area is characterized by mostly ferralitic, lateritic and clayey-sandy soils and sometimes hydromorphic in marshy areas. These soils are favorable for agriculture and the development of vegetation. Here, climate change is felt (heavy rains, soil leaching, prolonged drought) which degrades the soil. Regarding geomorphology, southern Cameroon is relatively flat with an average altitude between 600 and 700 m above sea level [15]. Altitude zones < 600 m are generally located in marshy U-shaped valleys which are interconnected in a dense dendritic hydrographic network. The terrains are generally asymmetrical with rounded tops and most of them are relics of the iron formations [16].

The geological location of Nkout belongs bto the Southern domain of the Pan-African Fold Belt in Cameroon. The southern domain or Yaounde domain corresponds of the Yaounde group [17]. It extends the Congo craton to the Adamawa-Yade domain. This domain continues into the Central african Republic, where it is known as the Oubanguides [18]. The southern domain has been the subject of several petrographic and structural study, including those by [17, 19, 20, 21, 22, 23, 24, 25]. All of these studies indicate that this part of the Pan-African Fold Belt consist of basic metamorphic rocks; Amphibolite, Amphibole and biotite gneisses, biotite gneisses, garnet-bearing gneisses, quartzite and micaschist. The garnet bearing gnesses have been interpreted as high-grade metamorphic platform deposits or as sedimentary deposit from margin of the Congo craton. The formation of the Yaounde group contain rock corresponding

to former evaporitic and/or saline sediments that recrystallized during the Pan=African orogeny under granulite facies conditions [22].

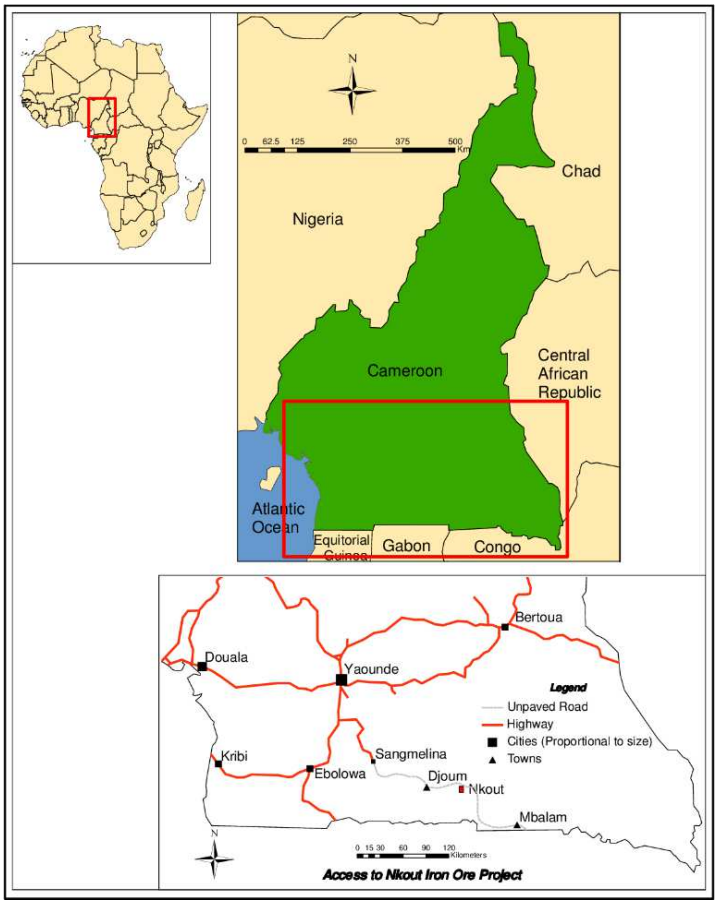

Figure 1: Location of the Nkout area

3. Methodology

The k-NN method is a statistical prediction technic previously introduced in the 50s [26] but neglected for several years only after [27] really developed it mathematically. The k-Nearest Neighbors (or k-NN) algorithm is one of the most widely used learning algorithms due to its simplicity. It is a lazy, non-parametric learning algorithm. It uses data with multiple classes to predict the classification of the new sample point. The k-NN is nonparametric because it makes no assumptions about the data being studied that is the model is distributed from the training data. The principle of the k-NN classifier algorithm described in [28] is as follows: the choice of the appropriate distance calculation method, the specification of the training data, and the calculation of the distances between the d learning and the objective data and the selection of the data

closest to the objective data according to the factor k. Several technics can be use to calculate the distance between values, but in this article, three main ones will be used:

The Minkowski distance [29] is a metric for real-valued vector spaces. It can only be computed in a normed vector space, that is, a space where distances can be represented as a vector that has a length and lengths cannot be negative.

$$d(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^c\right)^{1/c}$$

Manhattan distance [30] is also known as taxi distance or city block distance, it is due to the manner in which this distance is calculated. The distance between two points is the sum of the absolute differences of their Cartesian coordinates.

$$d(x,y) = \sum_{i=1}^{n} x_i - y_i \vee$$

Euclidean distance is a measurement of the true straight-line distance between two points in Euclidean space. It can be used by defining the value of p equal to 2 in the Minkowski distance metric.

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i| - y_i)^2}$$

The flowchart in Fig. 2**,** shows the methodology of data processing.

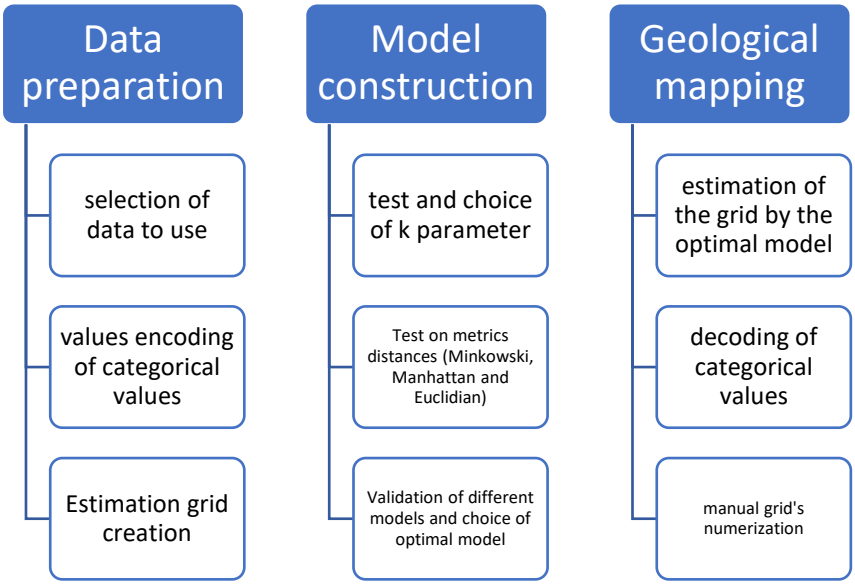

Figure 2 : k-NN mapping methodology

4. Results and discussion

4.1. Petrographic, mineralogical and structural analysis of the study area

The work carried out in the field as part of this work made it possible to collect petrographic and structural data on the various outcrops encountered. The petrographic and structural descriptions covered several outcrops and samples. The criteria for selecting the sampling locations were based on the petrographic type, the geographical position, and its surface condition (degree of alteration). Sampling was done on the freshest possible rock outcrops in order to obtain sound samples. When this was not possible rock core samples were sampled. Indeed, the outcrops encountered were previously refreshed and several representative samples of average diameter (Φ=10cm) were sampled from each outcrop. The rock samples collected made it possible to make an in-situ macroscopic description. This description focused on a visual identification of the characteristics of the rock, namely: its mode of outcrop, its petrographic type, its color, its structure, its degree of alteration (thickness of the alteration patina), its mineralogical composition (identification of minerals visible to the naked eye). The structural study in the field consisted in identifying, describing and measuring the attitudes of the different geological structures encountered.

4.2. Petrography and Mineralogy

Despite limited exposure of outcrops in the studied area, different types of rocks such as granites, gneisses, amphibolite and cuirasses, associated with BIF and iron mineralization were discovered. The following descriptions are primarily based on observation and description of rock samples and drilled cores.

Plutonic rocks in the area consist of amphibole-biotite granite and biotite granite. The amphibole and biotite granite outcrops in the study area in the form of a slab (Fig. 3-a). A sample taken from this rock (Fig. 3-b) shows a greyish and brown color in some places, a low thickness (0.5 mm) of the dark brown weathering patina, a massive structure. The naked eye mineralogy shows minerals such as quartz, orthoclase, amphibole and biotite.

Two thin sections (AFF03, AFF13) were made in this granite. Microscopic observation shows that the amphibole and biotite granite have a grainy texture (Fig. 3-c). It consists of amphibole, biotite, quartz, plagioclase, orthoclase, microcline, chlorite and opaque minerals.

Amphibole is a green hornblende in sub-automorphic to xenomorphic crystals of variable size (1 to 6 mm). It constitutes about 5% of the volume of the rock. Some amphibole crystals contain inclusions of quartz, feldspars and opaque minerals (Fig. 3-c). Biotite (about 10%) forms dark brown to light brown flakes 2 mm long and 0.3 mm wide on average and is commonly associated with amphibole and opaque minerals. The biotite is transformed into chlorite in some places (Fig. 3-d).

Quartz represents 35% of the volume of the rock. It forms sub-automorphic to xenomorphic crystals of variable size (0.2 to 1.5 mm). It forms polycrystalline beaches with rolling extinction. Quartz is associated with feldspars (plagioclase, orthoclase and microcline). Plagioclase forms crystals of 3 mm long on an average. It constitutes 20% of the volume of the rock. It is automorphic to sub-automorphic and it is recognized by the polysynthetic twins which end in a bevel in the mineral (Fig. 3-c). Alkaline feldspars are formed from orthoclase and microcline. Orthoclase (about 10%) forms elongated crystals 4 mm long and 2 mm wide on an average. Orthoclase is sometimes perthitic and contains inclusions of quartz and opaque minerals. Orthoclase is associated with microcline. The microcline (about 10%) forms crystals of the same size as the

orthoclase. It is easily recognized by the polysynthetic twin of the pericline which gives it a shimmering grid.

Accessory minerals (about 5%) consist of apatite and opaque minerals.

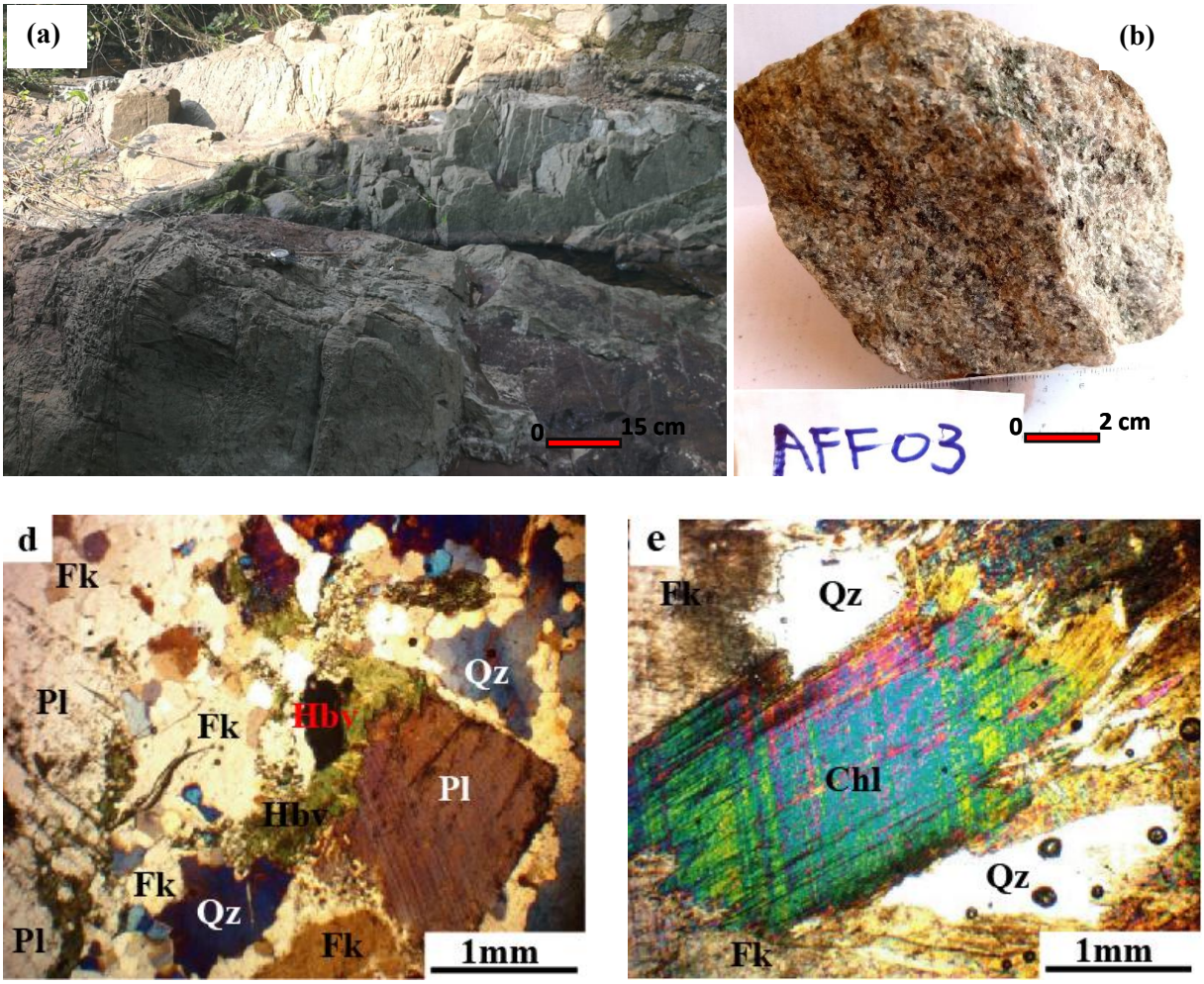

Figure 3: a- Outcrop of amphibole and biotite granite; b- sample; c- Grainy rock texture (LPA); d- Transformation of biotite into Chlorite (LPA).

The amphibole and biotite gneiss does not outcrop in the area; therefore, the macroscopic observation was made a rock core sample (Fig. 4-a). This sample has a white color, an absence of weathering patina, a massive structure. The strong white color of the sample is justified by the presence of large quantities of leucocratic minerals (quartz and feldspar). In addition to these minerals, we also note the presence of biotite macroscopically observable with the naked eye.

Two thin sections were made in the biotite granite: sections AFF01, AFF11. At the thin section scale, the biotite granite has a grainy texture (Fig. 4-b). It is made up of minerals

of biotite, quartz, plagioclase, orthoclase, microcline, chlorite and opaque minerals. Biotite (about 20%) appears in light brown flakes (Fig. 4-c) of variable size from 0.8 to 1.5 mm long. It has cleavages thickened by the concentration of opaque minerals and often contains quartz inclusions. Quartz represents 35% of the volume of the rock. It forms sub-automorphic to xenomorphic crystals of variable size (0.2 to 3 mm). It forms polycrystalline beaches with rolling extinction. Quartz is associated with feldspars (plagioclase, orthoclase and microcline). Plagioclase (20%) forms crystals 1 mm long on an average. Plagioclase crystals are sub-automorphic. It is associated with the other feldspars in the rock. Alkali feldspar consists of orthoclase and microcline. Orthoclase (15%) forms crystals of almost the same size as plagioclase crystals. It is sub-automorphic and is associated with opaque minerals. The microcline (8%) is recognized by its double intersecting twins which gives it a tissue appearance (Fig. 4-d). Chlorite (5%) forms at the expense of biotite and is also associated with plagioclase and opaque minerals in the rock. Opaque minerals (2%) are the only accessory minerals in the rock. Opaque minerals are found in inclusion in certain crystals of plagioclase and also associated with this mineral in the rock.

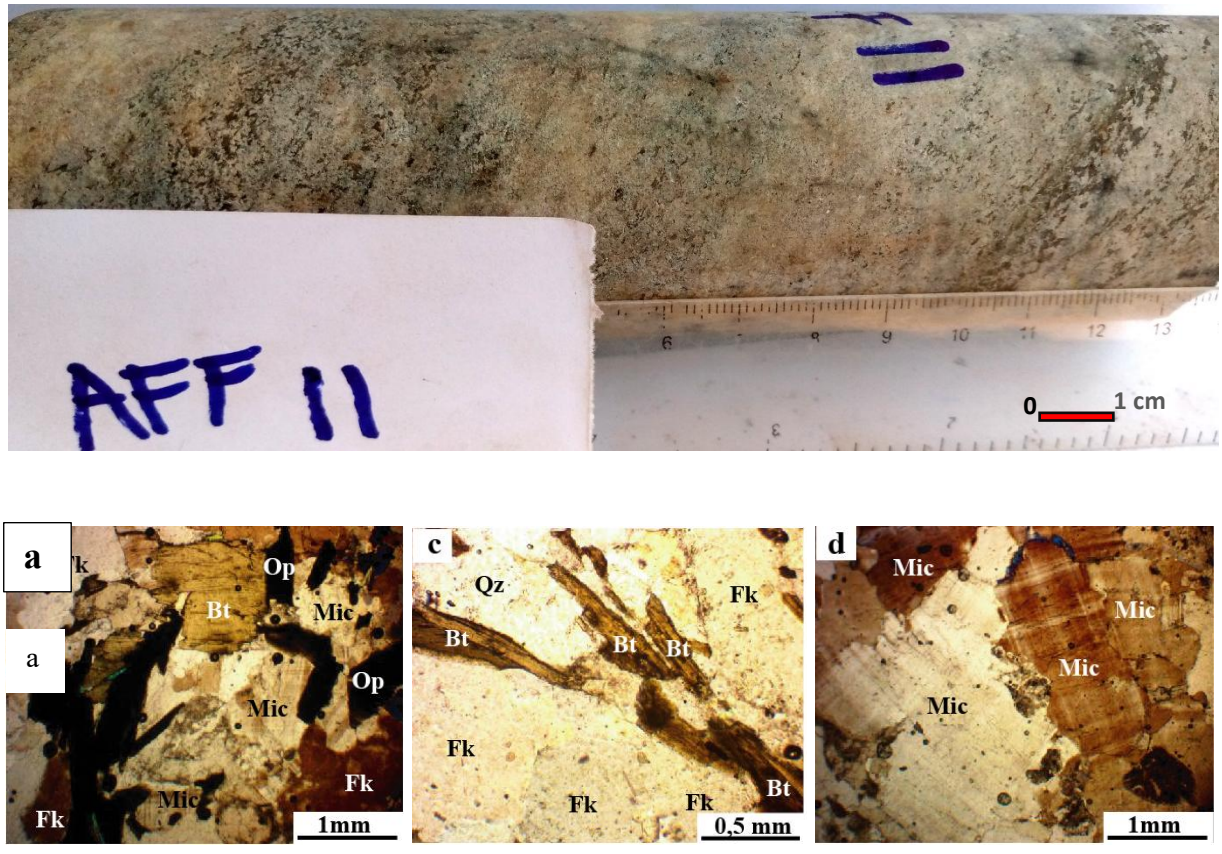

Figure 4: a- sample of a biotite granite core; b- grainy texture (LPA); c- Biotite flakes (LPNA); d- Cluster of microcline crystals (LPA).

Metamorphic rocks are the most abundant in our study area. They outcrop in places in blocks and slabs. These metamorphic rocks consist of biotite gneiss, amphibole and biotite gneiss, banded amphibolite and non-banded amphibolite.

The amphibole and biotite gneiss does not outcrop in the area due to the importance of the soil cover. So, the macroscopic observation of this rock was made on a core sample (Fig. 5-a). It presents a predominantly dark color with some mineral plans of white color. No weathering patina is observed. It has a foliated structure with alternating light and dark bed. The mineralogy to the eye presences of minerals such as quartz, amphibole and biotite. The AFF07 section was made from this gneiss. Under the microscope, the rock has a grano-nematoblastic texture (Fig. 5-b) and is made up of amphibole, biotite, plagioclase, orthoclase, quartz, chlorite and opaque minerals. Amphibole (20%) is a green hornblende of variable size from 0.2 to 1.5 mm long. It is associated with biotite which underlines the schistosity of the rock (Fig. 5-b and 5-c). Biotite (10%) forms dark brown sheets associated with amphibole. It is transformed in places into chlorite with concentration of opaque minerals in the cleavages. Plagioclase and orthoclase (30%) form the two feldspars of the rock (Fig. 5-d). Plagioclase crystals have an average size of about 1.5 mm long. It contains sub-microscopic inclusions which give a cloudy appearance to the rock. Orthoclase of identical size to plagioclase is associated with the latter and with quartz in the rock. The quartz (about 30%) forms polycrystalline beaches with rolling extinction and 1.5 mm long ribbons parallel to the schistosity of the rock (Fig. 5-d). It is associated with plagioclase in quartzofeldspathic beds. Opaque minerals and chlorite (about 10%) form the accessory minerals of the rock.

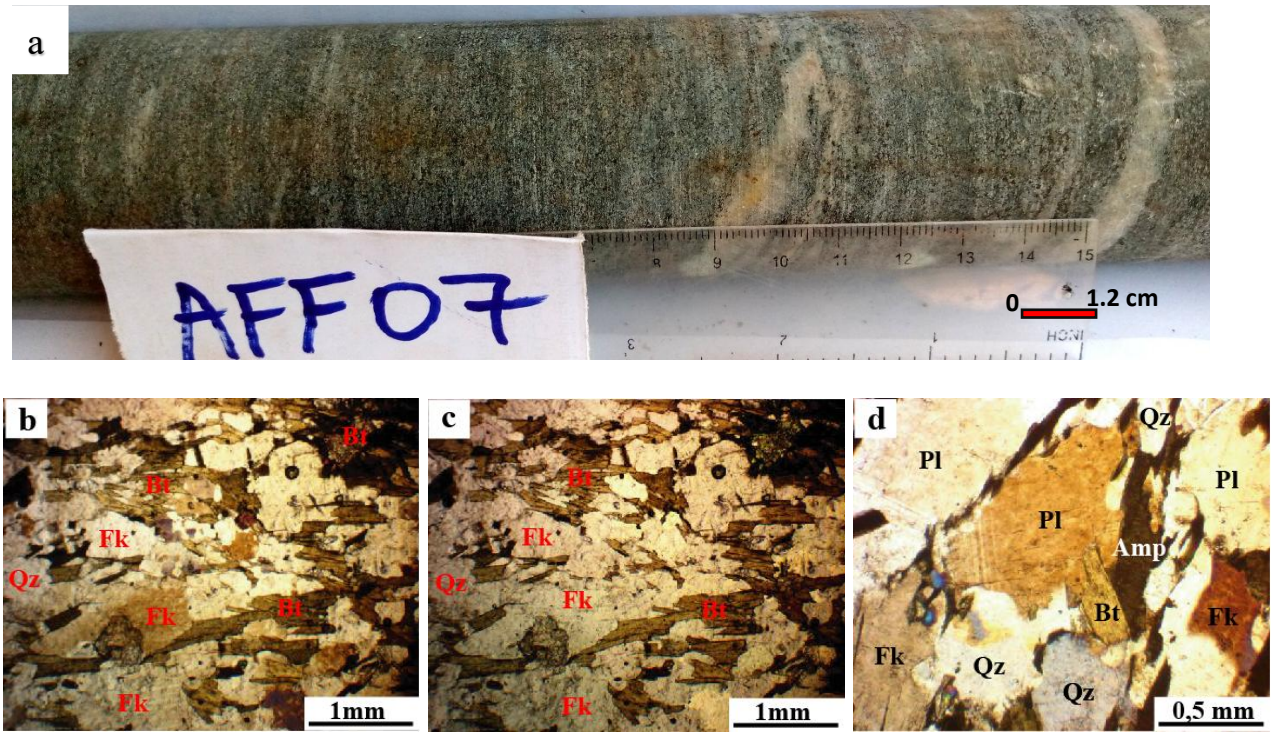

Figure 5: a- Sample of an amphibole and biotite gneiss core; b- Grano-nematoblastic textute (LPA); c- Grano-nematoblastic textute (LPNA); d- Alkaline Quartz Feldspar (LPA).

Biotite gneiss outcrops in the form of large blocks on the sides of a hill (Fig. 6-a) in the center of the study area. The outcrops of this rock show several fracture planes, signs of a brittle deformation phase. A sample of this rock (Fig. 6-b) shows a white to greyish coloration, a thin brown weathering patina, a more or less foliated structure marked by alternating light and dark beds. Macroscopically the mineralogy of the rock is made up of quartz, feldspar, and biotite.

Figure 6: sample of biotite gneiss (a-Outcrop; b-sample)

Banded amphibolite did not outcrop in the study area, hence the macroscopic description from a rock core. It has a greyish to blackish color. The banding is underlined by beds of quartzofeldspathic material in the rock. Consequently, the rock to rock has a foliated structure (Fig. 7-a). The naked eye mineralogy shows minerals such as: amphibole, quartz and feldspars.

The thin sections (AFF02, AFF08, and AFF12) were made in this rock. Under the microscope, the rock has a granonematoblastic to nematoblastic texture (Fig. 7-b). The rock is composed of amphibole, biotite, plagioclase, quartz, chlorite and opaque minerals.

Amphibole (50%) is a green hornblende (Fig. 7-c) of variable size which reaches 2 mm in length in some places. It is often corroded by quartz. The amphibole contains inclusions of zircon and opaque minerals. Biotite (10%) forms very fine light to dark brown sheets associated with amphibole. It turns into chlorite in some places.

Plagioclase constitutes 25% of the volume of the rock. Plagioclase is xenomorphic and forms crystals of variable size from 0.5 to 2 mm long. Plagioclase crystals are often affected by fracturing. Plagioclase contains submicroscopic inclusions and opaque minerals. Quartz (15%) forms crystals associated with plagioclase. In places, it appears in polycrystalline ribbons which are parallel to the schistosity of the rock (Fig.

7-c). Accessory minerals (10%) consist of opaque minerals. These opaque minerals are often found as inclusions in amphibole.

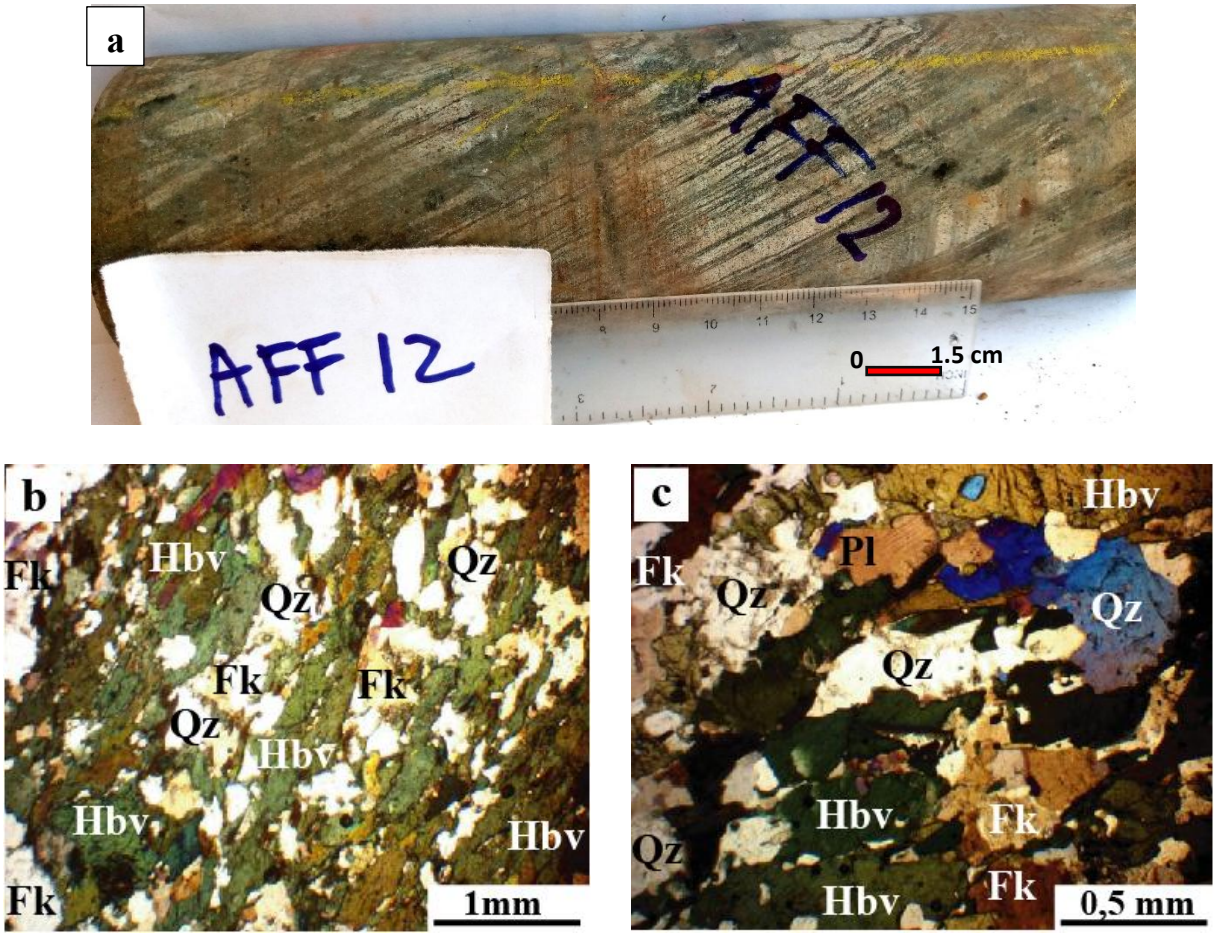

Figure 7: a- Sample of a banded amphibolite gneiss core; b- Grano-nematoblastic texture (LPA); c- Green Amphibole (Green Hornblende: Hbv) (LPA).

Non-banded amphibolite outcrops in the form of a slab near a stream (Fig. 8-a). These outcrops are more or less altered with water circulation. Nevertheless, a fresh sample could be sampled (Fig. 8-b). This sample has a dark coloration with a white color intrusion, a brown color weathering patina, a massive structure. The naked eye mineralogy of this rock shows a high concentration of amphibole.

Two thin sections (AFF04 and AFF09) were made in this rock. Under the microscope, the rock has a granoblastic texture with a nematogranoblastic tendency (Fig. 8-c). It consists of amphibole, biotite, plagioclase, zircon and opaque minerals. Amphibole (70%) is the most abundant mineral in the rock. It is a green hornblende (sub-automorphic to xenomorphic) (Fig. 8-d) of variable size from 1 to 4 mm long and 0.5 to

4 mm wide. Amphibole transforms into epidote at the edges of the mineral. The amphibole contains inclusions of quartz, zircon and opaque minerals (Fig. 8-d). Biotite (5%) forms dark brown sheets (Fig. 8-d) with a dimension of about 1 mm long on average. It is associated with amphibole in the rock. Plagioclase (15%) forms elongated crystals averaging 2 mm in length. It contains sub-microscopic inclusions of opaque minerals. The quartz (10%) forms range of variable size from 0.5 to 3 mm. It is sub-automorphic to xenomorphic and shows a rolling extinction. Quartz is associated with plagioclase and forms the quartzofeldspathic bed of the rock (Fig. 8-e). Accessory minerals (5%) consist of zircon and opaque minerals. Zircon is found as an inclusion in amphibole.

The magnetite-rich gneisses represent a melanocratic gneiss facies enriched in magnetite. Their black color is characteristic of the presence of iron oxides in the rock. They have a foliated structure. These gneisses show an alternation of light bands of quartz and dark bands with a predominance of iron oxides such as magnetite. The presence of magnetite in these rocks was tested using a magnet (**Fig. 9**).

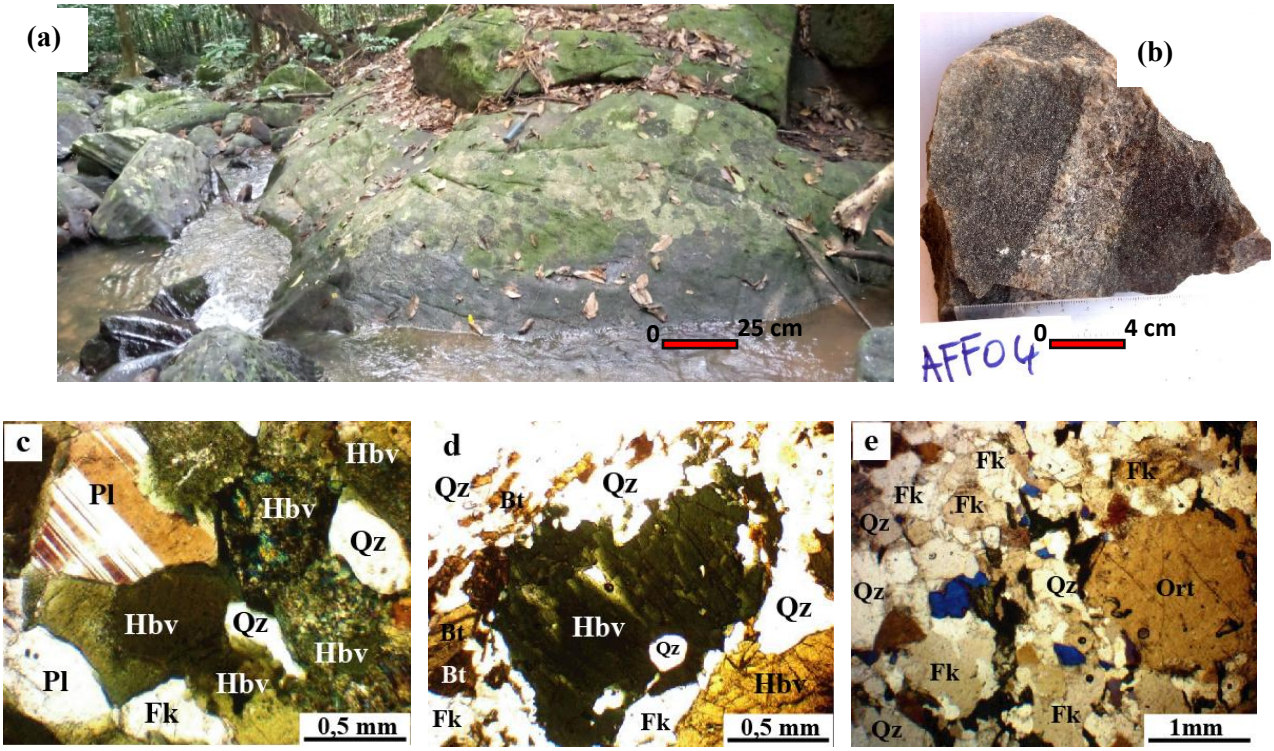

Figure 8: a- Outcrop of unbanded amphibolite; b- Rock sample; c- Granoblastic texture (LPA); d-Amphibole crystal with inclusion of quartz (LPNA); e- Quartzo-felspathic association (LPA).

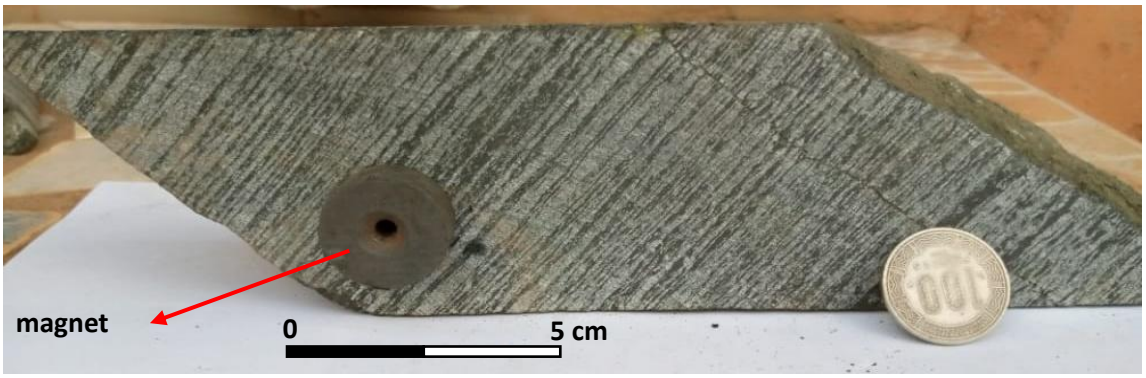

Figure 9: samples of magnetite gneiss.

These rocks generally have a very dark appearance and a black color, indicating the presence of iron oxides. They have a massive structure. These rocks are very rich in iron oxide and amphibole, and very little biotite. The dominant mineral here is magnetite, and sometimes hematite when weathered; because the alteration of magnetite by oxidation at the level of the fractures produces hematite. Magnetite amphibolites has a metallic appearance that refers to the more or less high content of iron (magnetite) they contain (Fig. 10). On cores, these amphibolites show sharp contacts with BIFs and gneisses.

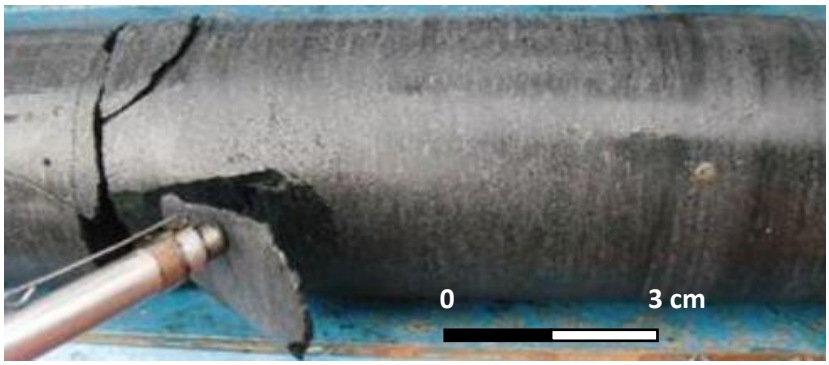

Figure 10: sample of amphibolite with magnetite

These are formations enriched in magnetite, characterized by alternating light bands of quartz and dark bands of ferric oxides. Sedimentation seems to be the main mode of emplacement of BIFs in the study area, although subsequently affected by tectonic phenomena. These rocks outcrop on almost all the hills of the massif in the study area. On trenches, BIFs can reach up to 12 m in height (Fig. 11a). However, weathering favors the emplacement of magnetite-rich rocky saprolites, as well as the transformation of magnetite into hematite.

On the surface, the rocks of the study area have undergone several phases of deformation, namely: ductile deformation and brittle deformation. Most of the deformations encountered in the area consist of foliations (Fig. 9, 11 and 12-a) and schistosities (Fig. 12-b).

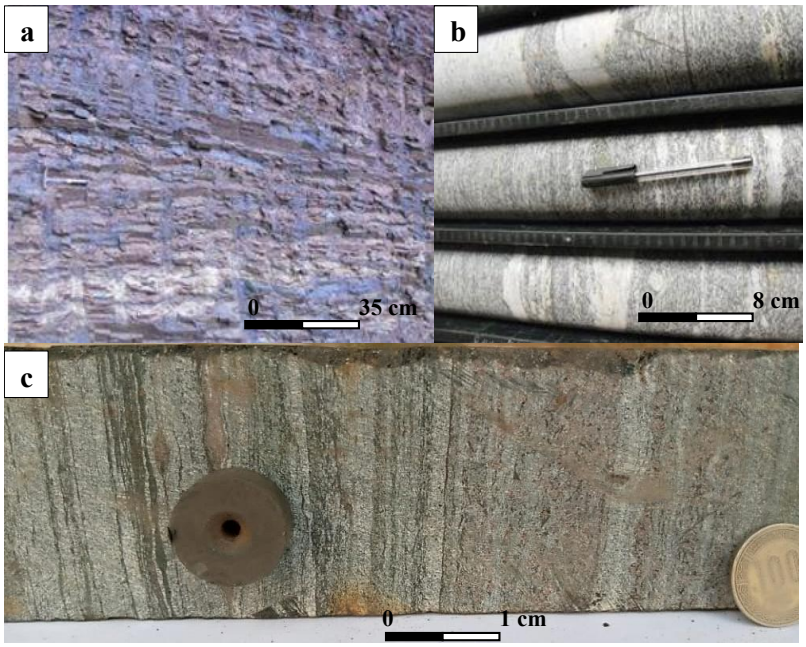

Figure 11: banded iron formations: weathering a-BIF observed on a trench; b, c-BIF with banded structure

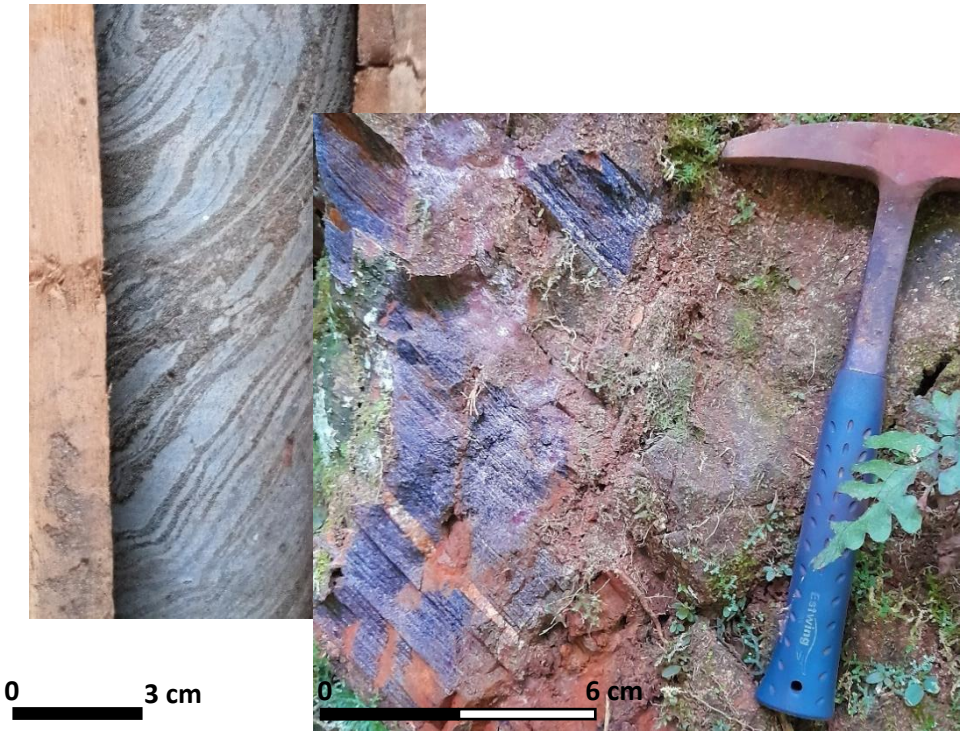

Figure 12: foliation planes of schistosity (a- foliation planes in BIF; b- schistosity planes in banded amphibolites)

Foliations, marked by alternating light and dark beds, have been observed in gneisses, banded amphibolites and banded iron formations. In the BIF, they are more accentuated, characterized by alternating dark bands of magnetite and light bands of quartz. In the gneiss, they are millimetric, characterized by a stacking of very tight thin planes with thicknesses between 0.3 and 1.5 mm (Fig. 9). The measurement of the attitudes of the foliation planes in the different formations encountered made it possible to obtain a major direction N150E (Fig. 13-a) with dip values between 10° and 64°. The density of concentration of the poles of the foliation planes (Fig. 13-b) makes it possible to highlight 2 main families of foliations with orientations between N45E and N60E for the first family, and N150E and N165E for the second family. These two families correspond to the strong concentration of the poles of the planes (Fig. 13-b).

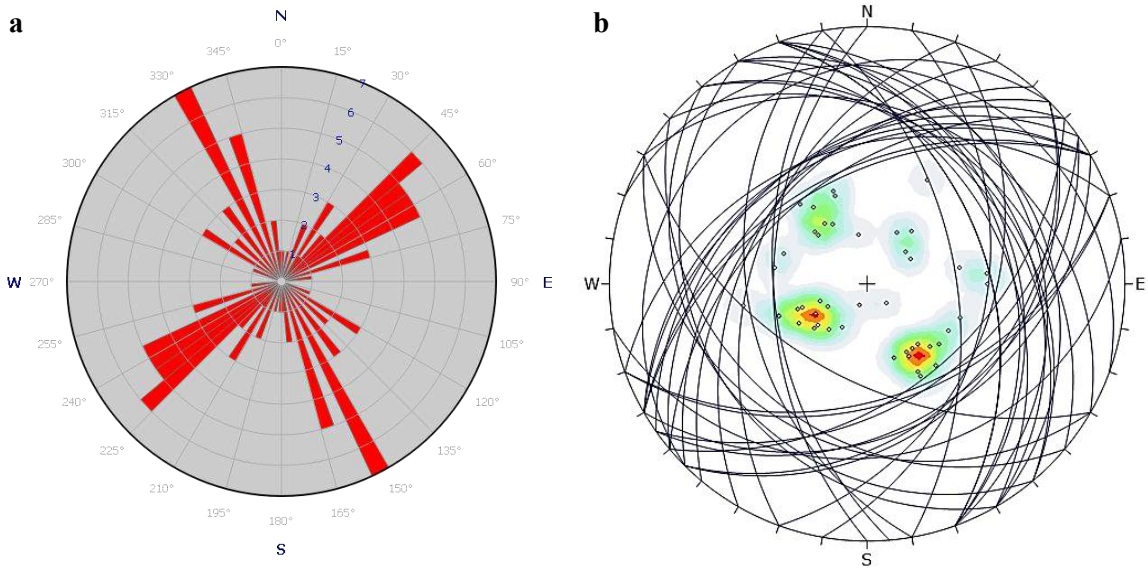

Figure 13: rosette and stereogram of the foliation planes (a-race of direction; b-stereogram showing the concentration density of the poles of the planes)

Marked by a stack of several planes of the rock thus giving it a stack of several sheets capable of individualizing and debiting from each other as shown in Figure 12-b. The measurement of the attitudes of the foliation planes in the different formations encountered made it possible to obtain a major direction N150E (Fig. 14-a) with dip values between 5° and 35°. The concentration density of the poles of the schistosity planes (Fig. 14-b) makes it possible to highlight a single family of schistosity with orientations between N105E and N150E.

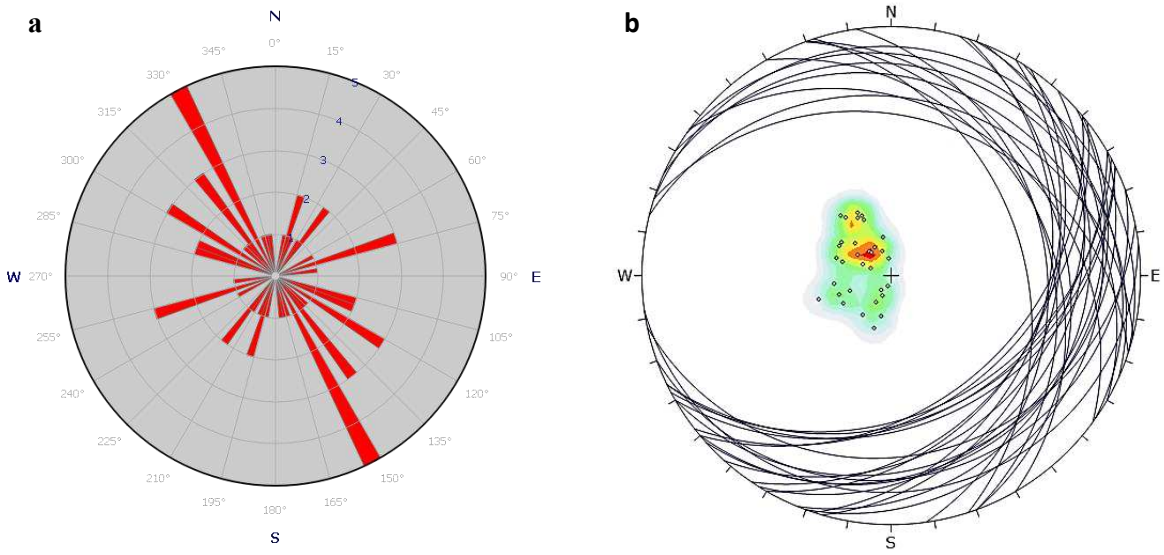

Figure 14: rosette and stereogram of the schistosity planes (a-directional rosette; b-stereogram showing the concentration density of the poles of the planes).

The brittle deformation in the study area is essentially marked by the presence of fractures. These are materialized by the dry diaclase or joints observed in practically all the formations (gneiss, amphibolite and BIF). These structures induce a fragmentation of the rocks into blocks of variable geometric shapes (cubic, square, lozenge-shaped) and of centimetric dimensions (Fig. 15). They are posterior to the put on place of ductile deformations. The measurement of the directions of 94 fracture planes made it possible to establish the direction rosette of the study area as shown in Fig. 16. Thus, from this rosette, it can be concluded that the fracture planes are mainly oriented following the direction N110E.

Figure 15: Joints on an amphibolite outcrop

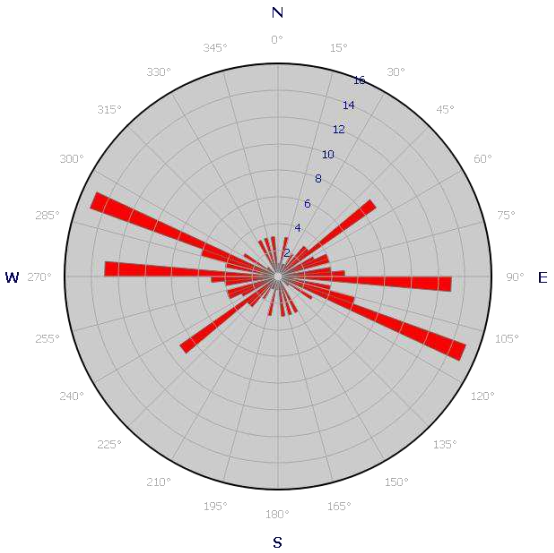

Figure 16: Fracture plane direction rosette

4.3. Geological mapping by K-NN (Data preparation)

The principle of data extraction consists in selecting the first layers of each borehole by excluding the lateritic domain of the zone defined by [31] . This method will therefore make it possible to have the real lithological cover of the zone at a reduced spacing, that is to say corresponding to the mesh used for the drillings, ie 50m x 50m. As indicated in the histogram of Fig. 17, the majority element is BIF with an estimated percentage of 72% and it is divided into three main types: Coarse banded BIF (CMB 8%), Haematite Magnetie BIF (HMB 48%) and Magnetite Haematite BIF (MHB 8%). The other elements are the majority, these are metasediments (MS 21%) and gneiss (Gn and Gnbq 12%) or minority, it is the case of amphibolites (Amp 2%), quartz veins (QB 1%) and shales (Sc 1%).

The distribution of lithologies over the 195 holes is shown in Fig. 18. The grid designed for the prediction of lithological types at unknown points has the characteristics of 100m long and 100m wide for a total of 31 rows and 20 columns equivalent to 620 estimation points.

An ordinal numerical encoding was carried out for each lithological category in alphabetical order from 0 to 8, ie 0 for amphibolite and 8 for schist. The type of machine learning chosen is classification to the detriment of regression in order to avoid poor interpolation between neighboring values of distant codes.

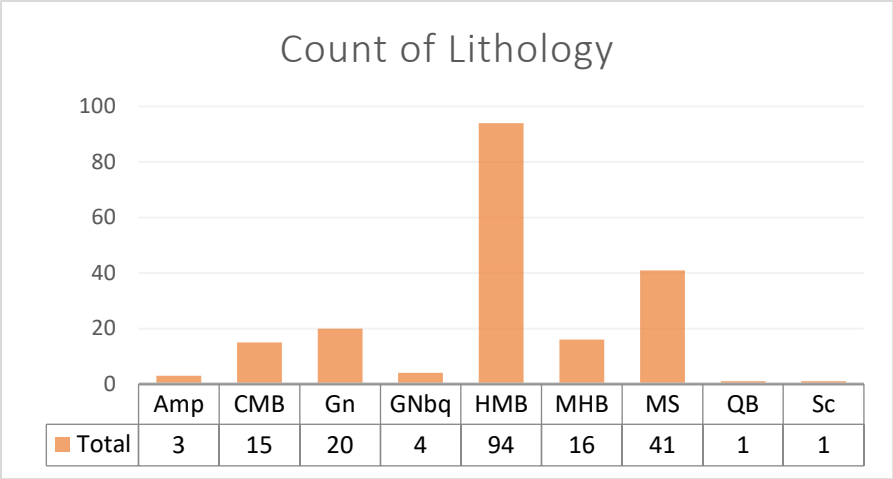

Figure 17: All lithologic data used

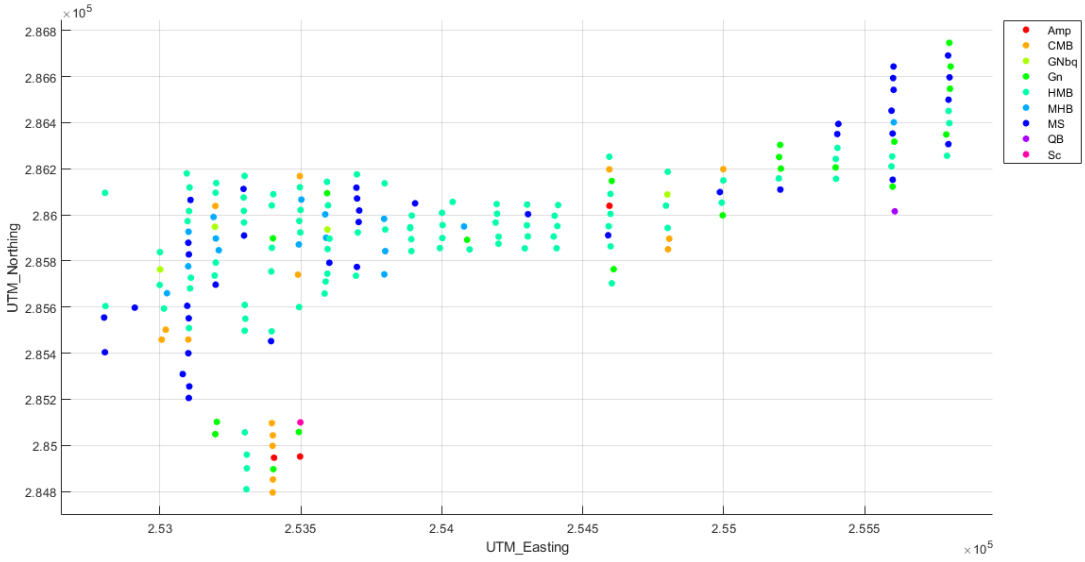

Figure 18: Spatial distribution of lithological types

4.4. Model development and validation

The first step in creating a machine learning model is choosing the parameter k. This choice is decisive especially in the case of classification because the performance of the classifier depends on the optimal choice of the number of neighbors k which is different from one data sample to another [32]. Moreover, when the value of k is very small (for example k=1), the result provides a low percentage of validation out of the data used for the modeling this due to the fact that the model takes into account only very little information when of learning therefore not requiring data structuring [33].

To develop the prediction model, we used the Classification Learner Toolbox module of Matlab [34]. The determination of the parameter k was done by optimization: 30 random values of neighbors were tested in order to find the optimal value (Fig. 19).

According to Figure 19 the value of k is optimal for the values 3, 5, 6 as well as in the interval [8-15] but the choice was put on the value k=3. This choice is justified by the low number of certain amphibolite, schist and quartz vein lithological types, the choice cannot however be made at k=1 to satisfy the last two types given the reason given in the previous paragraph.

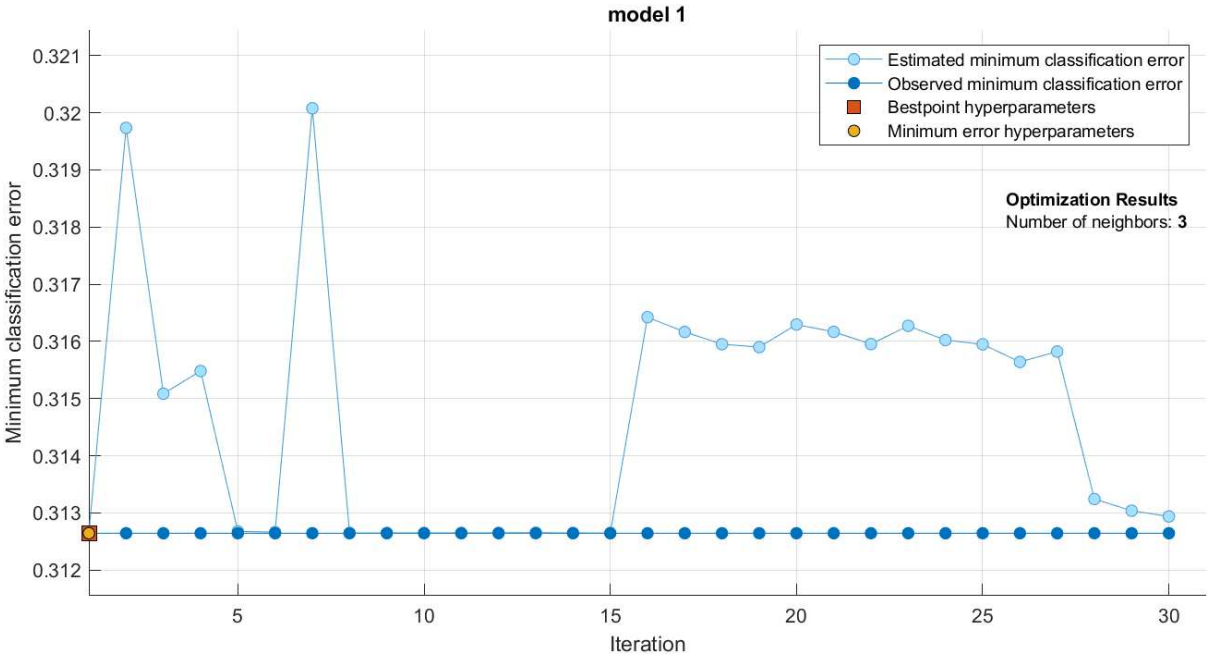

Figure 19: Finding the optimal k-NN model

After this step, the basic model was applied to the training data set using the three types of metric distances: Minkowski, Manhattan and Euclid. Three validation methods were applied to certify the reliability of the three models designed. First, we calculated the score of each model in order to determine the correlation coefficient between the true values and the predicted values. The Manhattan distance provides a performance r^2= 86% while the Euclidean distance gives r^2=91% and the Minkowski distance r^2=89%.

Then, a test using confusion matrices [35, 36] was performed for each metric distance measurement:

For Manhattan, three major classes of validation are to be raised. In the first, we note an elimination of the minority elements (Amp, Gnbq, QB and Sc) resulting in an absence of values on the diagonal between the true values and the predicted values. The intermediate class consisting of gneiss and coarsed BIF present an average validation of 67% and 52% respectively, the Gn being poorly estimated by the HMB and the CMB while the coarsed BIF is being replaced by mainly by the gneiss at 33 % and by amphibolites. The best predicted layers are progressively the 92% hematite BIF, the 93% metesediments and the 100% magnetite BIF.

For the Euclid distance, the validation of the Gnbq is null with a complete error. The CMB is average at 60% but largely substituted in the prediction by gneiss and amphibolites. Interaction which is confirmed with gneiss validated at 82%, the rest being predicted in favor of CMB. As with the previous case, the minority elements (Amp, Gnbq, QB and Sc) are absent and substituted while the MHB, HMB and MS provide good scores with a perfect score for the MHB.

The Markowski metric distance also eliminates the prediction of minority lithologies. The hematite BIFs are well estimated at 100%, we also note the poor Gbq score as in the Euclidean metric as well as the interactions in the prediction between the CMB and the Gn.

Another validation was done using the ROC curve and its AUC zone. This is a graph that represents a par model by tracking the curve of its sensitivity versus the complement to 1 of its specificity [37, 38]. On this curve, the best threshold, ie that of the optimal model, is the one that maximizes both specificity and sensitivity. For the three metrics:

- The Manhattan ROC curve has zero sensitivity with a specificity of 75%, which gives it an AUC of 99%. This makes it the least reliable of the three models.
- Then, the Euclidean ROC curve has a classifier closer to the angle with zero sensitivity and specificity at 0.94 as well as an AUC of 99% making it the best model.
- The model linked to the Minkowski distance has the second-best model given its ROC curve which presents the sensitivity-specificity pair equivalent to that of the Euclidean distance but rather with a lower AUC at 98%.

Table 1: Model performance comparison

			ROC CURVE		
	R^2	Confusion mean performance	TPR	FPR	AUC
Manhattan	0.86	80.80%	0.75	0	0.99
Euclidean	0.91	86.20%	0.94	0	0.99
Minkowski	0.89	85.40%	0.94	0	0.98

Table 1 provides a summary of the cross-validation operations performed on the three different models. The average performance of the confusion matrix is obtained by arithmetic mean of the performance of the five estimated lithological layers (CMB, Gn, HMB, MHB and MS) while the false positive rate (FPR) corresponds to the score of the sensitivity and the true positive rate (TPR) to that of the specificity.

In this table, Euclidian metric offers the best performance for the three cross-validation techniques. The Manhattan distance shows the second-best model and differs from the Manhattan metric on all validation tools used. It is therefore in our interest for the rest of this work to opt for the Euclidean distance prediction model for the mapping of the lithology in the Nkout Center iron deposit.

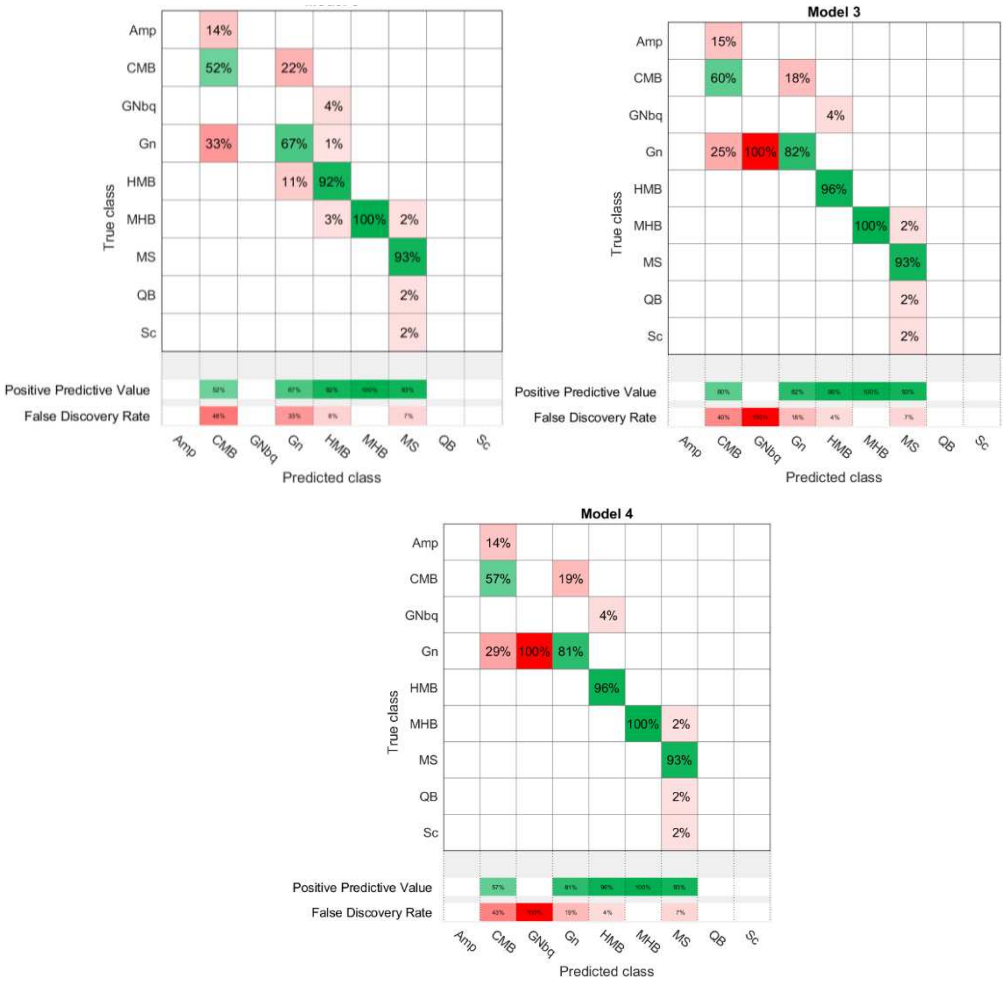

Figure 20: Validation by confusion matrices

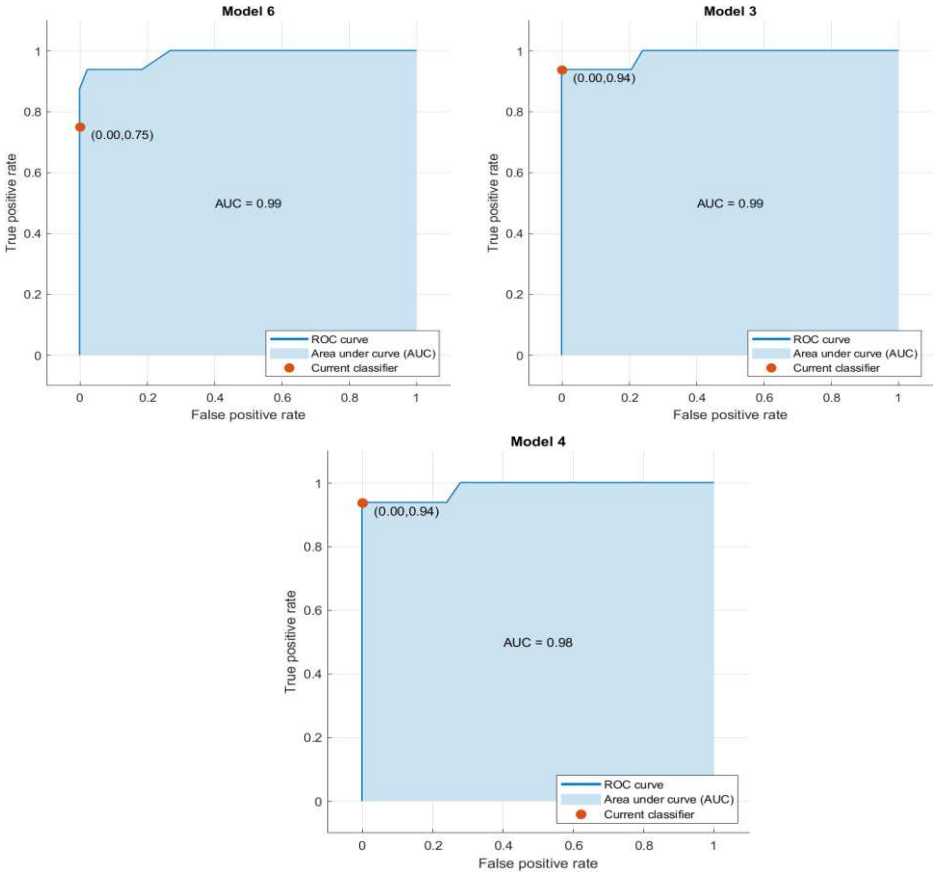

Figure 21: Validation by ROC curves

4.5. Geological mapping using 3-NN

The optimal model chosen being that of 3-NN by use of the metric distance of Euclid, the next step is that of the estimation of the points of the grid in order to have the corresponding lithology. The use of the prediction function provided the estimates distributed in Fig. 22. The triangle points represent the training data while the round ones are the estimated elements of the grid.

The first observation concerns the absence of certain minority elements on the estimated grid, these are Gnbq, QB and Sc. As for the amphibolites, they are found scattered predicted to the south of the mapped zone. Hematite BIFs are the most abundant elements with a percentage of 51.29 % and are mainly found in the center of the map, followed by Gneiss Gn, with a rate of 18.23 %. They are abundant in the southeast of the study area. The metasediments have a rate of 12 % and are located further north. Digitizing the grid gives the local geological map of the Nkout Center area (Fig. 23).

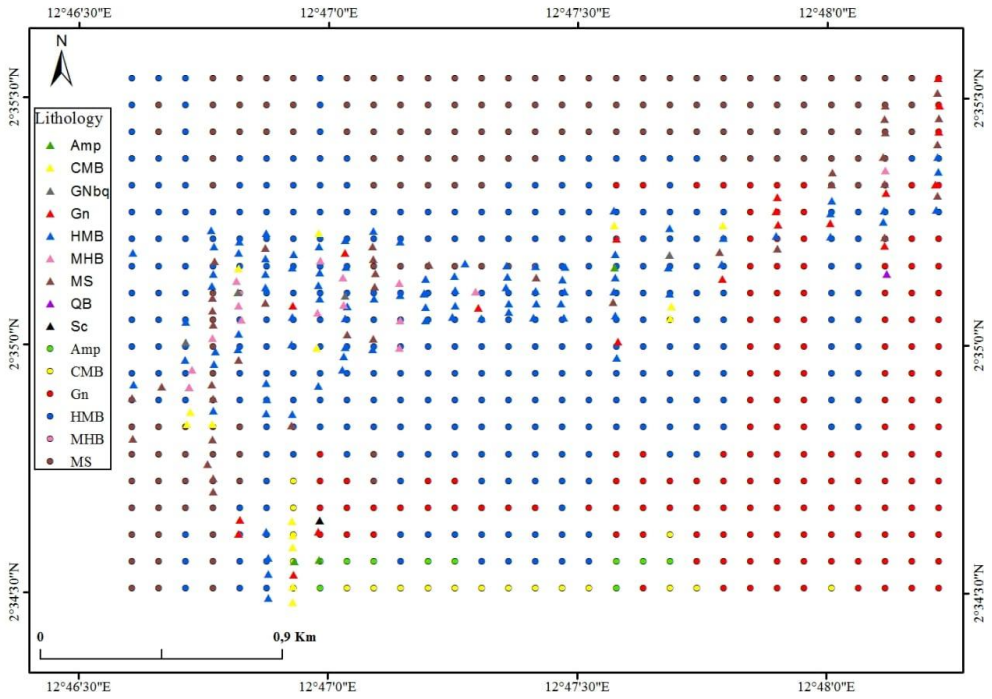

Figure 22: Estimated grid according to lithology (in triangle the base samples, in circles the predicted samples)

Like the lithology estimation grid, the geological map shows a large distribution of hematite BIF occupying 2.848 km² over the total area of the study area. Gneisses and metasediments are extended over an area of 1,581 Km2 and 1,324 km² respectively. This map, which determined the probability of the locations of the different lithological types used as input data to produce it, is validated in the field. It thus shows consistency with field data.

The visual validation of this map involves comparing it with a geological map previously drawn up in the study area. To do this, we compare the geological map drawn up by [39] with the map drawn up using the k-nearest neighbor method. This validation leads us to make the following observations:

- the geological map by [39] shows six lithological formations, including two pyroxene gneiss facies, the first to the north-east and the second to the south-east, and a broad cover of biotite gneiss to the north and south-west; the Nkout Mountains are mainly composed of magnetized garnet micaschist, BIF and a hematite cap at the end;

- unlike the map in Fig. 23-a, the geological map of Nkout-Centre produced in this study highlights 06 formations (amphibolite, metasediments, gneiss, magnetite BIF, hematite BIF, coarse-grained BIF);
- the geological map drawn up by k-NN has a larger scale and shows much more detail in terms of the distribution of formations and the differentiation of mineralized formations (magnetite-rich BIF, hematite-rich BIF, coarse-grained BIF);
- according to the map by [39], the spatial distribution of the BIFs is limited to the layout of the hills, contrary to the data from the borehole logs used to produce the geological map for this study. This map (Fig. 23-b) shows a wider spatial distribution of mineralized formations that does not follow the layout of the hills;
- although their distribution is small, k-NN mapping adds two lithological formations at Nkout center, namely amphibolite and metasediments. Their appearance is rather interesting because they were detected by the boreholes; it is assumed that they did not outcrop at the surface.

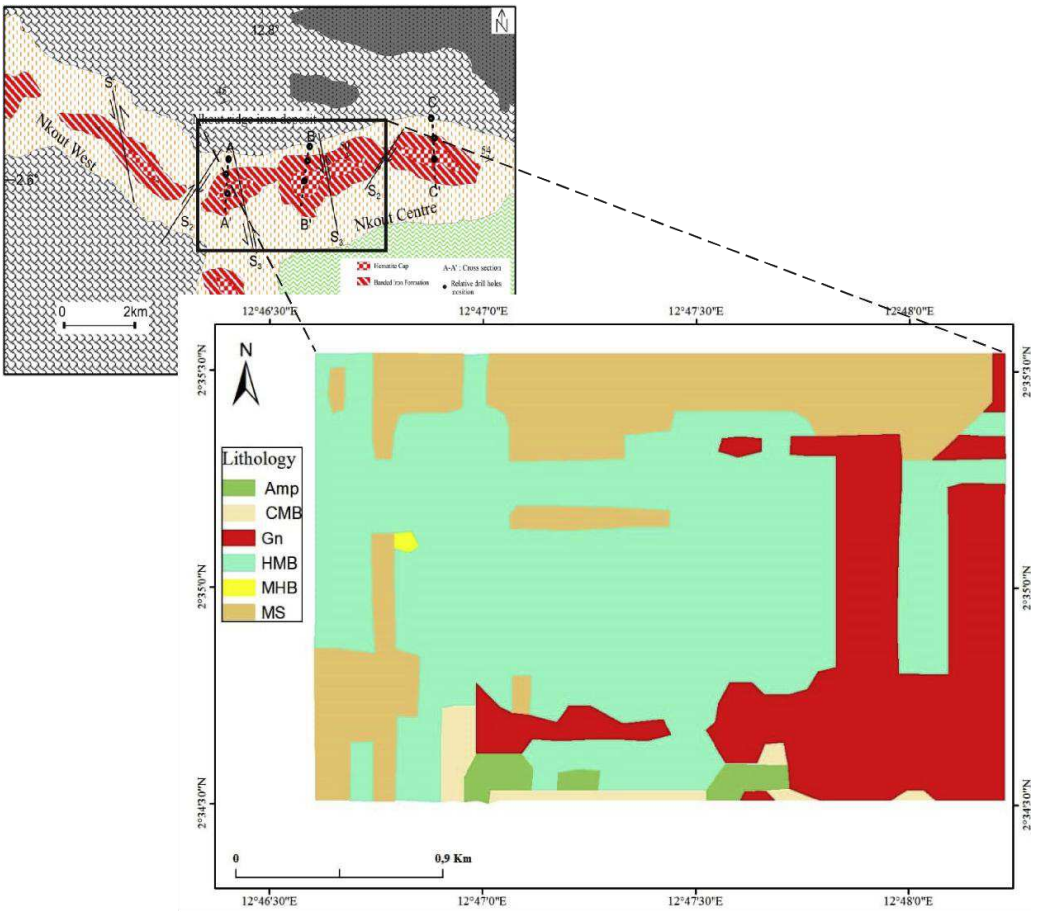

Figure 23: Geological map produced by k-NN

Conclusion

The Nkout iron deposit is located in granite, gneiss, and amphibolite-rich terrains of plutonic and metamorphic origin. Granular structures are typical of granites, but heterogranular granoblastic structures are found in metamorphic rocks. Compressive deformation (foliation and schistosity), as well as brittle deformation (fractures), had an impact on the research area. A compositional bedding, with the major attitude N150E, defines the foliation. Banded amphibolites were primarily impacted by schistosity. The fractures have a major N110E direction, and are posterior to the foliation and the schistosity. This study demonstrated the effectiveness of the k-Nearest Neighbours (k-NN) algorithm for modelling the nonlinear spatial correlations between geological data that already exist. With insufficient lithological data and high spatial dimensionality, this is for cartographic reasons. The evaluation of the machine learning findings shows that they perfectly align with field data. The cartography created allowed for the a priori targeting of regions that were conducive to the discovery of the iron deposit in the Nkout area. Therefore, our work can be seen as a tool that can drastically cut down on the risks, expenses, and delays related to exploration.

Acknowledgements

Thank to AcmeLab laboratory of Veritas office in Vancouver, Canada for the geochemical analyzes of the rock samples. The authors also thank the anonymous reviewers for improving the quality of work.

Authors contribution

In addition to AWB's coding of the prediction model, AWB and AKF's field work, AKD's petrological and mineralogical research, AA's geological mapping, and SKL's supply of drillhole data, there were other contributors. The NI oversaw the completion of every task.

Funding

No funding provided for this work.

Data availability

All data generated or analyzed in this study are included in the present paper.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that there is no competing interests related to this paper.

References

[1] An, P., Chung, C.-J.F., 1994. Neural network approach for geological mapping: technical background and case study. Can. J. Remote SensingJournal Can. Teledetection 20, 293–301.

[2] Yang, G., Collins, M., Gong, P., 1998. Multisource data selection for lithologic classification with artificial neural networks. Int. J. Remote Sens. 19, 3675–3680.

[3] Chakouri, M., El Harti, A., Lhissou, R., El Hachimi, J., Jellouli, A., 2020. Geological and mineralogical mapping in Moroccan central Jebilet using multispectral and hyperspectral satellite data and machine learning. Int J 9, 5772–5783.

[4] Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Comput. Geosci. 63, 22–33.

[5] El Fels, A.E.A., El Ghorfi, M., 2022. Using remote sensing data for geological mapping in semi-arid environment: a machine learning approach. Earth Sci. Inform. 1–12.

[6] Kumar, C., Chatterjee, S., Oommen, T., Guha, A., 2020. Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India. Int. J. Appl. Earth Obs. Geoinformation 86, 102006.

[7] Honarmand, M., Shahriari, H., 2021. Geological Mapping Using Drone-Based Photogrammetry: An Application for Exploration of Vein-Type Cu Mineralization. Minerals 11, 585.

[8] Sørensen, E.V., Guarnieri, P., 2018. Remote geological mapping using 3D photogrammetry: an example from Karrat, West Greenland. GEUS Bull. 41, 63–66.

[9] Bressan, T.S., de Souza, M.K., Girelli, T.J., Junior, F.C., 2020. Evaluation of machine learning methods for lithology classification using geophysical data. Comput. Geosci. 139, 104475.

[10] Costa, I.S.L., Tavares, F.M., de Oliveira, J.K.M., 2019. Predictive lithological mapping through machine learning methods: a case study in the Cinzento Lineament, Carajás Province, Brazil. J. Geol. Surv. Braz. 2, 26–36.

[11] Chattoraj, S.L., Prasad, G., Sharma, R.U., van der Meer, F.D., Guha, A., Pour, A.B., others, 2020. Integration of remote sensing, gravity and geochemical data for exploration of Cu-mineralization in Alwar basin, Rajasthan, India. Int. J. Appl. Earth Obs. Geoinformation 91, 102162.

[12] Thiele, S.T., Lorenz, S., Kirsch, M., Acosta, I.C.C., Tusa, L., Herrmann, E., Möckel, R., Gloaguen, R., 2021. Multi-scale, multi-sensor data integration for automated 3-D geological mapping. Ore Geol. Rev. 136, 104252.

[13] Radford, D.D., Cracknell, M.J., Roach, M.J., Cumming, G.V., 2018. Geological mapping in western Tasmania using radar and random forests. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 3075–3087.

[14] Grunsky, E.C., Mueller, U.A., Corrigan, D., 2014. A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: Applications for predictive geological mapping. J. Geochem. Explor. 141, 15–41. https://doi.org/10.1016/j.gexplo.2013.07.013

[15] Segalen, P., 1967. Les sols et la géomorphologie du Cameroun. Office de la Recherche Scientifique et Technique Outre-Mer.

[16] Ahanda, D.H.O., Ndzana, G.M., Bekoa, E., Abane, M.A.-A., Bitom, L.D., 2019. Morphological, geochemical and mineralogical studies of two soil profiles developed on the itabirites of Ntem Complex, southern Cameroon. J. Afr. Earth Sci. 153, 111–129.

[17] Nédélec A., Macaudiere J., Nzenti J. P., Barbey P., 1986. Evolution structurale et métamorphique des schistes de Mbalmayo (Cameroun). Implication sur la structure de la zone mobile panafricaine d´Afrique Centrale au contact du craton du Congo. *Comptes Rendus de l´Académie des Sciences,* 303, 75–80.

[18] Poidevin J. L., 1991. Les ceintures de roches vertes de la République Centrafricaine. Contribution à la connaissance du Précambrien du nord du craton du Congo. *Doctorat d'Etat, Université Blaise Pascal, Clermont-Ferrand (France)*.

[19] Nzenti J. P., Barbey P., Jegouzo P., Moreau C., 1984. Un nouvel exemple de ceinture granulitique dans une chaîne protérozoïque de collision : les migmatites de Yaoundé au Cameroun. *Comptes Rendus de l'Académie des Sciences, Paris,* 299, 1197-1199.

[20] Nzenti J. P., Barbey P., Macaudière J., Soba D., 1988. Origin and evolution of the late Precambrian high-grade Yaoundé gneisses (Cameroon). *Precambrian Research,* 38, 91–109.

[21] Penaye J., Toteu S. F., Van Schmus W. R., Nzenti J. P., 1993. U-Pb and Sm-Nd preliminary geochronologic data on the Yaoundé series, Cameroon: reinterpretation of

the granulitic rocks as the suture of a collision in the Centrafrican belt. *Comptes rendus de l'Academie des Sciences, Paris*, 317, 789-794.

[22] Nzenti J. P., Tchoua F. M., 1996. Les gneiss scapolitiques de la chaîne panafricaine Nord-Equatoriale au Cameroun : Témoins du précambrien d'une sédimentation évaporitique en bordure du craton du Congo. *Comptes Rendus de l'Académie des Sciences, Paris,* 323, 289-294.

[23] Vicat J. P., Pouclet A., Nkoumbou C., Seme Mouangue A. C., 1997. Le volcanisme fissural néoprotérozoïque des séries du Dja inférieur, de Yokadouma (Cameroun) et de Nola (RCA) - signification géotectonique. *Comptes Rendus de l'Académie des Sciences*, *Paris,* 325, 671-677.

[24] Vicat J. P., Moloto-A-Kenguemba G., Pouclet A., 2001. Les granitoïdes de la couverture protérozoïque de la bordure Nord du craton du Congo (Sud-Est du Cameroun et Sud-Ouest de la République Centrafricaine), témoins d'une activité magmatique post-kibarienne à pré-panafricaine. *Comptes Rendus de l'Académie des Sciences*, *Paris,* 332, 235-242.

[25] Nkoumbou C., Barbey P., Yonta-Ngouné C., Paquette J. L., Villiéras, F., 2014. Precollisional geodynamic context of the southern margin of the Pan-African fold belt in Cameroon. *Journal of African Earth Sciences,* 99, 245-260.

[26] Fix, E., Hodges, J.L., 1951. Nonparametric discrimination: Consistency properties. Randolph Field Tex. Proj. 21–49.

[27] Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27.

[28] Ganesan, K., Rajaguru, H., 2019. Performance analysis of KNN classifier with various distance metrics method for MRI images, in: Soft Computing and Signal Processing. Springer, pp. 673–682.

[29] Rodrigues, É.O., 2018. Combining Minkowski and Chebyshev: New distance proposal and survey of distance metrics using k-nearest neighbours classifier. Pattern Recognit. Lett. 110, 66–71.

[30] Lubis, A.R., Lubis, M., others, 2020. Optimization of distance formula in K-Nearest Neighbor method. Bull. Electr. Eng. Inform. 9, 326–338.

[31] Boroh, A., Sore-Gamo, K., Ayiwouo, N.M., Gbambie, M.I., Ngounouno, I., 2021. Implication of geological domains data for modeling and estimating resources from Nkout iron deposit (South-Cameroun). J. Min. Metall. Min. 57, 1–17.

[32] Hassanat, A.B., Abbadi, M.A., Altarawneh, G.A., Alhasanat, A.A., 2014. Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach. ArXiv Prepr. ArXiv14090919.

[33] Zhang, S., Li, X., Zong, M., Zhu, X., Cheng, D., 2017. Learning k for knn classification. ACM Trans. Intell. Syst. Technol. TIST 8, 1–19.

[34] Mathworks Inc., 2015. Matlab Classification Learner App.

[35] Townsend, J.T., 1971. Theoretical analysis of an alphabetic confusion matrix. Percept. Psychophys. 9, 40–50.

[36] Visa, S., Ramsay, B., Ralescu, A.L., Van Der Knaap, E., 2011. Confusion matrix-based feature selection. MAICS 710, 120–127.

[37] Gonçalves, L., Subtil, A., Oliveira, M.R., de Zea Bermudez, P., 2014. ROC curve estimation: An overview. REVSTAT-Stat. J. 12, 1–20.

[38] Hoo, Z.H., Candlish, J., Teare, D., 2017. What is an ROC curve? Emerg. Med. J.

[39] Ndime, E.N., Ganno, S., Soh Tamehe, L., Nzenti, J.P., 2018. Petrography, lithostratigraphy and major element geochemistry of Mesoarchean metamorphosed banded iron formation-hosted Nkout iron ore deposit, north western Congo craton, Central West Africa. J. Afr. Earth Sci. 148, 80–98. https://doi.org/10.1016/j.jafrearsci.2018.06.007