MICRO-MACRO MULTILEVEL ANALYSIS FOR MORPHOMETRIC DATA OF COMPUTED TOMOGRAPHY IMAGES: TWO ALTERNATIVE APPROACHES

Minerva Montero Díaz*[1], Roberto Rodríguez Morales**, Luis Antonio Rodríguez Sánchez***
*Department of Mathematics, Institute of Cybernetics, Mathematics and Physics, Havana, Cuba
**Department of Interdisciplinary Mathematics, , Institute of Cybernetics, Mathematics and Physics, Havana, Cuba
***Department of Neurology, General Teaching Hospital Roberto Rodríguez, Ciego de Ávila, Cuba

ABSTRACT
The computed tomography images data posse an intrinsic hierarchical structure; that is, when images (lower-level or micro-level units) are nested within individuals (higher-level or macro-level units). The problem with two-level data, where it is expected that a dependent variable measured on the individuals is influenced by variables measured on the images is the focus of this article. Methodology for the analysis of this type of scenario, known as micro-macro situation, has been little investigated. In this article, a latent variable modeling approach is proposed, where morphometric information of multiple images and characteristics of the patients, such as age and sex, are combined in a single logit model. Multilevel Factor Analysis is used to obtain latent variables at the macro level, which can be interpreted as synthetic indicators of morphometric differences in images between individuals. To illustrate the methodology, real data of patients with cerebral hemorrhage were used. The results are compared with those obtained by an alternative approach. The empirical study revealed a comparable performance between the two methodologies, although the new approach suggests a relative superiority in terms of predictive capacity.

KEYWORD: Computed tomography imaging, Image analysis, Multilevel Factor Analysis, Micro-macro analysis, Prognostic procedure.

MSC: 62P10

RESUMEN
Los datos de imágenes de tomografía computarizada poseen una estructura jerárquica intrínseca; es decir, las imágenes (unidades de nivel inferior o micro) se anidan dentro de los individuos (unidades de nivel superior o macro). El interés de este artículo se centra en un problema con datos de dos niveles, donde se espera que una variable dependiente medida sobre los individuos, esté influenciada por variables medidas sobre las imágenes. La metodología para el análisis de este tipo de escenario, conocido como situación micro-macro, ha sido poco investigada. En este artículo se propone un enfoque de modelación de variables latentes, donde se combina en un único modelo, información morfométrica de múltiples imágenes y características de los pacientes, tales como edad y sexo. Se utiliza Análisis Factorial Multinivel para obtener variables latentes a nivel macro, que pueden interpretarse como indicadores sintéticos de las diferencias morfométricas de las imágenes entre individuos. Para ilustrar la metodología, se utilizaron datos reales de pacientes con hemorragia cerebral. Los resultados se comparan con los obtenidos por un enfoque alternativo. El estudio empírico reveló un desempeño comparable entre las dos metodologías, aunque el nuevo enfoque sugiere una superioridad relativa en términos de capacidad predictiva.

PALABRAS CLAVES: Imágenes de Tomografía Computarizada, Análisis de imágenes, Análisis Factorial Multinivel, Análisis micro-macro, Procedimiento de Diagnóstico

1. INTRODUCTION

The morphometric examination of computed tomography (CT) images plays an important role in clinical practice (Rodríguez and Sossa, 2019). The development of modeling techniques for this type of information would facilitate the reduction of uncertainty in medical diagnosis and it will help to make better decisions.
In the neuroimaging field, there is a need of understanding the relative contribution of the morphometric analysis of the spontaneous intracerebral hemorrhage (ICH) in the survival of the patients. The presence of

[1] Corresponding author:
Minerva Montero Díaz,
E-mail: minerva@icimaf.cu

multiple sections of images (slices) for each patient (see Figure 1) could represent a problem in the interpretation of the hemorrhage magnitude. An accurate morphometric analysis of the hemorrhages and their role in the survival of the patients is still a challenge in the practice of specialists dedicated to the image analysis.

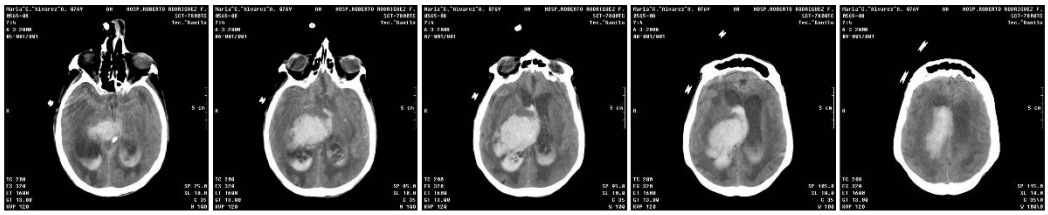

Figure 1. CT images of a spontaneous intracerebral hemorrhage (slides 5, 6, 7, 8 and 9)

Classical statistical modeling methods assumes that sampling are independent observations. In morphometric researches, however, the data belonged to images usually posse an intrinsic hierarchical structure in which images are grouped within individuals (see Figure 2). For this type of multilevel data, the assumption of independence among observations is not realistic, because it is reasonable to assume that the observations within an individual are more similar.

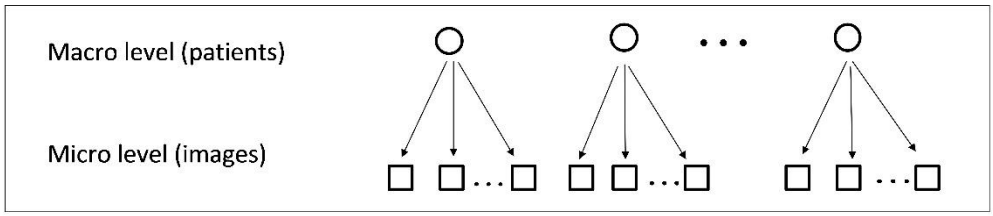

Figure 2. Hierarchical structure of CT neuroimaging data

It is very important to point out that, even when many variables are collected on the images, the interest in the present research is to find differences between individuals (or groups of individuals), instead of explaining the differences between the images. In this context, we assumed that independent variables measured on the images (micro-level) affect dependent variables measured on the individuals (macro-level). This type of scenario is referred as micro-macro situation (Raudenbush and Bryk, 2002) measured at different levels requires to consider an approach that allows analyzing the relationship between units at each level of the hierarchy, taking into account the two sources of variability: the variability within and between individuals (Snijders and Bosker, 2012). Multilevel analysis model is an appropriate methodological tool for analyzing data with hierarchical structure, however, most of the methods focus only on the so-called macro-micro models, where the response variable is measured at the lowest level of the model (Longford, 1993; Goldstein,1995; Kreft and De Leeuw,1998).

In a micro-macro analysis, data measured on the images need to be aggregated at the macro level, such that the aggregated values can be related to the patient response. On the other hand, the examination of the morphometric heterogeneity between the images of each patient can be very important for the diagnosis. -In the last decade, several methods were proposed in order to analyze micro-macro data (Croon and Van Veldhoven, 2007; Bennink et al.,2013; Bennink, 2014; Bennink et al., 2016; Croon and Kroon, 2016; Becker et al., 2018). These methods are based fundamentally on latent variable approaches (Muthén,1998) and are mainly motivated by psychological or sociological problems that include relationships between individuals and groups, and where the characteristics of individuals can affect some group-level response data.

How then to approach a problem with data at two levels where it is expected that an outcome variable related to the patient, measured at the macro level, is influenced by morphometric variables of CT images, measured at the micro level? To this end, Montero et al. (2017) proposed a latent variables modeling approach, where the multilevel structure of micro-macro relationships were appropriately handled. This approach is based on a strategy comprising two stages. In a first step, a Principal Component Analysis (Joliffe, 2002) carry out from the intra-individual covariance matrix of the morphometric variables for to produce factors (latent variables) that successively explain most of the total variance of the morphometric variables at the macro level (patients). These factors can be interpreted as synthetic indicators of the morphometric differences of the

images between patients. In a second step, a latent variable regression (LVR) model is adjusted, in which the values of the more important (in term of variance) principal factors at macro level (patients), obtained from the intra-individual PCA, are treated as predictors or explanatory variables for the response variable relative to the patient. We will refer to this strategy as PCA-LVR approach.
In the current article, we proposed an alternative approach. The purpose of this paper is to consider both methodologies, and investigate how they perform in practice, when applied to the micro-macro analysis of problems with morphometric data of TC images.
The difference of our proposal with respect to the PCA-LVR approach is in the method used to construct the latent variables at the macro level. Here, we propose the use of Multilevel Factorial Analysis (MFA) (Hox, 2010) to produce factors that explain most of the common variance of morphometric variables at the patient level, and that these factors can also be used as latent explanatory variables of the patient's response.
Both approaches are applied to a real data set and the results are compared. We conclude that, independently of the method (between-individual PCA or MFA) chosen to construct latent variables at the macro level, the two strategies discussed in this article can be considered as alternatives of a unified approach of micro-macro multilevel analysis of morphometric data of TC images that are related to a patient's clinical response.

2. MOTIVATION

In order to motivate and illustrate these different approaches, we take up the problem presented by Montero et al. (2017). This is about the study of the multiple factors that could be considered predictors of the survival of patients with ICH.
Today, despite having many researches there are not enough elements to reach consistent conclusions on this research issue (Morgenstern et al., 2010; Tellería 2006; Escudero et al., 2008; Ferrete et al., 2015). The incorporation of new analysis strategies, guided at building appropriate methods to analyze data from CT images, can lead to the improvement of the effectiveness in the diagnosis of this or other types of pathologies.
The examination of the morphometric heterogeneity of hemorrhage images within and between patients can be very important as part of a prognostic tool that allows us to assess the intensity of the neurological damage on the patient's life. The fundamental question is: What effect does the morphometry of the CT images have on the risk of not surviving the hemorrhage in certain groups of individuals? An obvious problem is: How to predict or explain variables at patient level from morphometric variables measured at the image level?

3. METHODOLOGY

A unified approach for the micro-macro multilevel analysis de CT image data can be summarized as follows:
In a first step, to use techniques of Factorial Analysis to reduce a large number of variables into fewer numbers of factors at the macro level. These technique (between-individual PCA or MFA) extracts maximum common variance from all morphometric variables and puts them into a common score (factors) at each level analysis. As an index of all morphometric variables, we can use this score for further analysis and can be interpreted as synthetic indicators that represent the morphometric differences of the hemorrhages between the patients.
In a second step, these latent variables are considered, together with covariates at the individual level, as predictors or explanatory variables for the response variable relative to the patient, in a single regression model at the macro level. For example, for a patient's dichotomous response, such as survival (0 = survived, 1 = not survived), the latent variables model can be formulated as a logistic regression model at the single level. The practical consequences of this strategy is to make inferences at the patient level using all the information available at the image level.
From a methodological point of view, the difference between approach PCA-LVR and that proposed in this article is lies in the Factorial Analysis technique considered in order to construct the latent variables at the macro-level. But the practical consequences of both strategies are to make inferences at the patient level using the information available at the image level.
We should be aware that, independently of the aggregation strategy that has been chosen to create summary measures in each level of analysis, the two strategies analyzed in this work allow the creation of latent variables that capture both the variability between images as between patients. This is possible from the decomposition of the total sample variance-covariance matrix of the morphometric measurements in within individual matrices (micro level) and between-individual matrices (macro-level).

Consider a study where there are *I* different individuals (macro-level) and for each individual *i* (*i*=1, 2,.. *I*) there are J_i images (micro-level). It is assumed that the values y_i of the individual *i* on a response variable *Y* at the individual level, can be explained or predicted by a set of *P* explanatory variables $Z_1, Z_2, \ldots, Z_P$ with values $z_{1i}, z_{2i}, \ldots, z_{Pi}$ for the individual *I* and *Q* morphometric explanatory variables $X_1, X_2, \ldots, X_Q$ with values $x_{1ij}, x_{2ij}, \ldots, x_{Qij}$ for the image j (j=1,2,…, J_i) of the individual *i*.

We can obtain a covariance matrix in the different levels from a multivariate model (Goldstein, 1995). For morphometric data of CT images, the model can be formulated in the following way:

Let x_{ijl} be the measurement of the lth morphometric variable (l=1,2,…,Q) in the *j*-th image ($j = 1,2, \ldots, J_i$) of the *i*-th individual ($i = 1,2, \ldots, I$). Although the morphometric indicators could be seen as the units at the lowest level, the model is defined at the image level and at the individual level. For each morphometric variable, an indicator variable $d_{ijq} = 1$ is created when $l = q$, $d_{ijq} = 0$ otherwise. Hence, the multivariate two-level model can be described by the following equation:

$$x_{ijl} = \sum_{q=1}^{Q} (\gamma_q + u_{ijq} + v_{jq}) d_{ijq} \quad (1)$$

The fixed parameters γ_q *(q=1,2,…,Q)* represent the estimates of the population means of the *Q* morphometric variables. The unobservable error vectors $\boldsymbol{u}'_{ij} = (u_{ij1}, u_{ij2,\cdots,} u_{ijQ})$ y $\boldsymbol{v}'_i = (v_{i1}, v_{i2,\cdots,} v_{iQ})$ and $\boldsymbol{v}'_i = (v_{i1}, v_{i2,\cdots,} v_{iQ})$ are assumed to be independent and identically distributed. More explicitly, it will be established that $\boldsymbol{u}_{ij} \sim N_{Q^+}(\mathbf{0}, \boldsymbol{\Sigma})$ *y* $\boldsymbol{v}_i \sim N_{Q^+}(\mathbf{0}, \boldsymbol{\Omega})$ for each *i* and *j*. The matrix of variance and covariance of the random errors of level-1, $\boldsymbol{\Sigma} = \mathrm{Cov}\ (\boldsymbol{u}_{ij})$, is called the intra-individuals covariance matrix, while the matrix of the errors of level-2, $\boldsymbol{\Omega} = \mathrm{Cov}\ (\mathbf{v}_i)$, is called the between-individual covariance matrix.

One of the advantages of this procedure is that the estimates of the covariance (or correlation) matrices at each level of analysis can be used directly and separately in subsequent analyzes (Goldstein, 1995); for example, in the PCA-LVR approach the set of factors or main components is estimated only at level-macro, this is, from a between-individual covariance matrix of the morphometric variables.

The approach proposed in this paper is based on using MFA to create one or more latent factors at each level of analysis (Croon and Van Veldhoven, 2007; Lawley and Maxwell, 1971). Here, the two covariance matrices are used to carry out a simultaneously factor analysis at both levels. Only the factors at individual-level are selected in order to represent the morphometric differences of the hemorrhages between the patients.

Beyond the method used to obtain the latent variables, the foundations of the strategy proposed in this article are based on the same ideas as the PCA-LVR approach. From a practical point of view, one advantage of both procedures is that these do not assume that there is a complete set of morphometric indicators per image; therefore, incomplete data due to the impossibility of performing any measurement on the image (missing data) can be accommodated without difficulty to produce the covariance matrices.

3.1. The proposed approach

We now can outline the procedure proposed, which we call MFA-LVR, in two-stage framework. Statistics techniques used in each step are briefly described in the following two subsections.

3.1.1 Multilevel Factor Analysis

In a first step, the idea is to conduct an MLFA in order to model the values of the set of *Q* morphometric explanatory variables, denoted by $\boldsymbol{X}_{ij}$= ($X_{1ij}, X \ldots, X_{Qij}$), as a function of factors at image level (or level 1) and at the individual level (or level 2), represented by $\boldsymbol{\eta}_{B_i}$ and $\boldsymbol{\eta}_{w_{ij}}$, respectively.

The model level image is given by:

$$\boldsymbol{X}_{ij} = \boldsymbol{\nu}_i + \boldsymbol{\Lambda}_w \boldsymbol{\eta}_{w_{ij}} + \boldsymbol{\varepsilon}_{ij}\,, \quad (2)$$

where, $\boldsymbol{\nu}_i$ is a vector of the mean responses of the individual *i* for each of the *Q* morphometric indicators for the population of images nested in the individual *i*; $\boldsymbol{\eta}_{w_{ij}}$ is a vector of the image *j* for the factors at the level image, with $E(\boldsymbol{\eta}_w) = \mathbf{0}$ and $Var(\boldsymbol{\eta}_w) = \boldsymbol{\Psi}_w$; $\boldsymbol{\Lambda}_w$ is a factor loadings matrix that describes the relationships between the image level factors, $\boldsymbol{\eta}_w$, and the variables, $\boldsymbol{X}_{ij}$; the $\varepsilon_{ij}{}'s$ are uncorrelated image-specific errors, with $E(\boldsymbol{\varepsilon}) = 0$ and $Var(\boldsymbol{\varepsilon}) = \Theta$. Typically, when the *X*'s are continuous variables, we assumed that the errors and factors are distributed normally, with all errors uncorrelated with each other and with the factors.

The between-individual model is given by:

$$\boldsymbol{\nu}_i = \boldsymbol{\gamma} + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{B_i} + \boldsymbol{\zeta}_i \,, \tag{3}$$

where, $\boldsymbol{\gamma}$ is a vector of means for the *P* morphometric indicators; $\boldsymbol{\eta}_i$ is a vector of the values of the individual *i* for the factors at the individual level, with $E(\boldsymbol{\eta}_B) = \mathbf{0}$ y $Var(\boldsymbol{\eta}_B) = \psi_B$; $\boldsymbol{\Lambda}_B$ is a factor loading matrix that describes the relationships between the individual level factors, $\boldsymbol{\eta}_{B_i}$, and the values of the individual level random intercept, $\boldsymbol{\nu}_i$; and $\boldsymbol{\zeta}_i$ is the error for the individual *i*, with $(\boldsymbol{\zeta}) = \mathbf{0}$ and $Var(\boldsymbol{\zeta}) = \boldsymbol{\sigma}$. Like the image-level model, the errors and factors are assumed to be distributed normally, with all the residues mismatched between themselves and with the factors.

Substituting equation (3) in (2), a single combined model is obtained:

$$\boldsymbol{X}_{ij} = \boldsymbol{\gamma} + \boldsymbol{\Lambda}_w \boldsymbol{\eta}_{w_{ij}} + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{B_i} + \boldsymbol{\zeta}_i + \boldsymbol{\varepsilon}_{ij}, \tag{4}$$

where, it is shown that the observed responses at the image level are specified as different effects of factors at the image level and at the individual level. In this model, the individual level has M_2 factors with corresponding factor loading matrix $\boldsymbol{\Lambda}_B$ and the image level has M_1 factors with corresponding factor loading matrix $\boldsymbol{\Lambda}_w$.

In this article, the vector of factors at the individual level $\boldsymbol{\eta}_{\boldsymbol{B_i}}$ represents a set of M_2 synthetic indicators of the morphometric differences of the hemorrhages between the patients, which should be considered as explanatory variables in a logistic regression model[23].

3.1.2 A model with latent variables for discrete response

For a dichotomous patient response, such as survival, the latent variable model can be formulated as a logistic regression equation:

Let y_i be the realization of a response variable Y measured at the individual level that can take the values one and zero with probability π_i and $1 - \pi_i$, respectively, with Bernoulli distribution with parameter π_i. Consider that the dichotomous answer Y_i can be explained or predicted by assuming that the logit of the underlying probability π_i is a linear function of multiple explanatory variables $Z_{1i}, Z_{2i}, \ldots, Z_{pi}$ at the individual level and a set of latent variables $\eta_{1B_i}, \eta_{2B_i}, \cdots, \eta_{M_2 B_i}$ at the individual level. The model can be expressed as:

$$logit(\pi_i) = log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \boldsymbol{\eta}_{\boldsymbol{B_i}}^{\intercal}\boldsymbol{\beta}_1 + \boldsymbol{Z}_i^{\intercal}\boldsymbol{\beta}_2 \tag{5}$$

It is important to note that $\boldsymbol{\eta}_{\boldsymbol{B_i}}$ is a vector of continuous latent variables, but the vector $\boldsymbol{Z_i}$ can be of any type and that the analysis of interactions between the unobserved variables of the vector $\boldsymbol{\eta}_{\boldsymbol{B_i}}$ and the observed variables of the vector $\boldsymbol{\eta}_{\boldsymbol{B_i}}$ is valid.

4. AN APPLICATION BASED ON DATA OF NEUROIMAGES AND CLINICAL PARAMETERS TO PREDICT SURVIVAL

By way of example, in this section we apply the strategy described in this paper to real data in order to explore what could be gained by using MFA-LVR approach in a real research problem. For this purpose, the same dataset previously described in Montero et al. (2017) was used (the online database is given in Montero (2019). The results of a logistic regression model were compared with those obtained in that publication, where the CPA-LVR approach was applied.

Once more, we considered that the data used in this study constitute a hierarchically ordered system, where the images are nested within individuals, establishing two different levels: the units of study in level-1 (micro level) are the images and units of study in level-2 (macro level) are of patients (Figure 2).

In summary, we analyzed data from 39 patients (13 women and 26 men) with a diagnosis of ICH and who were carried out a CT scan within the first 6 hours of initiating symptoms. Ages were in an interval between 48 and 87 years. Of the total patients, 11 died after the hemorrhage. The following morphometric measurements of each of the hemorrhage images obtained by CT are also part of the database: area, perimeter, elliptical shape factor, maximum diameter and minimum diameter (Rodríguez and Sossa, 2011). The morphometric variables were measured using the MADIP system (Rodríguez et al., 2001) and a transformation was applied to avoid the undue importance of disproportionate ranges caused by the different units of measure of some variables.

The response of interest in this study is the survival of the patient after the hemorrhage. The variable takes the value 0 if the patient survives and the value 1 if it does not survive. The risk factor is a single latent variable (named Factor) that synthesizes the morphometric characteristics of neuroimages. The explanatory variables of the patient's sex and age were also considered as predictors of survival, since it is hypothesized that men

and younger people are less likely to survive the hemorrhage. The female sex is taken as a reference category. Age was measured in number of years and was categorized into three groups according to the following age groups: less than 60 years, between 61 and 75 years and over 75 years. The group of patients between 61 and 75 years old is taken as reference. The interaction between the latent variable Factor and the variable Age was also included in the model, considering that the influence of the morphometry of the images on the survival of patients can be mediated by the influence of age.
The variables in level-1 (micro) refer to the morphometric characteristics of images; while the characteristics of patients represent the variables in level-2 (macro).

5. RESULTS AND DISCUSSION

Table 1 shows the values of the Intraclass Correlation Coefficient (ICC) (Koch, 1982) for each morphometric variable, which in this work refers to the proportion of the total variance in the variable of interest due to the differences between the individuals.
Between the variables that characterize the size of the hemorrhage, the highest estimated ICC (0.61) was for the Perimeter variable (Table 1). The smallest (0.58) was for the Area variable. Therefore, a significant percentage of the variability of morphometric measurements is attributable to differences between patients, rather than between images. However, considerable variability there is between the variables in terms of the proportion of variation explained between the images, especially for the case of the variable Elliptic Form Factor.

Table 1. Intraclass Correlation Coefficient (ICC) for each morphometric variable

Variables	Intraclass Correlation Coefficient
Area	0.58
Perimeter	0.61
Elliptical Shape Factor	0.15
Maximum diameter	0.45
Minimum diameter	0.59

Of particular interest in this paper is the variance decomposition of the morphometric variables with respect to within individual variation and between individual. Table 2 presents the corresponding within-individual and between-individual correlation matrices of the morphometric variables. As shown in this table, the correlations between the morphometric variables are different at the two levels. All the between-individual correlations are greater than the within individual correlations, except for the correlations of the variable Elliptic Shape Factor with the variables Maximum Diameter and Minimum Diameter. In particular, can been seen that the correlations at the image level are comprised in the range from -0.127 to 0.944. At the individual level, the correlations are greater with values from -0.333 to 0.986. These correlations are explained through a Multilevel Factorial Analysis of two factors, one in each level.

Table 2. Within and between correlation matrices of the morphometric variables

Within correlation matrix					
Area	1.00				
Perimeter	0.81	1.00			
Elliptical Shape Factor	0.17	-0.23	1.00		
Maximum diameter	0.83	0.70	0.52	1.00	
Minimum diameter	0.75	0.95	-0.34	0.56	1.00
Between correlation matrix					
Area	1.00				
Perimeter	0.98	1.00			
Elliptical Shape Factor	-0.49	-0.53	1.00		
Maximum diameter	0.97	0.97	-0.33	1.00	
Minimum diameter	0.98	0.98	-0.61	0.94	1.00

5.1. Multilevel Factorial Analysis of the morphometric variables

Here, we use Multilevel Factor Analysis in order to construct a latent variable that represents a summary measure of morphometric differences of the hemorrhage between patients. The results of a factorial model of two levels with a single factor in each level are presented in Table 3.

Table 3. Factor loads for the morphometric variables

Variable	Factor Loads: Within (micro level) Factor	Factor Loads: Between (macro level) Factor
Area	0.04	0.09
Perimeter	<u>1.00</u>	<u>0.98</u>
Elliptical Shape Factor	0.09	-0.30
Maximum diameter	<u>-0.87</u>	<u>0.77</u>
Minimum diameter	<u>-0.98</u>	<u>1.00</u>
Factor Variance	0.39	3.17

Note. Underlined figures represent loadings greater than 0.50.

The factor structure obtained from the within individual level analysis cannot be assumed to also hold at the between individual level analysis[27]. The interest of the results of the MLFA in this paper is in the estimations of factor loads at the macro level and in the values or "scores" of this factor. The loads associated to the related variables to the length of the image are great, both at the micro factor and at the macro factor, but the variance of the factor at the macro level is greater, and therefore, this factor is explaining a greater proportion of uncertainty. The loads of the factor at the macro level are all positive, except for the one corresponding to the variable Elliptic Shape Factor, and since it was already seen that all inter-individual correlations are also positive, with except for the correlations of each variable with the Elliptic Form Factor, it can be thought that this factor represents a general measure of the size of the factor as opposed to the form of the hemorrhage.
In Figure 3, a graph of the macro level factor values for the 39 patients is shown, ordered according to their rank. The graph suggests some evidence of how patients with extreme positive values for the factor could not survive the hemorrhage, and that patients with the extreme negative values of the factor survived the hemorrhage.

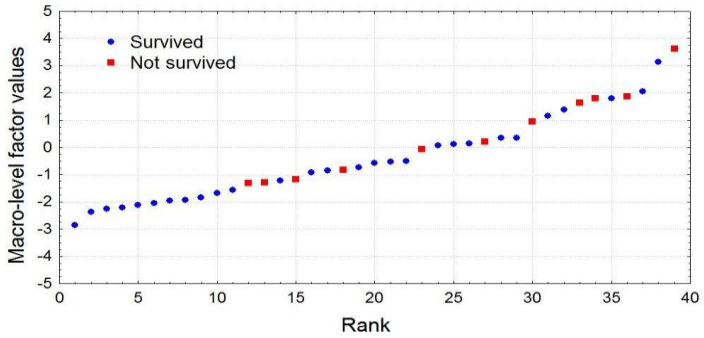

Figure 3. Plot of the values of the macro level factor for the 39 patients ordered according to rank

5.2. Logistic regression analysis

The scores of the macro level factor obtained by the MFA were considered as realizations of the latent explanatory variable Factor in a logistic regression model (MFA-LVR model) (see Table 4). The results obtained in this study (expressed as odds ratios and 95% confidence intervals) are compared with those obtained from a PCA-LVR approach, published in Montero et al. (2017). When the odds ratio (OR) exhibits a value greater than 1 and the confidence interval (CI) does not contain the value 1, the effect can be considered statistically significant.

The interpretation of the results of the logistic regression using the latent explanatory variable Factor obtained through the MFA-LVR approach are similar to those obtained through the CPA-LVR approach.

Table 4. Odds Ratio (OR) and 95% confidence intervals (CI), obtained for the logistic regression models of the survival of the patients, conditioned to variable Factor, constructed via PCA-LVR approach (PCA-LVR Model) and MFA-LVR approach (MFA-LVR Model) and the variables Age and Sex of the patients

	PCA-LVR Model			MFA-LVR Model		
	OR	95% CI		OR	95% CI	
		Left Limit	Right Limit		Left Limit	Right Limit
Sex	16.11	1.47	175.95	24.18	1.34	357.86
Age (≤ 60)	32.00	1.37	746.22	28.93	1.24	673.26
Age (≥ 76)	1.17	0.07	20.60	0.56	0.02	12.26
Factor	2.76	1.21	6.30	4.82	1.32	17.52
Factor x Age (≤ 60)	0.47	0.12	1.84	0.22	0.33	1.55
Factor x Age (≥ 76)	0.19	0.04	0.97	0.08	0.01	0.81

In both models, the indicator variables Sex (being male) and Age (less than 60 years) have a significant positive effect on the probability of dying after hemorrhage. The latent variable Factor also showed a positive effect on the reference group of the Age variable (patients between 61 and 75 years). On the other hand, the effect of the latent variable Factor is significantly higher in this age group in relation to those over 76 years of age. From here, it is interpreted that patients between 61 and 75 years with large hemorrhages are more likely to die after hemorrhage.

A probability of 0.5 was used as the cut-off point to classify the predicted values of the probability of occurrence of the event (survived or not survived) in two classes. The overall accuracy of the two models was determined by comparing the predicted values with the actual occurrence of the event. The results of this analysis are presented in Tables 5a (PCA model) and 5b (MLA model). The PCA model correctly classified 89.3% of the patients who survived and 63.6% of the patients who died (Table 5a). The MLA Model correctly predicted the result in 92.9% of the living patients, and in 63.6% of the deceased patients (Table 5b).

Table 5a. Classification Table (PCA model)

Observed		Predicted		
		Survival		Percentage of correctly classified cases
		Yes	No	
Survival	Yes	25	3	89,3 %
	No	4	7	63,6 %
Global Percentage				82,1 %

Table 5b. Classification Table (MLA model)

Observed		Predicted		
		Survival		Percentage of correctly classified cases
		Yes	No	
Survival	Yes	26	2	92,9 %
	No	4	7	63,6 %
Global Percentage				84,6 %

5.3. ROC curves analysis

The discriminatory power of the two models was analyzed using the area under the ROC curves (acronym of Receiver Operating Characteristic, or Receiver Operating Characteristic). The ROC curves (Figure 4) were constructed representing the ratio of true positives (patients who died and who the model predicted as deceased (sensitivity)) versus the reason for false positives (patients who survived and were incorrectly

classified as deceased (1 - specificity). The curve of the ACP model is very close to the curve of the MLFA model.

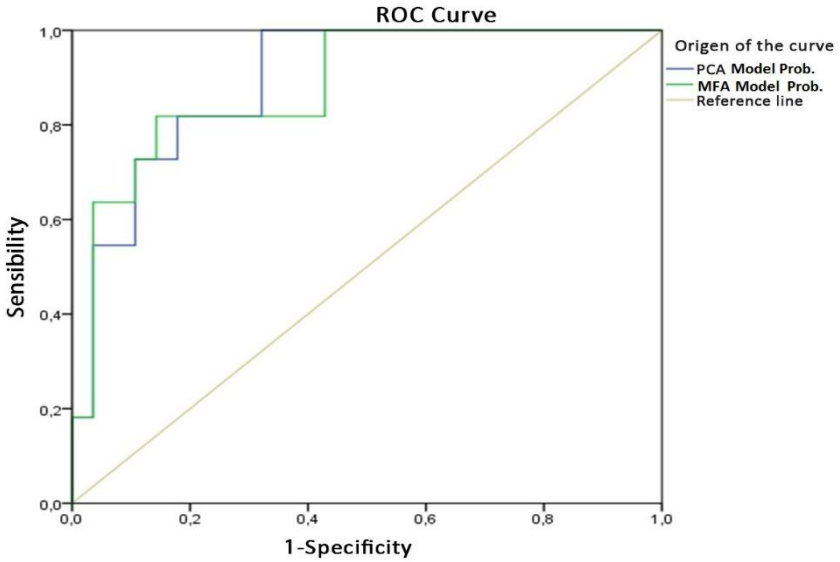

Figure 4. ROC Curves for the ACP model and the AFM mode

The area under the ROC curve (Table 6) for the logistic regression model (ACP-LVR approach) is 0.89 (0.79 to 0.99, 95% CI), while the area under the ROC curve associated with the model (MFA-LVR approach) is 0.88 (0.77 to 0.99). Hence, the MFA-LVR model is comparable to the ACP-LVR model to predict survival in this sample.

Table 6. Area under the curve ROC ($AREA_{ROC}$), Standard Error (SE), Signification (p) e 95% confidence interval (95% IC) obtained from the CPA model and the MLFA model

Variable(s) of resulted of the models	$AREA_{ROC}$	SE	P	95% confidence interval	
				Left Limit	Right Limit
Probability: CPA- LVR Model	0.893	0.051	0.000	0.792	0.994
Probability: .MFA-LVR Model	0.883	0.058	0.000	0.769	0.997

The MFA-LVR model fitted to predict the survival of patients with ICH showed an overall performance comparable to the PCA-LVR model and takes advantage in the correct classification of patients who survived in the sample studied. These results suggest a possible use of the MFA method to construct a latent typology of the images of the hemorrhages from morphometric variables at the image level. In addition, these results support the hypothesis that the MFA-LVR model would generate a more accurate classification of subjects than an ACP-LVR model using the same input morphometric variables.

6. CONCLUSIONS

In this article, we have described a novel strategy for analyzing micro-macro designs, where the multilevel structure between images and individuals were appropriately handled. This approach is especially useful for evaluate the effects of the morphometry of CT images on some clinical response of the patient.

In a micro-macro analysis, data at the micro level (morphometric variables) need to be aggregated at the macro level, such that the aggregate values can be related to the patient-related variable. The strategy propose in this paper allows the creation of variables that capture both the variability between images and between individuals, using one or more latent or unobserved factors at each level of analysis. The latent factors at the individual level (macro level) are then considered as explanatory variables in a single logistic model at individual level.

The statistical framework of this modeling scheme is based on the approach proposed by Montero et al. (2017). The difference of our proposal (MFA-LVR approach) with respect to the referred approach (ACP-

LVR approach) is in the method used in order to construct the latent variables at the macro level. Montero et al. (2017) proposed to employ a combination of two methods, Multivariate Multilevel Models and Principal Component Principal; while here recommend to use a particular technical: Multilevel Factorial Analysis. We show, through an example with real data, that to use this method would be likewise an appropriate way to obtain latent variables at the macro level, which also can be interpreted as synthetic indicators of the differences of the images between individuals.

Although the study suggests a better classification of the patients who survived the ICH, using the MFA-LVR approach, the interpretation of the data must take into account the relatively small sample size and the resulting limitations in terms of statistical power. Other potential limitations should also be considered, such as the non-inclusion in the model of other important clinical variables that may also be considered responsible for patient mortality. In this sense, future investigations with larger sample sizes and some numerical simulation could contribute to conclusions more consistent.

However, from this perspective, we think that by making explicit the assumptions on which method to use in order to construct the latent variables, both approaches discussed in this article could be integrated into a general strategy for analyzing micro-macro situations where the interest of the research lies combining morphometric characteristics of CT images and characteristics of patient to evaluate possible risks factors in the patient population. On the other hand, the study of the conditional probability distribution allows identifies groups or characteristics of the patient`s population where the morphometry of the images can influence of vital manner in the rates of outcome of the study population.

The decision of whether to choice one strategy or another will be mainly focus on the knowledge of the methods and the functional relationship between the outcome variable and the different covariates available in each study, as well as the available software programs.

In summary, we think that beyond the limitations presented, the micro-macro multilevel modeling strategies discussed here may enrich the statistical analysis of CT image data that will lead to improved efficacy in clinical diagnosis. In particular, we consider that the general methodology outline in this article has the following advantages: a) It is possible to include one parsimonious latent variables model that facilitates interplay between images and individuals. b) It allows making inferences at the individual level using the information available at the image level. c) Balanced designs are not a requirement. d) We can chose freely between two methods in order to construct the latent variables.

Further methodological developments are necessary in order to fit more sophisticated micro-macro models for TC data. However our strategy may be viewed as useful exploratory tool to measure the morphometric effects of multiple images on a clinical outcome variable (e.g. survived) of specific patients.

RECEIVED: SEPTEMBER, 2019
REVISED: NOVEMBER, 2019

REFERENCES

[1] BENNINK M., CROON, M.A., VERMUNT, J.K. (2013): Micro–macro multilevel analysis for discrete data: A latent variable approach and an application on personal network data. **Sociological Methods and Research**, 42, 431-457.

[2] BENNINK M. (2014): **Micro-macro multilevel analysis for discrete data**. Ridderkerk: Ridderprint.

[3] BENNINK M., CROON M.A., KROON B. and VERMUNT J.K. (2016): Micro-macro multilevel latent class models with multiple discrete individual-level variables. **Advances in Data Analysis and Classification**. 10 , 139-154.

[4] BECKER D., BREUSTEDT W., ZUBER C.I. (2018): Surpassing Simple Aggregation: Advanced Strategies for Analyzing Contextual-Level Outcomes in Multilevel Models. **Methods, Data, Analyses**, 12, 233-264.

[5] CROON M.A., VAN VELDHOVEN, M.J. (2007): Predicting group-level outcome variables from variables measured at the individual level: a latent variable multilevel model. **Psychological Methods** 12, 45-57.

[6] CROON M.A. and KROON B. (2016): Micro-macro multilevel latent class models with multiple discrete individual-level variables. In: **Advances in Data Analysis and Classification**. 10, 139-154.

[7] ESCUDERO D., MARQUÉS L. and TABOADA F. (2008): Actualización en hemorragia cerebral espontánea. **Med Intensiva**. 32, 282-95.

[8] FERRETE A.M., EGEA J.J., VILCHES Á., GODOY D.A. and MURILLO F. (2015): Predictores de mortalidad y mal resultado funcional en la hemorragia intra parenquimatosa espontánea grave: estudio prospectivo observacional. **Med Intensiva**. 39, 422-432.

[9] GOLDSTEIN H. (1995): **Multilevel Statistical Models**. 2nd. Ed. Halsted Press, New York.
[10] GOLDSTEIN H. and BROWNE W. (2002): Multilevel factor analysis modelling using Markov Chain Monte Carlo (MCMC) estimation. In: Marcoulides GA, Moustaki M, editors. **Latent variable and latent structure models.** Mahwah, NJ. Lawrence Erlbaum Associates Inc Publishers, 225–44.
[11] HOSMER D.W. and LEMESHOW S. (1989): **Applied Logistic Regression**. New York: Wiley.
[12] HOX, J.J. (2010): **Multilevel Analysis: Techniques and Applications** (2nd. ed.). NewYork, NY: Routledge.
[13] JOLIFFE I.T. (2002): **Principal component analysis.** Springer , New York.
[14] KOCH, G.G. (1982): Intraclass correlation coefficient. In Samuel Kotz and Norman L. Johnson (ed.). **Encyclopedia of Statistical Sciences.** 4. John Wiley & Sons, New York.
[15] KREFT IG DE LEEUW J. (1998): **Introducing Multilevel Modeling**. Sage, London.
[16] LAWLEY D.N. and MAXWELL A.E. (1971): **Factor Analysis as a Statistical Method.** 2nd edition. London.
[17] LONGFORD N. (1993): **Random Coefficient Models**. Oxford University Press, New York.
[18] MONTERO M., RODRÍGUEZ R. and RODRÍGUEZ L.A. (2017): Morphometric evaluation of neuroimaging as a prognostic factor of survival in patients with cerebral hemorrhage. **Imagen Diagnóstica.** 8, 37-74.
[19] MONTERO, M. (2019): Morphometric evaluation of neuroimaging as a prognostic factor of survival in patients with cerebral hemorrhage, **Mendeley Data**. Disponible en http://dx.doi.org/10.17632/y4h44mw8k7.1
[20] MORGENSTERN L.B., HEMPHILL J.C., ANDERSON C., BECKER K., BRODERICK J.P., CONNOLLY E.S. (2010) Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. **Stroke,** 41, 2108-2129.
[21] MUTHÉN B. (1998): Latent variable modeling in heterogeneous populations. **Psychometrika**.54, 557-85.
[22] RAUDENBUSH S.W, BRYK, A.S. (2002): **Hierarchical Linear Models. Applications and Data Analysis Methods**, Second Ed. Sage, Thousand Oaks, CA
[23] RODRÍGUEZ R., ALARCÓN T.E. and SÁNCHEZ L. (2001): MADIP. Morphometric Analysis by Digital Image Processing, Proceedings of the **IX Spanish Symposium on Pattern Recognition and Image Analysis**, Spain I. 291-298.
[24] RODRÍGUEZ R. and SOSSA. J.H. (2011): **Procesamiento y Análisis Digital de Imágenes**. Editorial Ra-Ma, España.
[25] RODRÍGUEZ R. and SOSSA J. H. (2019): Mathematical Techniques for Biomedical Image Segmentation. In In R. Narayan (Ed.), **Encyclopedia of Biomedical Engineering**, 3, 64–78.
[26] SNIJDERS, T.A.B., BOSKER, R.J. (2012): **Multilevel Analysis** (Second Edition). Sage, London.
[27] TELLERÍA A. (2006): Tratamiento e indicadores pronósticos del paciente con hemorragia intracerebral espontánea. **Rev. Neurol**. 42, 341-349.