

An Artificial Intelligence Approach to Modeling in Social Science

Juan Carlos Jesus Vazquez, Julio Javier Castillo, Leticia Edith Constable, Marina Elizabeth Cardenas, Juan Carlos Guillermo Vazquez

Information Systems Engineering Department of Cordoba Regional Faculty, National Technological University, Cordoba, Argentina

Email address:

jcvazquez@frc.utn.edu.ar (J. C. J. Vazquez), jcastillo@frc.utn.edu.ar (J. J. Castillo), lconstable@frc.utn.edu.ar (L. E. Constable), mcardenas@frc.utn.edu.ar (M. E. Cárdenas), jcg vazquez@frc.utn.edu.ar (J. C. G. Vazquez)

To cite this article:

Juan Carlos Jesus Vazquez, Julio Javier Castillo, Leticia Edith Constable, Marina Elizabeth Cardenas, Juan Carlos Guillermo Vazquez. An Artificial Intelligence Approach to Modeling in Social Sciences. *Journal of Health and Environmental Research*. Special Issue: *Health and the Environment as a Resource for the Reduction of Social Inequalities in Argentina*. Vol. 7, No. 1, 2021, pp. 58-68.

doi: 10.11648/j.jher.20210701.20

Received: January 4, 2021; **Accepted:** February 20, 2021; **Published:** March 17, 2021

Abstract: Computer Science has contributed to social sciences since decades ago: connecting people that build virtual communities where the interactions can be investigated, developing tools for statistically analytics, designing models that allow the analysis and simulation of the most diverse types, among many others. In this article, we describe an artificial neural network to model a theoretical framework for risk, housing, and health problematic, called DRVS (Diagnostic methodology for risk determination of urban housing for health), which uses a holistic approach for community and environmental health. The methodology also exposes digital clinic history for families and communities, developed to support the acquisition of necessary data. This software has advantages for the transference and application of the DRVS in different locations since it constitutes an expert system for the determination of local social indexes and supports the quantitative validation process for the underlying social theory. On the other hand, as many artificial intelligence techniques, it has constraints: unlike explicit logic inferences, artificial neural networks work as «black boxes», not explaining how they got the result; they have a strong dependency of the representativeness of training data and introducing new knowledge that may improve their results and performance is difficult (new data, addition or remotion of determining factors for the underlying social model, weighting factors, etc.). This article also shows some techniques and ideas on how to deal with the identified constraints.

Keywords: Artificial Neural Network, Family Clinic History, Risk, Health, Housing

1. Introduction

The concept of health has long been understood not only as the absence of health, but also including preventions, life quality, environment, public services among others. All this aspects are named as *holistic approach* in social science terminology.

Therefore, health is concern of health personnel (doctors, nurses, psychologists, specialists, paramedics, etc) and also involves to demographers, engineers, economists, architects, geographers, statisticians, teachers, politicians, social assistants, managers, among others. All these activities tackles health with the aims of to promote general welfare of the community.

Thus, when defining and planning public politics in health,

Integral Health Systems must be thought beyond hospital and diseases treatment systems. These systems must also include access to essential public services, education, communities strengthening and empowering, community and environmental health, among others. [1-4]

With this in mind, researchers from the Geo-historic Research Institute (IIGHI) of the National Council for Scientific and Technical Research (CONICET), under the supervision of Dra. Norma Meichtry and the direction of Dra. Maria del Carmen Rojas, developed a *Diagnostic methodology for risk determination of urban housing for human health (DRVS)* [5, 6]. Several factors are proposed in this methodology, using a community health approach, which are defined as social vulnerability determinants which includes: housing as the environment: available public

services, social and economic aspects of their inhabitants, emergency reaction capacity, and other social and infrastructure measures. These determining factors were selected from population census.

The social sciences research methodology usually involves the study of facts that already happened (history), collect qualitative and quantitative data, the development of detailed descriptions, build statistical reports, among others. This techniques are used to approach the reality and also to model the behavior of actors involved. Then, the hypothesis are proposed looking to establish causal relations between these actors, therefore being able to forecast the free evolution of the system or the results of modifying actions. In this sense, it is a great help to determine and evaluate adequate public politics in certain domains or situations. DRVS tries to go a step beyond this method.

In health politics, governments must decide how to assign the limited resources available to improve general welfare. Doing «social experiments» is hard (and ethically incorrect) since there are people involved, making it difficult to measure the impact on health of a modification in a specific risk factor, while keeping every other variable constant. However, decisions must be made, even under uncertainty and with the system in a dynamic state, and the scientific method can give some hints and tackle the uncertainty providing measures of the expected results.

Even if DRVS methodology proposes the incidence of certain factors considered non-measurable from demographic perspective, such as vulnerability, resilience, exposition, social fragility and others [6], social science experts from the IIGHI-CONICET believe these factors are tricky related and not linearly dependent.

The experts believes that these determining factors can be estimated in average, and thus they could infer the health risk.

These estimations are put together into a community work experience in order to strengthen the protective factors and reduce threatening ones, thus promoting the health of their people.

In 2005, the PhD Rojas, an social science expert, worked in collaboration to the Software Research Laboratory (Information Systems Engineering Department, Córdoba Regional Faculty of National Technological University) and, as a result, a prototype of a computational model was build. This model receives census data and can estimate the health risk following the DRVS methodology.

This article address the issues raised by DRVS from the computational science approach. Additionally, and considering the latest advances in machine learning, we suggest improvements and next steps.

The nexts sections are described as follow: i) the DRVS methodology is outlined, ii) is about the mapping between the DRVS methodology and a computational model proposed, an artificial neural network that can emulated an expert system, iii) a familiar clinic histories system is showed as a source of input data for the computational model, iv) application and communication, v) approach discussion, and

finally, vi) conclusions and future work are showed.

2. DRVS Methodology

DRVS methodology is inside the domain of Social, Community and Environmental Health in Latin America, and was first proposed in the PhD thesis named “Holistic Estimation of the seismic risk using complex dynamic systems” developed by the Colombian engineer Omar Dario Cardona Arboleda [7].

DRVS constitutes one of the conceptual elements for the development of a research line on environmental surveillance, with researchers working together from Argentina, Colombia, Paraguay, Brazil, Cuba, and other countries, as a part of the “Inter-American Healthy Housing Network” (VIVSALUD), related with the Pan American Health Organization (PAHO); this line is analogue in the IIGHI-CONICET [8].

For Social Science researcher at IIGHI, it is necessary to think in intervention spaces where systematic tools for protective and harmful processes can be valued in relation to health-disease conditions. In this sense, the study and measuring of housing and it's influence in their habitants is an interesting research subject.

The basic idea rises from the necessity of strengthening the national and local surveillance systems for risk and protection of health, associated with housing, through the design of new models that help generate alternatives for healthy and sustainable development in Latin America.

Thus the proposed approach treat the health risk as a non linear problem based on the dynamic process; this is, to think that the evolutionary changes that determine healthy status are carried out by a set of processes, and that they acquire different projections, related to the social constraints that convey in a specific space and time.

Therefore, we need an approach on housing not as a static reservoir for contaminants, threats, parasites, and disease vectors, but as a historically structured space where both benign and harmful consequences of social organization are revealed. In this context, processes are obliged mediators and social reproduction conditions are transformed in assets for promoting health, or in destructive forces promoting disease.

Dra. Rojas and her research group have proposed a model based on census information to assess the sociodemographic factors of a community, and have developed forms to acquire data, together with local governments, to evaluate the resilience (strengths to face adversities). The objective is to establish an urban housing topology in relation to human health, that has useful to model different social levels, but with a concrete and immediate application to marginal urban housing. Components of risk (physical housing risk), vulnerability (fragility and risk exposure) are also presented, calculated from housing factors that constitute risks for health and from social and demographic variables that influence vulnerability.

The proposed conceptual model for the determination of a health risk index, determined by the urban housing of an

specific community is showed in Figure 1.

This diagram shows that demographers have designed the model based on the biological neural network concepts, however the model does not fit in usual computational neural network architectures. Intermediate inputs, used for

calculating the physical risk, are included in the deep layers, and there are certain intermediate layers that are needed if nonlinear relationships are assumed, and are not visible in Figure 1.

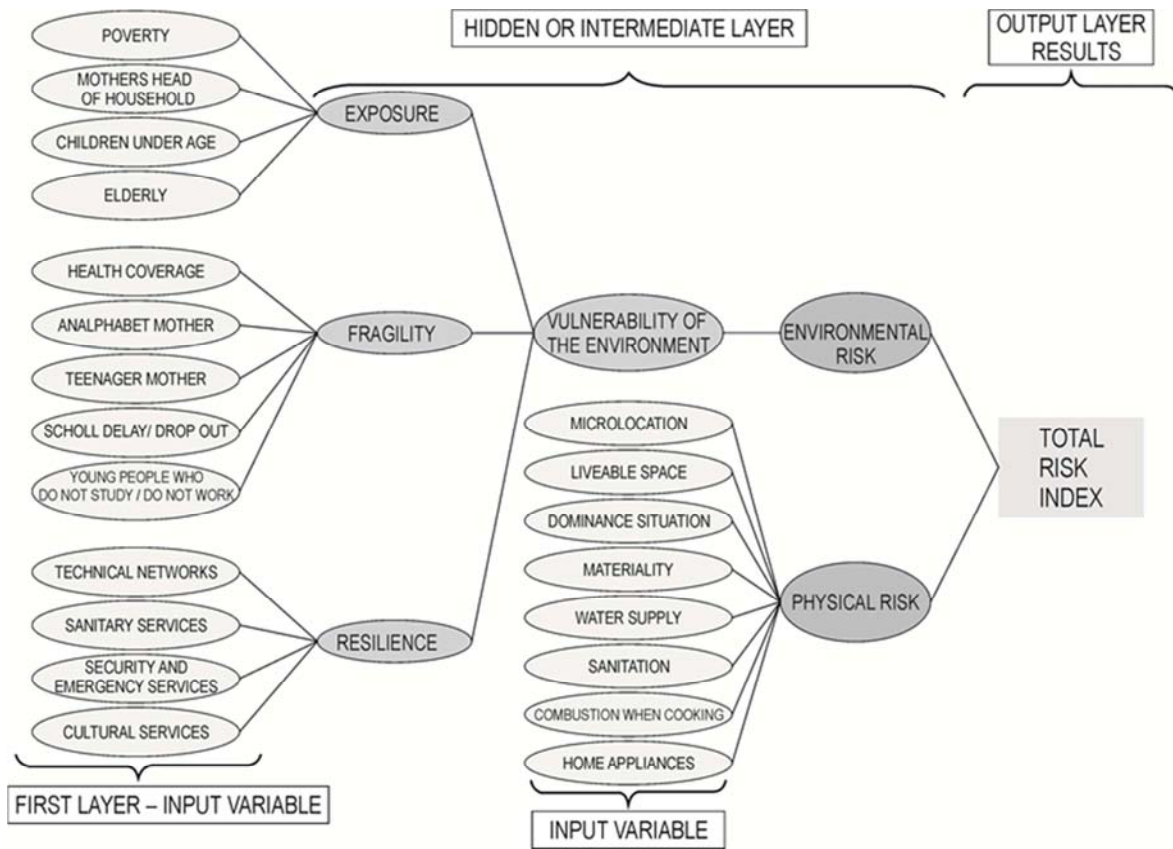


Figure 1. Conceptual Model Diagram for DRVS, proposed by the IIGHI-CONICET researchers.

3. Computational Model Supporting DRVS

To our knowledge there is not a proposed model in the literature that relates sociodemographic variables and the proposed factors.

Therefore, there is no set of algebraic or differential equations capable of modeling the problem, and under the assumption that the interactions are not linear and complex, we proposed to build an artificial neural network model with supervised learning.

The supervised network needs examples, consisting of known inputs and outputs, to train the model.

The IIGHI researchers provided these examples in different text files, generated from their field experience.

For the intermediate and final concepts, which are not measurable (resiliency, exposition, fragility, vulnerability, physical risk, and total risk), experts indicated they could create an approximation on a percentual estimation based on experience, according to the measurable features. However, these values are fuzzy as they provide ranges rather than individual values.

Therefore, we decided to establish a set of fuzzy triangular

sets for them (figure 3), based on the data provided.

3.1. Training Data

Not having domain expertise on demography, we made the questions of rigor about the proposed factors independence in each file and the data types (numeric, categorical, ranges, etc.) to be expected, without questioning the conceptual model itself.

Upon receiving the data, we carried out certain statistic studies measuring the correlations between the different factor inside the inputs, and between the input and output, and we observed certain linear relation between some of the files (see Figure 2 for an example). Even if the factors are not absolutely independent, IIGHI-CONICET experts considered that the mutual influence they may exist is included in the theoretical strategy by the interactions explained through the conceptual model, thus requesting to be considered valid.

Input values, that are required for the intermediate concepts, are always presented as ranges of possible values, since the number of houses that evaluate true for a certain feature is informed from the total of houses in the community. Therefore, for all the cases, the values range between 0 and 100 (example can be seen in Table 1).

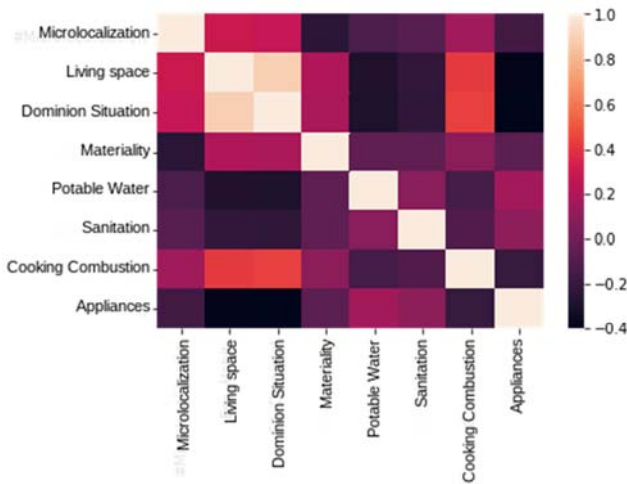


Figure 2. Heat plot of the correlation matrix for the physical risk concept (threat). It shows that there is some correlation between «living space» and «dominion situation»; there is also positive correlation between these two factors and «cooking combustion», and negative to «appliances».

Demographers assigned an importance index for each feature. To compare them with the data provided, we trained random forests [9-11] with each of the data files. This gave us an idea of the relative importance for each feature in the inputs for a sub network, in relation to the outputs.

These results are shown in Figure 4 and, with some minor differences, generally agree with the criteria of the IIGHI researchers.

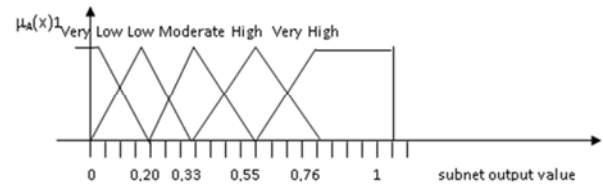


Figure 3. Fuzzy sets used to evaluate fuzzy results from the subnetworks. The ownership function $\mu_A(X)$ for a given X to the fuzzy set A , is obtained by vertically extending the value with the limits of the area for each of the five categories (Lowest, Low, Mid, High, Highest).

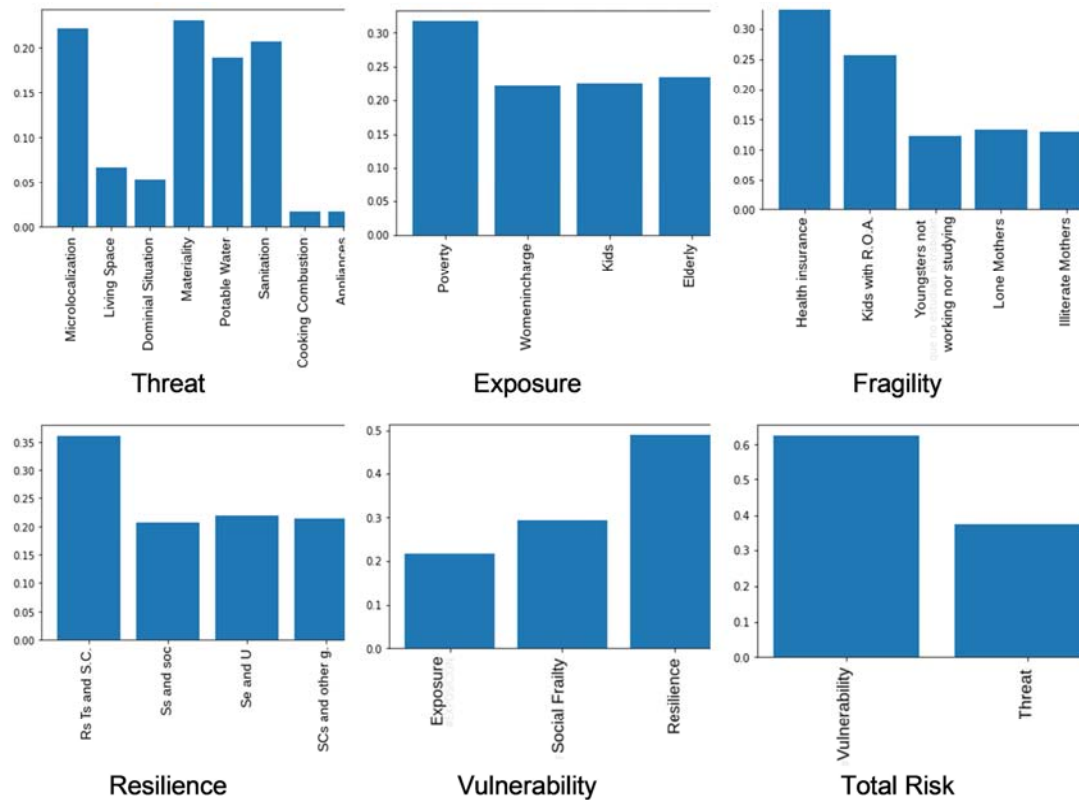


Figure 4. Relative Importance of features for each sub-network extracted for random forests from the training data.

Table 1. Example of the Data Table for training.

Microlocation	Living Space	Dominion Situation	Materiality	Water Supply	Sanitation	Combustion when cooking	Home Appliances	Threat
000-010	000-010	000-010	000-010	000-010	000-010	000-010	050-075	000-015
000-010	000-010	000-010	000-010	000-010	000-010	000-010	075-100	000-015

3.2. Conceptual Model Design

There is no defined mathematical model (algebraic, differentials, or other systems of equations, as stated above)

for this problem, since there is no knowledge of a formal relationship between the selected characteristics and the demographic sub-concepts, and between these sub-concepts and the total risk index. Thus, we proposed to model the

demographic conceptual network by developing an artificial neural network (ANN) with progressive structured communication between the layers, based on supervised learning.

This software artifact "learns" the relationships between the inputs and outputs of the examples, without the need to define them explicitly.

This decision was certainly based on the model proposed by the IIGHI-CONICET researchers and by the complexity and non-linearity of the problem domain, as expressed by the experts.

To that end, we designed experiments for two different alternatives, that were discussed in work sessions with Dra. Rojas and the development team:

a) Design a single neural network where every measurable feature were put on the same level, as an input layer, with a single output neuron representing the total risk (figure 5).

b) Generate several neural networks working independently. These networks generate values for each intermediate concept and would finally be integrated for the determination of total risk, each with its own intermediate layers (Figure 6).

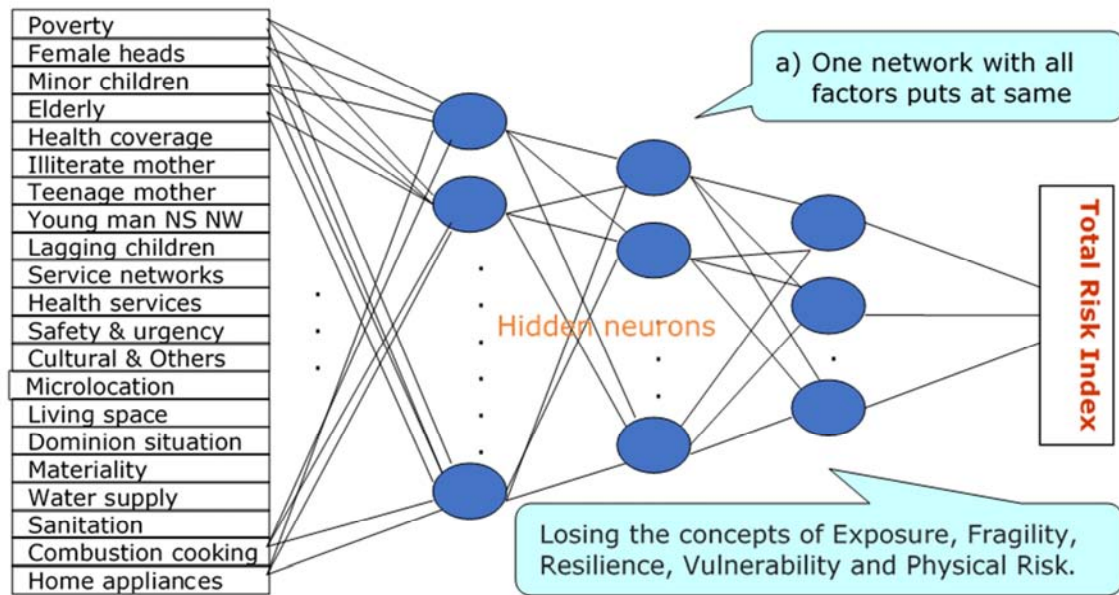


Figure 5. First proposal for the computational model: all features of the model in the first level.

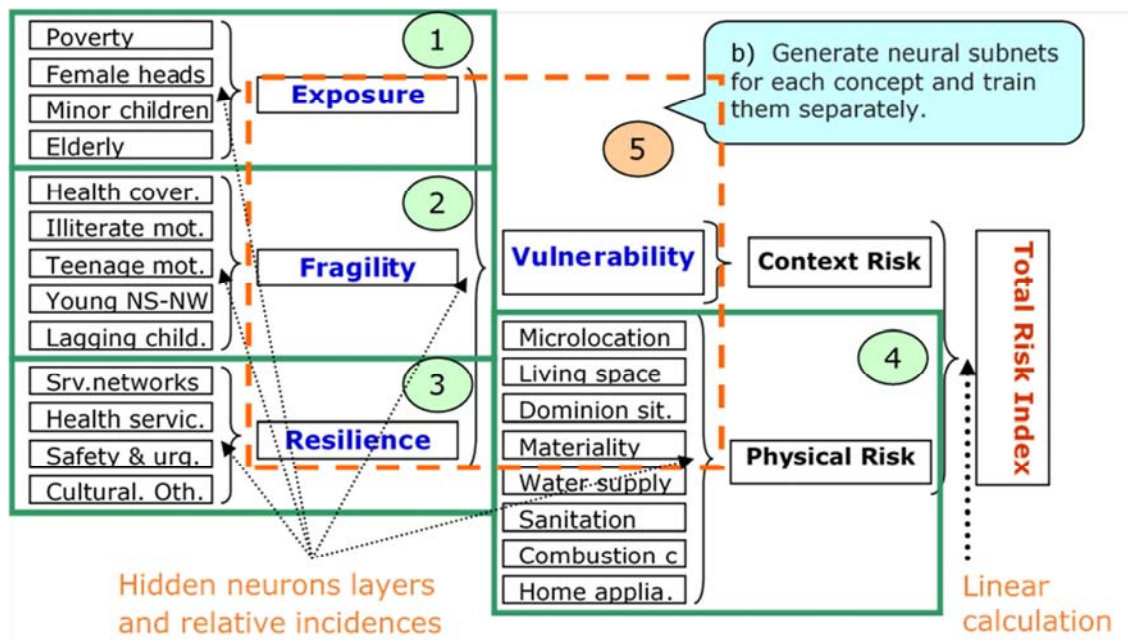


Figure 6. Second proposal for the computational model: five neural network working collaboratively. Selected.

The first option was discarded since the intermediate concepts (resilience, exposure, fragility, and vulnerability) were lost. These concepts were considered very important to

the IIGHI researchers and could also be used for further analysis and future research projects.

On the other hand, when generating numerical examples,

experts already had difficulty estimating the values of the non-measurable sub-concepts given a set of input values. Therefore, the estimation of the total risk considering all the characteristics at the same time would be even more complex and, thus, unreliable.

The second alternative (Figure 6) seemed more appropriate since it does not suffer from the previous difficulties and is conceptually similar to the theoretical model.

Consequently, we designed and built five multilayer perceptron (MLP) artificial neural networks, which were trained with the classical back propagation algorithm.

The final calculation of the risk index would be linear according to the weightings indicated by the experts.

The neural networks were designed as a pure multilayer perceptron (without momentum or other optimization techniques), with hidden layer neurons governed by a hyperbolic tangent and linear output neurons.

The elements of the input layer only normalized the input values so that they are within the real range of [0, 1] and that they would take the median of each proposed interval as a representative value for the feature.

The construction was carried out by the development team of the research and development project RNA-AC – 25/E078

of Cordoba Regional Faculty - UTN. The C# programming language was chosen as development language given the experience of the team and the availability in the Software Research Laboratory.

3.3. Modules Built in C#

The task of developing the tool was carried out in two stages. The first was the creation of the network configuration and network training module, and the second was the development of a redistributable module to use the pre-trained networks:

Training module: This module configures the neural networks using text configuration files, which allow the definition of the layers and numbers of neurons for each subnet. We conducted several experiments to find the best set of hyper parameters. The root mean square error tolerance and the number of epochs were also defined programmatically.

A descriptive file was generated, containing the network topology, the type of subnet, the training date, the number of examples used and the synaptic weights obtained at the end of the training process.

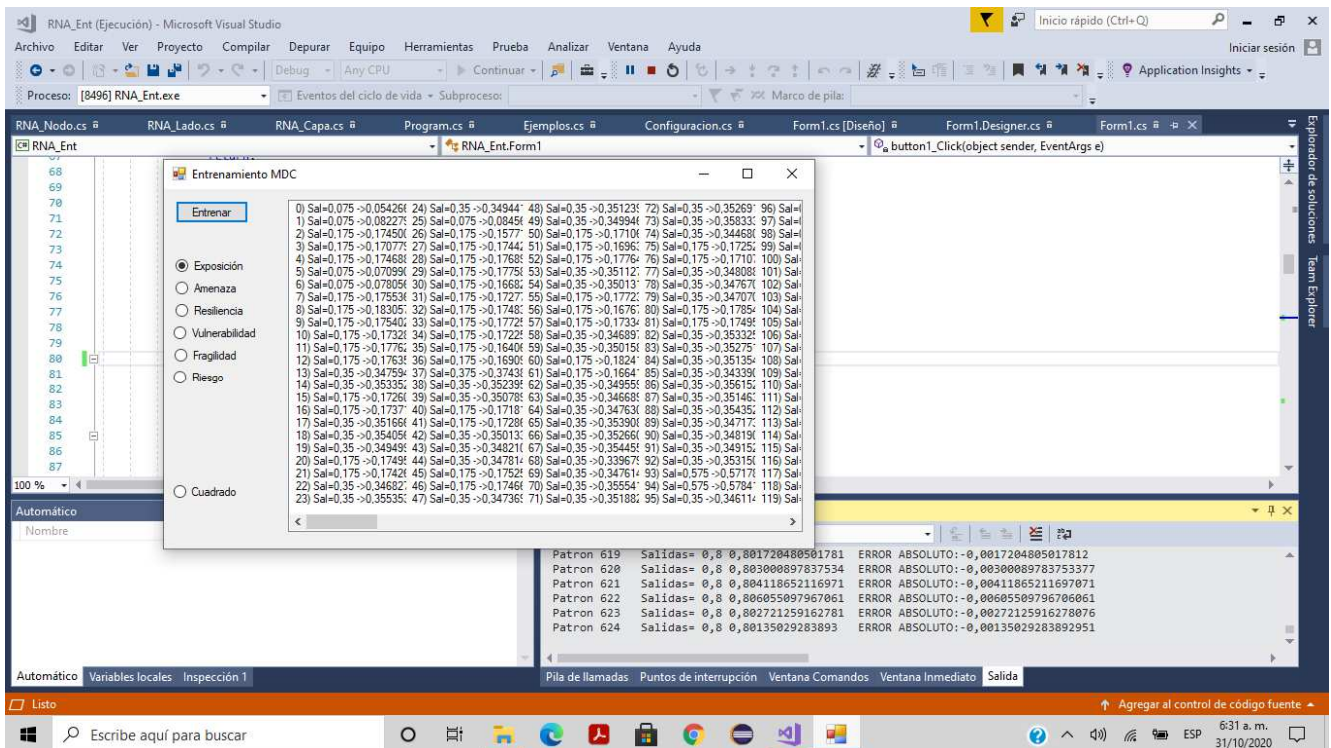


Figure 7. Example of neural networks training evolution of DRVS (Spanish).

Production module: this module (RVS v2.1) was designed to transfer the methodology to different population and community health research centers, and to be used intuitively by local governments. Reading the descriptive files generated, it configures the pre-trained neural networks in memory and allows entering real values for the studied populations. Then it calculates each concept and generates

the total risk index, recording the participation values according to the defined fuzzy sets. It also allows to export the inputs and calculations to Excel files for storage, sharing and later studying.

A sample of the interface developed for this module can be seen in Figure 8.

Entradas		Salidas	
Microlocalización	Índice	Salida calculada utilizando red neuronal entrenada por retropropagación de errores..	
0	0.20	0	
Espacio habitable	Índice	Muy Bajo	0 %
0	0.075	Bajo	0 %
Situación dominal	Índice	Moderado	0 %
0	0.075	Alto	0 %
Materialidad	Índice	Muy Alto	0 %
0	0.10		
Provisión de agua	Índice		
0	0.20		
Saneamiento	Índice		
0	0.20		
Combustión p/cocinar	Índice		
0	0.10		
Electrodomésticos	Índice		
0	0.05		
	Índice		
	Índice		

Figure 8. Normal user interface of RVS 2.1 module (Spanish or Portuguese). In this case, «physical risk» or «threat» (“Amenaza”).

4. Community and Familiar Clinic History System

DRVS Methodology and the associated software RVS v2.1 allow to determine the capabilities of a community to deal with unhealthy situations by measuring features represented by all the passive and actives of the human, social and physical capitals, which conform the basis for problem recovering or surpassing. In order to acquire and generate input data to RVS v2.1, we have developed a Community and Family Clinical History (HCFC for Spanish acronym) software, which allows increasing the available information that is necessary for the training of the artificial neural network.

This tool shows it is necessary to review the medical consultation register systems, even if there is health attention aspects to revisit. One of the perspectives of Collective and Familiar Medicine, based on McWinney’s proposal [12], suggests an approach patient oriented but considering his familiar and community context to improve the reach of the clinical method, gaining a more integral approach of the health-disease process. In order to be able to implement this process, called *Patient Centered Clinic Process*, a tool that supports gathering information is needed, considering the following aspects:

Prevention opportunities detection: In order to strengthen the protective health features and reduce the harmful ones, it is important to modify the reactive approach, focused on disease treatment, towards a proactive approach related to an integral patient attention.

Holistic attention: This aspect arises from an integral view of the person, including his interaction to their physical, social, and biological environment, with a focus on social vulnerability. This allows to contextualize the resources (active and passive) he has for the exploitation of the opportunity structures their current environment offers.

Disease and ailment research: Understanding that the ailment is the perception of the patient in relation to the expressed symptoms of his disease, the health professional must know and distinguish them to be able to answer the patient needs.

Attention Continuity: The continuity is focused not only to the diagnostic of a patient, but also allow to monitor its evolution, both in eventual and in chronic cases.

Quick access and updated information of a patient’s health problems: The quick identification of the health issues of both a patient as well as their familiar environment allows to analyze the problems he has had, as well as detect the ones that are not solved, in order to continue their treatment.

Continuous Monitoring of the attention quality: Several attention models can be studied by looking at the health professional logs. For example, an individual approach on the patient may be exposed by an absence of his familiar aspects.

Valid data sources for research: The registered data must be available to design studies or research to improve attention process and the health politics decision making.

Single tool to register all health data: It must allow to be a central registry for all information to be accessed for any health personnel.

Considering all these items, we developed a new tool, designed as a Community and Familiar Clinic History, with

an integral approach on the patient considering the social features from a social and environmental vulnerability focus [13]. The features to be built are based on intersectoral and multidisciplinary aspects, focusing on improving the citizen participation. This will allow to understand the existing relations between each of the components involved in the health-disease process, thus giving a broad information source that allows to understand health issues, decision making and the generation of intervention devices from a holistic perspective.

Considering the basic aspects of a standard Electronic Clinic History system, we included the key features to be used as input data by RVS.

These systems seeks to:

Facilitate integral medical care.

Research the ailment and disease of the patient.

Detect opportunities for prevention.

Facilitate continuity care.

Gain quick and updated access to an individual health issue.

Continuous monitoring of the attention quality.

Be a valid research data source.

Be a centered registry for all health personnel.

HCFC pretends to support decisions that allow to drive actions related to public and health politics, with two main focuses:

Medical Focus: Through the analysis of the data related to health parameters such as BMI (Body Mass Index), weight,

height, sex, age, as well as the ones related to pathologies, treatments, and studies.

Demographic Focus: Through the analysis of data related to the geographic location of the population and the social, human, and physical assets.

The methodology used for software development was *user oriented evolutive prototyping* since the system was developed iteratively based on periodic meetings with users.

We worked together with researchers from the medical team of the Department of Family Medicine of the Faculty of Medical Sciences – National University of Córdoba and with expert professionals from the IIGHI-CONICET to create the prototypes and elicit requirements.

Therefore, a multidisciplinary research team worked together on this research and development project, each helping with their knowledge and experience in the domain.

Each module was developed based on the information needs for both the health team and the RVS module, using the following structure (the data associated with the description concern the DRVS Methodology and the associated software RVS v2.1 with copyright of CONICET-UTN, registered on 25/8/2008 by Exp. N° 647178):

1. Basic personal data: General basic data is registered for both the person and the clinic history. This information includes clinic history number, document number and type, street address and birth date. Anthropometric information is also registered, such as weight, BMI, and abdominal circumference. (see figure 9 for sample).

Figure 9. User interface of the Clinic History of a patient (Spanish).

2. Human, social, and physical assets data: complementary modules were designed to group critical features by the type of asset, needed as inputs for RVS, keeping in mind the goal of making the system simple and fast.

a. Human asset: is the potential economic value of the

biggest productive capability for the individual, based on knowledges, abilities and attitudes acquired in school, university or by experience, improving the probabilities of a better and more stable future well being. This aspect was modeled through the

registration of the information related to the educational level and the work status of the person and his family group.

- b. Social asset: is a highly intangible asset based on the relations between people that constitute the social structure and that, beyond their functions, configure opportunity structures in the community. This asset was modeled in HCFC by registering data related to technical networks, community services and cultural spaces.
 - c. Physical asset: this asset is related to housing from the focus of being able to satisfy the housing needs related to health. The data registered are the micro-localization, the habitability, the ownership situation, the water supply, the basic cleaning infrastructure, and the cooking combustion in the house.
3. Preventive Practices: this represents the proactive focus of the patient attention process, since the allow that transform each visit in an opportunity to work with the patients and their family on the different health related issues, suggesting the implementation of pre-emptive practices related to the life cycle stage, through consecutive visits.
 4. Problems listing: the listing of issues constitutes a really useful tool in ambulatory practice to collect information related to the history of the patient and his family, identifying unsolved problems that need to be treated by the health team in successive visits. It is possible to follow each patient in HCFC through the identification of his issues, through the «Statistical Classification of Health Issues in Primary Attention» (CEPS-AP), which was an adaptation designed by the Health Ministry of Argentina and the PAHO of coding systems (CIP SAP – CIE 10) but adapted to the first level of attention.
 5. Prevalent Issues: HCFC allows to follow patients with chronic prevalent issues, such as HTA, DBT and nutritional problems (obesity) from a social vulnerability approach. The monitoring is not limited to the medical treatment data, but also to the lifestyle habits that influence the disease.
 6. Summarized reports: allows to carry out different analysis on the registered information on the epidemiologic profile for the consulting population, and on the resources related to the previously mentioned assets, additionally allowing to analyze the health professionals profile in relation to their attention, identifying several quality metrics.

It should be noted that, in order to facilitate the recording and monitoring of the available information, the basic patient and the medical visit data was set as priority, and the community data were set as complementary so that the health user may be completing it through consecutive visits, allowing a personal profile of each patient.

It should be noted that, in order to facilitate the recording and monitoring of the available information, the basic patient and the medical visit data was set as priority, and the community data were set as complementary so that the health user may be completing it through consecutive visits,

allowing a personal profile of each patient.

5. Application and Communication

The *DRVS* methodology and the *RVS 2.1* software has been used by the demographic researchers at IIGHI-CONICET using statistical data from census and real data collection in high, middle, and lower social classes in the cities of *Cordoba* and *Resistencia* (Argentina), and in *Asuncion* (Paraguay) with positive results. We also work in Cuba, in the cities of *La Habana*, *Santa Clara* and *Santiago de Cuba*.

The *DRVS* methodology and the *RVS 2.1* software has been used by the demographic researchers at IIGHI-CONICET using statistical data from census and real data collection in high, middle, and lower social classes in the cities of *Cordoba* and *Resistencia* (Argentina), and in *Asuncion* (Paraguay) with positive results; work was also done in Cuba, in the cities of *La Habana*, *Santa Clara* and *Santiago de Cuba*.

It has also been transferred to research centers and universities in Latin America to be used, tested and measured. In Argentina, researchers of the IIGHI are currently working along municipal governments of “*Villa del Totoral*” and “*Salsipuedes*”, province of *Cordoba*, thanks to transference and assistance agreements established with their authorities.

The software and base methodology have been presented in different forums:

1. Argentine Association of Population Studies Conferences.
2. Iberoamerica Popular Habitat Seminar, Advanced Studies Center of the National University of Cordoba.
3. II Conference of the Latin American Association of Population Studies, Guadalajara, México.
4. First Mexican International Conference on Healthy Housing, Pachuca de Soto Hidalgo, México.
5. 54° International Conference of Americanists (ICA), Vienna, Austria.
6. three scientific communication books and academic, graphic and radio news.

It should be noted that the HCFC has been implemented in the Family Medicine Practices in the National Hospital of Clinics of the FCM-UNC and that it has been registered by CONICET-UTN-UNC, with registration number 814574 and denomination: “*HCFC v1.00 - Historia Clínica Familiar y Comunitaria versión 1.0*”.

Results of the dissemination of the *Family and Community Clinical History System* for obtaining primary data for the implementation of the *DRVS* methodology can be found in several articles by the authors. Presentations have also been made in 2012 and 2013, at the Conference of Young Researchers in Neuroscience organized by the National University of Cordoba.

On the other hand, the Inter-American Healthy Housing Network endorsed by the Pan American Health Organization has expressed interest to test the methodology and its associated software; the first transference was done to the

“Facultad de Arquitectura y Urbanismo of the Universidad Católica de Nuestra Señora de la Asunción”, in Paraguay (with the software being registered with joint ownership of CONICET and UTN-Cordoba Faculty); it was then transferred to the “Fundación Oswaldo Cruz” and to the “Universidad de la Amazonia”, both in Brazil, to the “Instituto Nacional de Higiene, Epidemiología y Microbiología” of Cuba, to the “Facultad de Medicina of the Universidad Nacional del Nordeste”, in Argentina, and to the “Universidad del Rosario de Colombia” in Colombia. We already mentioned the transference to the local governments of “Villa del Totoral” and “Salsipuedes” (Córdoba, Argentina); there is also an ongoing transference to the “Municipality of Benevides in Belém, Pará state, Brazil”. The HCFC was transferred to the Family Medicine Department of the National Hospital of Clinics of the National University of Córdoba.

In the studied cases, the software results agree with the expert opinions on the health risk. However, it is too early to feel successful, the methodology is statistically validating the forecasts, but the software must be taken to extreme situations to be tested for correctness, even the proposed examples are in constant revision. The use of the methodology by local governments is a strong metric of the usability of DRVS.

The RVS distributable module has been internationalized in its 3.0 version to translate to Portuguese for users in Brazil. The configurations and training module is being reviewed, considering the latest machine and deep learning advances.

6. Approach Discussion

As stated, the decision to use neural networks (2005-2008) to model the conceptual diagram of DRVS, was strongly influenced by the initial diagram, the absence of mathematical models (equations) that relate the measurable features to the sociodemographic concepts determined by them, and by the assumption of complexity and non linearity indicated by the experts.

Between 2010 and 2011, we also experimented a storage schema in a database using similarity search algorithm on the data [14], which did not reach a sufficiently well-defined product for transfer and distribution.

With the creation of our Research, Development and Transference Group on Machine Learning, Languages and Automata (GAALA, 2020), we are currently deepening our knowledge and designing new developments on diverse machine learning techniques, specially on the deep learning methods that have achieved unprecedented artificial intelligence success in the last years. This leads to a revision on the neural network configurations for DRVS and further experiments. These experiments do not alter significantly the results of the risk, vulnerability and resilience indexes, but we believe it could lead to a more robust and solid software.

We have been testing new data analysis tools (*Tensorflow*,

Keras, Pandas, etc.), in Python, that will allow to recreate the development of the training module improving its generalization and the training speed, allowing to adapt the software more easily to changes in the DRVS model that IIGHI demographers consider necessary.

On the other hand, these new techniques also apply to similarity searches through modern non-parametrical algorithms, which we hope will allow to recreate the work of 2011, creating an alternative adaptative approach that would work, in case of being correct, to improve the knowledge base of demographers without requiring training.

Finally, the new visualization techniques for learnt features in the intermediate layers of the neural network could allow to understand the calculations it is doing, which currently can not be explained.

7. Conclusions and Future Work

In Social Sciences, a dashboard providing information on the features involved and allows to guide the decision-making process is always useful, needing a group of experts to understand that information and decide on the preventive or corrective actions.

However, ¿what is to be done when the relationship between the features and their determinant is unknown? specially under a complex and not linear assumption.

Artificial neural network help in these cases, being able to learn the expert criteria through examples and, thus, be used by authorities that must decide on resource allocation, with the possibility of carrying out simulations on possible changes and their impact.

The field tests and the necessary corrections based on results that allow to have a methodology and associated software will be useful as another tool in this decision-making process for community and environmental health.

We are currently researching machine learning techniques [15-16] that could provide evolutive tools which do not need re-training when introducing new information, such as non-parametric search model through nearest neighbor and other techniques, common nowadays in Data Science.

The current knowledge on machine learning and neural networks explain some of the problems we had detected when developing the software, but which we could not explain. Additionally, powerful new tools and algorithms have been proposed in the last decade, allowing us to review the general computational approach. This is our current line of work.

Acknowledgements

The projects that have supported the development of the current tools have been financed by the *Secretary of Science, Technology and Postgraduate of the National Technological University, the Secretary of Science and Technology of the Cordoba Regional Faculty (UTN), of the Ministry of Science and Technology of the government of the Province of Cordoba, and of the Ministry/Secretary of Science and*

Technology of the National Government of Argentina.

The copyright on the RVS software and the DRVS methodology, as well as the work of the IIGHI researchers have been managed and financed by National Council for Scientific and Technical Research (CONICET).

The physical and technical infrastructure available where the different members of the research teams of UTN-FRC have developed their activities has been provided by the Department of Information Systems Engineering of UTN-FRC.

To every one of them goes our acknowledgements and thanks.

References

- [1] WHO, Environmental Health, Consulted on date 26/10/2020 https://www.who.int/topics/environmental_health/es/.
- [2] PAHO, Community Health in the XXI Century, Consulted on date 26/10/2020, https://www.paho.org/hq/index.php?option=com_content&view=article&id=13491:community-health-in-the-21st-century&Itemid=40283&lang=es.
- [3] Argentina.gob.ar, “Salud Ambiental”, Consulted on date 26/10/2020, <https://www.argentina.gob.ar/salud/ambiental>.
- [4] Argentina.gob.ar, “Salud Ambiental”, Consulted on date 26/10/2020, <https://www.argentina.gob.ar/salud/comunitaria>.
- [5] Rojas, M. C. “La vivienda precaria urbana marginal y su relación con la salud de la población en el proceso de sustentabilidad. Un enfoque teórico para la estimación del riesgo y la vulnerabilidad”. First Congress of the Latin American Population Association -ALAP-, CD-ROM. Brazilian Association of Population Studies -ABEP- [Presentation] Caxambu -MG- Brazil; 2004.
- [6] Rojas, M. C. “Población, vivienda salud y vulnerabilidad global. Propuesta teórico-metodológica para la estimación del riesgo de la vivienda urbana para la salud basada en el análisis de la vulnerabilidad sociodemográfica.” [doctoral thesis] Faculty of Economic Sciences of the National University of Cordoba; Cordoba, Argentina; 2006.
- [7] Cardona Arboleda, O. D. “Estimación holística del riesgo sísmico utilizando sistemas dinámicos complejos”, Barcelona, España, 2001. Consulted on date: 01/10/2020, [https://repositorio.gestiondelriesgo.gov.co/bitstream/handle/20.500.11762/19751/HolisticaRiesgoSismicoBogota\(Cardona_2001\).pdf?sequence=1](https://repositorio.gestiondelriesgo.gov.co/bitstream/handle/20.500.11762/19751/HolisticaRiesgoSismicoBogota(Cardona_2001).pdf?sequence=1).
- [8] IIGHI-CONICET, NRVAS Line of Research, Consulted on date: 01/10/2020, [https://iighi.conicet.gov.ar/nucleos/nrvas/#:~:text=La%20Red%20Interamericana%20de%20Vivienda%20Saludable%20\(Red%20VIVSALUD\)%20es%20una,la%20Regi%C3%B3n%20de%20las%20Am%C3%A9ricas](https://iighi.conicet.gov.ar/nucleos/nrvas/#:~:text=La%20Red%20Interamericana%20de%20Vivienda%20Saludable%20(Red%20VIVSALUD)%20es%20una,la%20Regi%C3%B3n%20de%20las%20Am%C3%A9ricas).
- [9] Chen, T., & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794); San Francisco, California, USA. August 2016. <https://doi.org/10.1145/2939672.2939785>.
- [10] Harris, C. R., Millman, K. J., van der Walt, S. J. et al., Array programming with NumPy; Nature 585, 357–362 (2020). <https://doi.org/10.1038/s41586-020-2649-2>.
- [11] Lundberg, S. M., Erion, G., Chen, H. et al., From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2, 56–67 (2020). <https://doi.org/10.1038/s42256-019-0138-9>.
- [12] McWinney, I. “Medicina de Familia”, Ed. Doyma. Barcelona, Spain; 2000.
- [13] Rojas, M., Vazquez J., Castillo J., Cardenas M. “Modelado del Riesgo de la Vivienda Urbana para la Salud y el empleo de Redes Neuronales Artificiales para su Estimación”, JAIIO 2010, Workshop CASI, Cordoba, Argentina, 2010.
- [14] Frittelli, V., Steffolani, F. A., Teicher, R. G., Rojas, M., Picco, J. E., Vazquez, J. C. “Búsqueda por Similitud aplicada en la Recuperación de Factores que inciden en el Cálculo del Índice de Riesgo para la Salud de la Vivienda Urbana”, 40 JAIIO – CASI, Cordoba, Argentina, 2011. Query at 01/10/2020: http://40jaiio.sadio.org.ar/sites/default/files/T2011/CAIS/CAI_S2011-28.pdf.
- [15] Goodfellow I., Bengio Y., Courville A., “Deep Learning”, MIT Press, NY, USA, 2016. (www.deeplearningbook.org).
- [16] LeCunn, Y., Bengio, Y., Hinton, G. “Deep Learning”; Nature 521, 436-444 (2015). doi: 10.1038/nature14539.