journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Risk groups in children under six months of age using self-organizing maps

A.O.C. Schilithz[a,*], P.L. Kale[b], S.G.N. Gama[c], F.F. Nobre[a]

[a] PEB/UFRJ, Rio de Janeiro, Brazil
[b] IESC/UFRJ, Rio de Janeiro, Brazil
[c] ENSP/FIOCRUZ, Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Received 22 July 2013
Received in revised form
22 January 2014
Accepted 20 February 2014

Keywords:
Infant growth
Artificial neural network
Self-organizing maps
Cesarean delivery
Preterm birth
Exclusive breastfeeding

ABSTRACT

Fetal and infant growth tends to follow irregular patterns and, particularly in developing countries, these patterns are greatly influenced by unfavorable living conditions and interactions with complications during pregnancy. The aim of this study was to identify groups of children with different risk profiles for growth development. The study sample comprised 496 girls and 508 boys under six months of age from 27 pediatric primary health care units in the city of Rio de Janeiro, Brazil. Data were obtained through interviews with the mothers and by reviewing each child's health card. An unsupervised learning, know as a self-organizing map (SOM) and a *K*-means algorithm were used for cluster analysis to identify groups of children. Four groups of infants were identified. The first (139) consisted of infants born exclusively by cesarean delivery, and their mothers were exclusively multiparous; the highest prevalences of prematurity and low birthweight, a high prevalence of exclusive breastfeeding and a low proportion of hospitalization were observed for this group. The second (247 infants) and the third (298 infants) groups had the best and worst perinatal and infant health indicators, respectively. The infants of the fourth group (318) were born heavier, had a low prevalence of exclusive breastfeeding, and had a higher rate of hospitalization. Using a SOM, it was possible to identify children with common features, although no differences between groups were found with respect to the adequacy of postnatal weight. Pregnant women and children with characteristics similar to those of group 3 require early intervention and more attention in public policy.

1. Introduction

Infant growth during the early months of life is affected by several factors, including ethnic differences, intrauterine development [1,2], genetic patterns, the consumption of foods and cultural, environmental, and social factors [1,3–5]. Unfavorable living conditions and their interactions with complications during pregnancy are known to be associated with poor fetal growth and child development in developing countries [6–8]. Additionally, there is evidence that rapid growth during the first six months of life, associated with several socioeconomic factors and maternal characteristics, is an important predictor of childhood obesity [9,10]. The assessment

* *Corresponding author*. Tel.: +55 2127521706.
E-mail address: arthur@peb.ufrj.br (A.O.C. Schilithz).

http://dx.doi.org/10.1016/j.cmpb.2014.02.011
0169-2607/

of growth is an important component for monitoring child development. The detection of changes in the rate of growth and weight gain is important markers of health in early childhood [11,12]. Anthropometric measurements, such as length and weight, are commonly used to evaluate the nutritional status of children under two years of age [13]. Atypical growth patterns may characterize children that are at nutritional risk, and the identification of those children may help prevent inadequate growth.

Several growth charts have been developed for monitoring child growth [14,15]. Recently, the World Health Organization [16] developed a growth chart using information from five continents through a multicenter study with children aged zero to five years, regardless of ethnic origin, socioeconomic status, or type of food. This multicenter study included data from Brazil, representing children in Latin America. Growth curves are primarily used for monitoring and assessing the health and development of children and infants, and they are known to be a cost-effective approach for identifying nutritional deficiencies. The associations of socioeconomic indicators, maternal characteristics and prenatal care with delays in development, including physical growth, are most often mediated by poor nutrition [17,18]. Thus, growth, which is a proxy for nutritional status, is correlated with socioeconomic status, maternal indicators, and prenatal care.

Most studies that investigate infant health and development assess growth failure and related factors by simply characterizing infant growth changes and by comparing the infant's growth with standard growth charts [19,20]. A recent study that evaluated the potential of geographically targeted nutrition programs to reduce the number of infants at nutritional risk concluded that the childhood nutritional status is determined by individual and household factors [21]. The study suggested that new interventions focusing on individual needs are required. Nutritional risk factors tend to cluster in individuals, and predicting which children will be at risk using factors such as socioeconomic indicators, maternal indicators, and prenatal care can help health services develop more efficient preventive measures. One approach for identifying groups of infants with similar risk factors is to use clustering techniques.

Clustering is based on the idea that infants that have similar risk factors should exhibit similar growth patterns. This relationship allows for the possibility that more than one pattern of infant growth will be established and associated with different profiles of socioeconomic and maternal indicators.

Among the available clustering techniques, the Kohonen network [22], which is an unsupervised artificial neural network, also known as a self-organizing map (SOM), is recognized as effective due to its ability to visualize multidimensional data in a low-dimensional space and to extract the essential features of a complex dataset by generating prototype vectors. This technique has been widely used in pattern recognition problems in engineering, and more recently, it has received attention in the field of epidemiology, with applications involving the clustering of patients with infectious diseases, such as dengue fever [23], and patients with chronic diseases [24–26].

The purpose of this study was to identify groups of children with different maternal characteristics, prenatal care factors, and socioeconomic indicators that might be of importance in characterizing growth development. The aim of this report also is to illustrate how a machine learning technique can provide a relatively simple framework with which to visualize and interpret multidimensional data relevant to child growth and nutritional problems.

2. Methods

2.1. *Sample design*

The data were obtained from a cross-sectional study that evaluated 1082 children under six months of age who required pediatric care in 27 primary health care units of the Brazilian Unified Health System (SUS) in the city of Rio de Janeiro from June to September 2007. This study was conducted by researchers from the National School of Public Health, Oswaldo Cruz Foundation.

The sampling plan included two stages. In the first stage, 27 health care units were selected based on both the Euclidean distance between the unit and the administrative center of the municipality of Rio de Janeiro and the cumulative frequency of the monthly average of pediatric consultations for children under six months of age during the first half of 2005. In the second stage, we used a systematic sampling strategy, selecting from the list of children enrolled for routine pediatric evaluation. From each health unit 40 children under six months of age were selected during the period of January to June 2007, resulting in a total of 1080 interviews.

2.2. *Collected variables*

The data were initially obtained from each child's health booklet and supplemented through interviews with the mother. The use of child health booklets was implemented by the Brazilian Ministry of Health in 2005, and these booklets contains information on events related to child health, including obstetric history and neonatal indicators of growth and development; information on breastfeeding; and clinical complications. The booklet is intended to be used for all those born in Brazil and is an appropriate instrument to monitor the health of the child. Once we had all the information recorded in a database, the database was linked to SINASC (Information system of live births in Brazil) to retrieve information on birth weight and gestational age. The SINASC is a national information system of all live births and includes additional information about maternal and infant characteristics, pregnancy, labor, and delivery (www.datasus.gov.br/sinasc). Seven variables related to maternal, pregnancy, and delivery history were collected. These measures include three continuous factors: maternal age, years of maternal schooling, and a socioeconomic attribute based on the presence of household assets using the HAI (household assets index). The HAI evaluates the frequency of the presence of an item in the household, giving more weight for rarer items [27]. Additionally, five categorical variables were considered: marital status (with or without live-in partner), satisfaction with prenatal care (excellent, good, fair/poor/very poor), parity (primiparous or multiparous), type of delivery (vaginal or cesarean), and the

Table 1 – Variables selected in the study by maternal and child characteristics.

Source	Variables	Categories
Maternal		
Demographic	Age	(Years)
	Schooling	(Years)
	Marital status	With live-in partner
		Without live-in partner
Socioeconomic	HAI	Assets in household index
Prenatal care	Adequacy of prenatal appointments	Appropriate
		Intermediary
		Not appropriate
	Evaluation of prenatal care	Excellent
		Good
		Fair/Poor/Very poor
Labor	Type of delivery	Vaginal
		Cesarean
	Parity	Primiparous
		Multiparous
Child		
Demographic	Infant age	(Years)
	Sex	Male
		Female
At birth	Prematurity	Yes
		No
	Weight at birth	Low (<2500 g)
		Eutrophic (2500–3999 g)
		Macrosomic (^{3}>4000 g)
Child care	Exclusive breastfeeding	Yes
		No
	Hospitalization	Yes
		No

modified Kotelchuck index (KI) [28], which also used three categories – no prenatal care and inadequate as "not adequate"; "intermediate"; and adequate and more than adequate as "adequate." The modified KI index incorporates the category "no prenatal care," and it includes in the category of "inadequate care" those women that start prenatal care after the fourth month and have more than 50% of the expected visits and women with fewer than 50% of the expected number of visits, even when starting the care during the fourth month of pregnancy. The infant attributes selected were age in months, sex (male or female), prematurity (born < 37 weeks of gestation or not), birthweight (low: <2500 g; eutrophic: between 2500 and 3999 g, macrosomic: >4000 g), type of feeding (exclusive breastfeeding or not), and previous hospitalization (yes or no). These maternal and infant variables were used to identify groups of infants with similar characteristics and are presented in Table 1.

The Z-score weight by age and sex was calculated to evaluate the appropriate infant weight at the time of the survey using the WHO Anthro software [29] (WHO, 2010). The children were classified as low or very low weight (Z score < −2), appropriate weight (−2 < Z score < 2), or high weight (Z score > 2), and the respective proportions were calculated for each infant group.

2.3. *Data analysis*

The available data for each infant were initially normalized to be in the interval 0–1 and then processed using a computational technique known as a self-organizing map or a Kohonen network [22,30].

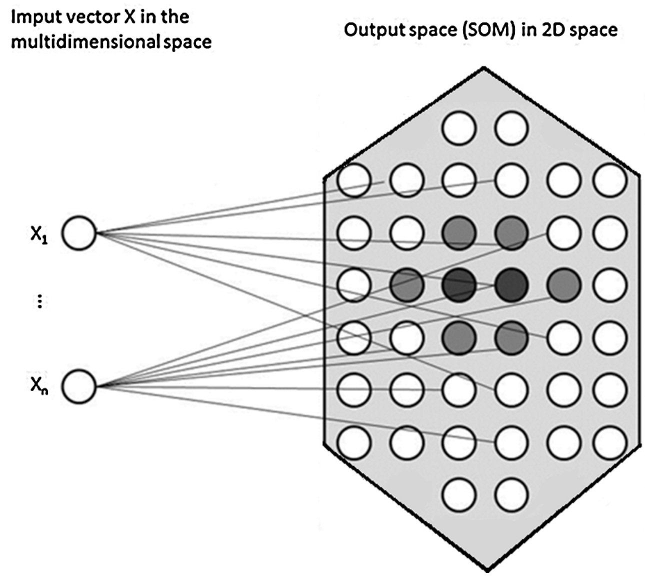

Fig. 1 – Architecture of a SOM. Darker units represent the best matching unit and its neighboring units. Adapted from Kohonen (1990).

A self-organizing map (SOM) is a single-layer artificial neural network that uses unsupervised competitive learning to form a nonlinear projection of a dataset in a reduced space while maintaining the original topology of the input space. Dimensionality reduction with topology preservation expands the capacity of the cluster analysis of the data. This neural network can be viewed as a two-dimensional grid in which each cell in the array has a processing unit, called a neuron. Neurons compete according to a learning rule known as "winner takes all." Fig. 1 shows a schematic of a two-dimensional Kohonen network, consisting of 48 cells, or an 8 × 6 map.

Inputs are represented by the vector $\mathbf{x} = [x_1, x_2, \ldots, x_m]^t$, where m is the dimension of the input vector that is fully connected to the neurons of the output layer of the network. The network is a two-dimensional array that maps the m-dimensional input space into a two-dimensional space. The weight vector of the neurons is represented by $\mathbf{w}_k = [w_{k1}, w_{k2}, \ldots, w_{km}]^t$, where k varies from 1 to N, the total number of neurons.

The pattern of the input vector $\mathbf{x}$ is mapped onto the network output using a three-stage training approach. In the competitive learning phase, a random sample is presented to the network, and the neuron with the most similar weight vector is declared the winning neuron, known as the BMU (best matching unit). The next step is the cooperation stage. This stage corresponds to the identification of BMUs' neighboring neurons. Typically, a Gaussian function is used to determine the neighborhood. As the final step, the weights of all neurons are updated according to the following formulas:

$$w_{km}(n) = w_{km}(n-1) + \Delta w_{km}$$
$$\Delta w_{km} = \eta(n) \cdot H(n) \cdot (x_m(n)) - w_{km}(n-1)) \quad (1)$$

At each step n of the training process, the region of cells identified by the function H surrounding the BMU is reduced, as is the learning rate η, which is the amount of adjustment of the weights. At the end of the training process, the SOM allows the visualization of the multivariate dataset in a two-dimensional graph, helping researchers to infer topological relationships in the original dataset.

To develop a SOM, one has to define the topology and size of the map (number of neurons). In this study, the selected neuron network topology was a hexagonal lattice. There is no direct method to define the size of the map. Here, we used two parameters to evaluate the quality of the map, the quantization error, and the topographical error. We implemented a series of SOM maps in which the number of neurons varied from 20 to 160 in increments of 5 neurons, and we evaluated the two types of error, choosing the map with the smallest error. Additionally, for each map size, we obtained the percentage of BMUs because it is known that not all neurons in the SOM map will be BMUs. It is desirable to have some neurons not be BMUs to help find clusters in the data, but it is not recommended to have a very large proportion of inactive neurons.

The quantization error is calculated as shown in formula (2), where N is the number of data-vectors and w_{km} is the BMU of the corresponding x_m data-vector. The quantization error measures the average distance between the BMU and each data vector and evaluates the fit of the SOM map to the data. The smaller the quantization error, the closer the data are to their BMU neuron. This error is calculated as follows:

$$qe = \frac{1}{N}\sum \|x_m - w_{km}\| \tag{2}$$

The topographical error measures the percentage of samples that have first and second BMUs that are not adjacent to each other. The lower the topographic error, the better the topology preservation of the SOM map. This error is calculated using the following formula:

$$te = \frac{1}{N}\sum_{m=1}^{N} u(x_m) \tag{3}$$

where $u(x_m) = 1$ if the first and second BMUs are adjacent and 0 otherwise. The resulting map after training the SOM shows the presence of groups of neurons with similar activation. These groups can be visualized using the unified distance matrix (U matrix), revealing potential clusters in the map. In this representation, regions with high values indicate frontiers between the groups, and units with low values suggest that there is greater similarity between the neighboring units.

The SOM algorithm creates a set of prototypes in a two-dimensional grid representing the original data and preserves the original topology, reducing data complexity. Various neighboring neurons may be modeling a cluster, and the U matrix allows an initial visualization of the groups. To facilitate the quantitative analysis of the map and the data, similar units should be clustered. The clustering technique used here was K-means, proposed by MacQueen in 1967 [31].

The method of K-means clustering is also an unsupervised learning algorithm. The algorithm randomly selects k samples, called centroids, and builds k groups by associating each sample to the centroid with the highest similarity. The average attributes in each group are calculated, generating k new centroids. The allocation process of the samples to the groups is repeated until the process convergences, that is, when no significant changes occur in the members of the groups.

To determine the number of clusters, the GAP validity index was used [32]. This index evaluates the relationship between the similarity within groups and the dissimilarity or heterogeneity among the groups. This index was calculated as the average of 10 replicates and was determined when varying the number of clusters k from 2 to 10. High values of the GAP index indicate good partitioning of the data.

After clustering the SOM neurons, the infants were classified according to their BMUs in each cluster and were analyzed to label each group based on the variables used.

The computational environment used was Matlab version 7.b (MathWorks), and the SOM map was implemented using the free Matlab SOM toolbox developed by Vesanto [33] and available at the web site www.cis.hut./projects/somtoolbox.

Significance testing was used to identify variables with possible associations between groups. For the continuous variables, ANOVA was performed, with Bonferroni and Duncan post hoc tests. For the categorical variables, Chi-square tests were performed.

2.4. *Ethical approval*

The study was approved by the Ethics Committee on Human Research of the Sérgio Arouca School of Public Health, part of the Oswaldo Cruz Foundation.

3. Results

3.1. *Characteristics of the children*

From the total of 1082 infants whose mothers were interviewed, 78 (7.2%) were excluded from the analysis because of missing information on some covariate of interest, resulting in a study population of 1004 children. The age of the children at the time of the interview varied from 0.13 to 5.95 months, with an average of 2.9 months (±1.4). The birth weight of those children was 3162 ± 513 g. There was no significant difference in the distribution of the children by sex; 496 (49.4%) were female, and 508 (50.6%) were male. The prevalences of low birth weight (<2500 g) and prematurity (<37 weeks of gestation) were, respectively, 8.3% and 6.6%. For every 100 infants, 92 had adequate weight for their age and sex at the time of the interview, and 34% were still exclusively breastfed. Nine percent of the children had a history of one or more hospitalization. The mean maternal age was 25.3 ± 6.3 years old, and the mothers had an average of 8 ± 2.8 years of schooling. Most mothers with live-in partners (86.4%) rated their prenatal care as good (47.4%), received appropriate prenatal care (59.3%), had vaginal deliveries (63%), and were multiparous (54.5%).

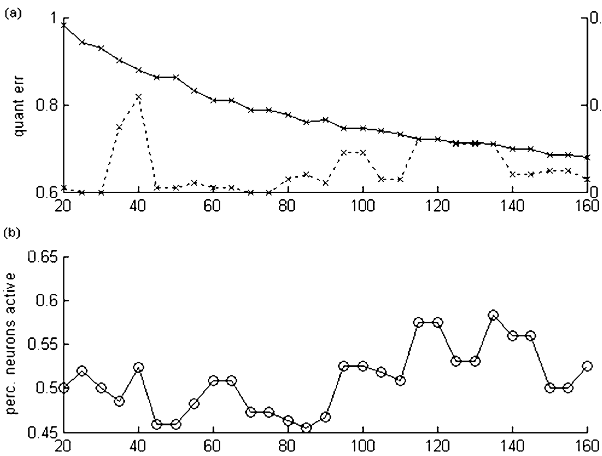

Fig. 2 – (a) Quantization (dashed line) and topography (dotted line) errors with varying numbers of map units and (b) the percentage of active neurons.

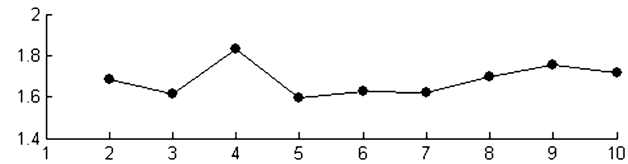

Fig. 4 – GAP index used to determine the number of clusters.

3.2. *Results from the self-organizing maps*

The best size for the SOM map was obtained by varying the number of units between 20 and 160 and calculating the quantization error, qe, and topological error, te. Additionally, we determined the percentage of neurons that were BMUs, i.e., neurons containing at least one instance of the dataset.

Fig. 2a shows the quantization error (dashed line) and topographical error (dotted line) as a function of map size. The percentage of neurons, pnb, that were BMUs is shown in Fig. 2b. The number selected was 30 map units because this size had small values for both qe and te, 0.93 and 0, respectively, and a reasonable proportion of active neurons, 0.5%.

The final SOM (after training) is displayed using the *U* matrix (unified distance matrix) in Fig. 3a as a grayscale image. This representation shows the distances between map units, where dark shades correspond to long distances and light tones to short distances between neurons. This representation enables, in a preliminary manner, the identification of up to six possible clusterings of units: the first and second groups are in the upper portion of the map, and the others are in the middle and lower portions. Fig. 3b shows the observed number of individuals for each BMU neuron.

To determine the best number of groups of neurons with similar features, we used a *K*-means cluster analysis and varied the number of clusters *k* from 2 to 10. Because the method is sensitive to the choice of the initial centroids, the process was replicated 10 times for each value of *k*. We used the cosine distance as a measure of similarity. Fig. 4 shows the results of the median of the 10 runs of the GAP index used to identify the number of clusters. The GAP index had a maximum value at 4, and thus, this number was chosen as the best grouping for the sample.

3.3. *Analysis of groups*

After clustering, the samples for each cluster were analyzed to obtain a main description of the groups with the significance tests. Table 2 presents a summary of the variables in the dataset for the three groups. The percentages of samples in the four groups were 13.8%, 24.8%, 29.7%, and 31.7%. Variables that exhibited statistical significance (p value <0.05) in discriminating the groups were age, maternal education, adequacy of prenatal care, type of delivery, parity, sex, gestational age, birthweight, exclusive breastfeeding, and hospitalization during the first months of life.

Group 1 contained 139 children, most of whom were females who were born by cesarean section. This group had the highest proportions of prematurity (12.9%), low birthweight (12.2%), and exclusive breastfeeding (56.8%). The mothers in this group had a higher mean age (29.5 ± 5.6 years) and were all multiparous.

The second group (249 children) was composed exclusively of girls and had the lowest proportions of prematurity

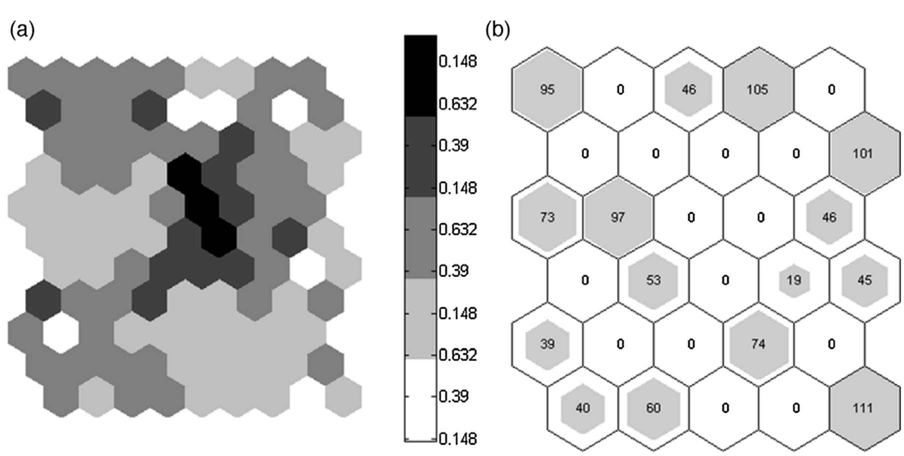

Fig. 3 – (a) The *U* matrix and (b) the number of individuals by BMU.

Table 2 – Characteristics of mothers, prenatal care, and children under six months of age who were seen at basic health units in the city of Rio de Janeiro in 2007 by group determined using a SOM algorithm.

		SOM categories				
		Group 1 ($n = 139$)	Group 2 ($n = 249$)	Group 3 ($n = 298$)	Group 4 ($n = 318$)	p value*
Maternal age (years) (mean)		29.5	26.0	21.2	26.7	<0.001
Schooling (years) (mean)		7.5	7.8	8.4	7.9	0.008
Marital status (%)	With live-in partner	89.9	88.4	83.9	85.5	0.255
	Without live-in partner	10.1	11.6	16.1	14.5	
HAI (mean)		1.46	1.28	1.40	1.43	0.110
Adequacy of prenatal Appointments (%)	Appropriate	57.6	62.7	54.1	62.3	0.181
	Intermediary	30.2	22.8	31.5	23.2	
	Not appropriate	12.2	14.5	14.4	14.5	
Evaluation of prenatal care (%)	Excellent	33.8	39.0	29.6	28.3	0.037
	Good	48.9	46.1	46.6	48.4	
	Fair/poor/very poor	17.3	14.9	23.8	23.3	
Type of delivery (%)	Vaginal	–	92.8	84.6	47.2	<0.001
	Cesarean	100.0	7.2	15.4	52.8	
Parity (%)	Primiparous	–	25.7	100.0	29.9	<0.001
	Multiparous	100.0	74.3	–	70.1	
Infant age (mean)		2.6	2.8	3.0	3.0	0.059
Sex (%)	Male	28.1	–	50.7	100.0	<0.001
	Female	71.9	100.0	49.3	–	
Prematurity (%)	Yes	12.9	4.8	6.4	5.3	0.010
	No	87.1	95.2	93.6	94.7	
Weight at birth (%)	Low (<2500 g)	12.2	6.4	9.7	6.6	0.005
	Eutrophic (2500–3999 g)	82.8	90.0	88.6	85.9	
	Macrosomic (^{3}>4000 g)	5.0	3.6	1.7	7.5	
Exclusive breastfeeding (%)	Yes	56.8	55.4	15.4	24.8	<0.001
	No	43.2	44.6	84.6	75.2	
Hospitalization (%)	Yes	5.8	5.6	11.4	11.6	0.021
	No	94.2	94.4	88.6	88.4	
Z score for weight for age at the time of the survey** (%)	Low and very low ($Z < -2$)	5.8	3.6	4.7	6.9	0.423
	Appropriate ($-2 \leqslant z \leqslant 2$)	92.8	93.2	93.0	89.9	
	High ($z > 2$)	1.4	3.2	2.3	3.1	

* ANOVA for continuous variables (with the Bonferroni correction); Chi-square for categorical variables.

** Indicator not included in the SOM formation.

(4.8%), low birth weight (6.4%), and hospitalization (5.6%); higher proportions of vaginal delivery (92.8%) and exclusive breastfeeding (55.4%); and predominantly multiparous mothers (74.3%).

Mothers in group 3 (298 children) were all primiparous with a lower mean age (21.2 ± 5.4 years), a higher mean years of schooling (8.4 ± 2.5 years), and a higher proportion of fair/poor/very poor prenatal care (23.8%). These mothers had the highest proportion of not living with a partner (16.1%) and a lower proportion of having an appropriate number of prenatal visits (54.1%), although the differences in these variables were not statistically significant. Children in this group exhibited no difference in relation to sex. In addition, this group had the lowest proportion of exclusive breastfeeding (15.4%) and a high proportion of hospitalizations (11.4%).

In group 4, the children (318) were male, and they had a higher proportion of macrosomia (7.5%). This group had the highest prevalence of hospitalizations (11.6%) and a breastfeeding prevalence of 24.8%. The mothers were predominantly multiparous (70.1%), and there was no difference in frequency between the types of delivery. Group 4 had a higher proportion of infants with an inappropriate weight for their age (Z score < −2 or Z score > 2) at the time of the interview (10%) than the other groups (mean 7%), although this difference was not statistically significant at the 5% level.

4. Discussion

Using a combination of SOM and K-means, we obtained four groups with different maternal, pregnancy, labor, and infant characteristics in this study.

The quantization and topological errors were used to help define the number of neurons for the SOM. It is recognized that the quantization error decreases with an increasing number of neurons. This trend is due to the fact that with a larger number of neurons, each sample tends to be closer to its BMU. However, a very small quantization error can affect the topological property of the SOM [34,35]. The topological error, in contrast, is expected to increase with a larger number of neurons, but it may have high values even for small maps. In this study, we selected a number of neurons that was a compromise between both types of error, and we chose to monitor the number of active neurons to preserve the original topology of the data.

One of the advantages of using a SOM followed by the K-means technique compared with other conventional cluster techniques is the ability to visualize the data using the U matrix of potential clusters, thereby obtaining an greater understanding of the structure of the data to support the selection of the number of groups to be explored using K-means. The neurons in the SOM map can be observed as prototypes representing subgroups of children [36]. The children represented by prototypes in the U matrix, grouped on the right, at the top, in the bottom left, and in the top left, pointed to the creation of the groups in this study.

One of the main challenges in data clustering is the definition of the number of groups that will best represent the available dataset. For the K-means algorithm used here to group the SOM neurons, we defined a range of possible groups based on the visual clues provided by the U matrix. Additionally, this technique can provide different results due to the random choice of the initial centroids, but the clustering tends to converge to the optimal solution when the process is repeated, minimizing the sum of distances within the group [37]. Here we used 50 runs to minimize this problem.

For group formation, maternal age, parity, type of birth, sex, and exclusive breastfeeding were the decisive descriptors in identifying the main characteristics of each group.

The group that had the best infant indicators and childbirth indicators was group 2: almost all births were by vaginal delivery, and there were smaller proportions of preterm births and low birthweight and a higher proportion of exclusive breastfeeding. Despite the prevalence of vaginal delivery in group 3, the other indicators exhibited trends opposite those of group 2 and were worse than those of other groups. Added to these risk factors, group 3 had the lowest average maternal age and was composed entirely of primiparous mothers.

Younger mothers and primiparous mothers, who do not have previous experience with breastfeeding, tend to introduce food to their children early, interrupting exclusive breastfeeding [38]. These features were present in group 3.

Cesarean section is associated with a low birthweight [39], prematurity [40], the delayed initiation of breastfeeding, and a shorter duration of breastfeeding [41–43]. In group 1 identified in this study, all infants were born by cesarean section. In addition, this group had the highest prevalences of low birthweight and prematurity among the groups. However, this group also had the highest rate of exclusive breastfeeding. Considering that exclusive breastfeeding has a protective effect against severe diseases during infancy [44–48], it may reduce the risk of hospitalization [49]. Shiva et al. (2007), comparing the frequency of hospitalization during the first six months of life with breastfeed children, showed that the protective effect of breastfeeding is not limited to developing countries [50]. An inverse relationship between exclusive breastfeeding and hospitalizations was observed in groups 1 and 2, which had the highest prevalences of exclusive breastfeeding and lower proportions of hospitalizations. The opposite relationship was observed in groups 3 and 4.

A result that also drew attention to group 4 was the high prevalence of macrosomia. Fetal macrosomia can be an indication for a cesarean section [51]. In our study, we found that the highest rates of cesarean section and macrosomia were in groups 1 and 4, respectively. In their study, Oliveira et al. [52] found a proportion of macrosomia similar to that observed in groups 1 and 4. The authors reported that the determining factors for macrosomia were multiparity and having had male children [52].

Regardless of the group, for every 100 children, 90 or more had the appropriate weight for age and sex in this survey. After the cluster analysis, we verified the adequacy of the infant weight at the time of the interview according to the groups formed using the index Z score for age and sex. Although there were no statistically significant differences, group 2 had a lower proportion of low and very low weight for age, whereas higher values of inadequate weight were found in group 4, followed by group 1. The lack of statistical evidence to indicate a difference between these proportions reinforces the idea that while the early months of life, infants who are small for gestational age (SGA) and/or who had IUGR (intrauterine growth restriction) have more difficulties to growth in the beginning but later tend to follow the eutrophic growth curve [53].

In summary, the common characteristics found by the SOM in this study that may be associated favorably with infant weight are good prenatal care and natural delivery [54,55], birth at term [5], adequate birthweight [56,57], exclusive breastfeeding [58], a lower frequency of hospitalization [59], and multiparity [60,61]. These features were present in group 2, which had the best infant and delivery indicators.

5. Conclusions

Using a combination of the SOM algorithm and the *K*-means technique for data analysis in this study, it was possible to cluster children with similar characteristics. In addition, this study demonstrates that artificial neural networks are an important tool for identifying individuals with better or worse health indicators. Using the groups identified in this study, we can infer, based on the characteristics of a given child, concerning socioeconomic aspects, prenatal and child care, the need of a more appropriate intervention for growth according to each child. Groups found to have worse health indicators may require special attention in public policies that affect populations with similar growth profiles.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

[1] M.E.F.A. Motta, G.A.P. da Silva, O.C. Araújo, P.I. Lira, M.C. Lima, Does birth weight affect nutritional status at the end of first year of life? J. Pediatr. (Rio J.) 81 (October) (2005) 377–382.

[2] A.D. Stein, M. Wang, R. Martorell, S.A. Norris, L.S. Adair, I. Bas, et al., Growth patterns in early childhood and final attained stature: data from five birth cohorts from low- and middle-income countries, Am. J. Hum. Biol. 22 (June) (2010) 353–359.

[3] G. Kac, L.B. Nucci, M.H.C. Spyrides, B.B. Duncan, M.I. Schmidt, Evaluation of the ability of a Latin-American gestational weight curve to predict adverse pregnancy outcomes, Int. J. Gynaecol. Obstet. 106 (September) (2009) 223–226.

[4] D.L.B. Ribas, A. Sganzerla, J.R. Zorzatto, S.T. Philippi, Child health and nutrition in a Teréna indigenous community, Mato Grosso do Sul, Brazil, Cad. Saúde Pública 17 (March) (2001) 323–331.

[5] M.H.C. Spyrides, C.J. Struchiner, M.T.S. Barbosa, G. Kac, Breastfeeding practices and infant growth: a longitudinal study in Rio de Janeiro, Brazil, 1999/2001, Cad. Saude Publica 21 (June) (2005) 756–766.

[6] M.A. Almeida, H.M. Novaes, G.P. Alencar, L.C. Rodrigues, Mortalidade neonatal no Município de São Paulo: influência do peso ao nascer e de fatores sócio-demográficos e assistenciais, Rev. Bras. Epidemiol. 5 (2002) 93–107.

[7] L.L. Hui, G.M. Leung, B.J. Cowling, T.H. Lam, C.M. Schooling, Determinants of infant growth: evidence from Hong Kong's "Children of 1997" birth cohort, Ann. Epidemiol. 20 (November) (2010) 827–835.

[8] C.G. Victora, J.P. Vaughan, F.C. Barros, A.C. Silva, E. Tomasi, Explaining trends in inequities: evidence from Brazilian child health studies, Lancet 356 (September) (2000) 1093–1098.

[9] L. Monasta, G.D. Batty, A. Cattaneo, V. Lutje, L. Ronfani, F.J. Van Lenthe, et al., Early-life determinants of overweight and obesity: a review of systematic reviews, Obes. Rev. 11 (2010) 695–708.

[10] E.M. Taveras, S.L. Rifas-Shiman, M.B. Belfort, K.P. Kleinman, E. Oken, M.W. Gillman, Weight status in the first 6 months of life and obesity at 3 years of age, Pediatrics 123 (April) (2009) 1177–1183.

[11] L.S. Adair, R. Martorell, A.D. Stein, P.C. Hallal, H.S. Sachdev, D. Prabhakaran, et al., Size at birth, weight gain in infancy and childhood, and adult blood pressure in 5 low- and middle-income-country cohorts: when does weight gain matter? Am. J. Clin. Nutr. 89 (May) (2009) 1383–1392.

[12] M. De Onis, C. Garza, A.W. Onyango, M.-F. Rolland-Cachera, WHO growth standards for infants and young children, Arch. Pediatr. 16 (January) (2009) 47–53.

[13] D. MS/Sisvan. Registro de informações do estado nutricional, 2011.

[14] P.V. Hamill, T.A. Drizd, C.L. Johnson, R.B. Reed, A.F. Roche, W.M. Moore, Physical growth: National Center for Health Statistics percentiles, Am. J. Clin. Nutr. 32 (March) (1979) 607–629.

[15] R.J. Kuczmarski, C.L. Ogden, S.S. Guo, L.M. Grummer-Strawn, K.M. Flegal, Z. Mei, et al., 2000 CDC Growth Charts for the United States: methods and development, Vital Health Stat. 11 (May) (2002) 1–190.

[16] WHO, WHO Child Growth Standards: Methods and Development: Length/Height-for-Age, Weight-for-Age, Weight-for-length, Weight-for-Height and Body Mass Index-for-Age, 2006 http://www.who.int/childgrowth/publications/technical_report_pub/en/index.html (accessed 14.04.13).

[17] C.A. Monteiro, I. França Jr., W.L. Conde, Evolução da assistência materno-infantil na cidade de São Paulo (1984–1996), Rev. Saúde Pública December (2000) http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S0034-89102000000700005 (accessed 05.01.13).

[18] R. Halpern, A.J.D. Barros, A. Matijasevich, I.S. Santos, C.G. Victora, F.C. Barros, Developmental status at age 12 months according to birth weight and family income: a comparison of two Brazilian birth cohorts, Cad. Saude Publica 24 (Suppl. 3) (2008) S444–S450.

[19] I.M.S. Engebretsen, T. Tylleskär, H. Wamani, C. Karamagi, J.K. Tumwine, Determinants of infant growth in Eastern Uganda: a community-based cross-sectional study, BMC Publ. Health 8 (2008) 418.

[20] C.A. Monteiro, M.H.D. Benicio, W.L. Conde, S. Konno, A.L. Lovadino, A.J.D. Barros, et al., Narrowing socioeconomic inequality in child stunting: the Brazilian experience, 1974–2007, Bull. World Health Organ. 88 (April) (2010) 305–311.

[21] B. Fenn, S.S. Morris, C. Frost, Do childhood growth indicators in developing countries cluster? Implications for intervention strategies, Public Health Nutr. 7 (October) (2004) 829–834.

[22] T. Kohonen, The self-organizing map, Proc. IEEE 78 (1990) 1464–1480.

[23] T. Faisal, M.N. Taib, F. Ibrahim, Reexamination of risk criteria in dengue patients using the self-organizing map, Med. Biol. Eng. Comput. 48 (March) (2010) 293–301.

[24] M.K. Markey, J.Y. Lo, G.D. Tourassi, C.E. Floyd Jr., Self-organizing map for cluster analysis of a breast cancer database, Artif. Intell. Med. 27 (February) (2003) 113–127.

[25] D. Soria, J.M. Garibaldi, F. Ambrogi, A.R. Green, D. Powe, E. Rakha, et al., A methodology to identify consensus classes from clustering algorithms applied to immunohistochemical data from breast cancer patients, Comput. Biol. Med. 40 (March) (2010) 318–330.
[26] V.-P. Valkonen, M. Kolehmainen, H.-M. Lakka, J.T. Salonen, Insulin resistance syndrome revisited: application of self-organizing maps, Int. J. Epidemiol. 31 (August) (2002) 864–871.
[27] C.L. Szwarcwald, F.I. Bastos, M.A.P. Esteves, State of animus among Brazilians: influence of socioeconomic context? Cad. Saúde Pública 21 (January) (2005) S33–S42.
[28] M.C. do Leal, S.G.N. da Gama, K.M.N. Ratto, C.B. da Cunha, Uso do índice de Kotelchuck modificado na avaliação da assistência pré-natal e sua relação com as características maternas e o peso do recém-nascido no Município do Rio de Janeiro, Cad. Saúde Pública (2004) 20 http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S0102-311X2004000700007&lng=en&nrm=iso&tlng=en (accessed on 16.11.12).
[29] WHO Anthro (version 3.2.2, January 2011) and macros – Software for assessing growth and development of the world's children, WHO, 2007 http://www.who.int/childgrowth/software/en/ (accessed on 23.09.12).
[30] S.S. Haykin, Neural Networks and Learning Machines, Prentice Hall, New York, 2009.
[31] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Comput. Surv. 31 (September) (1999) 264–323.
[32] R. Tibshirani, G. Walther, T. Hastie, Estimating the number of clusters in a data set via the gap statistic, J. Royal Stat. Soc: Ser. B (Stat. Methodol.) 63 (2001) 411–423.
[33] J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map, IEEE Trans. Neural Netw. 11 (2000) 586–600.
[34] K. Kiviluoto, Topology preservation in self-organizing maps, in: IEEE International Conference on Neural Networks, vol. 1, 1996, pp. 294–299.
[35] M. Pena, W. Barbakh, C. Fyfe, Topology-preserving mappings for data visualisation, in: A.N. Gorban, B. Kégl, D.C. Wunsch, A.Y. Zinovyev (Eds.), Principal Manifolds for Data Visualization and Dimension Reduction [Internet], Springer, Berlin, Heidelberg, 2008, pp. 131–150 http://rd.springer.com/chapter/10.1007/978-3-540-73750-6_5 (accessed on 15.04.13).
[36] A.M. Mora, J.J. Merelo, C. Briones, F. Morán, J.L.J. Laredo, Clustering and visualizing HIV quasispecies using Kohonen's self-organizing maps, in: F. Sandoval, A. Prieto, J. Cabestany, M. Graña (Eds.), Computational and Ambient Intelligence, Springer, Berlin, Heidelberg, 2007, pp. 940–947.
[37] F. Bação, V. Lobo, M. Painho, Self-organizing Maps as Substitutes for *k*-means Clustering. Computational Science – ICCS 2005, Springer, Berlin, Heidelberg, 2005, pp. 476–483.
[38] E.W. Kimani-Murage, N.J. Madise, J.-C. Fotso, C. Kyobutungi, M.K. Mutua, T.M. Gitau, et al., Patterns and determinants of breastfeeding and complementary feeding practices in urban informal settlements, Nairobi Kenya, BMC Publ. Health 11 (May) (2011) 396.
[39] C. Pedro, S. Fernanda, C. Maria, M. Sirlei, Perinatal outcomes associated with low birth weight in a historical cohort, Reprod. Health (2011) 2011.
[40] C.K. Shapiro-Mendoza, K.M. Tomashek, M. Kotelchuck, W. Barfield, A. Nannini, J. Weiss, et al., Effect of late-preterm birth and maternal medical conditions on newborn morbidity risk, Pediatrics 121 (February) (2008) e223–e232.
[41] R. Bahl, C. Frost, B.R. Kirkwood, K. Edmond, J. Martines, N. Bhandari, et al., Infant feeding patterns and risks of death and hospitalization in the first half of infancy: multicentre cohort study, Bull. World Health Organ. 83 (June) (2005) 418–426.
[42] C.S. Boccolini, M.L. de Carvalho, M.I.C. de Oliveira, M.C. do Leal, M.S. Carvalho, Factors that affect time between birth and first breastfeeding, Cad. Saúde Pública 24 (November) (2008) 2681–2694.
[43] K.G. Dewey, L.A. Nommsen-Rivers, M.J. Heinig, R.J. Cohen, Risk factors for suboptimal infant breastfeeding behavior, delayed onset of lactation, and excess neonatal weight loss, Pediatrics 112 (September) (2003) 607–619.
[44] S. Arifeen, R.E. Black, G. Antelman, A. Baqui, L. Caulfield, S. Becker, Exclusive breastfeeding reduces acute respiratory infection and diarrhea deaths among infants in Dhaka Slums, Pediatrics 108 (October) (2001) e67.
[45] V.R.G. Bachrach, E. Schwarz, L.R. Bachrach, Breastfeeding and the risk of hospitalization for respiratory disease in infancy: a meta-analysis, Arch. Pediatr. Adolesc. Med. 157 (2003) 237–243.
[46] C.S. Boccolini, M.L. de Carvalho, M.I.C. de Oliveira, P.M.M. de Boccolini, Breastfeeding can prevent hospitalization for pneumonia among children under 1 year old, J. Pediatr. (Rio J.) 87 (October) (2011) 399–404.
[47] M.A. Quigley, Y.J. Kelly, A. Sacker, Breastfeeding and hospitalization for diarrheal and respiratory infection in the United Kingdom Millennium Cohort Study, Pediatrics 119 (April) (2007) e837–e842.
[48] J.M.P. Talayero, M. Lizán-García, Puime ÁO, M.J.B. Muncharaz, B.B. Soto, M. Sánchez-Palomares, et al., Full breastfeeding and hospitalization as a result of infections in the first year of life, Pediatrics 118 (July) (2006) e92–e99.
[49] F.C. Barros, J.L.D. Rossello, A. Matijasevich, S.C. Dumith, A.J.D. Barros, I.S. Santos, dos, et al., Gestational age at birth and morbidity, mortality, and growth in the first 4 years of life: findings from three birth cohorts in Southern Brazil, BMC Pediatr. 12 (October) (2012) 169.
[50] F. Shiva, F. Ghotbi, M. Padyab, Infant feeding and hospitalization during the first six months of life, JPMA – J. Pak. Med. Assoc. 57 (12) (2007) 599.
[51] T.M. Sakae, P.F. Freitas, E. d' Orsi, Factors associated with cesarean section rates in a university hospital, Rev. Saúde Pública 43 (June) (2009) 472–480.
[52] L.C. de Oliveira, A.H.R.N. de Pacheco, P.L. Rodrigues, M.M. Schlüssel, M.H.C. Spyrides, G. Kac, Factors accountable for macrosomia incidence in a study with mothers and progeny attended at a Basic Unity of Health in Rio de Janeiro, Brazil, Rev. Bras. Ginecol. Obst. 30 (October) (2008) 486–493.
[53] M.K. Campbell, S. Cartier, B. Xie, G. Kouniakis, W. Huang, V. Han, Determinants of small for gestational age birth at term, Paediatr. Perinat. Epidemiol. 26 (November) (2012) 525–533.
[54] A.C.N.G. do Barreto, C.R.S. Maia, K.C. de Lima, H.S. de Maranhão, Postnatal growth restriction and predictors of nutritional outcome in very low birth weight infants fed human milk and assisted by the Kangaroo Mother Care method, J. Maternal-Fetal Neonatal Med. September (2012) 1–6.
[55] S.A. Romani, P.I. Lira, Determinant factors of infant growth, Rev. Bras. Saúde Mater. Infant. 4 (2004) 15–23.
[56] P.H. Casey, L. Whiteside-Mansell, K. Barrett, R.H. Bradley, R. Gargus, Impact of prenatal and/or postnatal growth problems in low birth weight preterm infants on school-age outcomes: an 8-year longitudinal evaluation, Pediatrics 118 (September) (2006) 1078–1086.
[57] M.B. Belfort, S.L. Rifas-Shiman, T. Sullivan, C.T. Collins, A.J. McPhee, P. Ryan, et al., Infant growth before and after term: effects on neurodevelopment in preterm infants, Pediatrics 128 (2011) e899–e906.
[58] M.S. Kramer, T. Guo, R.W. Platt, S. Shapiro, J-P. Collet, B. Chalmers, et al., Breastfeeding and infant growth: biology or bias? Pediatrics 110 (August) (2002) 343–347.

[59] C.G. Victora, F.C. Barros, B.L. Horta, R. Martorell, Short-term benefits of catch-up growth for small-for-gestational-age infants, Int. J. Epidemiol. 30 (2001) 1325.

[60] K.K.L. Ong, M.A. Preece, P.M. Emmett, M.L. Ahmed, D.B. Dunger, Size at birth and early childhood growth in relation to maternal smoking, parity and infant breast-feeding: longitudinal birth cohort study and analysis, Pediatr. Res. 52 (December) (2002) 863–867.

[61] A. Prentice, T.J. Cole, R.G. Whitehead, Impaired growth in infants born to mothers of very high parity, Hum. Nutr. Clin. Nutr. 41 (September) (1987) 319–325.