Coupling Discriminating Statistical Analysis and Artificial Intelligence for Geotechnical Characterization of the Kampemba's Municipality Soils (Lubumbashi, DR Congo)

Kavula Ngoy Elysée[1], Kasongo wa Mutombo Portance[2], Libasse Sow[3]* ⓘ, Ngoy Biyukaleza Bilez[4], Kavula Mwenze Corneille[5], Tshibwabwa Kasongo Obed[1]

[1]Civil Engineering Section, Higher Institute of Applied Techniques of Lubumbashi, Lubumbashi, Democratic Republic of Congo
[2]Department of Geology, Faculty of Science, University of Lubumbashi, Lubumbashi, Democratic Republic of Congo
[3]Civil Engineering Department, Higher Polytechnic School, Cheikh Anta Diop University of Dakar, Dakar, Senegal
[4]Department of Mines, Polytechnic Faculty, University of Lubumbashi, Lubumbashi, Democratic Republic of Congo
[5]Anvil Mining Congo, Kapulo Copper Project, Lubumbashi, Democratic Republic of Congo
Email: *libasse.sow@esp.sn

Abstract

This study focuses on the determination of physical and mechanical characteristics based on *in vitro* tests, by using field samples for the Kampemba urban area in the city of Lubumbashi. At the end of this study, we identified the soils according to their parameters, and established the geotechnical classification by determining their bearing capacity by the group index method using from the identification tests carried out. By using the AASHTO classification method (American Association for State Highway Transportation Official), the results obtained after our studies revealed five classes of soil: A-2, A-4, A-5, A-6, A-7 in a general way, and particularly eight subgroups of soil: A-2-4, A-2-6, A-2-7, A-4, A-5, A-6, A-7-5 and A-7-6 for the concerned area. The latter has given statistical analysis and deep learning based on multi-layer perceptron, the global values of the physical parameters. It's about: 31.77% ± 1.05% for the limit of liquidity; 18.71% ± 0.76% for the plastic limit; 13.06% ± 0.79% for the plasticity index; 83.00% ± 3.33% for passing of 2 mm sieve; 76.22% ± 3.2% for passing of 400 μm sieve; 89.07% ± 2.99% for passing of 4.75 mm sieve; 70.62% ± 2.39% passing of 80 μm sieve; 1.66 ± 0.61 for the consistency index; −0.67 ± 0.62 for the liquidity index and 8 ± 1 for the group index.

Keywords

Geotechnical Classification, Discriminant Factorial Analysis, Artificial

Intelligence, Deep Learning, Multi-Layer Perceptron

1. Introduction

To build a stable and sustainable road or civil engineering structures, it is imperative to treat the supporting soil in order to increase its bearing capacity. This stage of soil treatment can be accomplished only if the different physical characteristics of this soil are known. The study of the mechanical behavior of materials used in public works in general has been of interest to the scientific community for a very long time [1]-[9], among others. In geotechnical engineering, engineers use several systems to classify geomaterials by assigning them to each category by the similarity of their physical and/or mechanical properties a code.

The municipality of Kampemba consists of the Kafubu, Bel-Air 1 and 2, Bongonga, Industriel, Kigoma and Kapemba quarters, and has a total area of 9,741,644.63 m². This municipality is part of the square degree of the city of Lubumbashi city between the parallels 11˚36'5.60'' and 11˚44'44.91'' of Northing (South Latitude), and the meridians 27˚29'59.31'' and 27˚33'19.57'' of Easting (East Longitude).

The southern part of Katanga, which is part of our area of study, has a tropical climate with alternating two seasons, with a temperate and continental character linked to altitude but also to distance from the Indian and Oceanic Masses East and West 26, respectively [10] and annual rainfall is estimated at 1200 mm. The geology in place is described in Figure 1. This geological map is a necessary but not sufficient document to provide useful information to civil engineering professionals in their quest for good behavior soils to erect different structures. Hence, the detailed study of the soils and the alteration state of different parts of the massive rock in order to arrive at the presentation of the geotechnical map was the subject of this article. It will serve as a guide for the Kampemba's municipality civil engineers.

Several soil classification works have already been done for some municipalities in the city of Lubumbashi, including Kampemba [7] [11] [12] using different methods. This study was based on deep learning of inferential statistics and artificial intelligence based on the multilayer perceptron methods.

In the context of this article, the first objective is to arrive in the geotechnical classification of the Kampemba's municipality soils from the laboratory results. The second objective was to learn by deep learning the physical characteristics of the soils, and finally, the third objective was to identify the existing relationships between the physical parameters of the soils using statistical analyses and the perceptron multi-layer methods.

2. Materials and Methods

2.1. Sample Constitution

In situ sampling was based on the following criteria: soil color, moisture, consis-

tency, texture, structure and probable soil origin (natives? non-natives? nature of the bedrock?).

The sampling points were superimposed on the administrative background of Figure 2. Two hundred ninety-three (293) samples were collected. This sampling was based on the distribution of geological formations as shown on the geological map. For the learning of the system by artificial intelligence, 50% of samples were selected for learning and 50% for testing.

2.2. Determinant Predictive Variables

After the sampling campaign, there were followed by sieve analysis and consistency limit testing according to AFNOR standards (NF 1997-1, 2005) [13] to find all baseline data. Consistency or Atterberg limits are determined only for the fine elements of a soil, *i.e.* the fraction passing through the 400 μm sieve [13] or 420 μm [14], as these are the only elements on which water acts by changing the consistency of the soil [15]. In the case of this article, the sieve used was 400 μm. In practice, the liquidity limit (W_L) is defined as the water content from
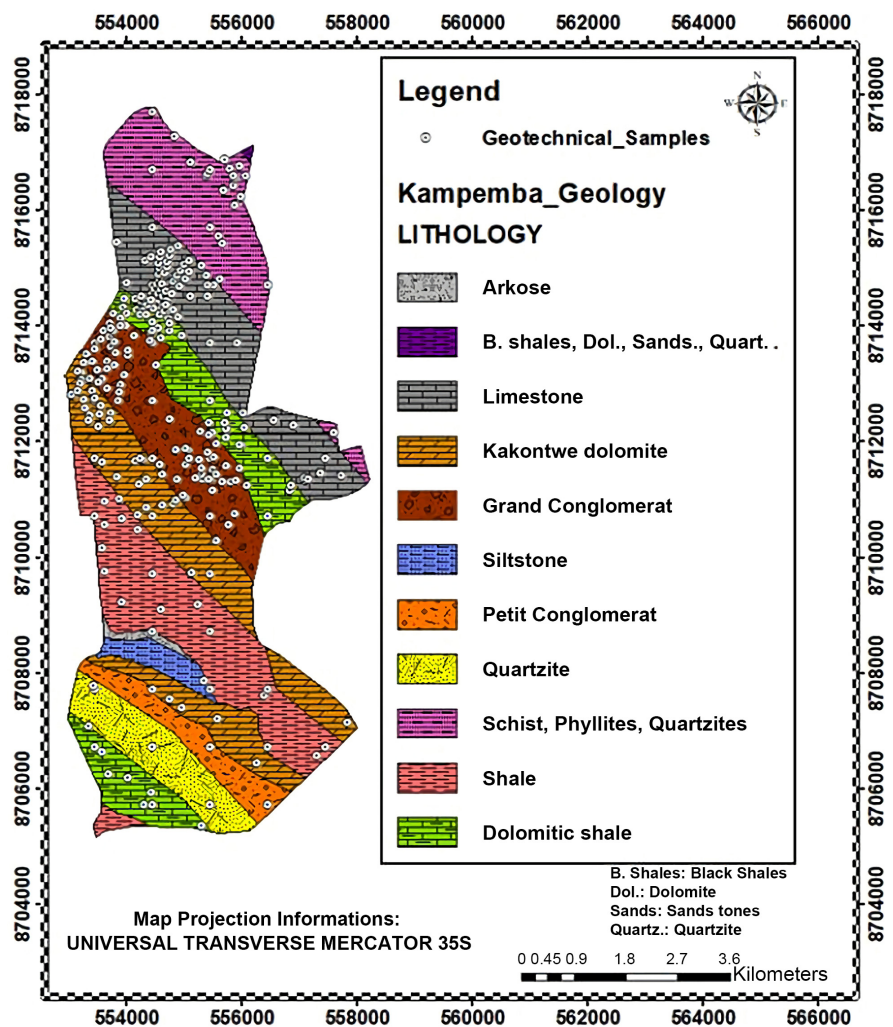

Figure 1. Geological map of the area of study showing representativity of sampling.
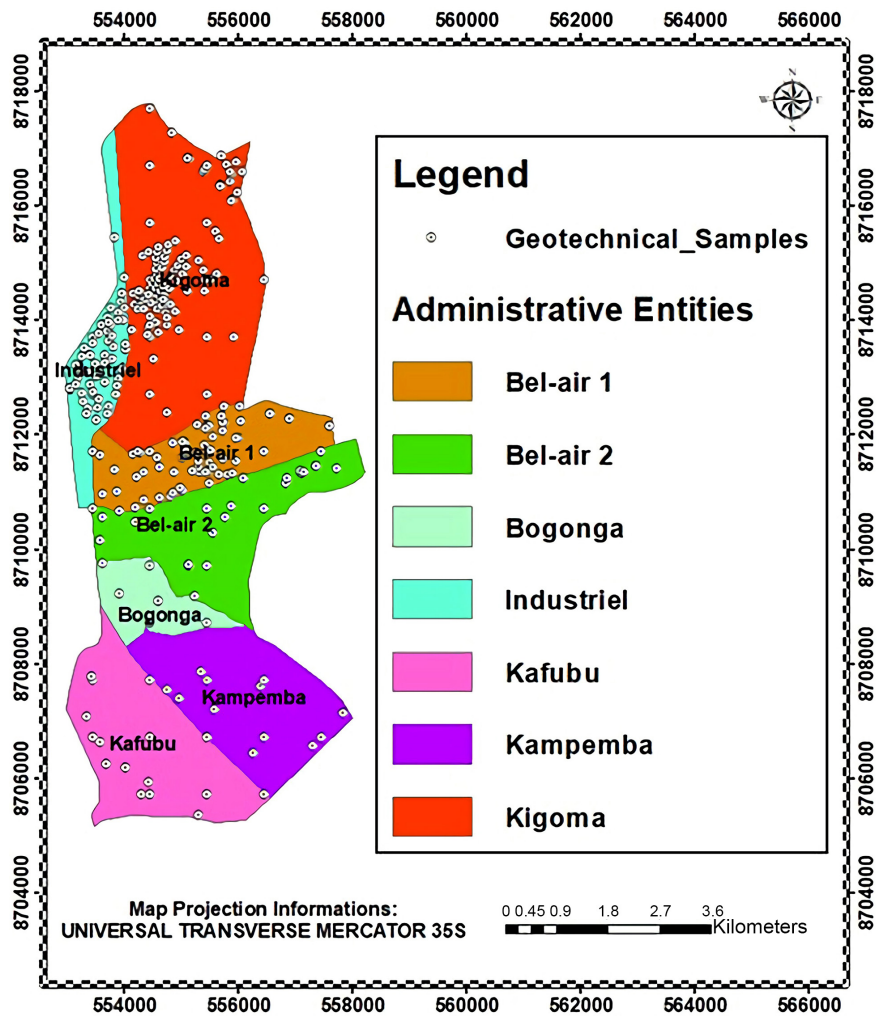
Figure 2. Map of sampling set out on administrative entities.

which a groove closes on 1 mm under 25 blows, the Casagrande cup was used for its determination. Its value was calculated by two methods:

- the analytical method (Equation (1) and Equation (2)),
- the graph based on the lower square right.

These technics were verified using deep learning based on the multilayer perceptron, used in artificial intelligence for regressions and predictions.

$$W_L = \omega \left(\frac{N}{25} \right)^{0.121}. \tag{1}$$

with: "ω" water content and "N" the number of blows.

The second analytical formula is that of the Washington State Highway Department [16] [17] in which the liquidity limit is determined by a single measure corresponding to a number of blows between 17 and 36.

$$W_L = \frac{\omega}{1.419 - 0.3 * \log_{10}^{N}}. \tag{2}$$

For this study, the results found with the two methods are almost similar, but

with a relative error of 1.97%.

In practice, the limit of plasticity (W_P) is the water content of the spindle that breaks into small sections of 1 to 2 cm long at the moment when its diameter reaches 3 mm. If the sample breaks at different diameters, several successive tests are made and the least square line is drawn to determine the water content corresponding to the 3 mm diameter.

The plasticity index "I_P", is the difference between the liquid limit and the plasticity limit. It gives an indication of the extent of the plasticity range. With this index, the soil can be classified according to its degree of plasticity.

The sieve analysis made it possible to calculate the following predictive variables: the content of fine particles (X_{80}), the sieve passing of 400 μm (X_{400}) and the sieve passing of 2 mm (X_2).

2.3. Geotechnical Classification of Soils

Burmisterin [16] has established the following classification for soils:
- $I_P = 0$: non-plastic soil;
- $1 \leq I_P < 5$: light plasticity soil;
- $5 \leq I_P < 10$: low plasticity soil;
- $10 \leq I_P < 20$: medium plasticity soil;
- $20 \leq I_P < 40$: high plasticity soil;
- $I_P > 40$: very high plasticity soil.

This rule was used in this work for the partial classification, as the materials are very heterogeneous and come from the alteration of several lithological and pedogenetics natures. For the global classification of soils, the method chosen was that of the AASHTO [18], which is based on sieve analysis, liquidity limit and plasticity limit such as:
- when the test results required for classification are available, the groups are examined from left to right by successive elimination. The first group as far as possible to the left, that corresponds to the group searched for;
- the plasticity index of subgroup A-7-5 is less than "$W_L - 30$" and the plasticity index of subgroup A-7-6 is higher than "$W_L - 30$"; note that this classification is completed by the group index method "I_g", which is calculated from the results of the sieve analysis, liquidity limit "W_L" and plasticity index "I_P" tests by the Equation (3).

$$I_g = \left(X_{200} - 35 \right)\left(0.2 + 0.005\left(W_L - 40 \right) \right) + 0.01\left(X_{200} - 15 \right)\left(I_P - 10 \right). \quad (3)$$

The calculation of this index defines the bearing capacity of a soil based on its identification tests. It can be used on the one hand to specify the classification of soils, and on the other hand, to evaluate the thickness of pavement sub-base layers according to the below formula of Steele in [16] without carrying out the CBR (Californian Bearing Ratio) tests [19]:

$$e\left[\text{cm} \right] = \frac{11}{4 * I_g} - \frac{1}{16 * I_g^2} . \quad (4)$$

According to Steele, the strength of a pavement foundation implies that its thickness depends on five factors [17]: nature of the subsoil, drainage, compaction, climate and the safety coefficient. The author gives the following classification as a function of the thickness of the subgrade:

- Null for a good subgrade (I_g = 0 or 1);
- 10 cm for a fair subsoil (I_g from 2 to 4);
- 20 cm for a bad subsoil (I_g from 5 and 9);
- 30 cm for a very bad basement (I_g from 10 to 20).

This thickness will have to be adapted to the conditions in Lubumbashi during the project study.

In order to avoid repetitive operations in data processing, the geotechnical mapping computer program based on deep learning: complex artificial neural networks [7] created according to the AASHTO classification was used, whose user interface is shown in **Figure 3**.

2.4. Analysis of the Data

For this article, two methods were used: statistical methods and artificial intelligence methods based on complex neural networks: Deep Learning.

2.4.1. Statistical Methods
1) Descriptive statistics

Figure 3. Software interface for geotechnical soil classification using artificial neural networks [7].

The statistical synthesizers used are: mean, first quartile, second quartile (median), third quartile, unbiased standard deviation, coefficient of variation, skewness, kurtosis and 95% confidence interval.

2) Inferential statistics

For this analysis, the following tests were used: the box plot test for detecting outliers, the Kolmogorov-Smirnov test for normality, and discriminant factor analysis to highlight trends. Explanation and prediction of the membership of individuals in several geotechnical classes based on the explanatory variables, quantitative or qualitative, was done using discriminant factor analysis (DFA). This method, which is both explanatory and predictive, can be used for:

- check on a two or three dimensional graph whether the groups to which the observations belong are distinct;
- identify the characteristics of the groups on the basis of explanatory variables;
- predict belonging group.

The analysis of covariance (ANCOVA), which is a method of modelling a quantitative dependent variable using quantitative and qualitative explanatory variables in the linear model, to model the correlations presented in the correlation and confusion matrices was used. Variables are said to be multi-collinear if there is a linear relationship between them. This is an extension of the simple case of collinearity between two variables.

While Discriminant Factor Analysis (DFA) can detect the presence of multi-collinearity in the data (the number of non-zero factors less than the number of variables indicates the presence of multi-collinearity), it does not identify the variables responsible for it.

In order to detect multi-collinearities and identify the variables involved, linear regressions of each of the variables in relation to the others must be carried out. It is then recommended to calculate:

- The R^2 of each of the models. If the R^2 is 1, then there is a linear relationship between the dependent variable of the model (the Y) and the explanatory variables (the X);
- The tolerance of each model is $(1 - R^2)$. It is used in several methods (linear regression, logistic regression, and discriminant factor analysis) as criteria for filtering variables. If a variable has a tolerance below a fixed threshold (the tolerance is calculated by taking into account variables already used in the model), it is not allowed to enter the model because its contribution is negligible and could lead to numerical problems;
- The VIF (Variance Inflation Factor) which is equal to the inverse of the tolerance.

It can be useful to detect multiple collinearities within a group of variables in particular in the following cases:

- To identify structures in the data and to derive operational decisions from them;
- To avoid numerical problems in some calculations.

2.4.2. Deep Learning Methods

A neural network is a mesh of several neurons organized by layers. The "S" neurons of a single layer are all connected to the "R" inputs. In this case, the layer is said to be fully connected. A weight "$w_{i,j}$" is associated to each connection. The first index "i" (row) designates the number of neuron on the layer, while the second index "j" (column) specifies the number of input. The set of weights in a layer forms a matrix "w" of dimensions "$S \times R$" (**Figure 4**). The mathematical model of an artificial neuron, shown in **Figure 4**, consists essentially of an integrator that performs a weighted summation of its inputs. The result "n" of this sum is then transformed by a transfer function "f" which produces the output "a" of the neuron. The "R" inputs of the neuron correspond to the vector $\boldsymbol{p} = \left[p_1 p_2 \cdots p_R \right]^{\mathrm{T}}$, while $\boldsymbol{w} = \left[w_{1,1} w_{1,2} \cdots w_{1,R} \right]^{\mathrm{T}}$ represents the vector of synaptic weight of the neuron [20] [21].

The output "n" of the integrator is given by the equations below:

$$\boldsymbol{n} = \sum_{j=1}^{R} \overset{*}{w}_{1,j} p_j - b. \tag{5}$$

with "b" the activation bias or threshold of the neuron and "n" the activation level that is then transformed by a transfer function "f" that produces the output "a" of the neuron (Equation (6)).

$$\boldsymbol{a}(t) = f\left(\boldsymbol{n}(t)\right) = f\left(w\boldsymbol{p} - b\right). \tag{6}$$

with "w" a matrix of synaptic weights and "t" time.

Several activation functions are used to solve different problems. Since this is more of a regression problem (approximation of functions) for this paper, the functions that have been used are the following:
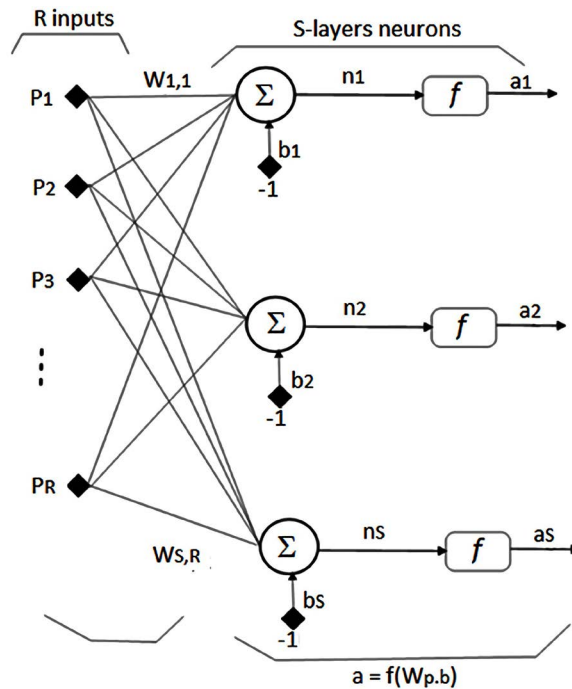

Figure 4. Layer with S neurons.

- linear;
- sigmoid;
- hyperbolic tangents.

For this type of problems, the multi-layer perceptron architecture was chosen. This is a static model, *i.e.* not considering time, because only the variables of soil's nature were used in the analysis. These variables are intrinsic characteristics of soils.

3. Results

3.1. Statistical Analysis

Before using advanced methods of analysis, it is first necessary to discover the data in order to identify trends, detect anomalies or simply to have essential information such as the minimum, maximum, or average of a sample of data [22]. **Table 1** and **Table 2** show the synthetic values founded. Without distinction of geotechnical classes, the soils of Kampemba have the following geotechnical properties given in **Table 1**. The Confidence Interval has been calculated according to student statistic at 95%.

Table 2 and **Table 3** provide details of all synthesizers of quantitative and qualitative variables.

The coefficient of variation of all variables is greater than 15%. This reflects significant variability in the sample (the materials are very heterogeneous); hence the mean alone is not a good summary of the whole sample. The distribution of values is asymmetric for all variables:

- the liquidity limit, the plasticity limit and all the sieve analysis parameters admit a left-hand asymmetry while;
- the plasticity and group index admit a right-hand asymmetry.
  With respect to the flattening of the curve:
- the liquidity limit, the plasticity limit, the plasticity index, the fine particle content, the sieve passing of 400 μm and 2 mm shows a platicurtic or hyponormal curve;
- the sieve passing of 4.75 mm shows a leptocurtic or hypernormal curve while;
- the group index shows a normal distribution.

Table 1. Synthesis of Kampemba's soil nature parameters.

Predictive variables	Average	Confidence Interval
W_L (%)	31.77	±1.05
W_P (%)	18.71	±0.76
I_P (%)	13.06	±0.79
X_{80} (μm)	70.62	±2.39
X_{400} (μm)	76.22	±3.2
X_2 (mm)	83.00	±3.33
$X_{4.75}$ (mm)	89.07	±2.99
I_g	8	±1

Table 2. Descriptive statistics of quantitative geotechnical variables.

Statistics	W_L	W_P	I_P	X_2 (mm)	X_{400} (µm)	$X_{4.75}$ (mm)	X_{80} (µm)	I_C	I_L	I_g
Number of observations	293	293	293	293	293	293	293	293	293	293
Number of missing values	0	0	0	0	0	0	0	0	0	0
Sum of weights	293	293	293	293	293	293	293	293	293	293
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	−8.394	−87.900	0.000
Maximum	57.810	38.460	45.700	100.000	100.000	100.000	100.000	88.900	9.394	45.000
Frequency of minimum	3	4	3	20	20	20	7	1	1	59
Frequency of maximum	1	1	1	55	21	103	1	1	1	1
Range	57.810	38.460	45.700	100.000	100.000	100.000	100.000	97.295	97.295	45.000
1st Quartile	26.570	14.780	9.350	78.991	66.533	95.548	63.352	0.833	−0.617	1.000
Median	32.750	18.600	13.210	98.672	86.862	99.786	75.000	1.214	−0.214	7.000
3rd Quartile	37.755	22.590	16.775	99.846	95.213	100.000	85.368	1.617	0.156	12.000
Sum	9307.781	5480.716	3827.065	24,317.702	22,332.138	26,098.290	20,692.495	487.004	−197.004	2328.000
Mean	31.767	18.706	13.062	82.996	76.219	89.073	70.623	1.662	−0.672	7.945
Variance (*n*)	82.781	43.334	47.075	834.030	773.649	675.135	432.516	28.661	28.637	48.161
Variance (*n* − 1)	83.064	43.483	47.236	836.886	776.298	677.447	433.997	28.759	28.735	48.326
Standard deviation (*n*)	9.098	6.583	6.861	28.880	27.815	25.983	20.797	5.354	5.351	6.940
Standard deviation (*n* − 1)	9.114	6.594	6.873	28.929	27.862	26.028	20.833	5.363	5.360	6.952
Variation coefficient	0.286	0.352	0.525	0.348	0.365	0.292	0.294	3.221	−7.959	0.873
Skewness (Pearson)	−0.437	−0.127	0.651	−1.831	−1.556	−2.804	−1.430	14.836	−14.850	1.145
Skewness (Fisher)	−0.439	−0.127	0.654	−1.840	−1.564	−2.818	−1.437	14.913	−14.926	1.151
Skewness (Bowley)	−0.105	0.022	−0.040	−0.887	−0.418	−0.904	−0.058	0.027	−0.042	−0.091
Kurtosis (Pearson)	0.773	0.742	1.939	2.223	1.454	6.619	2.057	237.893	238.181	2.615
Kurtosis (Fisher)	0.808	0.776	1.994	2.283	1.500	6.754	2.113	242.023	242.317	2.681
Standard error of the mean	0.532	0.385	0.402	1.690	1.628	1.521	1.217	0.313	0.313	0.406
Lower bound on mean (95%)	30.719	17.947	12.271	79.669	73.015	86.080	68.228	1.046	−1.289	7.146
Upper bound on mean (95%)	32.815	19.464	13.852	86.322	79.422	92.065	73.018	2.279	−0.056	8.745
Standard error of the variance	6.874	3.599	3.909	69.261	64.247	56.066	35.918	2.380	2.378	3.999
Lower bound on variance (95%)	71.080	37.209	40.421	716.147	664.301	579.711	371.384	24.610	24.589	41.354
Upper bound on variance (95%)	98.374	51.497	55.942	991.132	919.378	802.307	513.987	34.059	34.031	57.233

Table 3. Descriptive statistics of qualitative geotechnical variables.

Variables	Number of categories	Mode	Frequency mode	Categories
Classification according to argilosity and compressibility	7	Mineral clays with low plasticity	135	Mineral clays with low plasticity
				Mineral clays with medium plasticity
				Mineral clays with moderate plasticity
				Plastic mineral clays
				Mineral silts with intermediate compressibility
				Mineral silts with medium compressibility
				Mineral silts with low compressibility

Continued

Classification				
Classification according to plasticity	6	Medium plasticity flooring	172	Non-plastic soil
				Low plasticity soil
				Lightly plastic soil
				Medium plastic soil
				Highly plastic soil
				High plastic soil
Classification according to load-bearing capacity	4	Very bad basement	114	Good basement
				Bad basement
				Passable basement
				Very bad basement
Classification AASHTO	8	A-6	143	A-2-4
				A-2-6
				A-2-7
				A-4
				A-5
				A-6
				A-7-5
				A-7-6

The distribution of observations "box-plot" in Figure 5 represents the distance between Q1 and Q3 of the sample. The horizontal line inside the box represents the median and the "+" represents the mean. The vertical lines on each side of the box extend to the minimum and maximum values of the sample.

The limits at which data can be considered potentially outliers are represented by the lower limit $Q1 - 1.5(Q3 - Q1)$ and the upper limit $Q1 + 1.5(Q3 + Q1)$.

These box plots show almost no outliers in our statistical series. The analyzed variables are asymmetric due to a very high heterogeneity of the studied site's soils. Figures 6-8 show the frequency pie charts of the qualitative variables.

The clays contained in the soils of the Kampemba's municipality are mostly kaolinitic. Depending on the state of alteration and anthropic conditions, illites are also represented giving the soils a high plasticity. Note here the absence of montmorillonites in the soils of this sector. This mineralogical identification of the clays was carried out using the Casagrande abacus cited in numerous publications.

Figure 7 shows that the soils of Kampemba are very bad for foundation laying, they are the soils with medium plasticity. Only a few sites in this municipality with lateritic soils show good soils with low plasticity (Kigoma and Industrial quarters) (Figure 7).

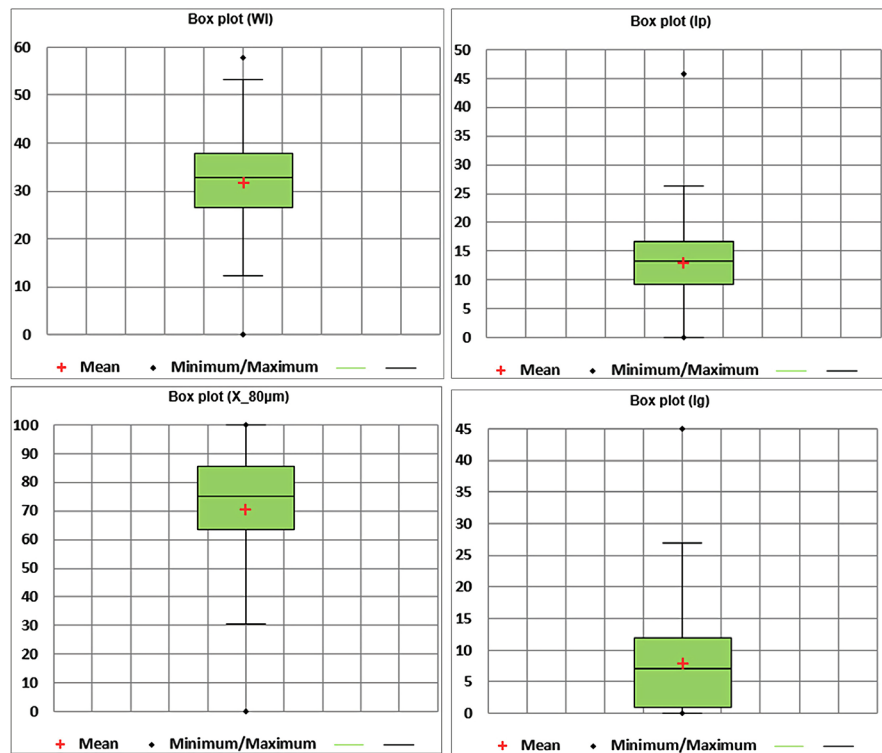According to the AASHTO classes, Kampemba's soils have the geotechnical

Figure 5. Mono variable box-plot diagrams showing mean, maximum and minimum values of used variables.

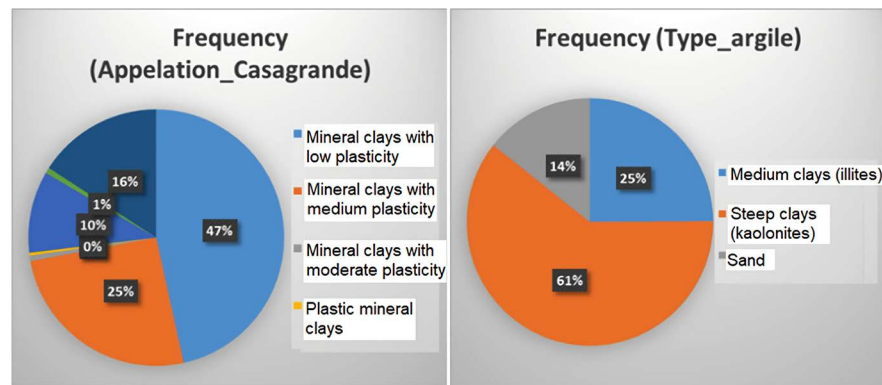**Figure 6.** Classification according to the nature of the clays contained in the soils of Kampemba.

properties given in Table 3. Soils of class A-6 constitute the modal class. Soils class of A-2 correspond to lateritic soils in this municipality.

Given the very high heterogeneity of these soils, a discriminant factor analysis (DFA) was carried out in relation to the AASHTO classification. The results are shown in Table 4.

With CI: Confidence Interval.

By geotechnical soil classes, some variables follow the normal law according to the Kolmogorov-Smirnov test:

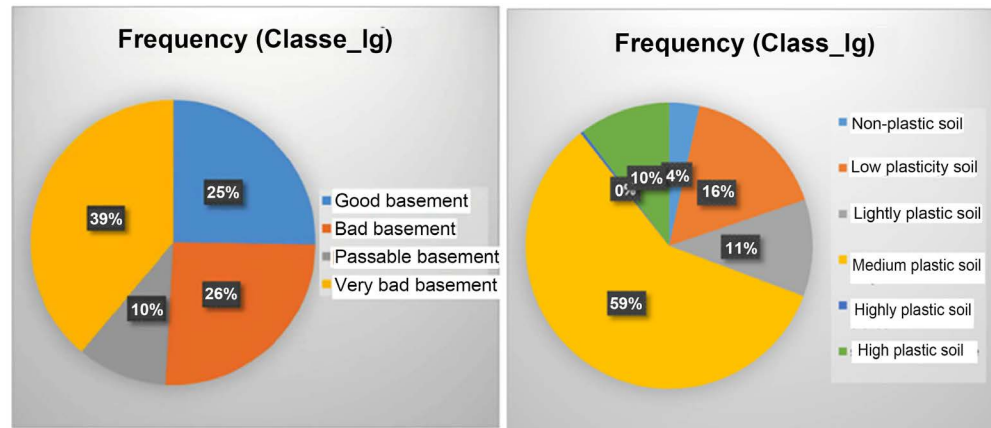o   Soils A-6: only the sieve analysis parameters do not follow the normal law;

Figure 7. Classification according to the plasticity and bearing capacity of Kampemba's soils.
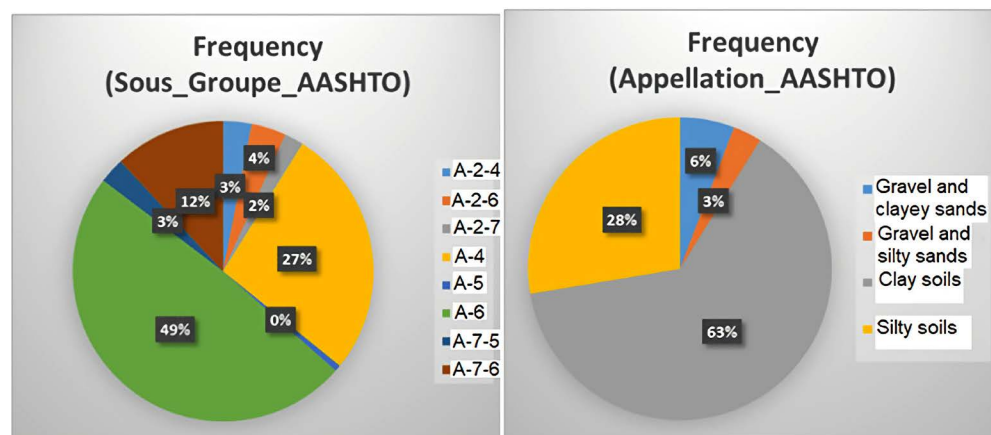

Figure 8. AASHTO classification of Kampemba's soils.

Table 4. Discriminant factor analysis by AASHTO class.

	W_L (%)		W_P (%)		I_P (%)		X_2 (%)		X_{400} (%)		$X_{4.75}$ (%)		X_{80} (%)		I_g	
	Value	CI	Value	CI	Value	CI	Value	CI	Value	CI	Value	CI	Value	CI	Value	CI
A-2-4	28.27	4.24	19.51	4.86	8.76	0.82	42.80	13.96	34.21	6.62	68.05	13.01	24.88	8.11	0	0
A-2-6	30.63	3.52	17.51	2.90	13.12	1.71	46.31	19.05	34.43	10.51	73.51	20.16	18.45	9.42	0	0
A-2-7	41.07	1.30	25.97	3.18	15.10	3.31	59.82	51.32	38.27	29.26	70.86	54.08	11.51	19.52	0	0
A-4	20.61	2.23	17.46	2.26	3.15	0.40	92.51	7.01	84.31	6.57	95.53	3.37	72.83	3.46	0	0
A-5	44.06	46.17	34.75	47.18	9.31	1.01	98.58	18.05	97.80	27.97	99.58	5.36	91.98	18.73	12	13
A-6	32.69	0.78	18.04	0.85	14.65	0.44	88.18	3.95	81.91	3.85	93.31	4.19	75.52	2.55	10	9
A-7-5	45.94	3.78	32.55	1.70	13.42	2.41	71.39	45.10	68.93	43.85	71.42	45.12	79.98	9.17	13	4
A-7-6	42.26	1.07	26.15	1.10	16.12	1.19	70.10	21.85	63.42	19.91	74.24	22.02	74.78	7.63	12	2

- Soils A-5, A-2-6, A-2-7: all parameters follow the normal law;
- Soils A-7-6: the parameters that follow Gauss law are W_P, I_P, X_{80} and I_g;
- Soils A-7-5: the parameters which follow the Gaussian law are W_L, W_P, I_P, X_{400}, X_{80} and I_g;

o Soils A-4 and A-2-4: no variable follows the normal law.

The first test carried out to detect correlations between the variables is the one based on the Pearson correlation matrix presented in Table 5.

Since the value of R^2 alone is not sufficient to demonstrate statistically significant correlations, deep learning is performed further down using artificial neural networks. These methods make it possible to approximate all the functions that may exist between the variables whatever the distribution law of each of them.

The results of the global statistics of multi collinearity are shown in Table 6. These results show the following collinear variables without distinction of geotechnical soil class:

- The passing of 2 mm and 400 μm (variance inflation factor VIF > 10);
- The liquidity and consistency index.

Multi-collinearity analysis by soil class shows that AASHTO classes with an index of 5 have several variables that are more self-correlated than others (Table 7).

3.2. Deep Learning of the Physical Characteristics of Kampemba's Soils

As defined in [20], a neuron is a bounded nonlinear function. This method has made it possible to test different functions. Only the function "Exponential Linear Unit elu: $f(x) = $ alpha $\times (\exp(x) - 1)$ for $x < 0$, $f(x) = x$ for $x \geq 0$" in the hidden layers and outputs gave a small error between known and predicted values (Table 8). By increasing the hidden layers and the number of neurons, performance deteriorate. The selected model is the one that used a hidden layer with eight neurons based on the AASHTO classification.

Table 5. Inter-variable correlation matrix using Pearson test.

	W_L	W_P	I_P	X_2 (mm)	X_{400} (μm)	$X_{4.75}$ (mm)	X_{80} (μm)	I_C	I_L	I_g
W_L	1									
W_P	0.660	1								
I_P	0.693	−0.084	1							
X_2 (mm)	−0.058	−0.080	0.000	1						
X_{400} (μm)	−0.059	−0.101	0.018	0.970	1					
$X_{4.75}$ (mm)	−0.081	−0.129	0.017	0.926	0.885	1				
X_{80} (μm)	0.047	−0.018	0.080	0.412	0.557	0.182	1			
I_C	−0.042	0.111	−0.162	−0.176	−0.171	−0.202	−0.014	1		
I_L	0.049	−0.105	0.165	0.177	0.172	0.204	0.013	−1.000	1	
I_g	0.674	0.100	0.798	0.203	0.286	0.082	0.572	−0.103	0.106	1

Table 6. Global multi-collinearity statistics.

X	Y	R^2	VIF	Tolerance
X_2 (mm)	X_{400} (μm)	0.94	17.20	0.06

Table 7. Multi-collinearity statistics by class AASHTO.

A-2-4				
X	Y	R^2	VIF	Tolerance
W_L	W_P	0.99	69.73	0.014
X_2 (mm)	X_{400} (µm)	0.90	10.02	0.100
A-2-6				
X	Y	R^2	VIF	Tolerance
X_2 (mm)	X_{400} (µm)	0.93	14.85	0.067
A-2-7				
X	Y	R^2	VIF	Tolerance
X_2 (mm)	X_{400} (µm)	0.93	14.26	0.070
A-4				
X	Y	R^2	VIF	Tolerance
X_2 (mm)	X_{400} (µm)	0.97	32.51	0.031
X_2 (mm)	$X_{4.75}$ (mm)	0.92	12.33	0.081
A-6				
X	Y	R^2	VIF	Tolerance
X_2 (mm)	X_{400} (µm)	0.93	14.46	0.069
A-7-5				
X	Y	R^2	VIF	Tolerance
W_L	I_P	0.95	19.89	0.050
X_2 (mm)	X_{400} (µm)	0.99	75.03	0.013
X_2 (mm)	$X_{4.75}$ (mm)	1.00	221,566.25	0.000
X_{400} (µm)	$X_{4.75}$ (mm)	0.99	75.31	0.013
A-7-6				
X	Y	R^2	VIF	Tolerance
X_2 (mm)	X_{400} (µm)	0.97	35.22	0.028
X_2 (mm)	$X_{4.75}$ (mm)	0.94	17.71	0.056
X_{400} (µm)	$X_{4.75}$ (mm)	0.92	11.79	0.085

Table 8. Regressions with multilayer perceptron.

Input (X)	Output (Y)	RMSE [%]	Model
W_L	W_P	49	Figure 9
W_L, I_P X_{80}	I_g	42: after 50 iterations 38: after 1500 iterations	Figure 10

The model in Figure 9 shows that: A-4 soils have inhibitory activity due from their very low to zero plasticity, which leads to an increase in bias; plastic soils (those with index 6 and 7) enhance synaptic activities.
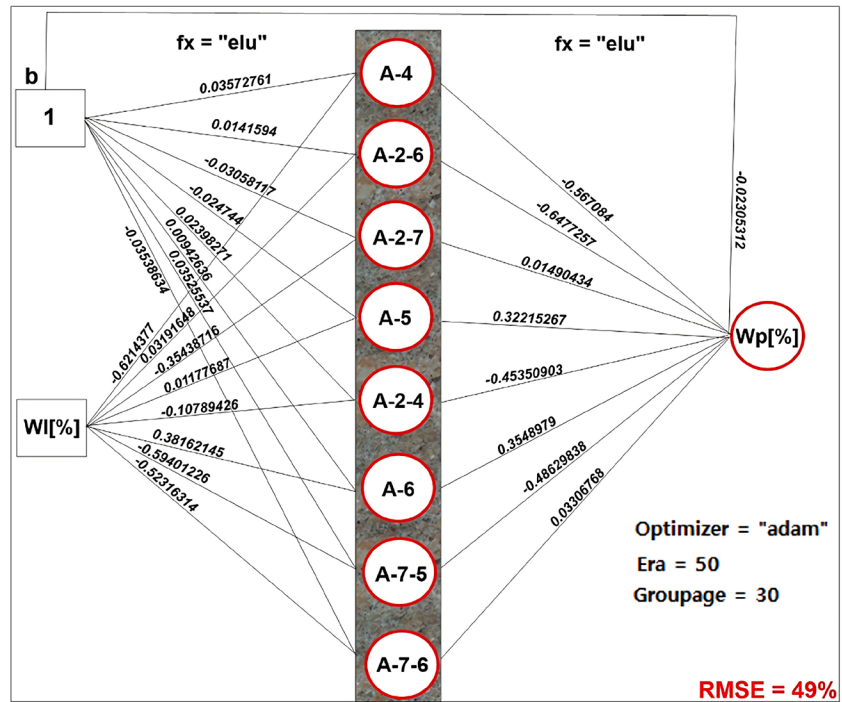
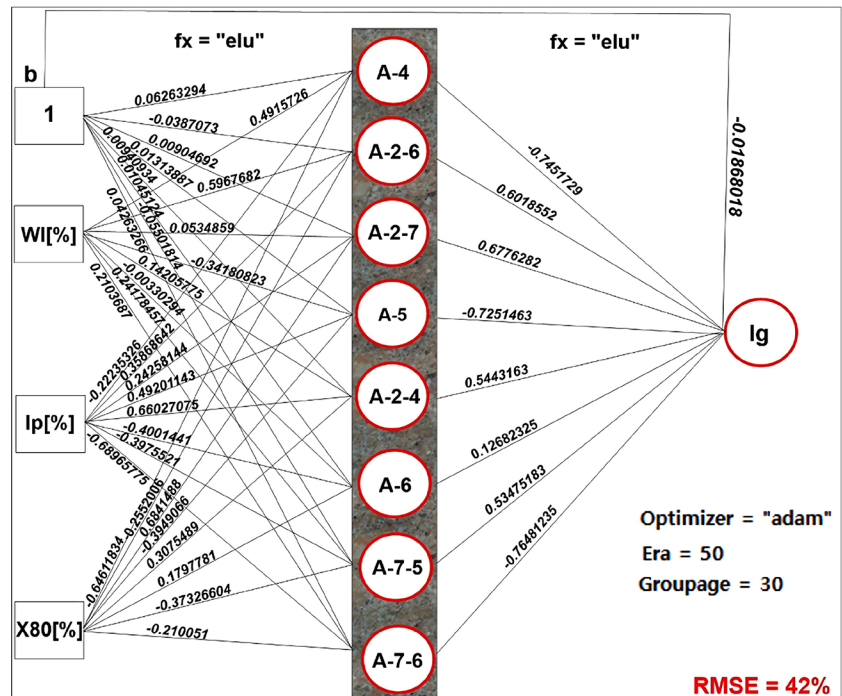**Figure 9.** Synaptic weights of the model $W_P = f(W_L)$.

Figure 10. Synaptic weights of the model $I_g = f(W_L, I_P, X_{80})$.

The model in **Figure 10** shows that: soil consistency characteristics have an important activity in the evaluation of the group index; granularity has a predominant effect on soil classes with an index of 4; network performance is improved by removing data from geotechnical classes that inhibit synaptic activities.

These methods made it possible to confirm the different correlations identified between the variables by the statistical methods.

3.3. Geotechnical Mapping of the Area

Figure 11 shows the final geotechnical map of the municipality of Kampemba after all statistical and deep learning analyses.

With the help of the AASHTO classification, the soils of Kampemba are divided into eight major groups, which are identified in detail in the following geotechnical classes arranged according to the increasing group index: The statistical analyses made it possible to determine the values of various parameters taken into account in the study as presented below:

- A-2-4 soils: this material is classified as silty gravels and sands (medium or steep clay type with the presence of illites and kaolinites). It is a soil with low plasticity and good subsoil for construction according to its group index, and having the following " W_L " as parametric values: 28.27% ± 4.24%; W_P: 19.51% ± 4.86%; I_P: 8.76% ± 0.82%; X_2 (mm): 42.80% ± 13.96%; X_{400} (μm): 34.21% ± 6.62%; $X_{4.75}$ (mm): 68.05% ± 13.01%; X_{80} (μm): 24.88% ± 8.11%; I_g: 0;
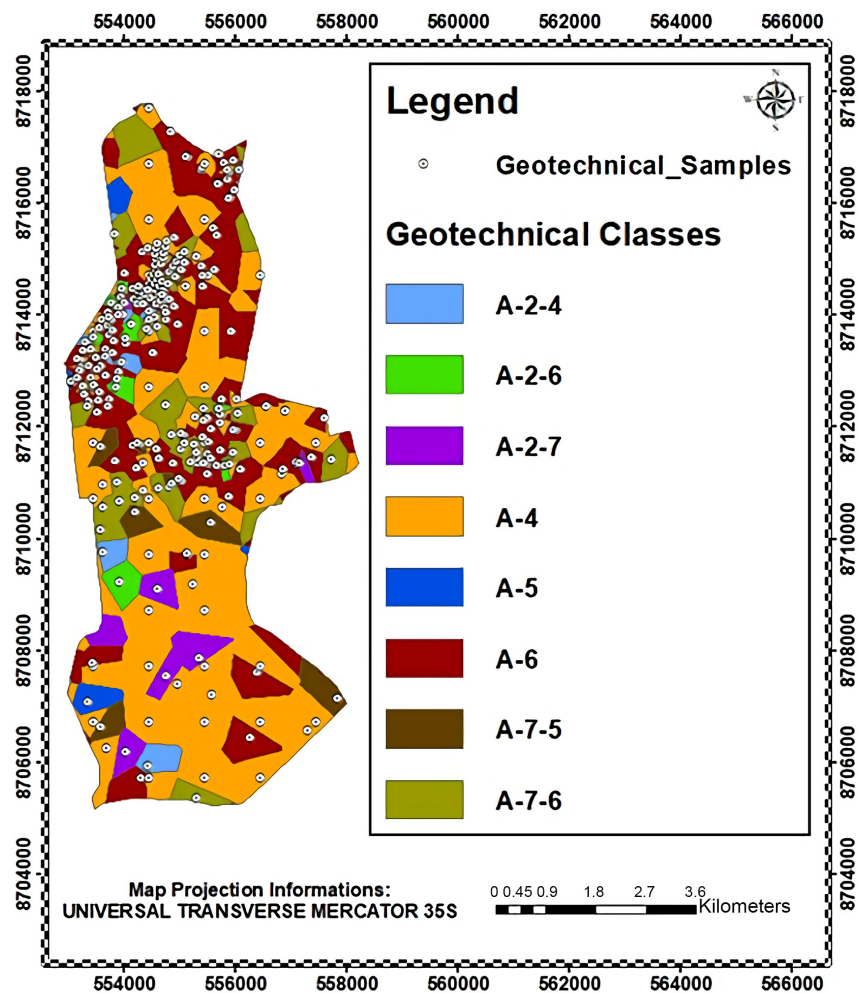

Figure 11. Final geotechnical map of the area of study.

- Soils A-2-6: this material is classified as clayey gravel and sand (a type of stiff clay with the presence of kaolinites). It is a soil with average plasticity and good subsoil according to its group index, and having the following W_L as parametric values: 30.63% ± 3.52%; W_P: 17.51% ± 2.90%; I_P: 13.12% ± 1.71%; X_2 (mm): 46.31% ± 19.05%; X_{400} (µm): 34.43% ± 10.51%; $X_{4.75}$ (mm): 73.51% ± 20.16%; X_{80} (µm): 18.45% ± 9.42%; I_g: 0;

- Soils A-2-7: this material is classified as clayey gravels and sands (medium or steep clay type with the presence of illites and kaolinites). It is a soil with low plasticity and good subsoil according to its group index, and having the following W_L as parametric values: 41.07% ± 1.30%; W_L: 25.97% ± 3.18%; I_P: 15.10% ± 3.31%; X_2 (mm): 59.82% ± 51.32%; X_{400} (µm): 38.27% ± 29.26%; $X_{4.75}$ (mm): 70.86% ± 54.08%; X_{80} (µm): 11.51% ± 19.52%; I_g: 0;

- Soils A-4: this material is classified as silty soils (consisting of sand and stiff clay with the presence of kaolinites). It is a soil with sometimes low plasticity (bad subsoil), sometimes light, and even non-plastic, considered as good subsoil according to its group index, and having W_L as parametric values: 20.61% ± 2.23%; W_P: 17.46% ± 2.26%; I_P: 3.15% ± 0.40%; X_2 (mm): 92.51% ± 7.01%; X_{400} (µm): 84.31% ± 6.57%; $X_{4.75}$ (mm): 95.53% ± 3.37%; X_{80} (µm): 72.83% ± 3.46%; I_g: 0;

- Soils A-5: this material is classified as silty soils (medium clay type with illites present). It is a soil with low plasticity and very poor subsoil according to its group index, and with the following parametric values: 44.06% ± 46.17%; W_P: 34.75% ± 47.18%; I_P: 9.31% ± 1.01%; X_2 (mm): 98.58% ± 18.05%; X_{400} (µm): 97.80% ± 27.97%; $X_{4.75}$ (mm): 99.58% ± 5.36%; X_{80} (µm): 91.98% ± 18.73%; I_g: 12 ± 13;

- Soils A-6: this material is classified as clayey soils (composed of medium or steep clay with the presence of illites and kaolinites). It is a soil with sometimes high plasticity (very bad subsoil) and sometimes average considered as bad subsoil according to its group index, and having W_L as parametric values: 32.69% ± 0.78%; W_P: 18.04% ± 0.85%; I_P: 14.65% ± 0.44%; X_2 (mm): 88.18% ± 3.95%; X_{400} (µm): 81.91% ± 3.85%; $X_{4.75}$ (mm): 93.31% ± 4.19%; X_{80} (µm): 75.52% ± 2.55%; I_g: 10 ± 1;

- Soils A-7-5: this material is classified as a clay soil (type of plastic clay with the presence of illites). It is a medium plasticity soil. They constitute very poor subsoil according to its group index, and have the following the parametric values: W_L: 45.97% ± 3.78%; W_P: 32.55% ± 1.70%; I_P: 13.42% ± 2.41%; X_2 (mm): 71.39% ± 45.10%; X_{400} (µm): 68.93% ± 43.85%; $X_{4.75}$ (mm): 71.42% ± 45.12%; X_{80} (µm): 79.98% ± 9.17%; I_g: 13 ± 9;

- Soils A-7-6: this material is classified as a clay soil (type of plastic clay like A-7-6). It is a soil of medium to high plasticity, considered as very bad subsoils according to its group index. They are characterized by: W_L: 42.26% ± 1.07%; W_P: 26.15% ± 1.10%; I_P: 16.12% ± 1.19%; X_2 (mm): 70.10% ± 21.85%; X_{400} (µm): 63.42% ± 19.91%; $X_{4.75}$ (mm): 74.24% ± 22.02%; X_{80} (µm): 74.78% ± 7.63%; I_g: 10 ± 2.

4. Conclusions

According to the AASHTO classification, the soils of the municipality of Kampemba are divided into five major groups (A-2, A-4, A-5, A-6 and A-7) which are further subdivided into eight subgroups with the following geotechnical characteristics:

o Soils A-2-4: this material is classified as gravelly sands with silt having a group index of $I_g = 0$;

o Soils A-2-6: this material is classified as gravelly sands to clays with a group index of $I_g = 0$;

o Soils A-2-7: this material is classified as gravelly sands with active clays with a group index of $I_g = 0$;

o Soils A-4: this material is classified as loamy to sandy soils with a group index of $I_g = 0$;

o Soils A-5: this material is classified as plastic loamy to sandy soils with a group index of $I_g = 12 \pm 13$;

o Soils A-6: this material is classified as clayey to sandy soils with a group index of $I_g = 10 \pm 1$;

o Soils A-7-5: this material is classified as an active clayey to sandy soil with a group index of $I_g = 13 \pm 9$;

o Soils A-7-6: as above, this material is also classified among active clayey to sandy soils with a group index of $I_g = 10 \pm 2$.

A-2 are good quality soils, *i.e.* good for supporting the foundations of the various civil engineering works. Soils of groups A-4 and A-5 are also good for laying foundations but are remarkably vulnerable to liquefaction, erosion and leaching, which will require treatment to ensure safe operation. However, soils of Groups A-6 and A-7 (A-7-5 and A-7-6) are not good soils for foundation, but may still have other advantages as they form impermeable substrates. It should be noted that A-6 soils are good substrates if their group index is low. Among other things, they can be considered as a usable reserve of clay for pottery, as a source for the production of ceramic materials, stabilised clay bricks and refractory materials.

Artificial intelligence methods based on deep learning using the multilayer perceptron have confirmed the inter-variable correlations identified by statistical methods. These techniques are widely used now in engineering, because they allow detecting any functions linking the variables. Statistical methods are often limited to linear correlations using the Pearson matrix, *i.e.* limited to two variables, one as input (*X*) and the other as output (*Y*). The artificial intelligence technic shows another advantage of identifying geomaterials from non-quantitative dataset based on labels of training samples.

The mechanical behavior of civil engineering infrastructures is correlated with the behavior of the materials making up the supporting soil. It is therefore very important to identify the supporting soil and to control its behavior. The present work is therefore an important scientific tool in this sense. One of the scientific

interests of this study is to contribute to the identification and choice of quality materials that can be used in public works in the area of the study.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Samb, F., Berthaud, Y., Ba, M., Fall, M. and Benboudjema, F. (2018) Nonlinear Mechanical Behavior Analysis of Flexible Lateritic Pavements of Senegal (West Africa) by FEM for M.-E. Pavement Design. *Geotechnical and Geological Engineering*, **36**, 2939-2956. https://doi.org/10.1007/s10706-018-0514-y

[2] Diedhiou, A., Sow, L. and Dione, A. (2020) Contribution to Comparative Study of Physical-Chemical Characteristics of Diack Basalt and Bandia Limestone for Use in Railway Engineering. *Geomaterials*, **10**, 25-34. https://doi.org/10.4236/gm.2020.102002

[3] Sow, L., Bernard, F., Kamali-Bernard, S. andKébé, C.M.F. (2018) Experiment-Based Modelling of the Mechanical Behaviour of Non-Hazardous Waste Incineration Bottom Ashes Treated by Hydraulic Binder. *MATEC Web of Conferences*, **149**, Article ID: 01038. https://doi.org/10.1051/matecconf/201814901038

[4] Sow, L., Bernard, F., Kamali-Bernard, S. and Kébé, C.M.F. (2018) Mesoscale Modeling of the Temperature-Dependent Viscoelastic Behavior of a Bitumen-Bound Gravels. *Coupled Systems Mechanics*, **7**, 509-524.

[5] Sow, L., Kamali-Bernard, S., Bartier, O., Mauvoisin, G. and Bernard, F. (2018) Experimental Estimation of the Elastic Modulus of Non-Hazardous Waste Incineration Bottom Ash Aggregates by Indentation Tests—Microanalysis of Particles by Scanning Electron Microscopy. *Advanced Materials Research*, **1145**, 80-84. https://doi.org/10.4028/www.scientific.net/AMR.1145.80

[6] Sow, L., Kamali-Bernard, S., Mauvoisin, G., Bartier, O. and Bernard, F. (2019) Original Experimental Campaign of Indentation Instrumented on Aggregates of Non-Hazardous Waste Incineration Bottom Ash to Study the Heterogeneity of Their Rigidity. *Key Engineering Materials*, **805**, 177-182. https://doi.org/10.4028/www.scientific.net/KEM.805.177

[7] Kasongo, W.M.P., Kavula, N.E., Sow, L. and Lunda, I.J.M. (2019) Cartographie Géotechnique par Deep Learning Approche par Réseaux de Neurones Artificiels. *European Scientific Journal*, **15**, 233-251. https://doi.org/10.19044/esj.2019.v15n12p233

[8] Sow, L. (2019) Ballasted Railways in Senegal—Characterization of Bandia Limestone and Diack Basalt for Use as Ballast Materials. *International Journal of Applied Engineering Research*, **14**, 3396-3405. https://www.ripublication.com/ijaer19/ijaerv14n15_10.pdf

[9] Sow, L. (2020) Study of the Behaviour of Senegalese Ballast Materials during Compaction with the C-Mould: Case of Bandia Limestone and Diack Basalt. *Key Engineering Materials*, **831**, 81-86. https://doi.org/10.4028/www.scientific.net/KEM.831.81

[10] François, A. (1987) Synthèse géologique sur l'arc cuprifère du Shaba (République du Zaïre). Société Belge de Géologie.

[11] Kasongo, W.M.P. (2017) Mise en place d'un système d'informations géotechnique.

Editions Universitaires Européennes, 192 p.

[12] Kasongo, W.M.P., Mukoko, K.G., Kipata, M.L. and Lunda, I.J.M. (2018) Élaboration de la carte géotechnique de Lubumbashi: Guide technique de sélection des sites d'implantation d'ouvrages du génie civil. *European Scientific Journal*, **14**, 407-431. https://doi.org/10.19044/esj.2018.v14n36p407

[13] AFNOR, Eurocode 7 (2005) Calcul géotechnique. Partie 1: Règles générales. NF EN 1997-1.

[14] Garcia-Gaines, A. and Frankenstein, S. (2015) USCS and the USDA Soil Classification System. US Army Corps of Engineers, 46 p. https://doi.org/10.21236/ADA614144

[15] Costet, J. and Sanglerat, G. (1975) Cours pratique de mécanique des sols, plasticité et calcul des tassements. Tome 1Dunod, 2ème Edition, 262 p.

[16] Verdeyen, J., Roisin, V. and Nuyens, J. (1968) La mécanique des sols. Tome 2Dunod, 508 p.

[17] Verdeyen, J., Roisin, V. and Nuyens, J. (1971) Application de la mécanique des sols. Tome 2 Dunod, 483 p.

[18] USDA (1987) Soil Mechanics Level I, Module 2: AASHTO (American Association of State Highway and Transportation Officials), Study Guide. Soil Conservative Service, 71 p.

[19] Kasongo, W.M.P., Kavula, N.E., Tshinkobo, K.G. and Ngoy, B.B. (2019) Geotechnical Reconnoitering of the KAMATANDA Platform-Building Site by Dynamic Cone Penetration Test (DCPT). *International Journal of Innovative Research in Advanced Engineering*, **6**, 573-581.

[20] Parizeau, M. (2004) Réseaux de neurones. GIF-21140 et GIF-64326, Université Laval, 117 p.

[21] Dreyfus, G.J., Martinez, M., Samuelides, M., Gordon, M.B., Badran, F., Thiria, S. and Hérault, L. (2004) Réseaux de neurones, Méthodologie et Applications. Deuxième Edition, Eyrolles.

[22] Addinsoft (2015) XLSTAT Tutorial. Paris, 1415 p.