This is a previous version of the article published in Soft Computing. 2020, 24: 11045-11057. https://doi.org/10.1007/s00500-019-04469-3

Quantification of cultural identity through artificial intelligence: a case study on the Waorani Amazonian ethnicity

Aldrin Espín-León[1], Antonio Jimeno-Morenilla[2], María Luisa Pertegal-Felices[3*], Jorge Azorín-López[2].

[1] Faculty of Sociology and Social Work. Central University of Ecuador, Ecuador
[2]Department of Computer Technology. University of Alicante, Spain
[3]Developmental and Educational Psychology department. University of Alicante, Spain
[*]Corresponding author email: ml.pertegal@ua.es

Abstract. Cultural identity represents a construct that is difficult to measure because the nature of the variables involved is complex and mostly subjective. Indigenous people belonging to communities in the Amazon have seen their identity drastically eroded in recent decades as a result of the process of Western acculturation. In this context, the quantification of this loss arises by identifying the most and least affected identity components of this process. This research work presents a quantitative method, based on artificial intelligence, which evaluates the cultural identity of an Amazonian indigenous community: the Woaroani. On the basis of an instrument developed in the field by the authors, this method automatically classifies the individuals and provides a subspace of it able to identify the weights of the subcomponents of this instrument in regard to its contribution to the Waorani identity. The systematic application of the instrument together with the AI-based system can provide decision-makers with valuable information about which aspects of their identity are most sensitive to change and thus help design development policies that minimally interfere with their ethnic identity.

Keywords: Cultural identity measurement, preservation of ethnic identity, Amazonia, Waorani community, artificial intelligence.

1 Introduction

Cultural identity is a broad and complex concept that addresses a multitude of factors, most of which are subjective in nature: law, beliefs, art, morals, customs, habits and acquired abilities [1]. It also considers ways of life and ways of living in society [2]. Identity is part of culture, and cannot be determined tangibly or through physical characteristics, nor is the individual considered in isolation from his or her community and is historically defined by various aspects of immaterial character such as language, social relations, ceremonies, rites, beliefs [3]. The culture and identity of peoples are constantly changing or merged by territorial heterogeneity and diversity of cultural expressions, making a system of indicators for analysis difficult and complex [4].

In Latin America, indigenous communities and nationalities have undergone accelerated cultural changes and mergers of identity as a consequence of colonyism by dominant groups [5]. The constant abolitions of their culture were established from the political power and the ethnic domination was maintained, excluding them of all type citizen participation, the state considered that they should be of an only culture, an only language and an only religion [6]. In short, indigenous people have suffered a strong process of acculturation as a result of colonialism, the main consequence of which has been the loss of their own identity [7].

In the case of the Amazonian Indians, there is widespread concern that the inhabitants will lose their identity or, at worst, disappear. The link that exists between individuals and nature is essential in Amazonian societies [8]. However, awareness of the problem has also been raised. Society is beginning to value indigenous cultural diversity and to sensitize Western societies through education, information and the provision of technical tools that will somehow recover lost ground [9].

Precisely, this study aims to present tools that allow us to discern the cultural state of an indigenous nationality in order to check whether it is being degraded as a consequence of a process of western acculturation.

The authors developed at an earlier stage an instrument that made it possible to identify factors of indigenous identity through variables and indicators obtained through qualitative information obtained from the indigenous people themselves. The objective of this research is classify populations with indigenous identity from populations with western identity and know which is the weight of the variables identified in the instrument in relation with the cultural identity. This identification would make it possible to know whether certain policies aimed at these peoples cause undesirable effects on their identity. In order to determine this influence, clustering analysis will be carried out using techniques based on artificial intelligence.

2 Background

The ethnic and socio-cultural backgrounds of individuals or minority groups in countries can be predictors of social attitudes. The accepted definition of ethnic group was enunciated by Max Weber [10]. In this context, we call "ethnic groups" those human groups that have a subjective belief in their common ancestry due to similarities of physical type or customs or both, or by memories of colonization and migration. It is precisely this subjectivity that is responsible for the fact that the measurement of identity is a complex process. In cultural statistics, when there is a subjective assessment or they have ideological components, alternatives must be proposed so that the assessments are objective and can be analyzed quantitatively [11]. If measuring identity has a large qualitative component, quantitative analysis minimizes intuitive and subjective conceptions.

The following is a survey of studies in two different approaches: first, studies that propose models and instruments to obtain, measure or classify ethnic identity information and second, studies that have used automatic tools based on Artificial Intelligence to classify or identify individuals with ethnic or social characteristics.

2.1 Models and instruments to measure ethnic identity

Ethnic identity is determined by four dimensions: cognitive, evaluative, affective, and behavioral [12]. Instruments often ask about a definition of self, about feelings of belonging to a group, about pride or intention of belonging to it, about the language, religion and knowledge of history of their own group [13].

In Western Europe, given the growing diversity of ethnicities and peoples from almost all over the world, Schneider and Heath [14] proposed, through statistical procedures, to classify and measure the validity of data from the ethnic and socio-cultural background of individuals. The work shows that ethnic origins can predict whether respondents identify themselves as belonging to an ethnic minority, whether they consider themselves in a discriminated group, national identity, and attitudes toward immigration. It also identified indigenous minorities, as well as those with migratory backgrounds. This study focused on the concept of background rather than current identity because some people with ethnic backgrounds no longer feel close to their ethnic group of origin. The criteria they use are: National origin, religion and language. They also include generational state criteria with six categories, only for those with a migration background.

Other studies focus on the adaptability of new individuals to a group. Carley [15] analyzed individual behavior based on the thesis that interaction leads to shared knowledge. He reported that some groups endure longer, are more stable, and have greater capacity than other groups to incorporate new members or ideas without losing their own character. He found that groups that are stable in the short term do not necessarily retain their distinctive character in the long term as new members enter or new ideas are discovered.

On the other hand, authors such as Cross and Helms [16] [17], developed theoretical models that contemplate three states referring to ethnic identity: 1. negative valuation of blacks and their preference toward another race (whites); 2. knowledge of and interest in them; and 3. international and group acceptance.

The ethnic identity seen from the acculturation includes questionnaires that use external and internal variables. External variables deal with language, media (tv radio), friendship relations and ethnic traditions, while internal variables contain cognitive, affective and moral issues [18]. Another model of ethnic identity proposed by Phinney and Rotheram [19], includes six components: Self-definition, attitudes towards one's own ethnic group, attitude towards oneself, interest of the ethnic group and commitment to ethnic identity.

Ethnic identity is also addressed from acculturation. John Berry [20] proposed a model between acculturation and ethnic identity considering the importance of customs: a) preserving the customs of the minority group, and b) the adoption of customs of the majority group, and whose variations of acculturation are known as: integration - preserving and adopting, assimilation - adopting without preserving, separation - conserving without adopting - and marginalization - without preserving or adopting [21].

Ethnic residential segregation is also analyzed in quantitative studies. Denton and Massey [22], using factor analysis, explored the relationships between segregation rates, identifying five distinct dimensions of segregation: uniformity, exposure, concentration, centralization, and clustering. Their work became the basis for segregation studies, and the approach they proposed was admitted almost entirely.

However, these models and components are focused from generalist points of view and lack items that specify aspects of indigenous ethnic identity and even less of the Amazon indigenous.

There are works that measure gender identity, related to sexual roles and the adaptation of gender identity towards stereotypes [23] [24]. Muradas and Rodríguez [25], on the other hand, use instruments with items to measure tangible culture such as churches, museums, parks, cultural centers, among others. These instruments would be limited to gender and would only make sense in populations with a Western structure.

Ethnic identity was also measured through an instrument applied in Costa Rica [26]. This instrument, which consisted of 12 items, was based on the work developed by Phinney which originally had 14 items. Although this instrument included items such as tradition, ethnic roots, music, food and participation, it did not cover specific components of Amazonian indigenous culture such as dance, hunting and fishing, housing, medicine or ancestral knowledge. Precisely, hunting is considered not only as a basic strategy for food but as an integral part of their cultures [27].

In the same country - Costa Rica - an ethnic identity study was also carried out on a group of 90 “Huetares de Quitirrisí de Mora” Indians, Indians who insist on maintaining their traditions and customs despite the accelerated processes of acculturation. This study used a quantitative methodology that included socio-demographic items, cultural elements regarding their identity and open questions grouped into categories. The results denoted little value in ethnicity such as traditional dress, language and customs, so these aspects were pointed out as a loss of culture of visible elements, although they were still proud to belong to the ethnic group.

A similar study on identity was conducted in the community of Chiapas, Mexico [28]. They worked on identity narratives using the autobiographical multi-methodology [29]. Three groups participated in the experiment: Chiapas indigenous people, students from a rural area, and from a Spanish city. The results obtained were clearly qualitative; however, this study takes the individual as a methodological reference and was applied to different regions with different identity characteristics.

2.2 AI-based tools for classifying and identifying individuals

IA-based technologies and tools have been widely used in the identification of characteristics and patterns for the recognition of individuals through morphological features [30] [31]. They have also addressed -although to a lesser extent- studies on the identification of individuals with non-morphological characteristics such as ethnic and social issues.

Information such as height, weight, age, gender, ethnicity and eye colour are generally not used in identification phases. Only when a trustworthy individual is erroneously rejected by the system a human operator intervene to verify the user's soft biometric traits [32]. In this sense, the same authors mention that soft biometric recognition systems, unlike traditional primary recognition systems that use fingerprint, face, iris or hand geometry, use age, gender, ethnicity or height.

Jain [33] proposed to use auxiliary information describing a framework for integrating the information provided by soft biometric indicators with the output of the primary biometric system, analyzing the performance by integrating auxiliary information such as gender, ethnicity and height with the output of a biometric fingerprint system. Used a database to classify 160 individuals, information containing 4 fingerprint prints with gender and ethnicity information. This classifier identifies the ethnicity of a test user as Asian or non-Asian with an accuracy of 96.3%. The use of ethnicity, gender and length information along with the fingerprint led to an approximate 5% improvement over the primary biometric system. Shakhnarovich et al. [34] presented a real time face detection and recognition system. The same structure is used for the extraction of demographic information, including gender and ethnicity. They defined the same categories of ethnicity, Asian and non-Asian, their system focuses on low resolution images (24 × 24) with weakly aligned face data. Its reported accuracy was 80%.

Recognition of Asian ethnicity was also proposed by Lu et al. [35]. He used a scheme based on linear discriminant analysis (LDA) on a set of balanced data so that the training process was not predisposed to a particular ethnicity. This also suggests that the previous probabilities of the two classes are almost equal (0.5). Experimental ethnic classification results were encouraging.

For the classification of gender and ethnicity Gutta et al.[36] proposed a hybrid classifier based on radial base function sets (RBF), inductive decision trees (DT) and support vector machines (SVM), using images of 64 × 72.

In the case of classifying members of a group into subgroups by proximity of individuals, Carley [37] and Freeman [38] used classification and optimization algorithms using graph partitions. Freeman relied on finding social groups given a matrix of social proximities based on the heuristics of a simple genetic algorithm. He proposed a new solution to the problem of dividing a matrix of social proximities into groups. He demonstrated that such an algorithm can be used to reveal the structure of the group without contradicting the descriptions of the ethnographers, and furthermore, the results could complement the ethnography.

Many research projects focused on classifying and identifying identities on social networks, so Otto et al.[39] tried to solve the problem of grouping millions of images and videos of unlabeled faces on the Internet into individual identities, many of them of the same person using different resolutions. They considered two methods based on the nearest neighboring K (K-NN), the method based on the recursive subdivision of characteristic space [40] and the random k-dimensional tree method (K-d, k-dimensional Tree). Both methods evaluated and compared their performance with the brute-force approach. The k-d Tree method achieved the best execution time and also the best grouping accuracy with 87%. While Gyongyi et al [41] studied the relationship of social identities of an individual and be able to distinguish whether it is a friend or the same user, since the same person may have several "identities" on several websites.

On the other hand, public spaces and urban identity are also addressed under the IA approach. Using a case study from the city of Zurich, Chang et al. [42] proposed a methodology using Principal Component Analysis (PCA) and the K-means approach to find from public space some characteristics of urban identity, as this can help the urban designer to identify which aspects are more important than others.

As a summary of these two subsections, indicate, on the one hand, that the ethnic models analyzed are far from being adapted to the present research. In the first place, because most of them have a qualitative nature and, secondly, because they do not analyze the factors that affect the reality of the Amazonian indigenous culture, since it has its own specific and unique characteristics in the world. On the other hand, the use of AI for the recognition of ethnic characteristics has focused almost exclusively on the field of image. Therefore, algorithms are limited to recognizing ethnic aspects that are reflected in the visible characteristics of humans through image analysis. There are intangible factors (faith, beliefs, ideologies...) that determine important aspects of a culture and that escape from the studies described above.

The paper has the following structure: section 3 will briefly describe the instrument developed by the authors. In section 4, the instrument is analyzed from an Artificial Intelligence point of view in order to automatically classify the individuals and to provide a subspace of it able to identify the weights of the subcomponents of this instrument in regard to its contribution to the Waorani ethnicity. Section 5 presents the main conclusions of the study and the lines of future work.

3 Instrument for the measurement and classifying of the Waoroani ethnic identity

The Ecuadorian indigenous population represents 7.0% in Ecuador, of which 78.5% are found in rural areas. The indigenous nationalities group together several peoples that have in common some cultural characteristics such as language and territory, but that differ in clothing, food and various types of cultural manifestations. Its population underwent significant changes, rising from 65% in the 18th century to 7.0% in 2010 [43]. However, these demographic data do not demonstrate the qualitative changes in their identity that they may have undergone.

The Waorani Nationality has approximately 13,000 inhabitants, distributed in 22 communities. The territory located in the Amazon (South America) shared by Ecuador and Peru is located in the provinces of Pastaza, Napo and Orellana (see Fig. 1). It is an extensive area of about 790,000 hectares whose boundaries range from the Napo River in the north to the Villano and Curaray Rivers in the south. Much of its territory is surrounded by the Yasuní National Park - protected by its exuberant biodiversity - and by an area called intangible in which the Tagaeri and Taromenane inhabit, uncontacted peoples who live in isolation with their original culture without any contact from the West or other cultures.

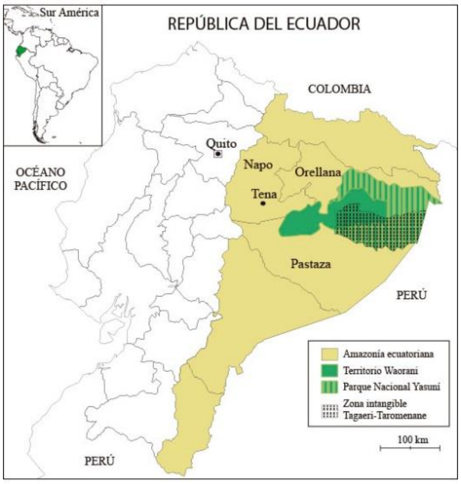

Fig. 1. Geographical location of the Waorani territory.

The instrument used in this study has been elaborated by the authors as a result of a previous research in which the dimensions were obtained from a qualitative study in which the *Atlas.ti* software was used for the identification of such dimensions. The technique for collecting qualitative data was based on the structure of levels [44] (see Table 1). Taking into account Guest levels, authorization from indigenous authorities and individual informed consent, data collection was conducted in two phases: the unstructured interview phase and the semi-structured interview phase.

Table 1: Levels for social data collection

Narratives	**Life Stories**
Interviews	Informal interviews
Focus groups	Unstructured participant observation
Participant observation	Qualitative systemic Elicitation
	Statistical questionnaires
	Observation

Source: Structure levels for social research according to Guest et al. [44]

In the first phase, it was intended that, through an open dialogue or conversation, the Waorani indigenous person would freely and spontaneously express stories on topics that the interviewee himself considered to be of interest. In this phase, 11 unstructured interviews were obtained. On the other hand, in the phase of semi-structured interviews, a series of concrete questions aimed at investigating cultural identity were carried out, totaling 14 semi-structured interviews (see table 2).

Table 2: Interviews with the Waorani population

Phase	**Interviews**	**Men**	**Women**	**Total**
1	Unstructured	4	7	11
2	Semi-structured	6	8	14
	Total	10	15	25

After the qualitative analysis carried out with the help of *Atlas.ti* software, the instrument consisted of 99 items that provided information about 30 subdimensions grouped in 5 dimensions (see Table 3).

Table 3. Dimensions and subdimensions of the Waorani Identity Instrument.

	Dimensions		Subdimensions	Items
1.	Economic (E)	9	Handicrafts (E1), Exchange of products (E2), Cultivation (E3), Tourism (E4), Work (E5), Mingas (E6) -mingas or nopos, indigenous words that means community work without payment-, Trade (E7), Hunting and fishing (E8), Breeding of animals (E9)	22
2.	Family and reproduction (F)	6	Education (F1), Care and upbringing of children (F2), Medicine (F3), Marriage (F4), Coexistence (F5), Reproduction (F6)	21
3.	Ideological (I)	4	Religion (I1), Beliefs (I2), Spirituality (I3), Rites (I4)	13
4.	Organization (O)	3	Community (O1), Justice (O2), Government (O3)	6
5.	Social (S)	8	Music, dance and songs (S1), Art (S2), Food (S3), Dress (S4), Housing (S5), Culture, ethnicity and identity (S6), Language (S7), Sports and recreation (S8)	37
Total Subdimensions		30	Total Items	99

Evidence of reliability was obtained through Cronbach's alpha statistic which gave an excellent overall result of 0.974 for the 99 items. In the case of the 5 constructs the evidences were: Economy 0.896 for 22 items; Family and reproduction 0.935 for 21 items; Ideology 0.906 for 13 items; Organization 0.602 for 6 items and Social 0.931 for 37 items. Evidence of validity was obtained through the judgment of 9 experts: 6 Ecuadorians (2 social workers, 1 economist, 1 social communicator, 1 psychologist and 1 sociologist), 2 Bolivian anthropologists and 1 Colombian political scientist. The validation test of the instrument was carried out by means of an alternative modified model called CVR' [45] according to Lawshe [46].

4 Experiments

4.1 Experimental setup

To determine whether the data for the three populations obtained in the instrument correspond to different groups and whether they can be separated by their identity characteristics, the technique of reducing spaces to lower dimensions by means of the Principal Component Analysis (PCA) was used.

The experiments have been carried out using classic classifiers in order to validate the ability of the instrument presented in the previous section to represent and classify the individuals. Specifically, the Self-Organizing Map (SOM) [47], Supervised Self-Organizing Map (SSOM) [48], Neural GAS (NGAS) [49], Linear Discriminant Analysis (LDA), k-Nearest Neighbour (kNN) [50] and the Support Vector Machines (SVM) [51] have been used. Moreover, a multiclassifier (MC) boosting designed from the above classifiers has been applied. The MC calculates from an input the most frequent class classified by the mentioned classic techniques.

The data of the instrument have been collected by interviewing 299 people: 88 Waorani individuals, 100 from Quito and 111 from the village of "El Tena". The samples were taken by simple random sampling using a systemic procedure from these three populations. These three populations represent 3 types of identities present in Ecuador: 1. The indigenous Waorani Amazonian Nationality is the target population of the study (it consists of 22 communities, the data collection was carried out in three of the 22 communities: Konipare, Menipare and Gareno). Quito (capital of Ecuadorian state) was selected because it is a totally westernized city and may represent the opposite extreme to the indigenous community within the country. The city of Tena is a westernized village closest to the Waoranis indigenous communities. Tena still retains vestiges of indigenous culture and has frequent commercial exchanges with various indigenous nationalities including communities of Waorani nationality, therefore the inhabitants of Ciudad de El Tena are supposed to be culturally placed between Amazonian Indians and western culture.

They conform the output of the classification process. Since the input data is imbalanced, the Synthetic Minority Over-Sampling Technique (SMOTE) [52] has been applied in order to balance the classes. Additionally, the 30 subdimensions of the Waorani instrument (showed in table 3) used as the inputs for each classifier have been normalized to the range (0 1). Each dimension of the instrument has been divided by the maximum value for each component.

A 10-fold cross validation has been performed obtaining Sensitivity and Specificity values, and the ROC curves in order to analyse the performance of classifiers and the instrument from an Artificial Intelligence point of view, to distinguish the class of the individuals (Waorani, Quito or El Tena).

The evaluation and sensitivity of the classifiers were carried out to identify the sub-dimensions that most discriminate the populations. We also determined the maximum probability of detection of a population with the sub-dimensions that most contribute to ethnic identification.

Finally, a study was carried out to identify those sub-dimensions of ethnic identity that discriminate the least. These components would indicate a super-position of characteristics that do not allow us to differentiate between a Waorani indigenous person and any other citizen, who in our research are the individuals of the cities of El Tena and Quito.

4.2 Results and discussion

Classification results with three main components denote good discrimination, a marked separation can be seen in Figure 2. No inhabitant of the Waorani indigenous population belongs to the set of Quito (0 of 88). However, there is only one inhabitant of the city of Quito who is in the Waorani population group (1 out of 100), while the inhabitants of the city of El Tena are more dispersed between the two population groups.

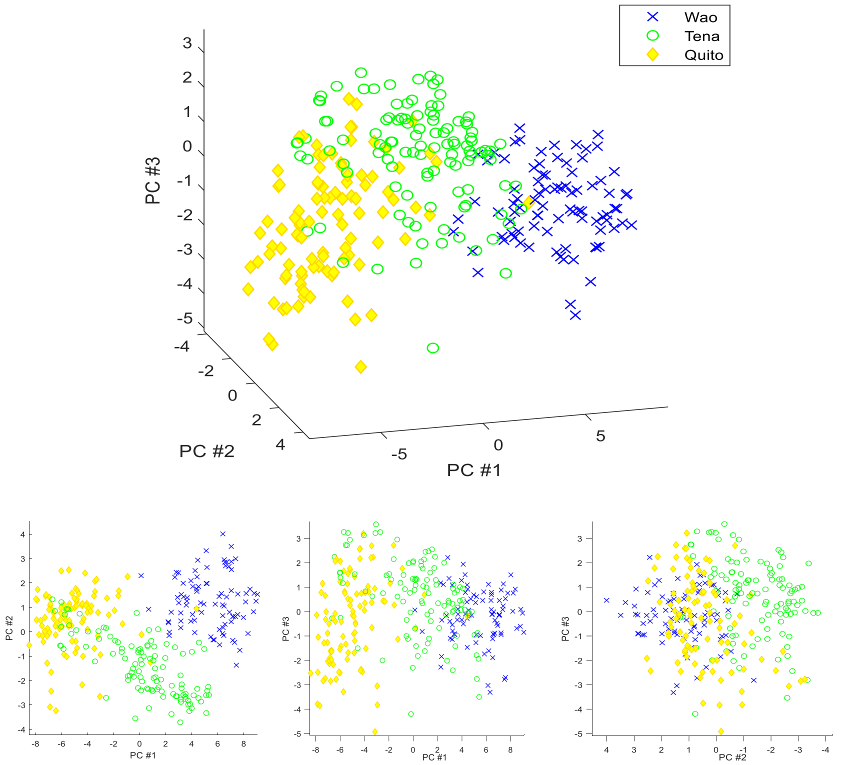

Fig. 2. Classification of the three populations (Waorani, El Tena and Quito) using three main PCA components.

Results of classification accuracy for each classifier of the subdimensions of the instrument are presented in Table 4. The performance criteria is related to the *Sensitivity* (correctly classified positive samples / true positive samples), *Specificity* (correctly classified negative samples / true negative samples) and *Accuracy* (correctly classified positive and negative samples / samples). The best results are achieved using the SVM classifier achieving a 97% of sensitivity, 98% of specificity and 98% of accuracy.

Table 4. Performance results according to different classifiers

Criteria	3-knn	LDA	SOM	SSOM	NGAS	SVM	MC
Sensitivity	0.9298	0.9632	0.9097	0.9331	0.9164	**0.9699**	0.9532
Specificity	0.9649	0.9816	0.9548	0.9665	0.9582	**0.9849**	0.9765
Accuracy	0.9532	0.9755	0.9398	0.9554	0.9444	**0.9799**	0.9688

In order to visually illustrate the performance of classifiers for the instrument, a ROC (receiver operating characteristic) curve is used in Fig. 3. It represents the relation between the probability of correctly detect the class of an individual (Sensitivity) and the probability of classify it in another class (1-Specificity). The ROC space shows the excellent performance for all classifiers that are able to correctly identify the ethnic identity above 91% and having less than 5% of false alarms.

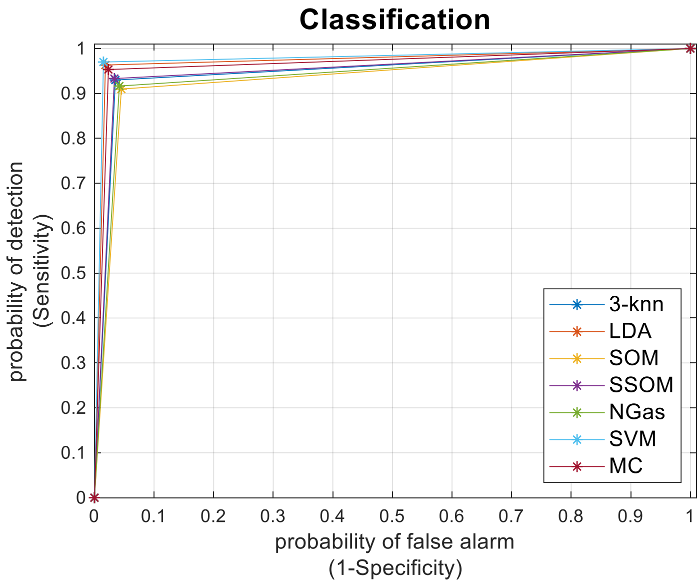

Fig. 3. Classification performance for the different methods to distinguish the classes "Quito, "El Tena" and "Waorani”.

Next, confusion matrix for the best classifier SVM is presented in Table 5. Matrix columns represent the actual classes, and rows represent the classifier prediction. As it can be observed, the Waorani individuals are almost perfectly detected with a 98.86% of probability (87/88 individuals). Moreover, only 1 person from Quito and El Tena was identified as Waorani (1/211 individuals), representing less than 0.5% of false Waorani identification.

Table 5. Confusion matrix for the SVM classifier

		Actual class		
		Waorani	El Tena	Quito
Pre-dicted	Waorani	87	0	1
	El Tena	1	109	5
	Quito	0	2	94

4.2.1 Most discriminating sub-dimensions

A study about a subset of the subdimensions of the instrument that are able to identify properly the classes has been performed. Fig. 4 shows the cumulative probability of detection of the different identities for the classifiers according to the subcomponent of the instrument. As it can be observed, the first subdimension I2 is able to identify about the 60% of the different identities (only the simple 3-knn is below the 50%). The more subdimensions are considered as input of the classifiers, the more probability of detection. However, after using the first 18 subdimensions represented in the figure (axis x), the performance is near constant.

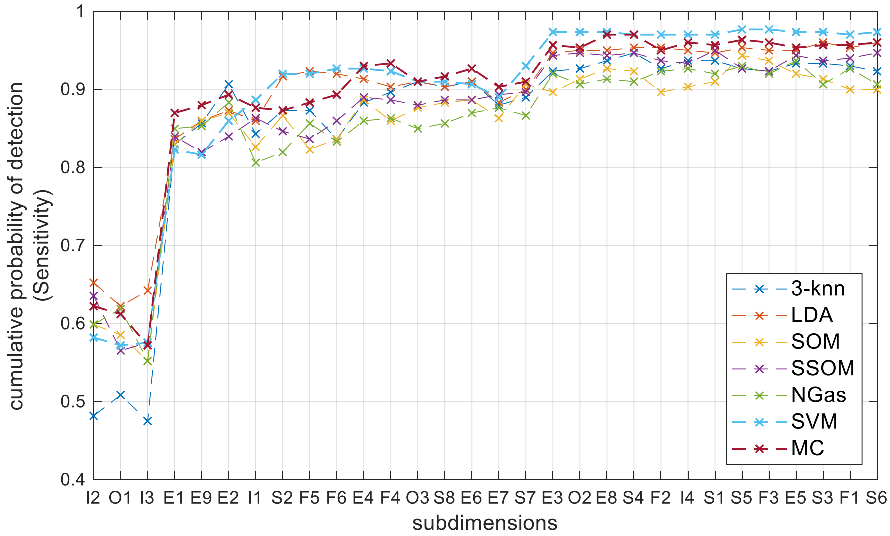

Fig. 4. Cumulative probability of detection of the different identities for the classifiers according to the subdimension of the instrument.

The study of the behavior of the cumulative probability of detection provides a subset of dimensions as shown in Table 6. In the table, the subdimensions that are able to increase the sensitivity of the classifiers are grouped by the dimensions. The best classifier, the SVM, is able to achieve a 97.7% of probability of detection and 1.2 % of false alarms using 9 components of 4 dimensions of the Waorani Identity Instrument. In this case, the Organization dimension is not relevant, being the most important the Economic and Social aspects.

Table 6. Subdimensions having an impact in the classification performance

Class.	Economic	Fam.	Ideo.	Org.	Soc.	Sens.	Spec.
3-knn	E1,E2,E3,E8,E9		I2	O1,O3,O2	S4	0.946	0.973
LDA	E1,E2,E3,E9	F5	I2	O2	S2,S4,S3	0.959	0.979
SOM	E1,E2,E4,E8,E9,E8		I2	O2	S7,S5	0.943	0.972
SSOM	E1,E3,E4,E7		I1,I2	O2	S1,S7	0.949	0.975
NGAS	E1,E2,E3,E5,E9	F2	I2,I4	O1	S5	0.936	0.968

SVM	**E1,E2,E3**	**F6**	**I1,I2**		**S2,S7,S5**	**0.977**	**0.988**
MC	E1,E2,E3,E4,E8,E9	F4	I2			0.969	0.985

4.2.2 Less discriminating sub-dimensions

The evaluation of classifiers to identify the subdimensions that least discriminate and to determine the items associated with those subdimensions is shown in Table 7. Note that these items refer to characteristics of ethnic identity that will not allow the three populations to be differentiated.

Table 7. Ethnic identity sub-dimensions that are the least discriminatory

'E7' 'S7' 'I1' 'F6' 'O3' 'S3' 'S1' 'O2' 'I2' 'E9' 'F1' 'E8' 'O1' 'E1' 'I4' 'F2' 'S8' 'S4' 'F5' 'F3' 'F4' 'E3' 'E5' 'S5' 'E4' 'E6' 'S6' 'S2' 'E2' 'I3'

Figure 5 shows the cumulative inverse probability of the 30 components. The first 7 components in SVM contribute practically nothing to the classification. It can also be seen that for the first 14 components ranging from E7 to E1, the detection probability curves for the classifiers remain almost constant at 95% (97% for SVM), of which the first 5 components in SMV (E7 to O3) have no variation. Then, with non-significant inputs the next 8 components up to E3 decrease to 94% (95% for SVM).

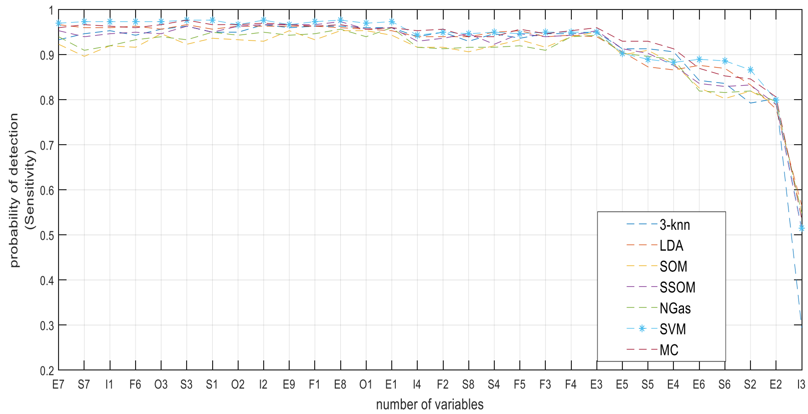

Fig. 5. Cumulative inverse probability of detection of Waorani, El Tena and Quito identities.

The items associated with each of the 14 least discriminating components (called sub-dimensions in the instrument) are shown in Table 8. When analyzing for each dimension, all dimensions have components that do not contribute to discrimination. The social dimension with items associated with characteristics of language, food, music and dance is the most numerous with 14 of 37 items (37.83%) representing 14.14% of the total of 99 items. While the dimension Organization that has items with characteristics of government, justice and organizational community has all the subdimensions of the model associated with 6 of 6 items (100%), these represent 6.06% of the total. All the items (42 of 99) of the sub-components that classify less, would not be influencing the classification of ethnic identity, therefore it can be deduced that in these items there would be a fusion of ethnic identity of the indigenous population with the culture of the other two cities of El Tena or Quito.

Table 8. Dimensions, subdimensions and items that contribute less to the ethnic identity classification of the Waorani indigenous people.

Dimension	Items	Subdimensions	Items	% Dim.	% Tot.
Economic (E)	22	E7 Trade	3		
		E9 Breeding of animals	1		
		E8 Hunting and fishing	3		
		E1 Handicrafts	2		
		Sub Total	9	40,9	9,1
Family and reproduction (F)	21	F6 Reproduction	6		
		F1 Education	1		
		Sub Total	7	33,3	7,1
Ideological (I)	13	I1 Religion	1		
		I2 Beliefs	5		
		Sub Total	6	46,1	6,1
Organization (O)	6	O3 Government	2		
		O2 Justice	3		
		O1 Community	1		
		Sub Total	6	100,0	6,1
Social (S)	37	S7 Language	3		
		S3 Food	6		
		S1 Music, dance, songs	5		
		Sub Total	14	37,8	14,1
Total	**99**	**Total**	**42**		

Table 9 shows the description of the 5 main sub-dimensions that contribute least to the classification (sub-dimensions taken from Table 8) together with the items that most influence or affect from the two westernised cultures (cities of El Tena and Quito) towards the Waorani indigenous people.

Table 9. Main sub-dimensions that contribute least to the classification together with the item descriptions

Subdimensions	Associated items
E7 Trade	▪ Sell products produced in their community (yucca, banana, bushmeat, handicrafts) ▪ How important it is for you to grow cocoa or products that are not part of your tradition ▪ Prefers to grow products (such as cocoa, flowers) rather than traditional products (such as cassava, banana, corn, handicrafts, hunting)
S7 Language	▪ You consider that you speak and write Spanish ▪ In the community you speak your language ▪ You prefer your children to speak your language

I1 Religion	▪ It is better to believe in the religions and gods that the missionaries brought from the city or the west.
F6 Reproduction.	▪ How many children do you have, like, or would you like to have children?
O3 Government	▪ The election of community representatives is done by vote.
	▪ Currently, the election of community representatives is as follows

5 Conclusions

Cultural identity is a complex concept affected by a multitude of variables, most of which are subjective in nature, making it difficult to analyse them quantitatively. In the case of the Amazonian indigenous cultures it is urgent to carry out this analysis because, due to the strong process of western acculturation to which they are subjected, their own identity is endangered and even feared for their disappearance.

This paper analyses las components de un instrument specifically designed for the measurement of the Waoroani ethnic identity from an Artificial Intelligence (AI) point of view. 30 subcomponents have been used as the input of several classifiers in order to analyze the performance of an AI system to automatically classify individuals.

The results show a high performance both the sensitivity and the specificity of the classifiers being able to provide with high accuracy the ethnic identity of an individual. Moreover, this paper analyses the components of the instrument as a feature selection problem obtaining that a subset of about 10 variables are able to achieve the same performance results.

In particular, the SVM classifier obtained the best results being able to use 4 of the 5 components of the instrument to properly classify the identity of the individuals. Particularly, it used 9 of the 30 original subcomponents representing only the 30% of them. It allows us to reduce the dimensions and practically the number of questions needed to identify them.

The quantification of each factor is one of the most important contributions of this work. On the one hand, the identity sub-components that contribute most to the classification were identified. These factors are the most influential in acculturating or affecting the ethnic identity of the Waorani indigenous people and therefore, from a socio-political point of view, more important to protect. Of the total of the sub-components, the sub-components of the Social dimension (S) those that most affect indigenous identity with 14.1% followed by the Economic dimension (E) with 9.1%. The Organization (O) dimension affects with 6.1%, although it influences with all its sub-components.

On the other hand, it is also very important to identify the least influential factors, since from a socio-political point of view they would be the dimensions on which policies could be carried out to help these peoples with the guarantee of minimally affecting their current cultural identity. The items of the 7 components that contribute least to ethnic classification have characteristics related to trade, language, religion, reproduction, government, food and music. These characteristics of ethnic identity would be the cause of the acculturation of the Waorani people and that, in some way, the citadine cultures of the West (as in the case of the cities of Quito and El Tena) would have penetrated with greater force on the identity of the indigenous people.

The contribution of the AI to the research is twofold. First, it allow us to provide a machine learning method able to automatically identify the identify factors of the Waorani indigenous identity, showing a high performance even compared to close individuals as the as the inhabitants of "La Tena" village. Second, the machine learning method is able to determine the most important dimensions of the instrument to reduce the number of variables needed to identify properly the Waorani indigenous identity and, in consequence, improve the instrument initially provided.

In consequence, as future line a detailed study of the variable selection is proposed. It also will allow us to design distance function (using the subcomponents) able to determine what level of identity is in an individual and whether it can be considered that the original identity of the indigenous nationality has been preserved.

Compliance with Ethical Standards

Aldrin Espin-Leon declares that he has no conflict of interest. Antonio Jimeno-Morenilla declares that he has no conflict of interest. María Luisa Pertegal-Felices declares that she has no conflict of interest. Jorge Azorin-Lopez declares that he has no conflict of interest.

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

[1] W. H. Goodenaugh, J. KAHN, A. KROEBER, B. Malinowski, E. B. TYLOR, y L. A. WHITE, *El concepto de cultura: textos fundamentales*. Anagrama, 1975.

[2] UNESCO, «Declaración Universal sobre la Diversidad Cultural una visión una plataforma conceptual un semillero de ideas un paradigma nuevo». 2002.

[3] I. González-Varas, *Conservación de bienes culturales: teoría, historia, principios y normas*. Ed. Catedra, 1999.

[4] S. Carrasco-Arroyo, «Medir la cultura: una tarea inacabada», pp. 1-2, 2006.

[5] A. Quijano, «Colonialidad y modernidad/racionalidad», *Perú Indígena*, vol. 13, n.º 29, pp. 11–20, 1992.

[6] Codenpe, *Construyendo el Estado Plurinacional*. Quito, 2011.

[7] D. Racines Izquierdo, «La occidentalización de los waoranis», B.S. thesis, Quito: USFQ, 2012, 2012.

[8] de C. Vivieros, «Perspectivismo y multinaturalismo en la América indígena», en *Tierra Adentro: Territorio indígena y percepción del entorno*, Lima: International Work Group for Indigenous Affairs, 2004.

[9] O. Espinosa de Rivero, «Ciudad e identidad cultural.?` Cómo se relacionan con lo urbano los indígenas amazónicos peruanos en el siglo XXI?», *Bull. Inst. Fr. Détudes Andin.*, n.º 38 (1), pp. 47–59, 2009.

[10] M. Weber, *Economy and society: An outline of interpretive sociology*, vol. 1. Univ of California Press, 1978.

[11] S. Carrasco, «Indicadores culturales: una reflexión», *Univ. Valencia Recuperado*, vol. 21, pp. 3-4, 1999.

[12] H. Tajfel, *Human groups and social categories: Studies in social psychology*. CUP Archive, 1981.

[13] M. Verkuyten, «Self-esteem and the evaluation of ethnic identity among Turkish and Dutch adolescents in the Netherlands», *J. Soc. Psychol.*, vol. 130, n.º 3, pp. 285–297, 1990.

[14] S. L. Schneider y A. F. Heath, «Ethnic and cultural diversity in Europe: validating measures of ethnic and cultural background», *J. Ethn. Migr. Stud.*, 2018.

[15] K. Carley, «A theory of group stability», *Am. Sociol. Rev.*, pp. 331–354, 1991.

[16] W. E. Cross Jr, «The Thomas and Cross models of psychological nigrescence: A review», *J. Black Psychol.*, vol. 5, n.º 1, pp. 13–31, 1978.

[17] J. E. Helms, «An update of Helm's White and people of color racial identity models.», en *Versions were presented at the Psychology and Societal Transformation Conference, U Western Cape, South Africa, Jan 1994, and at a workshop entitled" Helm's Racial Identity Theory," Annual Multicultural Winter Roundtable, Teachers Coll–Columbia U, New York, Feb 1994.*, 1995.

[18] J. V. Espín, M. Á. Marín, M. Rodríguez, y F. Cabrera, «Elaboración de un cuestionario para medir la identidad étnica y la aculturación en la adolescencia», *Rev. Educ. Madr.*, n.º 315, pp. 227–249, 1998.

[19] J. S. Phinney y M. J. Rotheram, *Children's Ethnic Socialization: Pluralism and Development.* ERIC, 1986.

[20] J. W. Berry, «Understanding individuals moving between cultures», *Appl. Cross-Cult. Psychol.*, vol. 14, n.º 1, p. 232, 1990.

[21] D. L. Sam y J. W. Berry, «Acculturation: When individuals and groups of different cultural backgrounds meet», *Perspect. Psychol. Sci.*, vol. 5, n.º 4, pp. 472–481, 2010.

[22] N. A. Denton y D. S. Massey, «The dimensions of residential segregation», *Soc. Forces*, vol. 67, n.º 2, pp. 281–315, 1988.

[23] A. I. Vergara y D. Páez, «Revisión teórico-metodológica de los instrumentos para la medición de la identidad de género», *Rev. Psicol. Soc.*, vol. 8, n.º 2, pp. 133–152, 1993.

[24] V. C. Vega, «Adaptación argentina de un inventario para medir identidad de rol de género», *Rev. Latinoam. Psicol.*, vol. 39, n.º 3, 2007.

[25] I. M. Muradas y J. R. O. Rodríguez, «Diseño de un instrumento de medida para el análisis comparativo de los recursos culturales tangibles», *Número Patrocinado Por*, p. 245, 2005.

[26] V. S. Castro, «La Escala de Identidad Etnica Multigrupo (EIEM) en el contexto costarricense 1», *Actual. En Psicol.*, vol. 18, n.º 105, pp. 47–67, 2002.

[27] E. De la Montaña Andrés, «Cacería de subsistencia de distintos grupos indígenas de la Amazonía ecuatoriana», *Rev. Ecosistemas*, vol. 22, n.º 2, pp. 84–96, 2013.

[28] M. Esteban y I. Vila, «Modelos culturales y retratos de identidad. Un estudio empírico con jóvenes de distintos contextos sociodemográficos», *Estud. Psicol.*, vol. 31, n.º 2, pp. 173–185, 2010.

[29] A. Bagnoli, «Researching identities with multi-method autobiographies», *Sociol. Res. Online*, vol. 9, n.º 2, pp. 1–15, 2004.

[30] V. E. Duró, «Evaluación de sistemas de reconocimiento biométrico», *Dep. Electrónica Automática Esc. Univ. Politécnica Mataró*, 2001.

[31] J. Daugman, «How iris recognition works», en *The essential guide to image processing*, Elsevier, 2009, pp. 715–739.

[32] L. Hong, A. K. Jain, y S. Pankanti, «Can multibiometrics improve performance?», en *Proceedings AutoID*, 1999, vol. 99, pp. 59–64.

[33] A. K. Jain, S. C. Dass, y K. Nandakumar, «Soft biometric traits for personal recognition systems», en *International conference on biometric authentication*, 2004, pp. 731–738.

[34] G. Shakhnarovich, P. A. Viola, y B. Moghaddam, «A unified learning framework for real time face detection and classification», en *Proceedings of Fifth IEEE international conference on automatic face gesture recognition*, 2002, pp. 16–23.

[35] X. Lu y A. K. Jain, «Ethnicity identification from face images», en *Proceedings of SPIE*, 2004, vol. 5404, pp. 114–123.

[36] S. Gutta, J. R. Huang, P. Jonathon, y H. Wechsler, «Mixture of experts for classification of gender, ethnic origin, and pose of human faces», *IEEE Trans. Neural Netw.*, vol. 11, n.º 4, pp. 948–960, 2000.

[37] K. M. Carley, «Artificial intelligence within sociology», *Sociol. Methods Res.*, vol. 25, n.º 1, pp. 3–30, 1996.

[38] L. C. Freeman, «Finding groups with a simple genetic algorithm», *J. Math. Sociol.*, vol. 17, n.º 4, pp. 227–241, 1993.

[39] C. Otto, D. Wang, y A. K. Jain, «Clustering millions of faces by identity», *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, n.º 2, pp. 289–303, 2018.

[40] J. Chen, H. Fang, y Y. Saad, «Fast approximate kNN graph construction for high dimensional data via recursive Lanczos bisection», *J. Mach. Learn. Res.*, vol. 10, n.º Sep, pp. 1989–2012, 2009.

[41] Z. I. Gyongyi, M. Heymans, M. Lopyrev, y H. Roehrig, «Social identity clustering», ene-2014.

[42] M.-C. Chang, P. Bus, y G. Schmitt, «Feature extraction and k-means clustering approach to explore important features of urban identity», presentado en Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017, 2018, vol. 2018-January, pp. 1139-1144.

[43] A. M. Enrique, «Lucha política y origen de los partidos políticos en el Ecuador, Corporación Editora Nacional, 2da», *Ed. Quito Ecuad.*, 1982.

[44] G. Guest, E. E. Namey, y M. L. Mitchell, *Collecting qualitative data: A field manual for applied research*. Sage, 2012.

[45] A. Tristán-López, «Modificación al modelo de Lawshe para el dictamen cuantitativo de la validez de contenido de un instrumento objetivo», *Av. En Medición*, vol. 6, n.º 1, pp. 37–48, 2008.

[46] C. H. Lawshe, «A quantitative approach to content validity», *Pers. Psychol.*, vol. 28, n.º 4, pp. 563–575, 1975.

[47] T. Kohonen, «The self-organizing map», *Proc. IEEE*, vol. 78, n.º 9, pp. 1464–1480, 1990.

[48] S. Papadimitriou, S. Mavroudi, L. Vladutu, G. Pavlides, y A. Bezerianos, «The supervised network self-organizing map for classification of large data sets», *Appl. Intell.*, vol. 16, n.º 3, pp. 185–203, 2002.

[49] T. M. Martinetz, S. G. Berkovich, y K. J. Schulten, «“Neural-gas” network for vector quantization and its application to time-series prediction», *IEEE Trans. Neural Netw.*, vol. 4, n.º 4, pp. 558–569, 1993.

[50] T. M. Cover y P. E. Hart, «Nearest neighbor pattern classification», *IEEE Trans. Inf. Theory*, vol. 13, n.º 1, pp. 21–27, 1967.

[51] C.-W. Hsu y C.-J. Lin, «A comparison of methods for multiclass support vector machines», *IEEE Trans. Neural Netw.*, vol. 13, n.º 2, pp. 415–425, 2002.

[52] N. V Chawla, K. W. Bowyer, L. O. Hall, y W. P. Kegelmeyer, «SMOTE : Synthetic Minority Over-sampling Technique», vol. 16, pp. 321-357, 2002.