Paraconsistents artificial neural networks applied to the study of mutational patterns of the F subtype of the viral strains of HIV-1 to antiretroviral therapy

PAULO C.C. DOS SANTOS[1,2], HELDER F.S. LOPES[2], ROSANA ALCALDE[3], CLÁUDIO R. GONSALEZ[3], JAIR M. ABE[2,4] and LUIS F. LOPEZ[1,5]

[1]Universidade de São Paulo, Faculdade de Medicina, Instituto Oscar Freire - Medicina Legal, Disciplina de Informática Médica da Faculdade de Medicina (DIM), Laboratório de Investigação Médica em Informática Médica (LIM01-HCFMUSP), Av. Dr. Arnaldo, 455, Cerqueira César, 01246-903 São Paulo, SP, Brasil
[2]Universidade de São Paulo, Instituto de Estudos Avançados, Rua Praça do Relógio, 109, bloco K, 5º andar, Cidade Universitária, 05508-050 São Paulo, SP, Brasil
[3]Universidade de São Paulo, Faculdade de Medicina, Laboratório de Investigação Médica em Dermatologia e Imunodeficiência (LIM56-HCFMUSP), Av. Dr. Enéas de Carvalho Aguiar, 255, 3º andar, Cerqueira César, 05403-900 São Paulo, SP, Brasil
[4]Universidade Paulista de São Paulo, Departamento de Engenharia de Produção, Instituto de Ciências Exatas e Tecnologia (ICET), Rua Dr. Bacelar, 1212, Vila Clementino, 04026-002 São Paulo, SP, Brasil
[5]Núcleo de Aplicações em Redes Avançadas (NARA-ANSP), Instituto Oscar Freire - Medicina Legal, Av. Dr. Arnaldo, 455, Cerqueira César, 01246-903 São Paulo, SP, Brasil

ABSTRACT

The high variability of HIV-1 as well as the lack of efficient repair mechanisms during the stages of viral replication, contribute to the rapid emergence of HIV-1 strains resistant to antiretroviral drugs. The selective pressure exerted by the drug leads to fixation of mutations capable of imparting varying degrees of resistance. The presence of these mutations is one of the most important factors in the failure of therapeutic response to medications. Thus, it is of critical to understand the resistance patterns and mechanisms associated with them, allowing the choice of an appropriate therapeutic scheme, which considers the frequency, and other characteristics of mutations. Utilizing Paraconsistents Artificial Neural Networks, seated in Paraconsistent Annotated Logic Eτ which has the capability of measuring uncertainties and inconsistencies, we have achieved levels of agreement above 90% when compared to the methodology proposed with the current methodology used to classify HIV-1 subtypes. The results demonstrate that Paraconsistents Artificial Neural Networks can serve as a promising tool of analysis.

Key words: Artificial Neural Networks, HIV, genotyping, paraconsistent logic, Paraconsistents Artificial Neural Networks, pattern recognition.

INTRODUCTION

Since the first reports in 1981 (Gotlieb et al. 1981), the AIDS pandemic continues to advance with an incidence of HIV infection still high. According to the WHO (World Health Organization) and UNAIDS (Joint United Nations Program on HIV/AIDS) in 2014, were accounted for around 34 million infected (WHO 2014, UNAIDS-AIDS 2014).

These alarming figures driving the scientific community to develop anti-HIV strategies, which aim to prevent new infections and treat infected individuals.

Correspondence to: Paulo Cesar Costa dos Santos
E-mail: pccs@usp.br

Currently, there are two types of AIDS virus known as HIV-1 and HIV-2, which are closely related. HIV-2 is endemic in West Africa and is spreading throughout India. However, the majority of AIDS cases worldwide are caused by HIV-1, the most virulent (Grant and Cock 2001).

The High variability of HIV-1, as well as the lack of efficient repair mechanisms during the stages of viral replication, contribute to the rapid emergence of HIV-1 strains resistant to antiretroviral drugs. The selective pressure exerted by drugs leads to fixation of mutations that confer varying degrees of resistance. The presence of these mutations is the main factor in the failure of therapeutic response to antiretroviral. Thus, it is paramount to understand the resistance patterns and mechanisms associated with them, allowing the choice of appropriated treatment to consider the frequency and other characteristics of mutations (Dos Santos 2010).

The HIV-1 protease is an aspartic protease, which is composed of two identical 99 amino acid monomers with each other non-covalently associated. Against these 99 amino acids can be made comparison and classification of virus subtypes (Dos Santos 2010).

In recent years, research and experiences gained great momentum thanks to the interest of the Ministry of Health, through the Department of SUS-DATASUS Informatics, to establish at the national level, a standardized computer system for collecting and processing of clinical and administrative data originated in each contact with the care system (DATASUS 2009). Specifically, the computer systems developed by DATASUS are used by the National STD/AIDS, to their management, using the notification of AIDS cases generated universal and compulsory registration in the System at Notification Diseases Information (SINAN) and the number of deaths registered in the Mortality Information System (SIM), and the own National STD/AIDS manages four specific systems: the Prevention of Inputs Monitoring System (PREVINI), the Logistics Management Systems of Medicines (SICLOM), the Laboratory Tests Control System Count Network lymphocytes CD4+/CD8+ and viral load (SISCEL), and the Information System for Genotyping Network (SISGENO), all with web query (SISCEL-SISGENO 2009).

These systems are relevant databases for assistance and conducting research. However, because they were not prepared with the basic purpose of promoting studies and the need to ensure the confidentiality of individuals, it is virtually impossible to link the records of these systems, recovering the history of assisted patients (Campos et al. 2006).

In order to this information be used by an intelligent data storage process in a structured database, you must map it in an appropriate format, the information contained in the various medical forms related to patients treated in dermatology clinics (Hospital of Clinical of University of São Paulo). In this context, medical computing, Department of Medical Informatics (DIM-USP) has, today, tools and instruments to support the administrative organization of medical visits, capture, storage and processing of patient information, the generate diagnoses, therapeutic counseling and access to information. To perform this process, we apply a new class of artificial neural networks: Paraconsistent Artificial Neural Networks (PANN).

Therefore, in this paper we suggest a methodology to create a database to store amino acid sequences of HIV-1 subtypes and compare them to positive serum samples from patients to determine the virus subtype in this sample.

BACKGROUND

PARACONSISTENT ARTIFICIAL NEURAL NETWORK

PANN was introduced in fundamental of logical annoted (Abe 1992). Its basis leans on

paraconsistent annotated logic $E\tau$ (Abe 1992). Let us present it briefly.

The atomic formulas of the logic $E\tau$ are of the type $p(\mu, \lambda)$, where $(\mu, \lambda) \in [0, 1]^2$ and $[0, 1]$ is the real unitary interval (p denotes a propositional variable). $p(\mu, \lambda)$ can be intuitively read: "It is assumed that p's favorable evidence is μ and contrary evidence is λ." (Abe 2010) Thus:

- $p(1.0, 0.0)$ can be read as a true proposition.
- $p(0.0, 1.0)$ can be read as a false proposition.
- $p(1.0, 1.0)$ can be read as an inconsistent proposition.
- $p(0.0, 0.0)$ can be read as a paracomplete (unknown) proposition.
- $p(0.5, 0.5)$ can be read as an indefinite proposition.

We introduce the following concepts (all considerations are taken with $0 \leq \mu, \lambda \leq 1$):

- Uncertainty degree (Eq. 1);
- Certainty degree (Eq. 2);

$$G_{un}(\mu, \lambda) = \mu + \lambda - 1 \qquad (1)$$

$$G_{ce}(\mu, \lambda) = \mu - \lambda \qquad (2)$$

An order relation is defined on $[0, 1]^2$: $(\mu 1, \lambda 1) \leq (\mu 2, \lambda 2) \Leftrightarrow \mu 1 \leq \mu 2$ and $\lambda 1 \leq \lambda 2$, constituting a lattice that will be symbolized by t.

With the uncertainty and certainty degrees, we can get the following 12 output states (Table I): extreme states, and non-extreme states.

TABLE I
Extreme and Non-extreme states.

Extreme states	Symbol	Non-extreme states	Symbol
True	V	Quasi-true tending to Inconsistent	Qv→T
False	F	Quasi-true tending to Paracomplete	Qv→⊥
Inconsistent	T	Quasi-false tending to Inconsistent	Qf→T
Paracomplete	⊥	Quasi-false tending to Paracomplete	Qf→⊥
		Quasi-inconsistent tending to True	T→v
		Quasi-inconsistent tending to False	T→f
		Quasi-paracomplete tending to True	⊥→v
		Quasi-paracomplete tending to False	⊥→f

Some additional control values are:

- V_{scct} = maximum value of uncertainty control = Ft_{ct}
- V_{scc} = maximum value of certainty control = Ft_{ce}
- V_{icct} = minimum value of uncertainty control = $-Ft_{ct}$
- V_{icc} = minimum value of certainty control = $-Ft_{ce}$

All states are represented in the next figure (Figure 1).

THE MAIN ARTIFICIAL NEURAL CELLS

In the PANN, the certainty degree G_{ce} indicates the 'measure' falsity or truth degree. The uncertainty
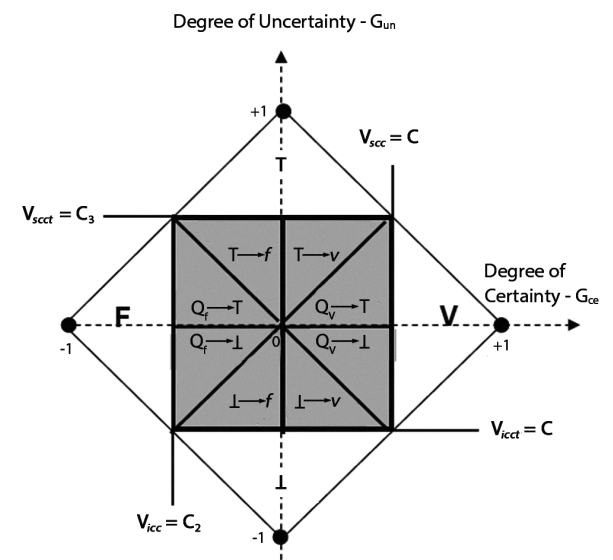

Figure 1 - Extreme and Non-extreme states.

degree G_{un} indicates the 'measure' of the inconsistency or paracompleteness. If the certainty degree is low or the uncertainty degree is high, it generates an indefinition (Figure 2).
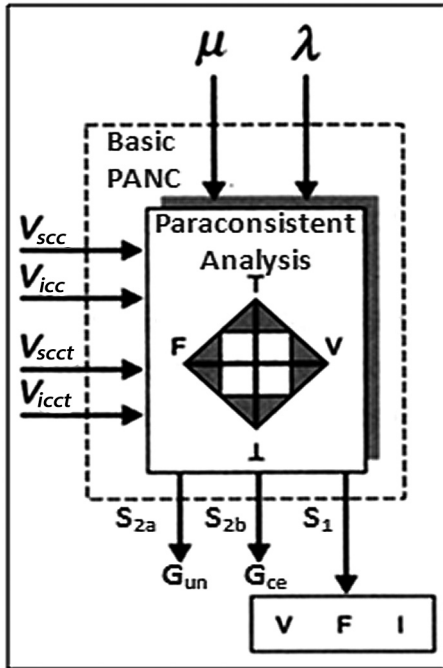

Figure 2 - Basic cell of PANN.

The resulting certainty degree G_{ce} is obtained as follows:

*   If: $V_{icc} \leq G_{un} \leq V_{scc}$ or $V_{scct} \leq G_{un} \leq V_{icct}$ $\Rightarrow$ G_{ce} = Indefinition
*   For: $V_{cpa} \leq G_{un} \leq V_{scct}$
    If: $G_{un} \leq V_{icc} \Rightarrow G_{ce}$ = False with degree G_{un}
    $V_{scct} \leq G_{un} \Rightarrow G_{ce}$ = True with degree G_{un}

A Paraconsistent Artificial Neural Cell – PANC – is called *basic* PANC when given a pair (μ, λ) is used as input and resulting as output:

* $S_{2a} = G_{un}$ = resulting uncertainty degree
* $S_{2b} = G_{ce}$ = resulting certainty degree
* $S_1 = X$ = constant of Indefinition.

Using the concepts of *basic* Paraconsistent Artificial Neural Cell, we can obtain the family of PANC considered in this work: Analytic connection (PANCac), Maximization (PANCmax), and Minimization (PANCmin) as described in Table II below:

PARACONSISTENT ARTIFICIAL NEURAL CELL OF ANALYTIC CONNECTION – PANCac

The Paraconsistent Artificial Neural Cell of analytic connection cell (PANCac) is the principal cell of all PANN, obtaining the certainty degree (G_{ce}) and the uncertainty degree (G_{un}) from the inputs and the tolerance factors.

This cell is the link which allows different regions of PANN perform signal processing in distributed and through many parallel connections (Da Silva Filho and Abe 2001).

The different tolerance factors certainty (or contradiction) acts as inhibitors of signals, controlling the passage of signals to other regions of the PANN, according to the characteristics of the architecture developed.

TABLE II
Paraconsistent Artificial Neural Cell.

PANC	Inputs	Calculations	Output
Analytic connection: PANCac	m, l, $Ft_{ct,}$, Ft_{ce}	$\lambda_c = 1 - 1$ G_{un} G_{ce}, $\mu_r = (G_{ce} + 1)/2$	If $|G_{ce}| > Ft_{ce}$ then $S_1 = \mu_r$ and $S_2 = 0$; If $|G_{un}| > Ft_{ct}$ and $|G_{un}| > |G_{ce}|$ then $S_1 = \mu_r$ and $S_2 = |G_{un}|$, if not $S_1 = \frac{1}{2}$ and $S_2 = 0$
Maximization: PANCmax	m, λ	G_{ce} $\mu_r = (G_{ce} + 1)/2$	If $\mu_r > 0.5$, then $S_1 = \mu$ If not $S_1 = \lambda$
Minimization: PANCmin	m, λ	G_{ce} $\mu_r = (G_{ce} + 1)/2$	If $\mu_r < 0.5$, then $S_1 = m$ if not $S_1 = \lambda$

PARACONSISTENT ARTIFICIAL NEURAL CELL OF
MAXIMIZATION – PANCmax

The Paraconsistent Artificial Neural Cell of
maximization cell (PANCmax) allows selection of
the maximum value among the entries.

Such cells operate as logical connectives OR
between input signals. For this is made a simple
analysis, through the equation of the Degree of
Evidence (Table II) which thus will tell which
of the two input signals is of greater value, thus
establishing the output signal (Da Silva Filho and
Abe 2001).

PARACONSISTENT ARTIFICIAL NEURAL CELL OF
MINIMIZATION – PANCmin

The Paraconsistent Artificial Neural Cell of
maximization cell (PANCmin) allows selection of
the minimum value among the entries.

Such cells operate as logical connectives AND
between input signals. For this is made a simple
analysis, through the equation of the Degree of
Evidence (Table II) which thus will tell which
of the two input signals is of smaller value, thus
establishing the output signal (Da Silva Filho and
Abe 2001).

PARACONSISTENT ARTIFICIAL NEURAL UNIT

A Paraconsistent Artificial Neural Unit (PANU) is
characterized by the association ordered PANC,
targeting a goal, such as decision-making, selection,
learning, or some other type of processing.

When creating a PANU, one obtains a data
processing component capable of simulating the
operation of a biologic neuron (Da Silva Filho
2001).

PARACONSISTENT ARTIFICIAL NEURAL SYSTEM

Classical systems based on binary logic are
difficult to process data or information from
uncertain knowledge. These data are captured or
received information from multiple experts usually

comes in the form of evidence that bring many
contradictions.

Paraconsistent Artificial Neural Systems
(PANS) modules are configured and built
exclusively by PANU, whose function is to provide
the signal processing similar to processing that
occurs in the human brain.

MORPHOLOGICAL ANALYSIS OF HIV GENOTYPES

The process of morphological analysis of HIV
genotype is to compare the profile of a given
sample against the database of reference and
thereby determine how the sample is closer to the
reference sequences stored.

Before starting any processing using PANN,
is necessary to understand the workings and
characteristics of data that can be used for that
PANNP understand such characteristics.

To perform this procedure, first carried out a
conversion value of the DNA sequences, which are
alphanumeric to numeric. Thus we have: "?" takes
the value 0, *A* takes the value of 1, *B* takes the value
of 2…. *Z* has a value of 26.

With the sequences converted, using an PANN
that processes the data generated by three expert
systems in order to generate a profile based on some
characteristics of the sample: Specialist 1 - Number
of known mutations; Specialist 2 - Number of
unknown mutations; Specialist 3 - Intensity of the
mutations.

In each analysis, it generates a profile of each
sample versus reference stored. Elects itself as the
most similar comparison reference to return most
mi (m) and the lowest lambda (l) generated by the
analysis of PANN.

At the end of the process of analyzing the
PANN, it uses the value of (m) and (l) resulting
to subject them to a grid of paraconsistent logic
for making the final decision. This lattice will be
responsible for classifying the sample as "sample

with similarities in the database of reference" (the region of truth) or as "Sample without similarity in the database of reference" (region of falsehood).

EXPERT SYSTEM 1 - NUMBER OF KNOWN MUTATIONS

This specialist has the function to quantify the known mutations in the sample using the following equation (Eq. 3):

$$E_1 = \frac{\sum c}{n} \tag{3}$$

Where: E_1 is the value of expert system 1; d is the sum of positions greater than 0 (different of "?") and; n is total element sequence; x is the value of the position in the sequence of the sample, y is the value of the position in the sequence.

EXPERT SYSTEM 2 - NUMBER OF UNKNOWN MUTATIONS

This specialist has the task of quantifying the unknown mutations in the sample using the following equation (Eq. 4):

$$E_2 = \frac{\sum d}{n} \tag{4}$$

Where: E_2 is the value of expert system 2; d is the sum of positions equal 0 (equal of "?") and ; n is total element sequence; x is the value of the position in the sequence of the sample; y is the value of the position in the sequence.

EXPERT SYSTEM 3 - INTENSITY OF EXISTING MUTATIONS

Like any calculation comparing samples and references is accomplished using a string converted as explained above, this expert has the object to quantify (in principle, without any biological inference), the size of the mutation occurred, i.e. the difference between changing the existing value in the sample and reference. This measurement is used the following calculation (Eq. 5):

$$E_3 = \frac{\sum \frac{|x - y|}{a}}{n} \tag{5}$$

Where: E_3 is the value of expert system 3; x is the value of the position in the sequence of the sample; y is the value of the position in the sequence of reference; is the maximum amplitude of the samples (a = Z = 26); n = total number of elements of the sequence.

THE PARACONSISTENT ARTIFICIAL NEURAL NETWORK ARCHITECTURE

The architecture of the PANN used in decision-making is based on the architecture of Paraconsistent Artificial Neural System for Treatment of Contradictions (Figure 3).

This Artificial Neural System Paraconsistent receives three input signals and presents as a result, a value that represents the consensus between three information. The contradictions between the two values are added to the third value, so that the output, the value proposed by the dominant majority. The analysis is done on the fly carrying the entire real-time processing, similar to the operation of biological neurons.

In the final process of analysis undertaken by the PANN, the resulting values of output μr (value of evidence favorable) and λr (value of contrary evidence). After, submit these resulting values to Paraconsistent Logic, using a lattice for the final decision-making (Figure 4).

To achieve the final configuration of the lattice of decision-making were held some batteries of tests, using data from samples and references (20 samples and five references) in a controlled way (double blind), with the aim to discover the regions of falsehood (when the controlled tests were performed with samples really distinct from the references), and the regions of truth (when the controlled tests were performed with samples similar and identical). The limits of the areas of screening, to decision-making, can be seen in Table III.

MATERIALS AND METHODS

ETHICAL ASPECT

This study was analised and approved by the Comissão de Ética em Experimentação Animal (CEEA) and by the Comissão de Ética em Pesquisa com Seres Humanos (CEPSH) of the CEP-ICB (Protocol 274/2008).

EXEMPLIFYING THE METHODOLOGY

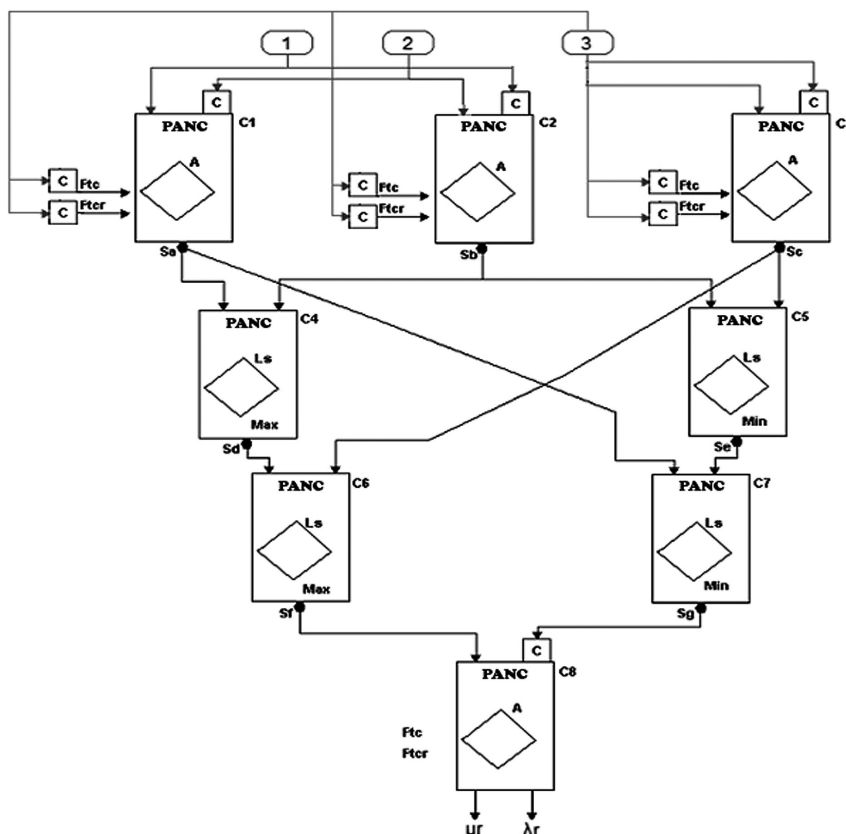The following is an example of the recognition process, which will consider three sequences

Figure 3 - The architecture for morphological analysis. Three expert systems operate: PA, for check the number of wave peaks; PB, for checking similar points, and PC, for checking different points: The 1st layer of the architecture: C1-PANC which processes input data of PA and PB; C2-PANC which processes input data of PB and PC; C3-PANC which processes input data of PC and PA. The 3rd layer of the architecture: C4-PANC which calculates the maximum evidence value between cells C1 and C2; C5-PANC which calculates the minimum evidence value between cells C2 and C3; C4 and C5 constitute the 2nd layer of the architecture; C6-PANC which calculates the maximum evidence value between cells C4 and C3; C7-PANC which calculates the minimum evidence value between cells C1 and C5. The 4th layer of the architecture: C8 analyzes the experts PA, PB, and PC and gives the resulting decision value. PANC A = Paraconsistent artificial neural cell of analytic connection. $PANCLs_{Max}$ = Paraconsistent artificial neural cell of simple logic connection of maximization. $PANCLs_{Min}$ = Paraconsistent artificial neural cell of simple logic connection of minimization. Ft_{ce} = Certainty tolerance factor; Ft_{ct} = Contradiction tolerance factor. S_a = Output of C1 cell; S_b = Output of C2 cell; S_c = Output of C3 cell; S_d = Output of C4 cell; Se = Output of C5 cell; Sf = Output of C6 cell; Sg = Output of C7 cell. C = Complemented value of input; μ_r = Value of output of PANN; λ_r = Value of output of PANN.
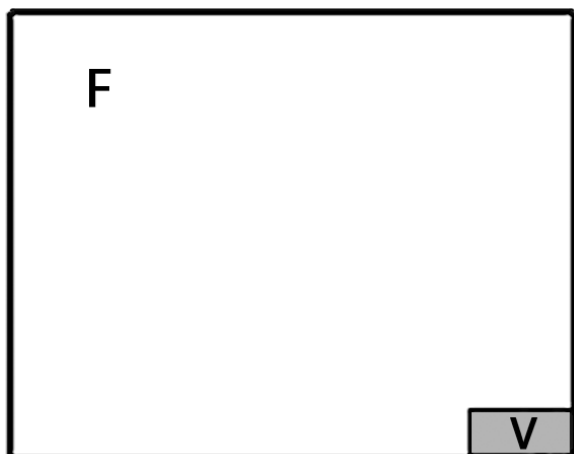
Figure 4 - Lattice for decision-making used in morphological analysis used after making PANN; F: logical state false (it is interpreted as referece not similar); V: logical state true (it is interpreted as reference similar).

TABLE III
Limits of the areas of lattice for decision-making used in morphological analysis.

Limits of the areas of lattice	
True (V)	Fe >= 0.75 Ce <= 0.15
False (F)	Fe < 0.75 Ce > 0.15

Ce: Contrary evidence; Fe: Favorable evidence.

(Figure 5) of 20 elements, with maximum amplitude of 11 points (00-10) and hypothetical values (Table IV). This example is intended to explain in detail and didactic, the pattern recognition process using this methodology.

The Sample is the sequence that will be submitted for recognition to PANN. The Reference 1 and Reference 2 are two sequences that were previously stored in the database control (normal range).

To be able to process the PANN sequence analysis, it is necessary that each entry of PANN is properly calculated. These input variables are called expert systems; they are related to specific routines for extracting information.

The first expert system is responsible for quantifying the known mutations (Table V) by comparing the sample and reference, according to the formula presented earlier.
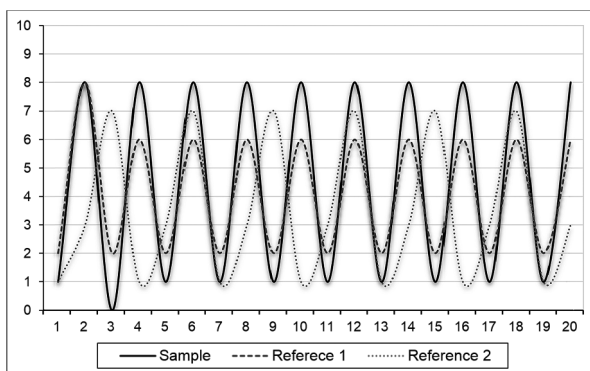
Figure 5 - Visual example of sequences used in morphological analysis.

TABLE IV
Values of the example of the process of recognition of sequences.

Sample	Reference 1	Reference 2
1	2	1
8	8	3
0	2	7
8	6	1
1	2	3
8	6	7
1	2	1
8	6	3
1	2	7
8	6	1
1	2	3
8	6	7
1	2	1
8	6	3
1	2	7
8	6	1
1	2	3
8	6	7
1	2	1
8	6	3

TABLE V
Verification the quantity of known mutations.

Sequence	Sample x Reference 1	Sample x Reference 2
Knows mutations	18	16
Number of mutations normalized by the number of elements of the sequence. Expert system 1 (E_1):	0.90	0.8

Performing the comparison between sequences, we have:

The second expert system is responsible for quantifying the unknown mutations (Table VI) by comparing the sample and reference, according to the formula presented earlier.

TABLE VI
Verification the number of unknown mutations.

Sequence	Sample x Reference 1	Sample x Reference 2
Unknows mutations	1	1
Number of mutations normalized by the number of elements of the sequence. Expert system 2 (E_2):	0.05	0.05

Performing a comparison between the sequences, we have:

The third expert system is responsible for quantifying the size of the changes wrought by

comparing the sample and reference, according to the formula presented earlier (Table VII and VIII).

Performing the comparison between sequences, we have:

The following are the values of each expert system that will be used as input values for the PANN (Table IX below).

In practical terms, one can say that, by analyzing the sequences of these characteristics, we are doing with the PANN "sees" the profile of each sample sequence. Combining such information, as has the sequence is similar to one another.

This procedure is always performed by comparing a sample of all references in the database. It is voted the most similar to the reference sample, which had the highest and lowest mi resulting lambda resulting from the processing of PANN.

After the analysis of expert systems and PANN, the values of favorable evidence (the

TABLE VII
Comparison between the sample and reference 1.

Sample	Reference 1	Difference in module	Normalization of the maximum amplitude difference
1	2	1	0.04
8	8	0	0.00
0	2	2	0.08
8	6	2	0.08
1	2	1	0.04
8	6	2	0.08
1	2	1	0.04
8	6	2	0.08
1	2	1	0.04
8	6	2	0.08
1	2	1	0.04
8	6	2	0.08
1	2	1	0.04
8	6	2	0.08
1	2	1	0.04
8	6	2	0.08
1	2	1	0.04
8	6	2	0.08
1	2	1	0.04
8	6	2	0.08
Sum of normalized differences: 1.16			
Expert system 3 (E_3): Normalized by the total of element: 0.06			

TABLE VIII
Comparison between the sample and reference2.

Sample	Reference 2	Difference in module	Normalization of the maximum amplitude difference
1	1	0	0.00
8	3	5	0.19
0	7	7	0.27
8	1	7	0.27
1	3	2	0.08
8	7	1	0.04
1	1	0	0.00
8	3	5	0.19
1	7	6	0.23
8	1	7	0.27
1	3	2	0.08
8	7	1	0.04
1	1	0	0.00
8	3	5	0.19
1	7	6	0.23
8	1	7	0.27
1	3	2	0.08
8	7	1	0.04
1	1	0	0.00
8	3	5	0.19
Sum of normalized differences: 2.65			
Expert system 3 (E_3): Normalized by the total of element: 0.06			

TABLE IX
Summary of results of expert systems.

Case	Expert system 1 (E_1)	Expert system 2 (E_2)	Expert system 3 (E_3)
Sample X Reference 1	0.90	0.05	0.06
Sample X Reference 2	0.80	0.05	0.13

highest resultant m) and contrary evidence (the smaller the resulting l) are submitted to the lattice of logic states which will set its output logic state, i.e. the similarity between the sequences is true or not.

DATA

Three hundred and eight samples from region sequences of the protease enzyme of the pol gene (polymerase) of HIV-1 subtype F, B and BF recombinants, with different therapeutic regimen, including protease inhibitors and reverse transcriptase, obtained from the database regarding HIV resistance to antiretroviral drugs and Stanford University, California "*Stanford University HIV Drug Resistance Database*" (HIVdb 2009). The reference sequences (consensus) used for analysis were obtained from the database of HIV sequences from the Los Alamos National Laboratory, USA (HIV DATABASES 2009).

RESULTS AND DISCUSSION

In the preliminary test carried out with three hundred and eight protease sequences for subtypes F, B and BF of HIV-1 database at Stanford, tested by the program, showed a high level of agreement (Coefficient Kappa 0.92) as can be seen in Table X.

In Table XI, we can see that the classification for subtypes F and Non-F showed a high level of agreement (sensitivity 92% and Specificity 100%).

For subtypes B and Non-B can also see a high rating (sensitivity 93% and specificity 100%) as can be seen Table XII.

TABLE X
General comparison of subtypes of HIV-1 sequences between Stanford and PANN.

		Stanford Database					
	Subtypes	A	B	C	D	F/BF	Total
PANN	A	0	1	0	0	5	6
	B	0	29	0	0	18	47
	C	0	0	0	0	0	0
	D	0	1	0	0	0	1
	F/BF	0	0	0	0	254	254
	Total	0	31	0	0	277	308

TABLE XI
Comparison of f and Non-f subtypes of HIV-1 sequences between Stanford and PANN.

		Stanford Database		
	Subtypes	F/BF	Non-F/BF	Total
PANN	F/BF	254	0	254
	Non-F/BF	23	31	54
	Total	277	31	308

TABLE XII
Comparison of B and Non-B subtypes of HIV-1 sequences between Stanford and PANN.

		Stanford Database		
	Subtypes	B	Non-B	Total
PANN	B	29	20	49
	Non-B	0	259	259
	Total	29	279	308

Several configurations were tested for analysis until it has obtained the best configuration of the architecture of artificial neural network paraconsistent, prevailing until the moment the configuration with the best sensitivity and specificity.

All analysis of the sequences were performed by means of double-blind trials (using control samples, not included in batteries of tests, i.e., the diagnostic validation was not released until the best configuration of PANN had been chosen, using as criterion the correlation between results and clinical diagnosis.

Comparing the clinical correlation obtained in this study with others in the literature, we can see a promising advantage over the levels of processing

methods. While studies use ANNs (Artificial Neural Networks Systems) Classic combined with other mathematical tools to arrive at a clinical correlation of 90%, the methodology of this study has a clinical correlation value using only one type of analysis.

The methodology of pattern recognition using morphological analysis showed to be effective, achieving recognize patterns of reference similar to patterns stored in the database, allowing quantifications and qualifications of the blood samples infected with HIV to be used by PANN in their process analysis of examination.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous referees providing useful comments to improve this version of the paper.

RESUMO

A elevada variabilidade do HIV-1, bem como, a ausência de mecanismos eficientes de reparo durante os estágios da replicação viral, contribuem para a rápida emergência de cepas de HIV-1 resistentes aos antirretrovirais. A pressão seletiva exercida pelas drogas, leva à fixação de mutações capazes de conferir graus variados de resistência. A presença dessas mutações constitui um dos fatores mais importantes na falha da resposta terapêutica aos medicamentos. Assim, é de fundamental importância compreender os padrões de resistência e os mecanismos a eles associados, possibilitando a escolha de um esquema terapêutico apropriado que considere a frequência e outras características das mutações. Utilizando a Rede Neural Artificial Paraconsistente, assentada na Lógica Paraconsistente Anotada Eτ que tem a capacidade de mensurar incertezas e inconsistências, obtivemos níveis de concordância acima de 90% quando comparado à metodologia proposta com a metodologia atual empregada para classificar os subtipos do HIV-1. Os resultados obtidos demonstram que a Rede Neural Artificial Paraconsistente pode servir como ferramenta promissora de análise.

REFERENCES

ABE JM. 1992. "Fundamentos da lógica anotada". Tese de Doutorado - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, Brasil, 135 p.

ABE JM. 2010. "Redes Neurais Artificiais Paraconsistentes e Análise de Distúrbio de Aprendizagem". Tese de Livre-Docência – Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brasil, 177 p.

CAMPOS DP, LISBOA CSV, MATZENBACHER LA, GRISZTEJN B, VELOSO VG, RIBEIRO SR, BRAGA EB AND JASHAR E. 2006. Banco de dados de indivíduos HIV positivos para fins de pesquisa clínica: elaboração e atualização. In: Conference on informatics in health (SIBS-2006), 10, Florianópolis, SC. Annals. Florianópolis: SBI, p. 31.

DA SILVA FILHO JI. 2001. Fundamentos Paraconsistentes. Publishing VillaPress, São Paulo, Brasil, 272 p.

DA SILVA FILHO JI AND ABE JM. 2001. "Fundamentos das Redes Neurais Paraconsistentes - Destacando Aplicações em Neurocomputação", Transl. (in Portuguese). Editora Arte & Ciência, 247 p.

DATASUS. 2009. Available: http://www2.datasus.gov.br/DATASUS/, accessed March 10, 2009.

DOS SANTOS PCC. 2010. "Banco de Dados Inteligente e Ferramentas Associadas de Sequências, Mutações e Resistências aos Antirretrovirais do Vírus HIV". Tese de Doutorado - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, Brasil, 129 p.

GOTLIEB MS, SCHROFF R, SCHANKER HM, WEISMAN JD, FAN PT, WOLF RA AND SAXON A. 1981. Pneumocystis carinii pneumonia and mucosa candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. New Engl J Med 305: 1425-1431.

GRANT AD AND COCK KMD. 2001. ABC of AIDS – HIV infection and AIDS in the developing world. BMJ 322: 1475-1478.

HIV DATABASES. 2009. Los Alamos National Laboratory HIV Sequence Database. Available: http://www.hiv.lanl.gov/content/sequence/HIV/, accessed February 15, 2009.

HIVDB. 2009. Stanford University HIV Drug Resistance Database. Available: http://hivdb.stanford.edu/, accessed February 15, 2009.

SISCEL-SISGENO. 2009. Available: http://www.portal.saude.gov.br/portal/, accessed March 12, 2009.

UNAIDS-AIDS. 2014. Epidemic Update. Available: http://www.unaids.org/, accessed January 10, 2015.

WHO - WORLD HEALTH ORGANIZATION. 2014. Progress Report on the Global Plan. Available: http://www.who.int/en/, accessed January 10, 2015.