

**UFRRJ**  
**INSTITUTO DE CIÊNCIAS EXATAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E**  
**COMPUTACIONAL**

**DISSERTAÇÃO**

Aplicação de Modelos ARTMAP na Predição de Resistência do HIV-1  
aos Inibidores de Protease.

**ALESSANDRO DOS SANTOS CALIN**

**2014**



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO**  
**INSTITUTO DE CIÊNCIAS EXATAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E**  
**COMPUTACIONAL**

**Aplicação de Modelos ARTMAP na Predição de Resistência do**  
**HIV-1 aos Inibidores de Protease.**

**ALESSANDRO DOS SANTOS CALIN**

*Sob a Orientação do Professor*

**Dr. Robson Mariano da Silva**

*e Co-orientação do Professor*

**Dr. Rafael Bernardo.**

Dissertação submetida como requisito parcial para obtenção do grau de Mestre em Modelagem Matemática e Computacional, no Curso de Pós-graduação em Modelagem Matemática e Computacional, Área de concentração em Inteligência Computacional e Otimização.

Seropédica, RJ  
Setembro de 2014

006.3

C154a

Calin, Alessandro dos Santos, 1981-

T

Aplicação de modelos ARTMAP na predição de resistência do HIV-1 aos inibidores de protease / Alessandro dos Santos Calin. - 2014.

74 f.: il.

Orientador: Robson Mariano da Silva.

Dissertação (mestrado) - Universidade Federal Rural do Rio de Janeiro, Curso de Pós-Graduação em Modelagem Matemática e Computacional, 2014.

Bibliografia: f. 70-74.

1. Redes neurais (Computação) - Teses.  
2. Infecções por HIV - Tratamento - Inovações tecnológicas - Teses. 3. Agentes antirretrovirais - Teses. 4. Inteligência computacional - Tese. I. Silva, Robson Mariano da, 1963- II. Universidade Federal Rural do Rio de Janeiro. Curso de Pós-Graduação em Modelagem Matemática e Computacional. III. Título.

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO  
INSTITUTO DE CIÊNCIAS EXATAS  
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E  
COMPUTACIONAL

ALESSANDRO DOS SANTOS CALIN

Dissertação submetida como requisito parcial para obtenção do grau de Mestre em Ciências, no Curso de Pós-graduação em Modelagem Matemática e Computacional, área de concentração em Inteligência Computacional e Otimização.

Dissertação aprovada em \_\_\_\_/\_\_\_\_/\_\_\_\_

---

Robson Mariano da Silva. Prof. Dr. UFRRJ - RJ  
(Orientador)

---

Francisco Bruno Souza Oliveira. Prof. Dr. UESC - BA

---

Angel Ramon Sanchez Delgado. Prof. Dr. UFRRJ - RJ

Dedico este trabalho aos meus pais Ismael e Noemi, meus irmãos Anderson e Franklin, e as minhas sobrinhas: Rayanne, Raissa e Brenda pelas palavras de carinho, compreensão e companheirismo para vencer todas as dificuldades nessa etapa importante da minha vida.

---

## Agradecimentos

---

Primeiramente tenho que registrar meus agradecimentos ao meu Deus, pois sem Ele nada sou, Ele é a justificativa por toda a minha força em desenvolver este trabalho. Obrigado meu Deus em me iluminar nos momentos difíceis e por me dá sabedoria na medida certa.

Aos meus pais e a toda minha família! A vocês, que estiveram ao meu lado nas horas que chorei e nas horas que sorri, nas horas que me lamentei e nas horas que de uma forma ou de outra demonstrei total alegria. Agradecer pelo sorriso diário, sem mágoas nem rancores, agradeço de peito aberto, de alma explosiva. Vocês fazem e farão sempre parte de minha história!

Ao Professor Dr. Robson Mariano da Silva, meu orientador, pela colaboração, paciência e seus conhecimentos repassados durante todo o desenvolvimento da pesquisa, além da grande amizade formada. “Se deres um peixe a um homem faminto, vais alimentá-lo por um dia. Se o ensinares a pescar, vais alimentá-lo toda a vida”. Muito obrigado grande mestre e amigo de todos os momentos!

Ao Programa de Pós-Graduação em Modelagem Matemática e Computacional da UFRRJ, em especial a Professora Dr. Priscila Machado Vieira de Lima pela sua dedicação no início desse curso.

Existem pessoas em nossas vidas que nos deixam felizes pelo simples fato de terem cruzado o nosso caminho. Agradeço aos meus amigos de turma, mas não posso deixar de registrar em especial aos meus amigos: Fabio, Arnaldo, Eloã, Italo e Diogo.

Aos amigos Felipe, Denise e a secretaria do PPGMMC, Janaina Gama. Serão sempre inesquecíveis os diversos momentos que me apoiaram nessa jornada.

As minhas Tias Ruth, Salete e Eliude pelo amor incondicional. A todos que me apoiaram direta e indiretamente. Meu muito obrigado!

## RESUMO

CALIN, Alessandro dos Santos. **Aplicação de Modelos ARTMAP na Predição de Resistência do HIV-1 aos Inibidores de Protease**. 2014. 74 f. Dissertação (Mestrado em Modelagem Matemática e Computacional). Instituto de Ciências Exatas, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2014.

Durante a última década as Terapias Antirretrovirais (TARV) reduziram a mortalidade em pacientes portadores do HIV-1. No entanto, esta diminuição não conseguiu impedir totalmente o surgimento de novas formas virais resistentes, causadas principalmente pela elevada taxa mutacional do HIV-1. O desenvolvimento de resistência do HIV-1 aos antirretrovirais (ARV) é um fator limitante para o sucesso da TARV. Pacientes com deficiência virológica, normalmente, necessitam de alterações em seus esquemas antirretrovirais, desta forma, técnicas que possam apoiar na previsão de resistência aos ARV possibilitam minimizar as falhas terapêuticas e, em consequência, evitam o aumento da carga viral dos pacientes. Em virtude desses fatos, desenvolvemos o presente estudo com o objetivo de elaborar dois modelos computacionais: um baseado em Redes Neurais ArtMap e outro em Redes Neurais Fuzzy ArtMap. De modo a investigar a resistência na terapia antirretroviral do HIV-1 aos inibidores de protease (IPs) para os subtipos B e C. Para aplicar a metodologia utilizamos os dados obtidos do Laboratório de Virologia Molecular da Universidade Federal do Rio de Janeiro (UFRJ, Brasil) e de uma base pública cedida pela Universidade de Stanford (SU, Estados Unidos). Antes de dividirmos o conjunto de teste (30%) e treino (70%), foi feito um pré-processamento analisando a frequência da ocorrência de mutações em todas as posições da protease e verificou-se que haviam posições com taxas de mutação muito baixas, as quais não teriam relevância para a categorização das amostras presentes na base geral de dados, e assim, foi considerado para o modelo apenas as posições com um total igual ou maior que 7,5% de mutações. Os resultados obtidos foram significativos em ambos os modelos, principalmente nos agrupamentos aos pacientes resistentes ao Lopinavir, ao Nelfinavir e aos pacientes não resistentes aos antirretrovirais. A análise dos resultados foram feitas usando o conceito de especificidade, sensibilidade e acurácia.

Palavras chaves: HIV-1, Redes Neurais Artificiais (RNAs), Terapia Antirretroviral.

## ABSTRACT

CALIN, Alessandro dos Santos. **Model Application ARTMAP in the Prediction of resistance to HIV-1 protease inhibitors.** 2014. 74 f. Dissertation (Master of Science in Mathematical and Computational Modelling). Instituto de Ciências Exatas, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2014.

Through the last decade, the Antiretroviral Therapies (ART) reduced the mortality of patients with HIV-1. Nevertheless, this decrease could not completely prevent the emergence of new resistant viral forms, mainly caused by the high mutational rate of HIV-1. The development of resistance of HIV-1 to antiretroviral (ARV) is a limiting factor for the success of ART. Because, in addition to not respond adequately to treatment, patients with resistant virus can transmit these mutant viruses, representing a serious public health problem. Patients with virological failure, usually, require changes to their antiretroviral regimens, thus, techniques that can assist in the prediction of resistance to antiretroviral therapies allow to minimize the gaps and, consequently, prevents the increase of the viral load of patients. In view of these facts, we developed this study with the goal of developing two computational models: one based on ArtMap Neural Networks and other Fuzzy ArtMap Neural Networks. In order to investigate resistance to antiretroviral therapy for HIV-1 protease inhibitors (IPs) for subtypes B and C. To apply the methodology we use the data obtained from the Laboratory of Molecular Virology at Universidade Federal do Rio de Janeiro (UFRJ, Brazil) and a public base courtesy of Stanford University (SU, United States). Prior to divide the test set (30%) and training (70%), was made a preprocessing analyzing the frequency of occurrence of mutations in all positions of the protease. And it was found that some positions had very low mutation rates, which would have significance for the categorization of samples in the general database, and thus, was considered for the model only the positions of a total equal to or greater than 7.5% of mutations. The results were significant in both models, especially in groupings to patients resistant to Lopinavir, Nelfinavir and the non-resistant antiretroviral patients. The results were made using the concept of specificity, sensitivity and accuracy.

Keywords: HIV-1, Artificial Neural Networks (ANN), Antiretroviral Therapy



---

## Lista de Figuras

---

2.1.	Representação esquemática da classificação do HIV .....	08
2.2.	Representação esquemática da estrutura do genômica do HIV-1 .....	09
2.3.	Ciclo de vida resumido da replicação viral do HIV-1 .....	11
2.4.	Classificação dos Antirretrovirais .....	14
2.5.	Mutações no gene da PR associadas com a resistência aos IPs .....	18
2.6.	Família ART .....	23
2.7.	Arquitetura da Redes Neurais ArtMap .....	25
2.8.	Etapas de um sistema nebuloso .....	28
2.9.	Arquitetura da Redes Neurais Fuzzy ArtMap .....	29
3.1.	Resumo das Metodologias da Pesquisa .....	31
3.2.	Fluxograma da metodologia proposta no modelo computacional ArtMap com a base de dados do laboratório de virologia molécula da UFRJ .....	32
3.3.	Fluxograma da metodologia proposta no modelo computacional Fuzzy ArtMap com a base de dados do laboratório de virologia molécula da UFRJ .....	33
3.4.	Fluxograma da metodologia proposta no modelo computacional Fuzzy ArtMap com a base de dados híbrida .....	34
3.5.	Distribuição da Amostra da Base de Dados do Subtipo B - Base de Dados RENAGENO .....	36
3.6.	Distribuição da Amostra da Base de Dados do Subtipo B - Base de Dados RENAGENO .....	37
3.7.	Distribuição da Amostra da Base de Dados Híbrida do Subtipo B .....	39
3.8.	Distribuição da Amostra da Base de Dados Híbrida do Subtipo C .....	40
3.9.	Programa de validação de Resistência de Stanford .....	43

---

## Lista de Tabelas

---

2.1.	Classificação dos aminoácidos .....	06
3.1.	Posições selecionadas com a regra de cortes na Base de Dados .....	37
3.2.	Posições selecionadas com a regra de cortes na Base de Dados .....	38
3.3.	Posições selecionadas com a regra de cortes na Base de Dados Híbrida	38
3.4.	Posições selecionadas com a regra de cortes na Base de Dados Híbrida	39
3.5.	Escala de hidrofobicidade com Valores atribuídos a cada aminoácido e sua respectiva classificação quanto ao valor de sua hidrofobicidade, pela escala de Kyte e Doolittle.....	41
3.6.	Escala de hidrofobicidade normalizada .....	42
4.1.	Distribuição da Amostra da Base B .....	46
4.2.	Agrupamento com o Parâmetro de Vigilância 0,5 – Rede ArtMap Subtipo B .....	47
4.3.	Análise dos Resultados do Agrupamento 1 (resistente ao LPV/r) .....	47
4.4.	Análise dos Resultados do Agrupamento 2 (não resistentes) .....	48
4.5.	Análise dos Resultados do Agrupamento 3 (resistente ao NFV) .....	48
4.6.	Análise dos Resultados do Agrupamento 4 (resistente ao APV) .....	48
4.7.	Análise dos Resultados do Agrupamento 5 (resistente ao IDV) .....	48
4.8.	Análise dos Resultados do Agrupamento 6 (resistente ao ATV) .....	48
4.9.	Análise dos Resultados do Agrupamento 7 (resistente ao SQV) .....	49
4.10.	Distribuição da Amostra da Base C .....	49
4.11.	Agrupamento com o Parâmetro de Vigilância 0,5 – Rede ArtMap Subtipo C .....	50
4.12.	Análise dos Resultados do Agrupamento 1 (não resistentes) .....	50
4.13.	Análise dos Resultados do Agrupamento 2 (resistente LPV/r) .....	50
4.14.	Análise dos Resultados do Agrupamento 3 (resistente ao IDV) .....	51
4.15.	Análise dos Resultados do Agrupamento 4 (resistente ao NFV) .....	51
4.16.	Análise dos Resultados do Agrupamento 5 (resistente ao ATV) .....	51
4.17.	Análise dos Resultados do Agrupamento 6 (resistente ao APV) .....	51
4.18.	Análise dos Resultados do Agrupamento 7 (resistente ao SQV) .....	51
4.19.	Agrupamento com o Parâmetro de Vigilância 0,3 – Rede Fuzzy ArtMap Subtipo B .....	52

4.20. Análise dos Resultados do Agrupamento 1 (não resistente) .....	53
4.21. Análise dos Resultados do Agrupamento 2 (resistente ao ATV) .....	53
4.22. Análise dos Resultados do Agrupamento 3 (resistente ao NFV) .....	53
4.23. Análise dos Resultados do Agrupamento 4 (resistente ao IDV) .....	53
4.24. Análise dos Resultados do Agrupamento 5 (resistente ao LPV/r) .....	54
4.25. Análise dos Resultados do Agrupamento 6 (resistente ao SQV) .....	54
4.26. Agrupamento com o Parâmetro de Vigilância 0,4 – Rede Fuzzy ArtMap Subtipo C .....	54
4.27. Análise dos Resultados do Agrupamento 1 (resistente ao ATV) .....	55
4.28. Análise dos Resultados do Agrupamento 2 (resistente ao LPV/r) .....	55
4.29. Análise dos Resultados do Agrupamento 3 (resistente ao NFV) .....	55
4.30. Análise dos Resultados do Agrupamento 4 (não resistentes) .....	55
4.31. Análise dos Resultados do Agrupamento 5 (resistente ao APV) .....	55
4.32. Análise dos Resultados do Agrupamento 6 (resistente ao SQV) .....	56
4.33. Análise dos Resultados do Agrupamento 7 (resistente ao IDV) .....	56
4.34. Distribuição da Amostra da Base B - Base Híbrida .....	56
4.35. Agrupamento com o Parâmetro de Vigilância 0,4 – Fuzzy ArtMap Subtipo B – Base Híbrida .....	56
4.36. Análise dos Resultados do Agrupamento 1 (não resistentes) .....	57
4.37. Análise dos Resultados do Agrupamento 2 (resistente ao IDV) .....	58
4.38. Análise dos Resultados do Agrupamento 3 (resistente ao NFV) .....	58
4.39. Análise dos Resultados do Agrupamento 4 (resistente ao LPV) .....	58
4.40. Análise dos Resultados do Agrupamento 5 (resistente ao ATV) .....	58
4.41. Análise dos Resultados do Agrupamento 6 (resistente ao SQV) .....	58
4.42. Análise dos Resultados do Agrupamento 7 (resistente ao APV) .....	59
4.43. Distribuição da Amostra da Base C - Base Híbrida .....	59
4.44. Agrupamento com o Parâmetro de Vigilância 0,4 – Fuzzy ArtMap Subtipo C – Base Híbrida .....	60
4.45. Análise dos Resultados do Agrupamento 1 (resistente ao NFV) .....	60
4.46. Análise dos Resultados do Agrupamento 2 (resistente ao IDV) .....	61
4.47. Análise dos Resultados do Agrupamento 3 (não resistentes) .....	61
4.48. Análise dos Resultados do Agrupamento 4 (resistente ao APV) .....	61
4.49. Análise dos Resultados do Agrupamento 5 (resistente ao LPV/r) .....	61

4.50.	Análise dos Resultados do Agrupamento 6 (resistente ao ATV) .....	61
4.51.	Análise dos Resultados do Agrupamento 7 (resistente ao SQV) .....	62
5.1.	Análise dos Resultados da Base de Dados RENAGENO – Subtipo B .....	64
5.2.	Análise dos Resultados da Base de Dados RENAGENO – Subtipo C .....	65
5.3.	Análise dos Resultados da Base de Dados Híbrida – RNA Fuzzy ArtMap	66

---

## Lista de Abreviaturas

---

AIDS	Síndrome da imunodeficiência adquirida (do inglês, <i>Acquired Immunodeficiency Syndrome</i> )
APV	Amprenavir
ARV	Antirretroviral
ATV	Atazanavir
AZT	Zidovudina
cDNA	DNA complementar
CRF	Forma recombinante circulante
CV	Carga viral
DNA	Ácido desoxirribonucleico
DST	Doenças sexualmente transmissíveis
HAART	Terapia antirretroviral altamente eficaz (do inglês, <i>Highly Active Antiretroviral Therapy</i> )
HIV	Vírus da imunodeficiência humana (do inglês, <i>Human Immunodeficiency Virus</i> )
HIV-1	Vírus da imunodeficiência humana tipo 1
ICTV	Do inglês, <i>International Committee on Taxonomy Viruses</i>
IDV	Indinavir
IN	Integrase
IP	Inibidor de protease
IPs	Inibidores de protease
LPV	Lopinavir
mRNA	RNA mensageiro
MS	Ministério da Saúde
NFV	Nelfinavir
PR	Protease
r	Ritonavir
RNA	Ácido ribonucleico
RTV	Ritonavir

SOM	redes neurais auto-organizáveis (do inglês, <i>Self-Organized-Maps</i> )
SQV	Saquinavir

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	<b>01</b>
<b>1.1. Objetivos</b> .....	<b>04</b>
1.1.1. Objetivo Geral .....	04
1.1.2. Objetivos Específicos .....	04
<b>2. REFERENCIAL TEÓRICO</b> .....	<b>04</b>
<b>2.1. Conceitos Biológicos</b> .....	<b>05</b>
<b>2.2. HIV-1</b> .....	<b>07</b>
2.2.1. Estrutura do HIV-1 .....	08
2.2.2. Ciclo de Replicação do HIV-1 .....	10
<b>2.3. Terapia Antirretroviral</b> .....	<b>12</b>
<b>2.4. Resistência as Drogas</b> .....	<b>15</b>
2.4.1. Resistência aos Inibidores de Protease .....	17
<b>2.5. Redes Neurais Artificiais</b> .....	<b>18</b>
2.5.1. Redes Neurais ART .....	20
2.5.2. Redes Neurais ArtMap .....	24
2.5.3. Sistema Fuzzy .....	26
2.5.4. Redes Neurais Fuzzy ArtMap .....	28
<b>3. MATERIAIS E MÉTODOS</b> .....	<b>31</b>
<b>3.1. Metodologia</b> .....	<b>31</b>
<b>3.2. Base de Dados</b> .....	<b>35</b>
3.2.1. Regra de Corte na Base de Dados .....	36
3.2.2. Base de Dados RENAGENO – Com o Corte e Posições Selecionadas .....	36
3.2.2.1. Subtipo B .....	36
3.2.2.2. Subtipo C .....	37
3.2.3. Base de Dados Híbrida – Com o Corte e Posições Selecionadas .....	38
3.2.3.1. Subtipo B .....	38
3.2.3.2. Subtipo C .....	39
<b>3.4. Codificação da Base de Dados</b> .....	<b>40</b>
<b>3.5. Validação das Resistências</b> .....	<b>43</b>

<b>3.6. Aplicação do Modelo .....</b>	<b>43</b>
<b>4. RESULTADOS .....</b>	<b>45</b>
<b>4.1. Resultados da Base de Dados RENAGENO .....</b>	<b>46</b>
4.1.1. Redes ArtMap – Subtipo B .....	46
4.1.2. Redes ArtMap – Subtipo C .....	49
4.1.3. Redes Fuzzy ArtMap – Subtipo B .....	52
4.1.4. Redes Fuzzy ArtMap – Subtipo C .....	54
<b>4.2. Resultados da Base de Dados Híbrida .....</b>	<b>56</b>
4.2.1. Redes Fuzzy ArtMap – Subtipo B .....	56
4.2.2. Redes Fuzzy ArtMap – Subtipo C .....	59
<b>5. DISCUSSÃO DOS RESULTADOS .....</b>	<b>63</b>
<b>5.1. 1ª Etapa – Discussão dos Resultados da Base de Dados     RENAGENO - Subtipo B .....</b>	<b>64</b>
<b>5.2. 2ª Etapa – Discussão dos Resultados da Base de Dados     RENAGENO - Subtipo C .....</b>	<b>65</b>
<b>5.3. 3ª Etapa – Discussão dos Resultados da Base de Dados Híbrida     .....</b>	<b>66</b>
<b>6. CONCLUSÕES.....</b>	<b>68</b>
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>70</b>



---

# Capítulo 1

## INTRODUÇÃO

---

De acordo relatório global do UNAIDS 2014, das 35 milhões de pessoas portadoras do HIV-1 (Human Immunodeficiency Virus), aproximadamente, 54,3% não sabem que têm o vírus. Esses indivíduos quando descobrem que possuem o HIV-1, na busca de aumentar a qualidade e o tempo de vida, passam a utilizar os antirretrovirais para impedir a multiplicação do vírus no organismo. Na África subsaariana, quase 90% das pessoas portadoras do HIV-1 estão em tratamento com antirretrovirais. Essas pesquisas mostram que, na mesma região, 76% dos indivíduos com acesso aos antirretrovirais alcançaram a supressão do vírus – o que significa que a possibilidade de transmissão para seus parceiros sexuais é pouco provável. Em 2013, mais 2,3 milhões de pessoas passaram a ter acesso aos medicamentos, elevando o número global de indivíduos recebendo antirretrovirais para cerca de 13 milhões.

No que se refere ao Brasil, nos últimos 10 anos, a taxa de detecção de AIDS sofreu uma elevação de cerca de 2%. Em 2012, foram declarados 11.896 óbitos por AIDS no Brasil, que corresponde a um coeficiente de mortalidade de 5,5 por 100.000 habitantes, conforme o Boletim Epidemiológico de AIDS e DST 2013<sup>1</sup>.

Estimativas realizadas pelo Departamento de DST, AIDS e Hepatites Virais - Secretaria de Vigilância em Saúde do Ministério da Saúde, em 2014, aproximadamente 718 mil pessoas vivem com HIV-1/AIDS no Brasil, onde dentre esses existem um número de indivíduos que não são conhecedores de ter a doença, além disso, mostra que a taxa de prevalência da infecção pelo HIV-1 na população jovem apresenta tendência de aumento. Os dados acumulados até junho de 2013 mostram que no Brasil, temos um total de 686.478 casos de HIV-1/AIDS, dos quais 445.197 (64,9%) são do sexo masculino e 241.223 (35,1%) do sexo feminino (Boletim Epidemiológico de AIDS e DST 2014).

No intuito de minimizar a mortalidade e a morbidade, desde de 1991 o Governo Brasileiro, adota uma política que visa garantir o acesso universal à terapia

---

<sup>1</sup> Documento com dados epidemiológicos de DST, HIV-1/AIDS acumulados até junho de 2012

com antirretrovirais (ARV) para indivíduos portadores do HIV-1, segundo critérios definidos por comitês técnicos assessores.

No mundo inteiro, todos os profissionais que pesquisam HIV-1 lidam com um grande desafio: cresce a resistência do HIV-1 às drogas da terapia antirretroviral (TARV). Essa problemática traz como consequência a falha do tratamento, o que exige resgate terapêutico. A enorme variabilidade genética e antigênica do HIV-1, resultantes as suas elevadas taxas de mutação, é uma das maiores problemáticas para que os pacientes infectados criem resistência aos medicamentos a longo prazo.

Essa falha terapêutica, ao longo do tempo acarreta na coexistência de diferentes mutações em um único vírus, levando a padrões complexos de mutações e consequentemente, a resistências cruzadas às drogas antirretrovirais.

Desta forma, testes para a resistência as drogas foram rapidamente incorporadas ao tratamento padrão dos pacientes portadores do HIV-1, sendo classificados em dois tipos de ensaios: fenotipagem e genotipagem. Os testes de resistência fenotípica fornecem medição mais direta do medicamento antirretroviral, porém é dispendiosa e demorada (SILVA, 2009). Os testes de resistência genotípica são mais simples e acessíveis, entretanto, a interpretação de resistência aos medicamentos, utilizando somente informações do genoma viral é complexa e, muitas vezes, exige análise técnica. Para facilitar e acelerar a interpretação dos resultados de genotipagem, métodos *in silico* podem ser aplicados para prever a resistência aos antirretrovirais em pacientes com falha terapêutica. (BONET *et al*, 2007).

Modelos computacionais com a aplicação de métodos estatísticos e algoritmos de aprendizagem de máquina (*Machine Learning*, ML - programas computacionais que melhoram o desempenho automaticamente por meio da experiência), vêm se destacando em pesquisas para o desenvolvimento de modelo de predição do fenótipo de resistência do HIV-1 às drogas, a partir do genótipo. Nesse contexto as redes neurais e outras técnicas de inteligência computacional podem ser úteis na identificação de novas mutações e combinações que causam a resistência aos medicamentos, assim como, na classificação de subtipos do HIV-1.

Dentre as técnicas de ML citadas na literatura, destaca-se a proposta de WANG e LARDER (2003) que utilizou as redes neurais artificiais para a previsão de resistência ao inibidor de protease lopinavir (LPV). Nesse estudo os autores desenvolveram dois modelos de redes neurais. No primeiro utilizou-se, 11 posições

de mutações da sequência da protease descrita na literatura que promovem resistência para o lopinavir; e o segundo foi baseado nas 28 posições de mutação do gene da protease resultante da análise de prevalência no conjunto de dados.

DRAGHICI e POTTER (2003) analisaram a predição da resistência a dois inibidores de protease (IPs): Indinavir e Saquinavir, avaliando por duas perspectivas. Primeiramente, foi construído um modelo baseado nas características estruturais do complexo protease do HIV-1 – droga com ação inibidora. Posteriormente, um classificador foi construído baseando-se na sequência genotípica com as mutações de resistência à droga. Em ambos os casos, as redes neurais auto-organizáveis (SOM) foram utilizadas para extrair características importantes e agrupar os padrões de entrada de forma não-supervisionada, seguido da identificação caracterizada pelo IC90, resultado do teste fenotípico que representa a quantidade de droga necessária para reduzir em 90% a replicação viral.

DEFORCE *et al.* (2007) propõem a aplicação de redes neurais bayesianas, objetivando visualizar as relações entre o tratamento, mutações de resistência e a presença de polimorfismos para os inibidores de protease indinavir (IDV), saquinavir (SQV) e nelfinavir (NFV). Os resultados obtidos permitiram identificar as posições de mutação 30N, 88S e 90M para o NFV, 90M para SQV e 82A/T para IDV como as principais mutações de resistência.

PASOMSUB E., SUKASEMC., SUNGKANUPARPH S. (2009) propuseram uma aplicação de redes neurais artificiais para previsão de resistência às drogas, avaliando e comparando com outros sistemas de interpretação.

Outro potencial teórico nessa pesquisa foi o trabalho de SILVA (2009) que teve como objetivo propor um modelo computacional híbrido baseado na utilização de algoritmos genéticos (AGs) e no classificador *Kernel* Discriminante de Fisher (KDF), intitulado (AG/KDF), para identificar possíveis novas mutações de resistência no gene da protease do genoma do HIV-1, bem como predizer a resistência em pacientes em falha terapêutica no Brasil, para os inibidores de protease (Saquinavir, Nelfinavir e Lopinavir).

SOUZA (2010) propõem uma rede neural sem peso capaz de categorizar os diferentes Subtipos do HIV-1 e também de identificar a existência de mutações de resistência a este tipo de droga. Os resultados obtidos nessa pesquisa mostraram que

as redes sem peso possuem excelente desempenho para o reconhecimento dos Subtipos.

Ainda em 2010, SANTOS, desenvolveu um banco de dados “inteligente” de sequência, mutações e resistências aos antirretrovirais empregados contra o HIV-1, baseado na utilização de redes neurais artificiais. Foram processados 358 pacientes pelo método proposto e analisadas estatisticamente com o conceito de sensibilidade e especificidade.

Outro embasamento é a proposta de FERNANDES (2010) que utilizou o SOM na prévia organização dos dados em vetores protótipos e posterior utilizou a técnica de k - médias na separação de sequências da protease do HIV-1 de indivíduos infectados em grupos naturais.

## **1.2 - Objetivos**

### 1.2.1 – Objetivo Geral

Elaborar modelos computacionais estruturado em Redes Neurais ArtMap e Fuzzy ArtMap de modo a investigar a resistência na terapia antirretroviral do HIV-1, aos inibidores de protease para os subtipos B e C.

### 1.2.2 – Objetivos Específicos

- Analisar o agrupamento de mutações nas posições da sequência da protease e a relação na resistência aos retrovirais;
- Avaliar a capacidade de uma rede neural ArtMap e Fuzzy ArtMap, com aplicação da hidrofobicidade, em classificar por clusters os inibidores de protease e os respectivos Subtipos B e C;
- Analisar as tendências de mutações em posições clássicas e avaliar o princípio da estabilidade e plasticidade das redes ART.

---

# Capítulo 2

## Fundamentação Teórica

---

### 2.1 – Conceitos Biológicos

As proteínas são as moléculas orgânicas mais abundantes e importantes nas células e são encontradas em todas as partes das células, uma vez que são fundamentais sob todos os aspectos da estrutura e função celulares. Existem muitas espécies diferentes de proteínas, cada uma especializada para uma função biológica diversa. Além disso, a maior parte da informação genética é expressa pelas proteínas.

As proteínas, por sua vez, são formadas por subunidades chamadas aminoácidos. Elas são formadas a partir de um grupo de 20 aminoácidos, que formam as proteínas através de ligações covalentes numa sequência linear. Como cada um destes 20 aminoácidos possui uma característica diferente, as propriedades de uma dada proteína dependem diretamente de quais aminoácidos estão presentes na sua estrutura.

Muitos dos aminoácidos receberam nomes triviais, dependendo de como foram descobertos. Eles também possuem um código de abreviação, conforme a tabela 2.1.

Quando um aminoácido se junta a outro através da ligação uma nova molécula é formada chamada peptídio. Se é formada por 2 aminoácidos, é chamada um dipeptídio; se por 3, um tripeptídio, e assim por diante. Os polímeros de aminoácidos são chamados de peptídios quando contém até cerca de 50 aminoácidos. A partir daí, estes polímeros são chamados de proteínas.

A vida depende da realização de inúmeras reações químicas que ocorrem no interior das células e também fora delas. As substâncias do grupo de proteínas que catalisam essas reações biológicas são chamadas de enzimas. O poder catalítico de uma enzima relaciona a velocidade das reações com a energia despendida para que elas aconteçam.

Tabela 2.1 - Classificação dos aminoácidos (Tutorial sobre estrutura de proteínas - Prof. Bayardo Torres:  
<http://www.iq.usp.br/wwwdocentes/bayardo/software/proteina/index.html>)

Nome	Símbolo	Abreviação
Glicina ou Glicocola	Gly, Gli	G
Alanina	Ala	A
Leucina	Leu	L
Valina	Val	V
Isoleucina	Ile	I
Prolina	Pro	P
Fenilalanina	Phe ou Fen	F
Serina	Ser	S
Treonina	Thr, The	T
Cisteína	Cys, Cis	C
Tirosina	Tyr, Tir	Y
Asparagina	Asn	N
Glutamina	Gln	Q
Aspartato ou Ácido aspártico	Asp	D
Glutamato ou Ácido glutâmico	Glu	E
Arginina	Arg	R
Lisina	Lys, Lis	K
Histidina	His	H
Triptofano	Trp, Tri	W
Metionina	Met	M

Muitas proteínas atuam como enzimas, que são catalisadores biológicos de reações químicas específicas. A Protease são enzimas que catalisam a hidrólise<sup>2</sup> das ligações peptídicas de outras proteínas, clivando-as em fragmentos de aminoácidos menores.

Proteases ocorrem naturalmente em todos os organismos e correspondem a 1-5% de seus conteúdos genéticos. Essas enzimas estão envolvidas numa grande variedade de reações metabólicas, da simples digestão de proteínas do alimento a cascatas altamente reguladas (por exemplo a da coagulação, o sistema complementar, as vias de apoptose, e a cascata ativadora da profenoloxidase nos invertebrados). (HEDSTRON, 2002).

O HIV-1 depende diretamente da protease no seu ciclo reprodutivo, pois algumas proteínas virais são codificadas em uma longa cadeia peptídica, sendo libertadas por proteases, só então assumindo sua conformação ideal e função. A inibição dela poderia ser útil no controle da doença. Por isso que nossa pesquisa visa prever possíveis falhas terapêuticas relacionadas a esses inibidores.

<sup>2</sup> Quebra de ligação covalente com participação de uma molécula de água.

## 2.2- HIV-1

A descoberta clínica do HIV ocorreu no início da década de 80 no Centro Norte-Americano de controle e prevenção de doenças (CDC). A observação do aumento de infecções oportunistas, principalmente em homens homossexuais, como a pneumonia por *Pneumocystis Carinii* associada a um conjunto padrão de sinais clínicos levou à classificação deste quadro como uma imunodeficiência característica da AIDS (RAMBAUT *et al*, 2004). O primeiro isolamento do vírus ocorreu em 1983 (GALLO *et al*, 1983, BARRÉ-SINOUSI *et al*, 1983).

Na literatura são descritos dois tipos de HIV, denominados HIV-1 e HIV-2. O vírus da imunodeficiência humana do tipo 1 (HIV-1) é classificado como um retrovírus pertencente à família *Retroviridae* do gênero *Lentivirus* e seu conteúdo genético está disposto em uma fita simples de RNA. Essa definição foi feita pelo ICTV (International Committee of Viral Taxonomy), através de normas internacionais baseadas principalmente nas propriedades estruturais do vírus (LEVY *et al*, 1984).

A classificação do HIV é feita por meio da análise filogenética de sequências nucleotídicas dos vírus. A classificação atual é hierárquica e consiste em tipos, grupos, Subtipos, sub-Subtipos e formas recombinantes (Figura 2.1).

O HIV-1 é subdividido em 4 grupos: grupo M (do inglês, *major* ou majoritário), grupo N (do inglês, *new* ou *non-M, non-O*, ou novo, não-M, não-O), grupo O (do inglês, *outlier*) o mais divergente dentre os grupos, e ainda o grupo P. A maioria das infecções ocorre com HIV-1 do grupo M, o qual é diferenciado em Subtipos (A, B, C, D, F, G, H, J e K). Os Subtipos A e F, por sua vez, são subdivididos em A1, A2, A3, A4 e A5, e em F1 e F2, respectivamente. Quando um indivíduo é portador de uma infecção mista, composta por dois ou mais vírus de linhagens (Subtipos) diferentes, pode ocorrer a transferência de material genético entre eles, dando origem às formas recombinantes (RF, do inglês *recombinant forms*).

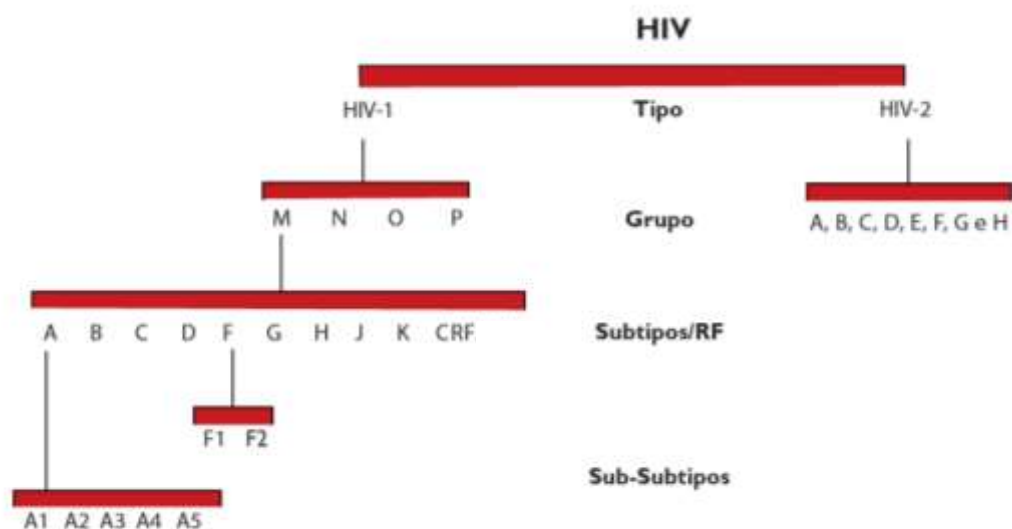


Figura 2.1 - Representação esquemática da classificação do HIV.  
 Fonte: Manual Técnico para o Diagnóstico da Infecção pelo HIV do Ministério da Saúde

O subtipo C do HIV-1 é o subtipo de maior ocorrência no mundo todo e é o principal subtipo encontrado em alguns países do Oriente Médio, da Ásia e do Sul e Leste Africano. Estas são regiões muito populosas e com altas taxas de infecção pelo HIV-1. Nos países centrais, Europa e Estados Unidos, o subtipo circulante predominante é o subtipo B.

No Brasil, os subtipos virais mais freqüentes são o B e F, seguido do subtipo C e das formas recombinantes, mas já foram relatados casos isolados do subtipo D (MORGADO *et al.*, 1998, SANCHES *et al.*, 2007). Existem diferenças regionais na prevalência dos subtipos virais no território nacional, sendo o subtipo B relatado em diversos estudos como o predominante no país, porém com a co-circulação de pelo menos mais um subtipo não-B por região geográfica (POTTS *et al.*, 1993; MORGADO *et al.*, 1994).

### 2.2.1 – Estrutura do HIV

Baseado no Manual Técnico para o Diagnóstico da Infecção pelo HIV do Ministério da Saúde, podemos definir que o genoma do HIV inclui três principais genes que codificam as proteínas estruturais e enzimas virais: *gag*, *env* e *pol*. A nomenclatura das proteínas virais utiliza a abreviação “gp” para glicoproteína ou “p”



para proteína, seguida de um número que indica o peso molecular em kilodaltons (kd). O gene *gag* codifica a p55, a partir da qual quatro proteínas estruturais do capsídeo são formadas: p6, p9, p17 e p24. O capsídeo que circunda o ácido nucleico viral contém p24, p6 e p9, enquanto a p17 se encontra em uma camada entre o núcleo proteico e o invólucro, denominada matriz proteica, a qual reveste a superfície interna da membrana viral.

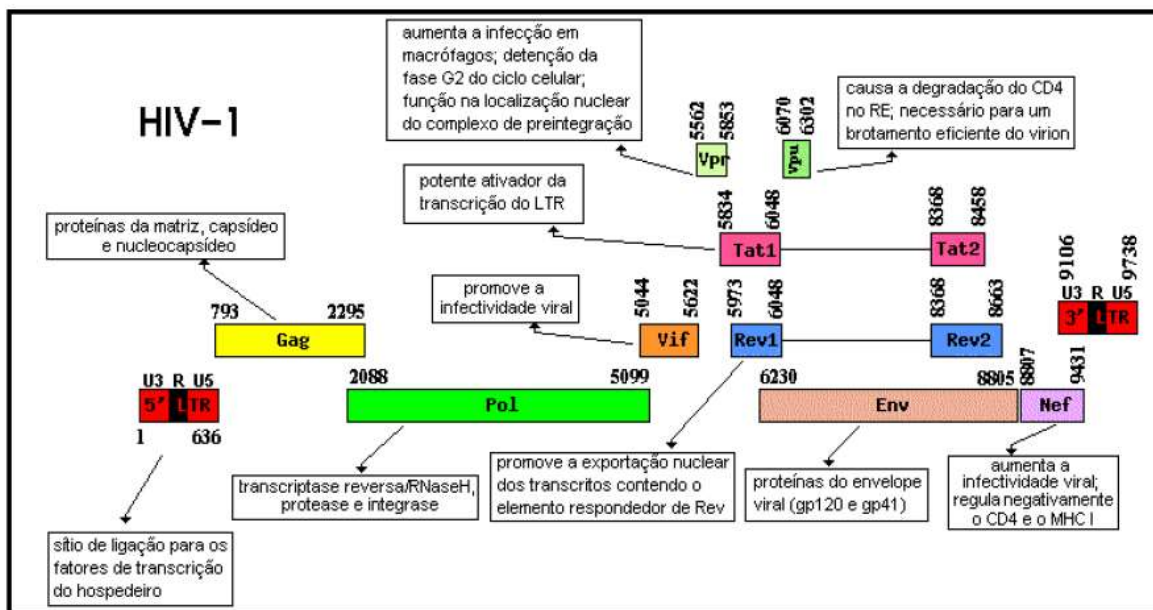


Figura 2.2: Representação esquemática da estrutura do genoma do HIV-1 (adaptado de www.nature.com) – Extraído de SILVA (2004)

O gene *env* codifica as glicoproteínas gp160, gp120, e gp41, que são encontradas no envelope viral. A gp160 é uma proteína precursora, que é clivada para formar a gp120 e gp41. A gp120 se projeta na superfície viral na forma trimérica, enquanto a gp41 é uma glicoproteína transmembrana e se associa à gp120. Ambas gp120 e gp41 estão envolvidas na fusão e ligação aos receptores de HIV nas células do hospedeiro.

O terceiro gene estrutural, *pol*, codifica as enzimas p66 e p51, que compõem a enzima transcriptase reversa (RT), necessária à replicação do HIV. Essas proteínas são subunidades da transcriptase reversa p31, ou integrase, a qual medeia a integração do DNA viral no genoma das células do hospedeiro, e a p10, uma protease que cliva precursores proteicos em unidades ativas menores. A proteína p66 também está envolvida na degradação do RNA original do HIV. Essas proteínas estão localizadas no núcleo, sendo associadas ao RNA do HIV.

Vários outros genes no genoma do HIV codificam produtos com função reguladora ou acessória. Embora esses produtos não sejam parte integrante da estrutura viral, eles atuam no controle da replicação viral e infectividade. O gene *tat* (transativador) codifica a p14, uma proteína reguladora que ativa a transcrição de genes provirais do HIV. O gene *rev* (que regula a expressão das proteínas do vírion) codifica a p19, uma proteína que transporta o RNA viral para a tradução no citoplasma. O gene *nef* codifica a p27, a qual apresenta múltiplas funções, incluindo a modificação da célula hospedeira para aumentar a replicação viral e torná-la menos suscetível a ser destruída pelo sistema imune do hospedeiro. O gene *vpu* (proteína viral “U”) codifica a p16, uma proteína com múltiplos papéis, incluindo a montagem de forma eficiente e brotamento dos vírions para fora das células hospedeiras infectadas e a promoção da morte da célula hospedeira. O gene *vpr* (proteína viral “R”) codifica a p15, que auxilia na integração do DNA do HIV no núcleo da célula hospedeira. O gene *vif* codifica a p23, que atua como um fator de infecciosidade viral, estabilizando o DNA HIV recém-sintetizado e facilitando o seu transporte para o núcleo.

Os principais componentes virais com utilidade diagnóstica incluem as proteínas do envelope viral (gp160, gp120 e gp41), as proteínas codificadas pelo gene *gag* (p55, p24 e p17) e as proteínas codificadas pelo gene *pol* (p66, p51, p31).

### 2.2.2 – Ciclo de Replicação do HIV-1

De acordo com Manual de Treinamento para TR DPP® - HIV-1/2 - SSP, o HIV é um vírus com genoma RNA, pertencente ao gênero retrovírus. Este gênero é caracterizado pela presença de uma enzima denominada transcriptase reversa que é responsável pela transcrição do RNA viral para uma cópia de DNA complementar (cDNA). No desenvolvimento do ciclo, o cDNA pode, então, se integrar ao genoma da célula hospedeira, geralmente as células do sistema fagocítico mononuclear (SMF) (<http://bio.fiocruz.br>). Então o vírus tem um núcleo central que consiste num revestimento de proteínas que encerra o RNA e as enzimas necessárias à replicação viral. O centro é rodeado por uma membrana externa, a partir da qual projeta as fitas que originam as proteínas (proteínas de envelope). Na figura 2.3 temos o ciclo de replicação do HIV.

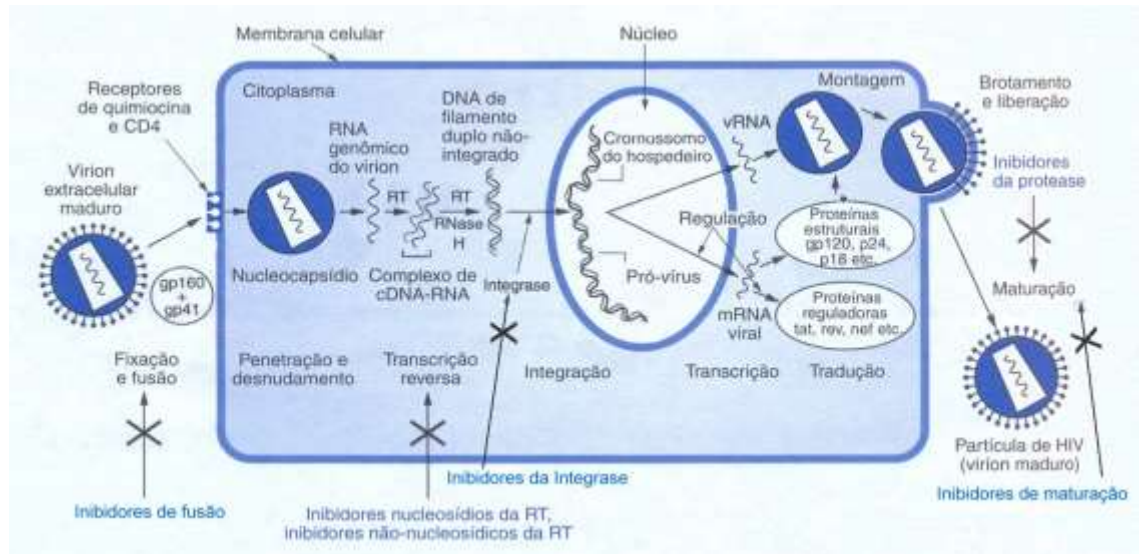


Figura 2.3– Ciclo de vida resumido da replicação viral do HIV-1  
 Fonte: Ministério da Saúde Secretaria de Vigilância em Saúde Departamento de DST, e AIDS

Como relatado acima as proteínas de envelope do vírus ligam-se à célula hospedeira, desencadeando a fusão das membranas celular e viral. Ao invadir a célula hospedeira, as enzimas virais e o RNA são libertados para o citoplasma, chegando à corrente sanguínea e atacando algumas células, sobretudo, os linfócitos T.

As proteínas de envelope prendem-se a dois receptores (CD4<sup>+</sup> e um co-receptor colocado próximo) na superfície destas células. A interação das proteínas de envelope com os dois receptores desencadeia a fusão da membrana do vírus com a da célula hospedeira (FERNANDES, 2010).

Dentro da célula a proteína central é dissolvida pelas enzimas da célula hospedeira, libertando RNA e enzimas virais (transcriptase reversa, integrase e proteases) para o citoplasma.

A transcriptase reversa permite que o código genético do HIV se junte ao DNA da célula hospedeira, produzindo uma réplica do RNA original em DNA de dupla cadeia. Para que os genes do HIV entrem no DNA da célula hospedeira, o RNA viral tem primeiro de ser convertido em DNA.

A Integrase faz a integração entre o DNA humano e a dupla cadeia de DNA viral. Dirigindo a síntese do mensageiro viral de RNA, que abandona o núcleo e entra no citoplasma, com instruções para produzir proteínas virais.

A célula infectada com HIV contém DNA viral bem como DNA humano e ambos são copiados para mRNA (RNA mensageiro). O mRNA deixa o núcleo e entra no citoplasma.

O mRNA derivado do DNA viral usa os mecanismos celulares da célula hospedeira para produzir proteínas de vírus<sup>3</sup>.

Em sequência ocorre o desenvolvimento viral, onde na membrana da célula, os vírus imaturos libertam-se da célula e entram na corrente sanguínea. As células CD4<sup>+</sup> são destruídas pelo HIV, o que pode resultar em imunodeficiência profunda. Os componentes virais juntam-se na membrana da célula e está começa a fabricar cápsulas que, eventualmente, saem da célula levando consigo todas as proteínas virais e o RNA necessário para formar partículas de vírus.

As novas partículas ainda estão imaturas quando entram na corrente sanguínea e, nesta fase, são incapazes de infectar outras células, tendo de passar por um processo de amadurecimento para se tornarem infecciosas. As células CD4<sup>+</sup> não sobrevivem, habitualmente, à invasão do HIV. Desintegram-se devido ao elevado número de vírus germinados ou porque o sistema imunológico do corpo reconhece as proteínas de envelope viral na membrana da célula e destrói as células danificadas. Como as células CD4<sup>+</sup> são uma parte essencial do sistema imunológico, a sua destruição pode provocar imunodeficiência profunda.

Depois do novo vírus deixar a célula inicia-se o processo de maturação viral onde outra enzima, a protease, corta a molécula que contém as proteínas centrais do HIV. As proteínas individuais libertadas são remontadas para formar um vírus estruturado e maduro. Este vírus pode agora infectar outras células.

### **2.3 – Terapia Antirretroviral**

Os antirretrovirais (ARV) são fármacos usados para o tratamento de infecções por retrovírus, principalmente o HIV-1. A terapia altamente eficaz com antirretrovirais é conhecida como TARV.

O advento dos ARVs trouxe significativas mudanças na história da epidemia de HIV-1/AIDS, criando condições para a redução da incidência de infecções oportunistas, da incidência da AIDS (ao diminuir a disseminação do vírus à medida

---

<sup>3</sup> Proteínas centrais, proteínas de envelope, enzimas e proteínas reguladoras essenciais para a replicação do HIV.

que reduz a carga viral), assim como, da mortalidade dela decorrente. Portanto, como consequência da introdução dos ARVs, tem-se verificado aumento da sobrevivência e melhora da qualidade de vida do paciente (Palella et al. 1998, Sepkowitz 2001).

A melhoria do acesso à terapia de antirretrovirais (TARV) tem contribuído para a redução da mortalidade relacionada ao HIV principalmente em países subdesenvolvidos.

Na década de 1990, o Brasil foi pioneiro, entre os países em desenvolvimento, a adotar uma política pública de acesso universal ao tratamento antirretroviral. O Ministério da Saúde adquire os medicamentos antirretrovirais de forma centralizada, respaldado na Lei nº 9.313/96. Os medicamentos para as infecções oportunistas são adquiridos pelos estados, de acordo com pactuações estabelecidas entre as três esferas de governo que compõem o SUS.

Aproximadamente 280.000 indivíduos HIV+ estão em tratamento gratuito na rede pública de saúde brasileira, o que representa um elevado percentual dos pacientes diagnosticados com HIV e elegíveis para o tratamento. Os medicamentos antirretrovirais estão disponíveis gratuitamente na rede pública de saúde, que supre também pacientes da rede privada.

O uso dos antirretrovirais é feito de acordo com o ciclo de replicação do HIV-1, o qual apresenta diversos eventos exclusivamente relacionados a componentes virais, que podem ser utilizados como alvos para intervenção quimioterápica. Os compostos disponíveis atualmente como fármacos anti-HIV estão divididos em quatro classes:

1. **Inibidores de fusão:** esses medicamentos impedem o vírus de se alojar nas células CD4 ao aderir a proteínas que ficam do lado de fora do vírus. Até agora apenas uma droga da categoria, o Fuzeon, chegou ao mercado.
2. **Inibidores Nucleosídeos da Transcriptase Reversa:** impedem o vírus de fazer cópias de seus próprios genes. Para isso, criam versões defeituosas dos nucleosídeos, unidades básicas dos genes.
3. **Inibidores Não-Nucleosídeos da Transcriptase Reversa:** também afetam o processo de replicação do HIV, ao aderir à enzima que controla o processo, conhecida como transcriptase reversa.

4. **Inibidores de Protease:** essas drogas atingem outra enzima envolvida no processo de multiplicação do vírus, a protease.

(PORTAL EDUCAÇÃO

<<http://www.portaleducacao.com.br/farmacia/artigos/7829/terapia-anti-retroviral#ixzz386Kl4hT2>> acessado em 01/06/2014)

Uma ampla variedade de combinações de drogas antirretrovirais encontra-se em uso atualmente. Estas várias combinações de medicamentos potentes tornou-se o padrão usual de tratamento para os indivíduos infectados pelo HIV-1. Na tabela 2.4 temos a lista de classificação dos inibidores de acordo com a enzima que atua.

Inibidores da Transcriptase Reversa Análogos de Nucleosídeos	Inibidores da Transcriptase Reversa Não-Análogos de Nucleosídeos	Inibidores de Protease (IP)
<ul style="list-style-type: none"><li>• Abacavir (ABC);</li><li>• Didanosina (ddl);</li><li>• Estavudina (d4T);</li><li>• Lamivudina (3TC);</li><li>• Zidovudina (AZT);</li><li>• Tenofovir (TDF).</li></ul>	<ul style="list-style-type: none"><li>• Efavirenz (EFZ);</li><li>• Nevirapina (NVP);</li><li>• Delavirdina.</li></ul>	<ul style="list-style-type: none"><li>• Fosamprenavir (FAPV);</li><li>• Atazanavir (ATV);</li><li>• Darunavir (DRV);</li><li>• Indinavir (IDV);</li><li>• Lopinavir (LPV);</li><li>• Nelfinavir (NFV);</li><li>• Ritonavir (RTV);</li><li>• Saquinavir (SQV).</li></ul>

Figura 2.4 - Classificação dos Antirretrovirais.

O primeiro medicamento antirretroviral utilizado para o tratamento da infecção pelo HIV-1 em adultos foi a Zidovudina (AZT), um análogo de inibidor de nucleosídeo da transcriptase reversa (INRT), o qual foi capaz de promover uma diminuição da mortalidade e da frequência de infecções oportunistas em pacientes com AIDS, prolongando a sobrevivência dos indivíduos, embora, somente, por alguns meses. No entanto, a eficácia da monoterapia prolongada com AZT foi questionada devido à rápida emergência de isolados virais que apresentavam resistência ao antirretroviral, dada à presença de uma ou mais mutações não silenciosas no gene da transcriptase reversa (FERNANDES, 2010).

Os resultados obtidos com a monoterapia baseada em outros INRTs disponíveis demonstraram-se inferiores aos alcançados com a terapia dupla, em relação a um aumento nas contagens absolutas e percentuais de células CD4+, na redução da replicação viral, das infecções oportunistas e da progressão da doença, além de uma eficácia reduzida quanto à manutenção destas alterações e a atenuação da emergência de resistência viral (RAMBAUT *et al.*, 2004). Além dos INRTs, outro grupo de inibidores da RT tem sido extensivamente estudado: os inibidores análogos não nucleosídeos da transcriptase reversa (*Non Nucleoside Reverse Transcriptase, INNRTs*).

A inclusão dos inibidores da enzima protease à terapia antirretroviral, na metade da década de 90, deu início a uma nova fase no tratamento de indivíduos infectados pelo HIV-1. Com o uso dos IP em combinação às outras drogas antirretrovirais, uma supressão viral muito mais profunda e sustentável e aumentos extremamente significativos nas contagens de células T CD4+ foram obtidos, em relação aos esquemas até então utilizados (FERNANDES, 2010). A primeira combinação de antirretrovirais amplamente utilizada no tratamento de indivíduos infectados pelo HIV-1 incluía um inibidor de protease e dois INRTIs.

## **2.4 – Resistência as Drogas**

A falha terapêutica de muitos pacientes portadores do HIV-1 está relacionada a não adesão ao tratamento e a replicação elevada de vírus mutantes resistentes aos alvos moleculares dos fármacos administrados na terapia (SHAFER, 2002). A resistência é simultaneamente causa e consequência da replicação do vírus na presença das drogas anti-HIV. A replicação residual, sob pressão seletiva das drogas antirretrovirais, cria um ambiente de seleção que, face ao surgimento de mutações associadas à resistência, numa pequena população viral, dá a esta variante viral uma vantagem seletiva. Esta pressão seletiva possibilita a resistência destas variantes, tornando-as predominantes (SILVA, 2009).

A suscetibilidade às drogas é maximizada pelo acúmulo de mutações de resistência, reduzindo progressivamente a potência dos componentes do esquema terapêutico. A replicação contínua do vírus na presença das drogas aumenta a resistência às mesmas. Sendo assim, esquemas terapêuticos impotentes, adesão

sub-ótima, absorção limitada e compartimentos corporais tratados de forma não efetiva podem permitir o surgimento de vírus resistentes; fato este que origina um ciclo vicioso de falha terapêutica, deixando o tratamento ainda mais difícil (RICHMAN *et al.*, 2003).

Frequentemente, na ausência da droga, os vírus mutantes apresentam um *fitness* reduzido, a ausência de uma pressão seletiva, motivada, por exemplo, pela interrupção do tratamento, levará a substituição dos vírus mutantes pelo vírus selvagem (*wild-type*), ou susceptível às drogas, de modo progressivo. No entanto, observa-se que estas populações virais persistem no plasma, mesmo que em populações virais minoritárias, chegando a não ser às vezes detectável nos testes de resistência (VANDAMME *et al.*, 2004).

A resistência aos ARVs subdivide-se em primária e secundária. Define-se como resistência primária a presença de mutações que conferem resistência aos antirretrovirais presentes no genoma viral em pacientes virgens de TARV.

Assim, programas para monitorar a prevalência de resistência primária em diferentes regiões são extremamente importantes para subsidiar os manuais de tratamento ARV, indicar o nível de sua eficiência e orientar os programas de prevenção do HIV-1 (Shafer & Schapiro 2008). A resistência primária emerge antes mesmo do uso da medicação e pode ocorrer pela transmissão de cepas resistentes ou mesmo pela geração espontânea de mutações.

A resistência secundária surge em decorrência da pressão seletiva exercida pelo próprio ARV. O surgimento de cepas resistentes selecionadas durante o tratamento pode comprometer o impacto positivo da terapia ARV, pois estas variantes escapam ao efeito dos ARV e podem ser transmitidas comprometendo a eficácia dos regimes terapêuticos de primeira linha (SHAFER, 2002).

Estudos demonstram que um significativo número de pacientes apresenta falência às combinações de drogas, em virtude do desenvolvimento da resistência secundária do HIV-1 (FERREIRA, 2011).



#### 2.4.1 – Resistência aos Inibidores de Protease

As mutações no gene da PR ocorrem por substituições de aminoácidos que emergem no sítio de ligação do substrato da enzima ou em sítios distantes. Estas alterações de aminoácidos modificam a afinidade e os pontos de contato da enzima com os IPs (CONDRA et al., 1995; KAPLAN et al., 1994). As mutações da PR nas posições 46, 54, 82 e 90 estão entre as mais frequentes nas mutações de resistência aos IPs (FERREIRA et al, 2011).

Dentre as mutações de resistência para IPs, as mutações primárias ou maiores ocorrem inicialmente, promovem redução substancial da sensibilidade ao IP e reduzem a capacidade replicativa viral. Geralmente ocorrem no sítio de ligação às drogas (HIRSCH et al., 1998). Mutações menores ou secundárias emergem posteriormente e isoladamente tem menor efeito na resistência, podem alterar a taxa de replicação viral e podem estar presentes como polimorfismos, principalmente em vírus de Subtipo não-B (JOHNSON et al., 2005). Normalmente, a resistência à IPs desenvolve-se gradualmente com o acúmulo de múltiplas mutações maiores e menores. O acúmulo de diversas mutações resulta em uma maior redução na susceptibilidade aos ARVs do que mutações únicas (SCHERRER et al., 2009).

Os inibidores de protease atuam no centro ativo enzimático e interagem com vários aminoácidos desta região de ligação, dependendo do tamanho de sua molécula. A forte ligação destes compostos ao substrato enzimático inibe a atividade da protease, impedindo a maturação de partículas virais infecciosas. Mutações relacionadas à resistência aos IPs têm sido relatadas para todos os compostos usados no tratamento de pacientes soropositivos. A distinção entre mutações primárias e secundárias é mantida para os IPs, uma vez que as mutações consideradas primárias apresentam maior efeito sobre a suscetibilidade às drogas quando comparadas às secundárias. As mutações secundárias ou acessórias não conferem, por si só, resistência aos IPs, mas contribuem para o grau de resistência quando associadas a outras mutações.

<b>Atazanavir +/- ritonavir</b>	L 10 I F V C	G 16 E R M I T Y	K 20 M I T Y	L 24 I I Q F Y	V 32 I I Q F Y	E 33 I I Q F Y	M 36 I L V	M 46 I L	G 48 V	I 50 L	F 53 L V M T A	I 54 L V M T A	D 60 E Y	I 62 V M V	I 64 L M V	A 71 V I S T A	G 73 C I S T A	V 82 A T F I	I 84 V V S	N 85 V V S	L 88 M	I 90 L M
<b>Darunavir/ ritonavir</b>	V 11 I			V 32 I F	L 33 I F			I 47 V	I 50 V	I 54 M L						T 74 P V	L 76 V	I 84 V	L 89 V			
<b>Fosamprenavir/ ritonavir</b>	L 10 F I R V			V 32 I				M 46 I V L	I 47 V	I 50 V L V M	I 54 L V M					G 73 S	L 76 V	V 82 A F S T	I 84 V	L 90 M		
<b>Indinavir/ ritonavir</b>	L 10 I R V	K 20 M R	L 24 I	Y 32 I	M 36 I			M 46 I L		I 54 V					A 71 V T	G 73 S	L 76 V	V 82 I A F T	I 84 V	L 90 M		
<b>Lopinavir/ ritonavir</b>	L 10 F I R V	K 20 M R	L 24 I	Y 32 I F	L 33 I F			M 46 I L A	I 47 V	I 50 V L V L A M T S	I 53 L V L A M T S	I 54 L V L A M T S			L 63 P	A 71 V T	G 73 S	L 76 V	V 82 A F T S	I 84 V	L 90 M	
<b>Nelfinavir</b>	L 10 F I		D 30 N		M 36 I			M 46 I L								A 71 V T	V 77 I A F T S	V 82 A F T S	I 84 V	N 85 D S	L 88 M	I 90 M
<b>Saquinavir/ ritonavir</b>	L 10 I R V		L 24 I					G 48 V		I 54 V L	I 62 V				A 71 V T	G 73 S	V 77 I A F T S	V 82 A F T S	I 84 V	L 90 M		
<b>Dipranavir/ ritonavir</b>	L 10 V V M R	I 13 M R	K 20 R		L 33 F G I	E 35 G I	M 36 I	K 43 T	M 46 L V	I 47 L V	I 54 A E M V	Q 58 E		H 69 K	T 74 P			V 82 L D V T	I 84 D V	L 90 M		

Figura 2.5 - Mutações no gene da PR associadas com a resistência aos IPs. A letra superior refere-se ao aminoácido do vírus selvagem sem mutação (wild type), a letra inferior ao aminoácido de substituição, a numeração em negrito indica as mutações primárias e a restante as mutações acessórias (adaptada de JOHNSON et al., 2009).  
 Extraída de: FERREIRA 2011

## 2.5 – Redes Neurais Artificiais

Redes Neurais Artificiais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência. São sistemas paralelos e distribuídos compostos por unidades de processamento simples que calculam determinadas funções matemáticas (Braga *et al* 2000).

Essa técnica computacional vem se destacando nos estudos referente ao reconhecimento de padrões, o qual estão se desenvolvendo rapidamente nos últimos anos com aplicações em diversas áreas científicas e tecnológicas. Outros exemplos importantes incluem: processamento de sinais, inteligência artificial e predição de séries temporais.

Os modelos computacionais baseados em redes neurais artificiais (RNAs) utilizam os conceitos derivados de pesquisas sobre a natureza do cérebro, utilizando para tarefas cognitivas, tais como aprendizado e otimização. O trabalho pioneiro, em redes neurais, é o de MCCULLOCH E PITTS - *A Logical Calculus of the Ideas Immanent in Nervous Activity* -1943.

Como já citamos uma das maiores vertentes das redes neurais é o reconhecimento de padrões que vem atualmente despertando o interesse dos pesquisadores. Com o princípio de fundamentar nosso trabalho analisamos na literatura pesquisas que utilizam redes neurais no reconhecimento de padrões de mutações genéticas relacionada à resistência aos fármacos no HIV-1. Dentre as pesquisas destacam-se os de DRAGHICI e POTTER (2003) que utilizaram redes neurais para categorizar amostras resistentes a Indinavir e Saquinavir e a pesquisa de WANG e LARDER (2003) que é focado na resistência em relação ao Lopinavir. Outro material que foca o assunto é o de SORIN (2000), o qual aplica Redes Neurais na problemática da resistência a drogas desenvolvida por pacientes infectados com o vírus HIV-1, utilizando as características da estrutura molecular mutante da protease Indinavir.

Todavia dentre todos os estudos correlatos destacamos como um importante contribuinte ao embasamento de aplicações de redes neurais o trabalho de SOUZA (2011) que aborda o problema da categorização do HIV-1, através do desenvolvimento de uma rede neural sem peso, capaz de discriminar amostras de acordo com o Subtipo do vírus e da existência ou não de resistência a algum medicamento.

Uma rede neural artificial é composta por várias unidades de processamento, onde essas unidades, geralmente são conectadas por canais de comunicação que estão associados a determinado peso. As unidades fazem operações apenas sobre seus dados locais, que são entradas recebidas pelas suas conexões. Um dos aspectos mais interessantes de uma rede neural é a habilidade de aprender e com

isso melhorar seu desempenho. Isso, geralmente, é feito através de um processo iterativo de ajustes aplicado a seus pesos, o qual chamamos de treinamento. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas.

O processo de aprendizado de uma rede neural é desenvolvido por algoritmos onde um conjunto de regras são definidas para a solução de um problema. Existem muitos tipos de algoritmos de aprendizado específicos para determinados modelos de redes neurais, estes se diferem entre si principalmente pelo modo como os pesos são modificados. O aprendizado supervisionado é utilizado quando um agente externo que indica à rede a resposta desejada para o padrão de entrada; e o aprendizado não supervisionado, quando não existe um agente externo indicando a resposta desejada para os padrões de entrada.

O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas. Existem diversos algoritmos para efetuar a aprendizagem de redes neurais artificiais (Regra de Hebb, Retropropagação, etc.). Assim como, vários tipos de RNA's (MLP - *Multi Layer Perceptron*, SOM - *Self Organizing Map*, ART, etc.) e suas variações sendo utilizadas nas mais variadas aplicações.

As aplicações de redes neurais são variadas de acordo com suas características de entrada, saída e treinamento. Alguns tipos de redes neurais são: perceptrons, perceptrons de multicamadas, redes sem peso, redes SOM, redes RBF e redes ART. Em nossa pesquisa iremos trabalhar com uma abordagem da família ART.

### 2.5.1 – Redes Neurais ART

A família de Modelos ART, que foi desenvolvida por Carpenter e Grossberg. A rede ART é uma rede com aprendizado baseado em semelhança. A principal característica dessas redes é sua habilidade para resolver o dilema da plasticidade-estabilidade, que possibilita que novos conhecimentos adquiridos pela rede não superponham conhecimentos prévios e que, também, sistemas de aprendizado possam preservar conhecimento previamente adquirido enquanto continua aprendendo novos conhecimentos. Desta forma esse tipo de rede mostra como um

sistema treinado pode permanecer adaptativo a cada nova informação e ainda permanecer estável em resposta a informações irrelevantes, preservando seu conhecimento já adquirido e ao mesmo tempo sendo flexível o suficiente para armazenar novas informações.

Esse dilema é resolvido com um mecanismo de vigilância que administra a inclusão de novas entradas em cada grupo. Uma regra de similaridade, que define onde agrupar um padrão é determinado por um grau de semelhança entre um padrão previamente armazenado.

Como no problema em estudo os dados podem se modificar ao longo do tempo, os padrões de entrada de uma RNA também se modificam para satisfazer as mudanças ocorridas. Então para que os padrões de uma rede não venham decair mediante as novas situações devemos adaptar novos padrões de entrada indefinidamente, o que está relacionado ao conceito de plasticidade. Todavia uma solução priori para solucionar esse problema seria treinar a rede com os novos padrões, porém isso acarretará em uma perda de informações aprendidas anteriormente. Então para preservar o conhecimento obtido inicialmente a nossa rede precisa ser também estável.

As primeiras redes ART utilizavam aprendizagem não supervisionada, em destaque as redes ART1, o qual foi desenvolvida para executar agrupamentos de padrões de valores lógicos binários, logo em seguida foi desenvolvida usando as mesmas metodologias da plasticidade e estabilidade, as redes ART2, para detectar regularidades nas sequências randômicas, porém essa arquitetura além de ter altos custos computacionais apresenta dificuldades na seleção dos parâmetros. O processo de classificação da rede ART consiste de quatro fases principais: reconhecimento, comparação, busca e treinamento (GROSSBERG, 1976), como a seguir:

- I. Reconhecimento: Cada neurônio do campo  $F1$  (camada de entrada) recebe sinais de baixo para cima do vetor de entrada e de cima para baixo do campo  $F2$  (camada de saída, que representa a categoria ativa). A atividade é calculada e o vetor resultante é comparado com todos os vetores de pesos já armazenados na rede (memória) para encontrar o que mais se assemelha ao padrão de entrada atual. O neurônio do campo  $F2$ , com maior valor de ativação, é selecionado como provável

categoria para armazenar o novo padrão. O neurônio selecionado envia ao campo  $F1$  seu protótipo;

- II. Comparação: Nesta fase acontece um mecanismo de reajuste, que é responsável por testar a similaridade entre o vetor de entrada e o vetor de comparação,  $z$ , que é o vetor resultante da atividade calculada no campo  $F1$  mediante o vetor protótipo dado pelo campo  $F2$  após a fase de reconhecimento. A maneira pela qual se define de que modo a comparação irá ocorrer é definida por um parâmetro chamado *vigilância* ( $\rho$ ), que determina se um padrão de entrada pode ser incluído em uma das categorias existentes. Se o valor da comparação for maior do que  $\rho$ , então o padrão de entrada é incluído na categoria ativa, caso contrário, a rede entra em fase de busca;
- III. Busca: Durante esta fase, a rede procura um novo neurônio do campo  $F2$  para representar o vetor de entrada atual. O neurônio da camada de saída rejeitado na fase de comparação anterior é desabilitado nesta seleção. O vetor de entrada é rerepresentado, e a rede entra novamente na fase de comparação, que termina com o teste no mecanismo de reajuste para determinar a similaridade entre o novo protótipo escolhido e o vetor de entrada atual. Este processo é repetido, desabilitando neurônios da camada de saída, até encontrar um neurônio de saída que melhor se assemelhe ao vetor de entrada corrente, dentro dos limites do parâmetro de vigilância  $\rho$ . Se nenhum neurônio de saída for encontrado, o vetor de entrada é então considerado de uma classe desconhecida, sendo alocado um neurônio de saída que não esteja associado ainda a nenhuma categoria para representá-lo;
- IV. Treinamento: O algoritmo de aprendizado da rede ART é não-supervisionado e pode ser ativado a qualquer momento, permitindo que a rede aprenda novos padrões continuamente.

Há dois tipos de treinamento para a rede neural ART: aprendizado rápido e aprendizado lento. No aprendizado rápido, os pesos de conexão são ajustados para seus valores ótimos em poucos ciclos, geralmente em apenas um ciclo de treinamento. No aprendizado lento, os pesos são ajustados lentamente em vários

ciclos de treinamento, possibilitando um ajuste melhor dos pesos da rede aos padrões de treinamento (GROSSBERG, 1976).

As redes ART são muito sensíveis a variações em seus parâmetros durante o treinamento. O parâmetro mais crítico é o parâmetro de vigilância ( $\rho$ ) que controla a resolução do processo de classificação. Se  $\rho$  assume um valor baixo, a rede permite que padrões não muito semelhantes sejam agrupados na mesma categoria de reconhecimento, criando poucas classes, e se for atribuído um valor alto (muito próximo a um), pequenas variações nos padrões de entrada levarão à criação de novas classes.

A família das redes ART está expressa na figura 2.6.

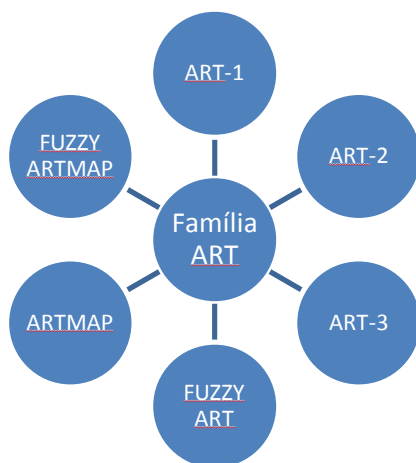


Figura 2.6 – Família ART

- ART 1 é a mais simples rede da família ART e aceita só dados binários;
- ART 2 suporta dados contínuos e binários. A chamada de entrada foi acrescida de várias unidades extras em relação a rede ART 1;
- ART 3 apresenta princípios organizacionais que permite o processo de busca em uma estrutura hierárquica;
- ARTMAP também conhecida como ART preditiva, combina duas redes ART ( $ART_a$  e  $ART_b$ ) e um módulo chamado *mapfiel* que mapeia (relaciona) os resultados das duas redes;
- Fuzzy ART implementa lógica *fuzzy* no reconhecimento de padrões e trabalha de forma não supervisionada;

- Fuzzy ARTMAP é um ARTMAP com unidades *fuzzy* ART. Essa rede pode trabalhar de forma supervisionada e não supervisionada.

Na atividade de reconhecimento de padrões, quando uma certa entrada não se assemelha a nenhum dos agrupamentos (*clusters*) existentes, um novo cluster é criado para tal entrada. Essa característica pode comprometer a estabilidade da rede, ou seja, quando uma nova aprendizagem ocorre, os padrões já treinados anteriormente são prejudicados podendo ocorrer a perda da capacidade de reconhecer os padrões antigos.

A rede ART é um sistema de aprendizagem competitivo não-supervisionado que auto organiza categorias em resposta a sequências arbitrárias de padrões de entrada, em tempo real, para reconhecimento de padrões. Tendo como característica importante a combinação de padrões, onde o padrão de entrada atual é comparado com uma representação de categoria selecionada.

As redes neurais da família ART apresentam algumas dificuldades operacionais: razoável sensibilidade aos parâmetros da rede (parâmetro de vigilância, etc.) e a precisão das análises. Os efeitos da sensibilidade paramétrica têm sido resolvidos, ou atenuados, através do emprego de novas concepções de treinamento e de arquitetura das redes neurais ART (CARPENTER et al., 1991a). Igualmente, várias propostas têm sido apresentadas na literatura especializada visando sanar o problema da imprecisão. Grande parte da imprecisão decorre em consequência, principalmente, do mecanismo de escolha de categorias e do teste de vigilância que necessita de aperfeiçoamento.

### 2.5.2 – Redes Neurais ArtMap

A rede neural ARTMAP constitui-se numa rede supervisionada, ou seja, o treinamento é realizado baseado em estímulos de entrada e de saída, sendo formada por duas sub-redes neurais ART: ARTa e ARTb interligadas por um módulo inter-ART que executa a função de estabelecer o “casamento” entre a entrada e a saída da rede neural. A rede neural ARTMAP nebulosa possui uma arquitetura robusta tal como a rede ART com características de estabilidade e plasticidade.



Esse tipo de rede é composta por duas sub-redes, ARTa, que recebe e processa um determinado padrão de entrada apresentado, e ARTb, que constitui a resposta desejada para o padrão apresentado à rede que são interligadas por um módulo de memória associativa, denominado Inter-ART, que realiza o mapeamento. Durante o treinamento supervisionado, a ARTa recebe um padrão de entrada  $a$  e ARTb recebe um padrão  $b$ , sendo que  $b$  é a predição correta de  $a$ . Quando uma predição em ARTa não é confirmada por ARTb, a memória associativa inibe o inter-ART, ativando o processo do match tracking. Este processo aumenta a vigilância em ARTa por uma quantidade mínima necessária para que o sistema ative outra categoria em ARTa que igual e a categoria em ARTb, ou para treinar uma nova categoria em ARTa. O mecanismo de controle interno do ARTMAP cria reconhecimento de categorias estáveis de medida ótima por maximização da generalização preditiva em operações em tempo real. A rede neural ARTMAP (figura 2.7) executa treinamento supervisionado como o retro propagação (backpropagation).

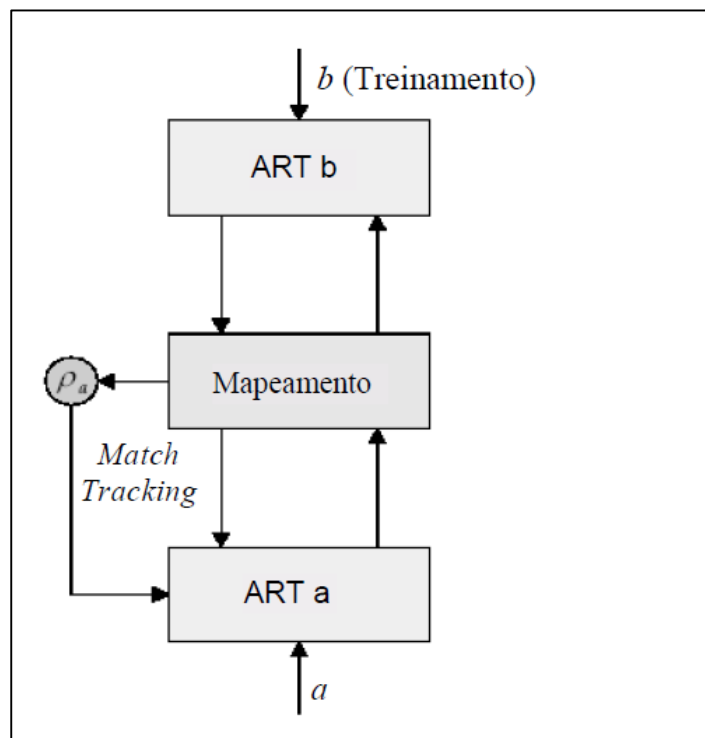


Figura 2.7 – Arquitetura da Redes Neurais ArtMap  
 Extraída de: MALANGE, 2010

O módulo inter-ART é responsável pela verificação se há casamento da entrada (ARTa) e da saída (ARTb). Todas as matrizes de pesos são iniciadas com

valores iguais a 1, ou seja, todas as atividades encontram-se inativas. Estas atividades são ativadas à medida que ocorre ressonância entre os padrões de entrada e de saída.

O módulo ARTa aprende a classificar os padrões de entrada apresentados em sua camada e o módulo ARTb desenvolve categorias referentes aos padrões de saída desejado.

Toda vez que os pares de entrada, associados aos módulos ARTa e ARTb, são confirmados, de acordo com o teste do match tracking (processo auto-regulador que maximiza a generalização e minimiza os erros), o par de treinamento será confirmado nas matrizes de pesos, caso contrário deve-se buscar um outro padrão de entrada, até que o critério seja satisfeito.

Após poucas apresentações do conjunto de treinamento, a rede se auto-organiza e obtém a saída esperada para cada entrada sem necessariamente efetuar todo processo de busca.

Baseada na observação que o treinamento da rede ARTMAP leva a diferentes pesos adaptativos e diferentes categorias de reconhecimento para diferentes ordenações de um dado conjunto de treinamento, introduz-se uma estratégia de votação (CARPENTER et al., 1992). A estratégia de votação usa um sistema ARTMAP que é treinado diversas vezes com diferentes ordenações para o mesmo conjunto de treinamento. A predição final é dada pelas soluções geradas com maior frequência. Como o conjunto de elementos que fazem predições errôneas varia de uma simulação para a próxima, esta estratégia cancela muitos erros que ocorrem na fase de análise (diagnóstico).

As primeiras redes ARTMAP foram utilizadas para classificar vetores de entradas binários, com cada elemento do vetor indicando a presença ou ausência de uma dada característica.

### 2.5.3 – Sistemas Fuzzy

O conceito de conjunto Fuzzy foi introduzido, em 1965, por Lotfi A. Zadeh (Universidade da Califórnia, Berkeley) com a finalidade de solucionar problemas de natureza industrial, biológica ou química, que compreendessem situações ambíguas,

não passíveis de processamento através da lógica computacional fundamentada na lógica booleana.

Em sistemas baseados em lógica fuzzy, ao contrário de sistemas lógicos clássicos, procura-se modelar a forma imprecisa de raciocínio, que desempenha um papel essencial na notável capacidade humana de tomar decisões racionais em ambientes de incerteza e imprecisão (ZUBEN – UNICAMP 2010).

Esta técnica entrou em evidência se tornando conhecida no final da década de 1980, quando o metrô de Sendai adotou um sistema baseado na LN - o *Automatic Train Operator* (ATO), em consequência surgiram várias empresas que tinham o objetivo de desenvolver e comercializar produtos baseados nesta tecnologia. A lógica fuzzy é a lógica baseada na teoria dos conjuntos fuzzy. Ela difere dos sistemas lógicos tradicionais em suas características e seus detalhes. Uma das características da lógica fuzzy é que o raciocínio exato corresponde a um caso limite do raciocínio aproximado, sendo interpretado como um processo de composição de relações nebulosas. Na lógica fuzzy, o valor verdade de uma proposição não é dado no formato binário, onde verdadeiro (1) e falso (0) e sim por um subconjunto fuzzy de qualquer conjunto parcialmente ordenado.

Os problemas complexos, normalmente, exigem soluções complexas que necessitam de tempo e esforço proporcional ao grau de complexidade. Os sistemas nebulosos empregam soluções simples para resolver estes tipos de problemas, pois, em vez de controlar e/ou modelar um processo através do uso de relações complicadas entre as variáveis, o sistema nebuloso leva em conta o controle e/ou modelagem do processo através do uso de um conjunto simples de vários tipos de regras de senso comum (LOESCH; SARI, 1996).

A teoria de conjuntos nebulosos tem sido agregada aos sistemas de redes neurais, chamadas de redes nebulosas ou sistemas neuro-*fuzzy*, aumentando, assim, a capacidade de aprendizado através de interface com dados numéricos (SILVEIRA et al., 2003). Informações imprecisas, vagas, incertas, qualitativas, comunicações verbais, capacidade de aprendizado e formulação de estratégias de tomadas de decisão, são características humanas. Portanto, a teoria *fuzzy* e as redes nebulosas são frequentemente referidas como “inteligentes”, em consequência do fato de emularem a inteligência humana.

O controle nebuloso segue as etapas mostradas a seguir (Figura 2.8.).

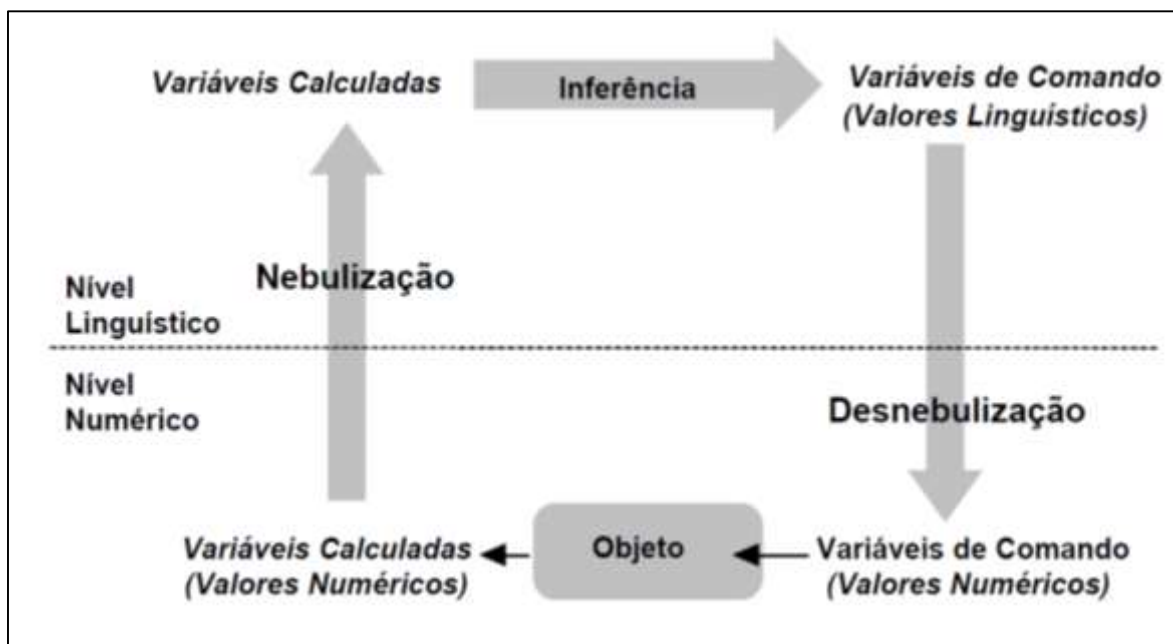


Figura 2.8 – Etapas de um sistema nebuloso  
Extraída de: MALANGE, 2010

#### 2.5.4 - Redes Neurais Fuzzy ArtMap

A rede Fuzzy ArtMap se diferencia da rede neural ARTMAP pela utilização em seus cálculos da teoria do conjunto nebuloso. Sendo capaz de aprender categorias estáveis em resposta a padrões de entradas tanto analógicos como binários.

Destinada à aproximação de funções não-lineares multidimensionais para classificar vetores de entrada cujos elementos assumem valores nebulosos, entre 0 e 1, indicando o nível de presença de cada característica.

Esta rede é composta por dois módulos ARTa nebuloso e ARTb nebuloso, que possuem a mesma estrutura da rede neural ART Nebulosa, exceto que na ARTa nebulosa inclui um campo de mapeamento controlando o treinamento de um mapa associativo de categorias de reconhecimento da ARTa para categorias de reconhecimento da ARTb (figura 2.8). O campo de mapeamento controla a regra match tracking, que aumenta o parâmetro de vigilância da rede ARTa por uma quantidade mínima necessária para corrigir um erro preditivo (MARCHIORI et al., 2002).

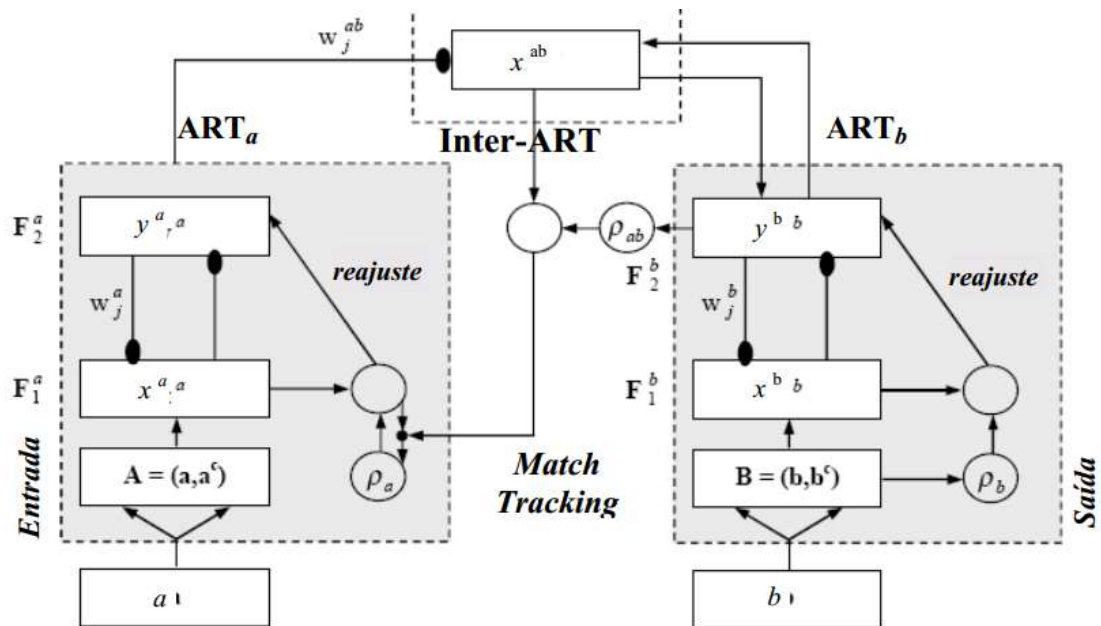


Figura 2.8 – Arquitetura da Redes Neurais Fuzzy ArtMap  
 Extraída de: MALANGE 2010

De acordo com MALANGE 2010 o algoritmo de execução de uma rede neural fuzzy ArtMap é descrito pelos passos abaixo:

- Passo 1: Normalizar os vetores de entrada ARTa e de saída ARTb, se necessário: Inicialmente, todos os valores dos neurônios devem ser normalizados se não estiverem entre 0 e 1;
- Passo 2: Codificar os vetores dos módulos ARTa e ARTb: Um novo padrão de entrada deve sofrer uma codificação complementar preliminar para se preservar a amplitude da informação;
- Passo 3: Iniciar os pesos e parâmetros dos módulos ARTa, ARTb e Inter-ART: deve-se iniciar os pesos (valor 1 = todas as categorias desativadas), taxa de treinamento ( $\beta$  entre 0 e 1), parâmetro de escolha ( $\alpha > 0$ ) e parâmetro de vigilância ( $\rho_a$ ,  $\rho_b$  e  $\rho_{ab}$  entre 0 e 1);
- Passo 4: Escolha da categoria para os módulos ARTa e ARTb: Se mais de um neurônio está ativo, é escolhido aquele com o maior índice de ordenação (maior valor).
- Passo 5: Teste de vigilância dos módulos ARTa e ARTb: A ressonância ocorre se é satisfeito o critério de vigilância. Caso contrário, se o critério de vigilância falhar para a categoria escolhida, ocorre o reset e um novo

índice é escolhido (retornar ao passo 4). O processo de busca se repete até que o índice escolhido satisfaça o teste de vigilância.

- Passo 6: Match tracking (teste de ressonância) entre os módulos ARTa e ARTb: Verificação se houve casamento da entrada com a saída. Se não houver casamento deve-se procurar outro índice que satisfaça o teste;
- Passo 7: Adaptação dos pesos: O vetor (camada F2) são atualizados com os novos pesos.
- Passo 8: Repetir passos 4 à 7 para todos os pares a serem treinados.

---

## Capítulo 3

### Materiais e Métodos

---

#### 3.1 – Metodologia

Um dos objetivos da aplicação do modelo proposto nesse trabalho é verificar agrupamento de pacientes portadores de HIV-1, de acordo com o regime terapêutico adotado, em um tempo específico de tratamento. Considerando o problema foram propostas metodologias usando duas técnicas computacionais: uma usando redes neurais ArtMap e outra usando redes neurais Fuzzy ArtMap. Usamos duas bases de dados distintas na aplicação da modelagem. Desta forma nossa pesquisa possui três metodologias conforme mostra o resumo representado na figura 3.1.

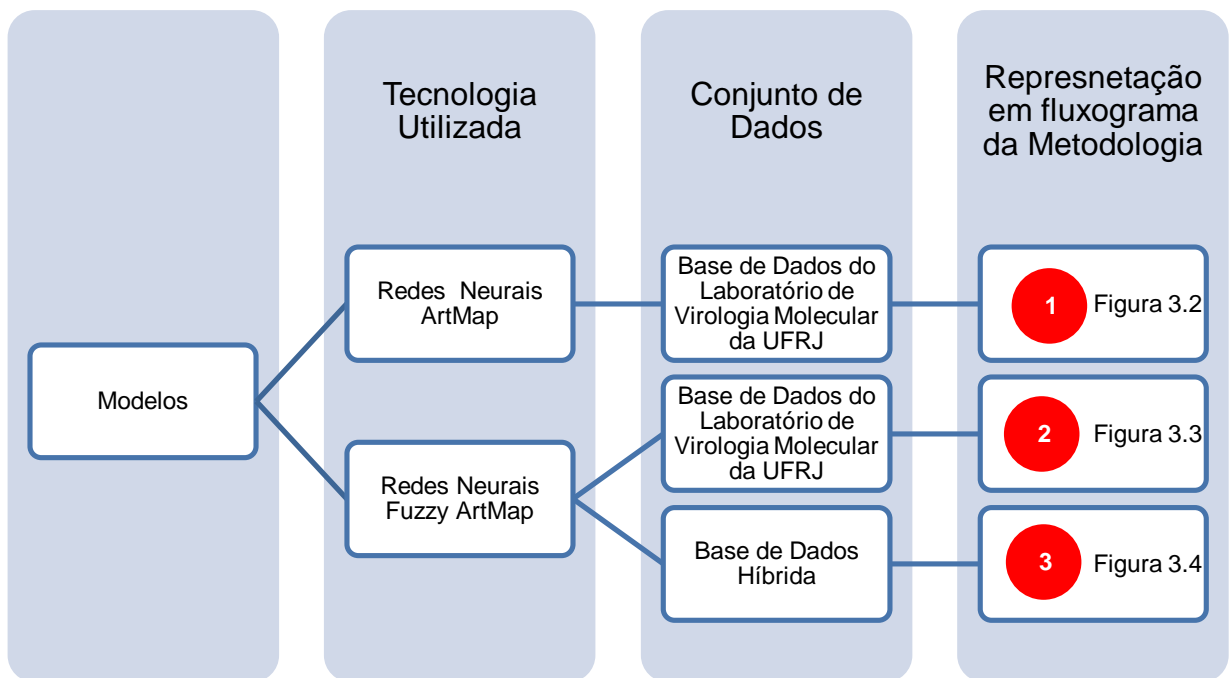


Figura 3.1 - Resumo das Metodologias da Pesquisa

Nas metodologias de número 1 e 2 utilizam a mesma base de dados e por isso foi possível fazer uma análise comparativa entre os dois tipos de redes. Na metodologia número 3 usamos uma base híbrida, o qual validamos as resistências de cada paciente por um programa disponível por Stanford. A seguir segue os fluxogramas das metodologias, nas figuras 3.2, 3.3. e 3.4.

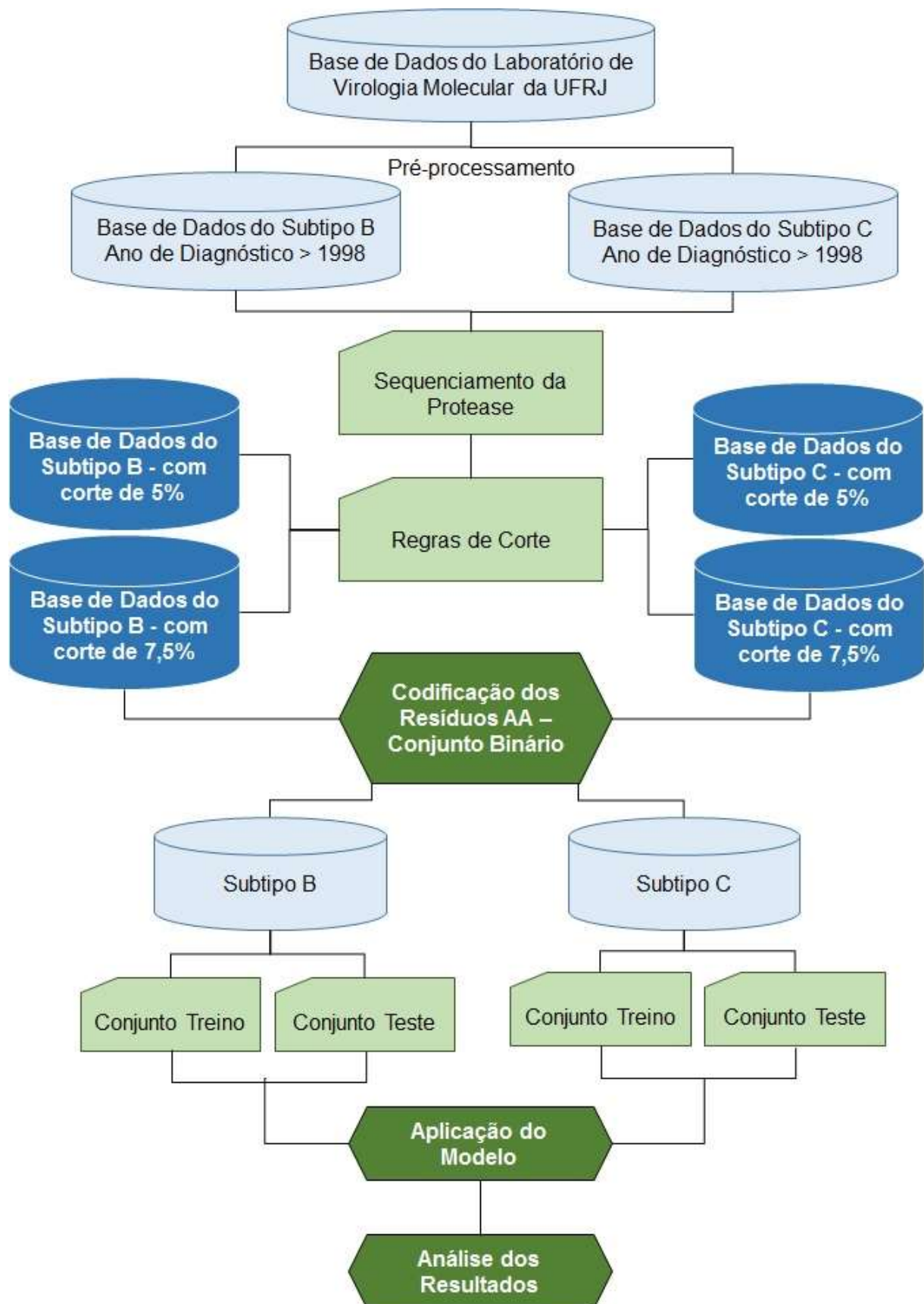


Figura 3.2 - Fluxograma da metodologia proposta no modelo computacional ArtMap com a base de dados do laboratório de virologia molecular da UFRJ



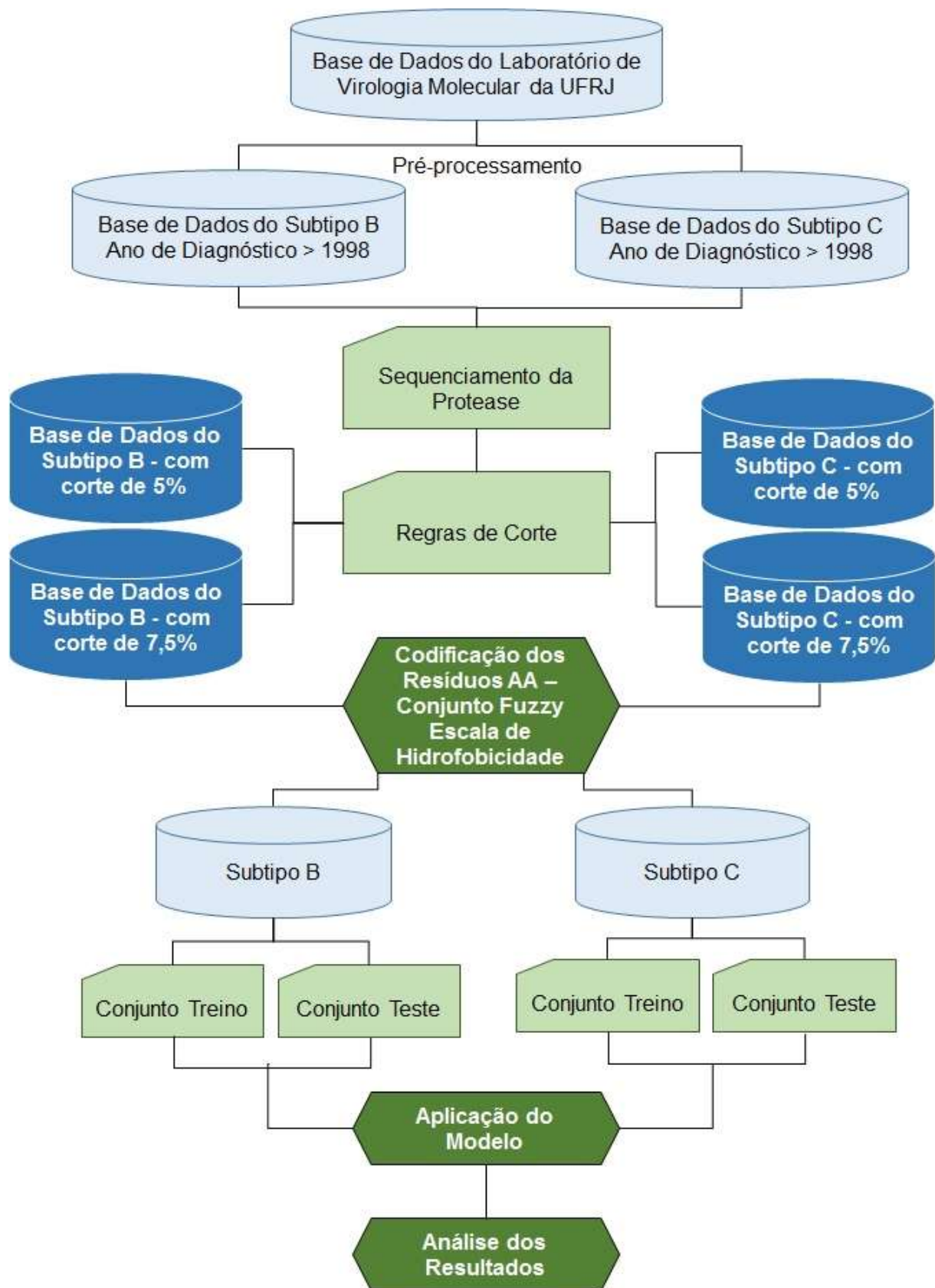


Figura 3.3 - Fluxograma da metodologia proposta no modelo computacional Fuzzy ArtMap com a base de dados do laboratório de virologia molécula da UFRJ

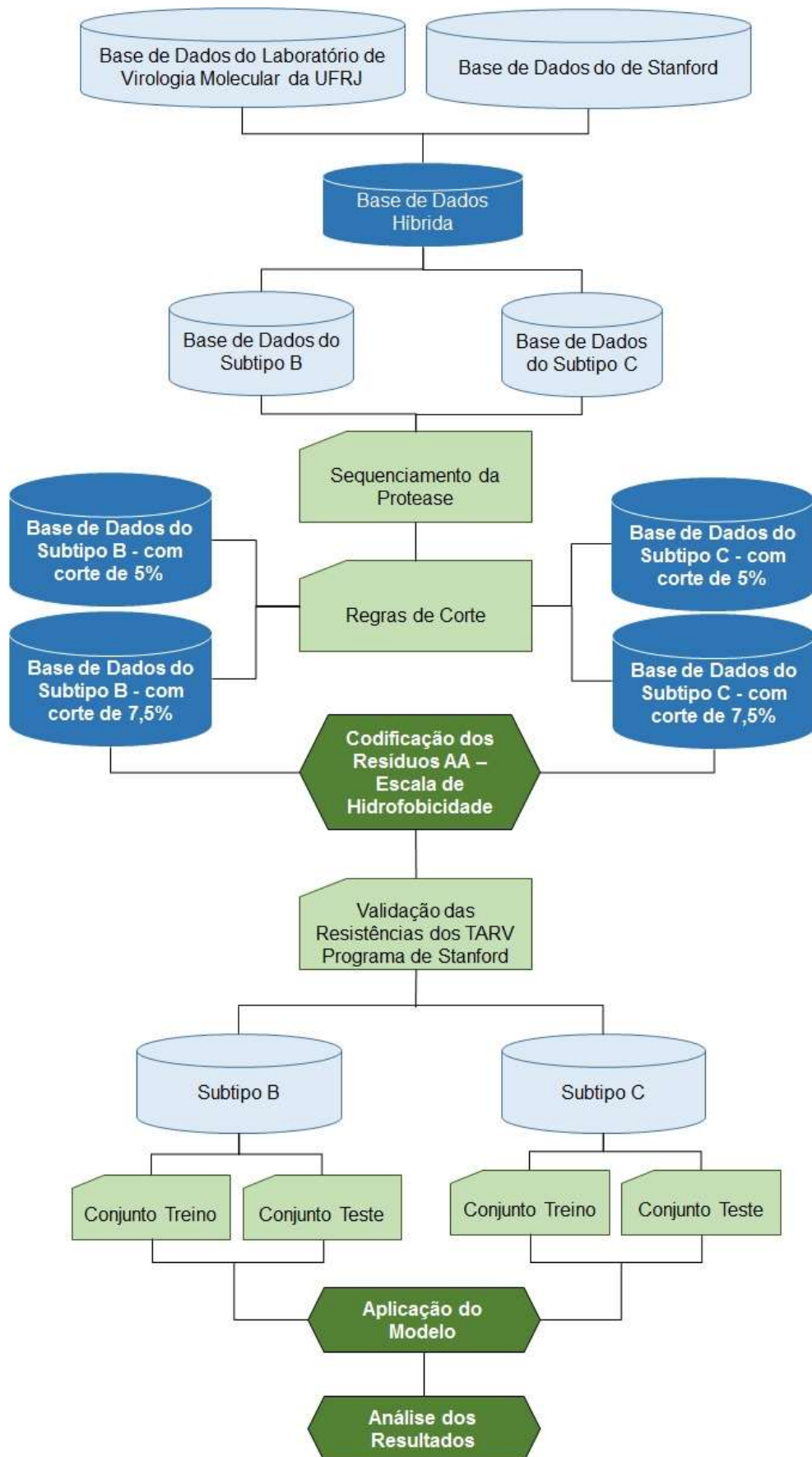


Figura 3.4 - Fluxograma da metodologia proposta no modelo computacional Fuzzy ArtMap com a base de dados híbrida

### 3.2 – Base de Dados

Conforme dito anteriormente para compor nossa amostra utilizamos duas bases de dados distintas, uma com dados obtidos do Laboratório de Virologia Molecular da UFRJ, que faz parte da rede de laboratórios destinados a realizar exames de genotipagem (RENAGENO) do Ministério da Saúde, em pacientes da rede pública de saúde infectados pelo HIV-1, e a outra com dados da base de Stanford, uma base de dados mantida pela Universidade de Stanford<sup>4</sup> como um benefício para a comunidade de pesquisa e educação, contendo vários diagnósticos de *HIV* RT e sequências de protease.

A primeira parte do estudo foi realizada usando os dados do Laboratório de Virologia Molecular da UFRJ, tendo como técnica computacional para o agrupamento as Redes Neurais ArtMap, com dados binários e, também, a redes neurais Fuzzy ArtMap, possibilitando assim uma análise comparativa dos seus resultados. Nessa base de dados os pacientes foram experimentados com inibidores da transcriptase reversa análogos de nucleosídeos (ITRN), inibidores da transcriptase reversa não-análogos de nucleosídeos (ITRNN) e inibidores de protease (IP),

Essa base de dados possui uma sequência de 99 aminoácidos, que formam a composição da protease, a contagem de CD4<sup>+</sup> e a informação acerca do experimento do paciente aos antirretrovirais. Podendo o vírus ser proveniente de um paciente que nunca foi submetido à terapia com uso de medicamentos antirretrovirais (*naive*) ou proveniente de um paciente que apresentou falha terapêutica por desenvolver resistência à droga que lhe estava sendo administrada.

A composição híbrida fora formada com dados fornecidos por Stanford e complementados com informações da base RENAGENO. Nesta formação temos apenas pacientes que foram experimentados com inibidores de Protease (IP) e paciente não resistentes, e neste caso utilizamos para obtenção dos resultados as redes Fuzzy ArtMap.

---

<sup>4</sup> Universidade de Stanford – Califórnia - EUA

### 3.2.1 – Regra de Corte na Base de Dados

Com a composição das bases classificadas pelo Subtipo aplicamos uma regra de corte nas seguintes etapas:

- 1º Corte (5%) - Foram descartadas as posições da sequência dos aminoácidos da protease que possuíam um número inferior a 5% resíduos de mutação, com relação ao total da amostra;
- 2º corte (7,5%) - Foram descartadas as posições da sequência dos aminoácidos da protease que possuíam um número inferior a 7,5% resíduos de mutação, com relação ao total da amostra.

Após o corte foi feita uma análise em cada tabela para verificar as posições clássicas encontradas na literatura, afim de que, com a regra de corte, não ocorra perda significativa dessas posições. Desta forma utilizamos o corte de 7,5% para todas as bases da pesquisa.

### 3.2.2 – Base de Dados RENAGENO – Com o Corte e Posições Seleccionadas

#### 3.2.2.1 – Subtipo B

Essa base de dados tem o total equivalente a 300 pacientes. Aplicando o primeiro pré-processamento, onde consideramos apenas pacientes com ano de diagnóstico superior a 1998, a base passa, após o filtro, para 130 pacientes e distribuída em conjunto de treino e teste, já que a técnica adotada é supervisionada, conforme mostra a figura 3.5.

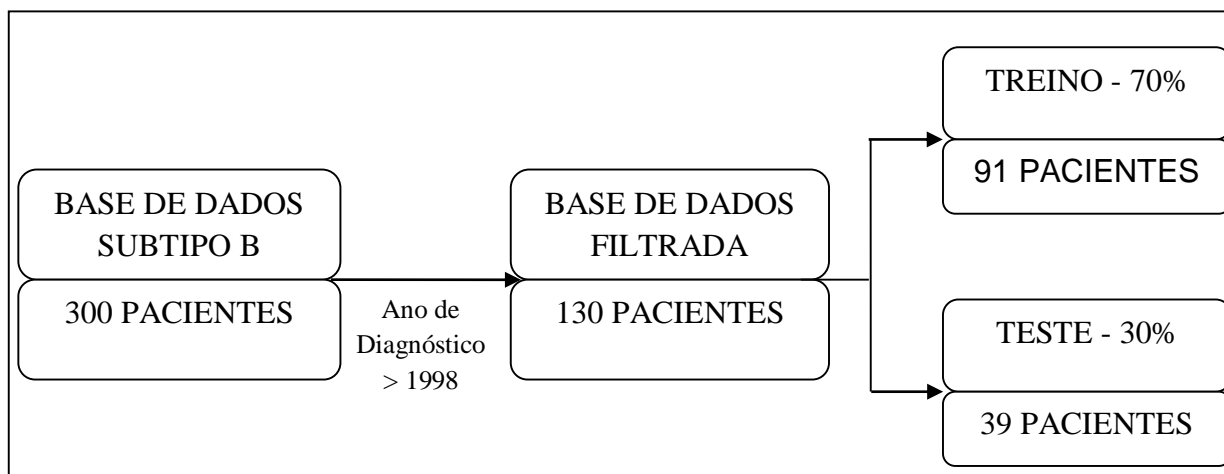


Figura 3.5- Distribuição da Amostra da Base de Dados do Subtipo B - Base de Dados RENAGENO

Após selecionados os 130 pacientes e, antes de dividir o conjunto de teste e treino, foi analisada a frequência da ocorrência de mutações em todas as posições da protease e verificou-se que havia posições com taxa de mutação muito baixas, as quais não teriam relevância para a categorização das amostras presentes na base de dados. Desta forma aplicamos as regras de cortes descritas anteriormente. Considerando as posições com um total igual ou maior que 5%, selecionamos os pacientes com mais de 7 mutações e tomando as posições com um total igual ou maior que 7,5%, selecionamos os pacientes com mais de 10 mutações. Na tabela 3.1 encontramos as posições selecionadas em cada corte realizado na base de dados.

Tabela 3.1 - Posições selecionadas com a regra de cortes na Base de Dados.

CORTE	POSIÇÕES	TOTAL
5%	L10, T12, I13, K14, I15, G16, G17, L19, K20, D30, L33, E35, M36, P39, R41, M46, K45, I54, Q58, D60, L63, I64, A71, I72, G73, T74, V77, V82, I84, N88, L90, I93	32
7,5%	L10, I13, K14, I15, K20, D30, L33, E35, M36, R41, M46, I54, D60, L63, I64, A71, I72, G73, T74, V77, V82, I84, N88, L90, I93	25

### 3.2.2.2 – Subtipo C

A base de dados com os pacientes do Subtipo C tem o total equivalente a 232 pacientes. Fizemos o mesmo pré-processamento que o anterior e nesse caso com o filtro a base passou a ter 120 pacientes, distribuída em conjunto de treino e teste, conforme mostra a figura 3.6.

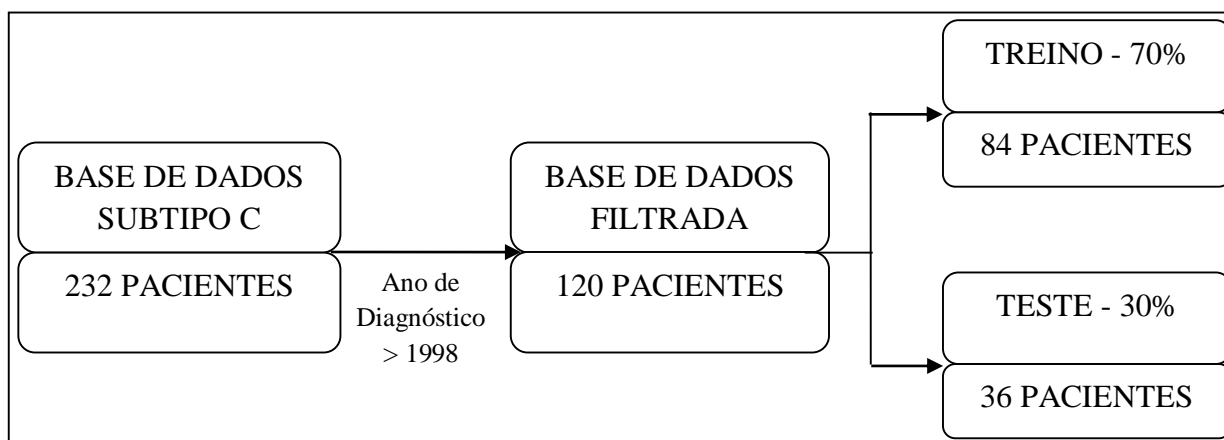


Figura 3.6 - Distribuição da Amostra da Base de Dados do Subtipo C - Base de Dados RENAGENO

Da mesma forma que na base do Subtipo B aplicamos as regras de corte, antes de dividir o conjunto de teste e treino. Com isso considerando as posições com um total igual ou maior que 5%, selecionamos os pacientes com mais de 6 mutações e considerando as posições com um total igual ou maior que 7,5% selecionamos os pacientes com mais de 9 resíduos de mutações. Na tabela 3.2 encontramos as posições selecionadas em cada corte realizado na base de dados.

Tabela 3.2 - Posições selecionadas com a regra de cortes na Base de Dados.

CORTE	POSIÇÕES	TOTAL
5%	V3, L10, T12, I13, K14, I15, G16, L19, K20, D30, L33, E35, M36, R41, k45, M46, I54, R57, D60, Q61, I62, L63, H69, A71, T74, V82, N88, L89, L90, I93	30
7,5%	L10, T12, I13, I15, G16, K20, D30, L33, E35, M36, R41, M46, I54, Q61, I62, L63, A71, T74, V82, N88, L89, L90, I93	23

### 3.2.3 – Base de Dados Híbrida – Com o Corte e Posições Selecionadas

#### 3.2.3.1 – Subtipo B

A base de dados híbrida é composta por informações obtidas nas duas fontes de dados (RENAGENO e STANFORD). Após essa seleção aleatória formamos um conjunto de dados para o Subtipo B composto de 320 pacientes. Aplicamos os cortes onde para um total igual ou maior que 5%, selecionamos as posições com mais de 16 mutações e para o corte de 7,5% consideramos as posições com mais de 24 mutações. Desta forma passamos a utilizar apenas as posições mostradas na tabela 3.3.

Tabela 3.3 - Posições selecionadas com a regra de cortes na Base de Dados Híbrida

CORTE	POSIÇÕES	TOTAL
5%	L10, T12, I13, K14, I15, L19, K20, D30, L33, E35, M36, P39, R41, M46, I54, R57, D60, I62, L63, I64, H69, A71, I72, G73, I74, V77, V82, I84, N88, L89, L90, I93	32
7,5%	L10, T12, K14, I15, L19, K20, D30, L33, E35, M36, R41, M46, I54, R57, D60, I62, L63, I64, A71, I72, V82, I84, N88, L89, L90, I93	26

Após selecionarmos as posições validamos a resistência aos antirretrovirais por um programa de Stanford e consideramos apenas os pacientes não resistentes e

com alta resistência. Sendo assim finalizamos a composição da amostra usando 150 pacientes escolhidos de forma aleatória.

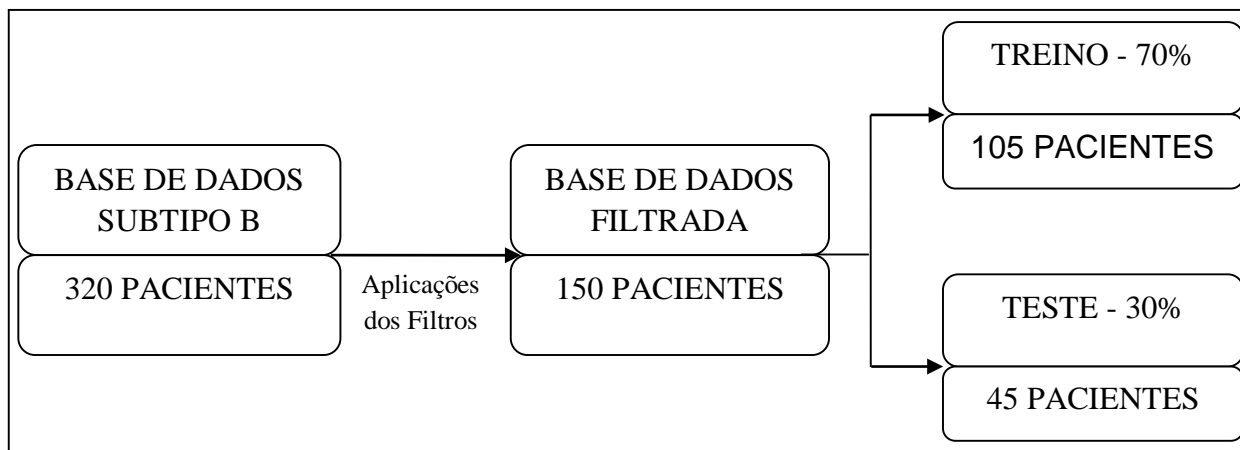


Figura 3.7- Distribuição da Amostra da Base de Dados Híbrida do Subtipo B

### 3.2.3.2 – Subtipo C

Com relação a base de dados híbrida para o Subtipo C obtemos um total de 280 pacientes. Com o corte de 5%, selecionamos as posições com mais de 14 mutações e usando o corte de 7,5%, consideramos as posições com mais de 21 mutações. Desta forma passamos a utilizar apenas as posições mostradas na tabela 3.4.

Tabela 3.4 - Posições selecionadas com a regra de cortes na Base de Dados Híbrida

CORTE	POSIÇÕES	TOTAL
5%	L10, T12, I13, K14, I15, G16, L19, K20, D30, E35, M36, P39, R41, K45, M46, D60, Q61, I62, L63, A71, I74, V77, V82, L89, L90, Q92 I93	27
7,5%	L10, T12, I13, I15, L19, K20, D30, E35, M36, P39, R41, K45, M46, D60, Q61, A71, I74, V77, V82, L89, L90, I93	22

Da mesma forma que anteriormente validamos a resistência aos antirretrovirais por Stanford e consideramos apenas os pacientes não resistentes, com alta resistência finalizando a composição da amostra com 100 pacientes escolhidos de forma aleatória.

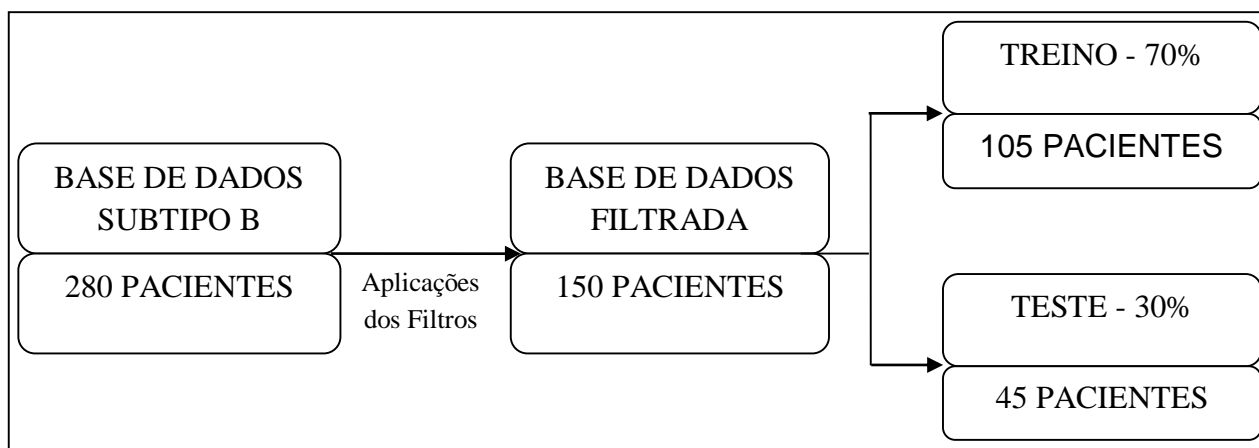


Figura 3.8- Distribuição da Amostra da Base de Dados Híbrida do Subtipo c

### 3.4 – Codificação da Base de Dados

Como temos duas metodologias diferentes usamos dois tipos de conversão de dados. Uma codificação foi usando o conceito de dados binários, o qual usamos na aplicação do modelo com redes neurais ArtMap e a outra codificação foi usando o conjunto Fuzzy para aplicação das redes neurais Fuzzy ArtMap.

O objetivo principal em usar o conjunto Fuzzy era considerar a relevância no tipo de mutação sofrida nas posições. Os resíduos foram codificados utilizando o fator de hidrofobicidade (hidrofóbicos ou hidrofílicos) dos aminoácidos (tabela 3.4.1), visto que a codificação será fundamental para formação desse conjunto e possibilita à observação apropriada das variações estequiométricas (espaciais) na molécula funcional da protease do HIV-1. Utilizando a escala de hidrofobicidade de KYTE e DOOLITTLE (1982), ponderada pelos respectivos pesos moleculares, através da equação (3.4.1), criamos uma tabela com os valores normalizados a serem utilizados na codificação dos dados tabela 3.5.



Tabela 3.5 - Escala de hidrofobicidade com Valores atribuídos a cada aminoácido e sua respectiva classificação quanto ao valor de sua hidrofobicidade, pela escala de Kyte e Doolittle.

Nome	Abreviação	Valor Hidrofobicidade	Categoria
Isoleucina	I	+4,5	Hidrofóbico
Valina	V	+4,2	Hidrofóbico
Leucina	L	+3,8	Hidrofóbico
Fenilalanina	F	+2,8	Hidrofóbico
Cisteína	C	+2,5	Hidrofóbico
Metionina	M	+1,9	Hidrofóbico
Alanina	A	+1,8	Hidrofóbico
Glicina	G	-0,4	Neutro
Treonina	T	-0,7	Neutro
Serina	S	-0,8	Neutro
Triptofano	W	-0,9	Neutro
Tirosina	Y	-1,3	Neutro
Prolina	P	-1,6	Neutro
Histidina	H	-3,2	Hidrofílico
Glutamina	Q	-3,5	Hidrofílico
Asparagina	X	-3,5	Hidrofílico
Ácido glutâmico	E	-3,5	Hidrofílico
Ácido aspártico	D	-3,5	Hidrofílico
Lisina	K	-3,9	Hidrofílico
Arginina	R	-4,5	Hidrofílico

$$matriz\_convers\tilde{a}o[i][j] = \frac{abs(hidroAA[i]*PM - hidroAA[j]*PM)}{1374} \quad (3.4.1)$$

Onde, hidroAA, corresponde ao valor de hidrofobicidade e PM ao valor do peso molecular do aminoácido.

Tabela 3.6 - Escala de hidrofobicidade normalizada

	I	V	L	F	C	M	A	G	T	S	W	Y	P	H	Q	N	E	D	K	R
I	0,000	0,071	0,067	0,093	0,209	0,223	0,313	0,451	0,490	0,491	0,563	0,601	0,563	0,791	0,802	0,766	0,804	0,768	0,844	1,000
V	0,071	0,000	0,005	0,021	0,138	0,152	0,241	0,380	0,419	0,419	0,492	0,529	0,492	0,719	0,730	0,694	0,733	0,697	0,773	0,928
L	0,067	0,005	0,000	0,026	0,142	0,156	0,246	0,384	0,423	0,424	0,496	0,534	0,497	0,724	0,735	0,699	0,737	0,702	0,777	0,933
F	0,093	0,021	0,026	0,000	0,116	0,130	0,220	0,358	0,397	0,398	0,470	0,508	0,471	0,698	0,670	0,673	0,711	0,676	0,751	0,907
C	0,209	0,138	0,142	0,116	0,000	0,014	0,104	0,242	0,281	0,282	0,354	0,392	0,354	0,582	0,592	0,557	0,595	0,559	0,635	0,791
M	0,223	0,152	0,156	0,130	0,014	0,000	0,090	0,228	0,267	0,267	0,340	0,378	0,340	0,567	0,578	0,543	0,581	0,545	0,621	0,776
A	0,313	0,241	0,246	0,220	0,104	0,090	0,000	0,139	0,177	0,178	0,250	0,288	0,251	0,478	0,489	0,453	0,491	0,456	0,531	0,687
G	0,451	0,380	0,384	0,358	0,242	0,228	0,139	0,000	0,039	0,039	0,112	0,150	0,112	0,339	0,350	0,315	0,353	0,317	0,393	0,548
T	0,490	0,419	0,423	0,397	0,281	0,267	0,177	0,039	0,000	0,001	0,073	0,111	0,073	0,301	0,312	0,276	0,314	0,278	0,354	0,510
S	0,491	0,419	0,424	0,398	0,282	0,267	0,178	0,039	0,001	0,000	0,073	0,110	0,073	0,300	0,311	0,275	0,314	0,278	0,354	0,509
W	0,563	0,492	0,496	0,470	0,354	0,340	0,250	0,112	0,073	0,073	0,000	0,038	0,000	0,228	0,238	0,203	0,241	0,205	0,281	0,437
Y	0,601	0,529	0,534	0,508	0,392	0,378	0,288	0,150	0,111	0,110	0,038	0,000	0,037	0,190	0,201	0,165	0,203	0,168	0,243	0,399
P	0,563	0,492	0,497	0,471	0,354	0,340	0,251	0,112	0,073	0,073	0,000	0,037	0,000	0,227	0,238	0,202	0,241	0,205	0,281	0,436
H	0,791	0,719	0,724	0,698	0,582	0,567	0,478	0,339	0,301	0,300	0,228	0,190	0,227	0,000	0,011	0,025	0,013	0,022	0,053	0,209
Q	0,802	0,730	0,735	0,709	0,592	0,578	0,489	0,350	0,312	0,311	0,238	0,201	0,238	0,011	0,000	0,036	0,003	0,033	0,043	0,198
N	0,766	0,694	0,699	0,673	0,557	0,543	0,453	0,315	0,276	0,275	0,203	0,165	0,202	0,025	0,036	0,000	0,038	0,003	0,078	0,234
E	0,804	0,733	0,737	0,711	0,595	0,581	0,491	0,353	0,314	0,314	0,241	0,203	0,241	0,013	0,003	0,038	0,000	0,036	0,040	0,196
D	0,768	0,697	0,702	0,676	0,559	0,545	0,456	0,317	0,278	0,278	0,205	0,168	0,205	0,022	0,033	0,003	0,036	0,000	0,076	0,231
K	0,844	0,773	0,777	0,751	0,635	0,621	0,531	0,393	0,354	0,354	0,281	0,243	0,281	0,053	0,043	0,078	0,040	0,076	0,000	0,156
R	1,000	0,928	0,933	0,907	0,791	0,776	0,687	0,548	0,510	0,509	0,437	0,399	0,436	0,209	0,198	0,234	0,196	0,231	0,156	0,000

### 3.5 - Validação das Resistências

A fim de evitar o uso de dados com erros de digitação ou de outro tipo de proveniência, durante o desenvolvimento da base de dados híbrida tivemos o cuidado de validar cada paciente com relação ao seu regime terapêutico, por um programa disponível na plataforma WEB da universidade de Stanford, representado na figura 3.5.1.

The image shows a web-based interface for validating drug resistance mutations. It is divided into three main sections: Reverse Transcriptase, Protease, and Integrase. Each section has a red header, an input field for 'Enter Mutation List', an 'OR' separator, and a grid of pull-down menus for selecting specific mutations. The mutations are listed in a grid format, with each mutation number followed by a pull-down arrow.

Reverse Transcriptase	Protease	Integrase
40, 41, 44, 62	10, 11, 13, 20	51, 66, 74, 92
65, 67, 69, 70	23, 24, 30, 32	95, 97, 114, 121
74, 75, 77, 90	33, 35, 36, 43	128, 138, 140, 143
98, 100, 101, 103	46, 47, 48, 50	145, 146, 147, 148
106, 108, 115, 116	53, 54, 58, 63	151, 153, 155, 157
118, 138, 151, 179	71, 73, 74, 76	163, 230, 263
181, 184, 188, 190	77, 82, 83, 84	
210, 215, 219, 221	85, 88, 89, 90	
225, 227, 230, 236	93	
238, 318, 348		

Figura 3.9 - Programa de validação de Resistência de Stanford

De acordo com o tipo de mutação sofrida no resíduo do aminoácido o programa classifica em: suscetível, resistência de baixo nível, resistência de alto nível, resistência intermediária e potencial de resistência de baixo nível.

### 3.6 – Aplicação do Modelo

O modelo computacional proposto foi implementado utilizando o *software* MATLAB (The Math Works; inc.: [HTTP://www.mathworks.com/](http://www.mathworks.com/)). As técnicas de redes neurais escolhidas para a redução de dimensionalidade são as redes neurais ArtMap e Fuzzy ArtMap. Os padrões de entrada na rede devem ser dados contínuos, logo, foi

necessário converter o conjunto de dados, aminoácidos representados por letras, em dados numéricos. Para isto, como já citamos anteriormente usamos o conjunto binário no treinamento da rede ArtMap e a escala de hidrofobicidade para as redes Fuzzy ArtMap. O processo de funcionamento das redes se resumem na leitura de um arquivo contendo a sequência dos resíduos de mutação de cada paciente. Os dados de entrada são ajustados no vetor de entrada, que inicialmente é composto pelo conjunto de treino. Ocorre o ajuste dos parâmetros da rede até atingirmos resultados de agrupamentos relevantes. E sobre os resultados foi inserido os elementos do conjunto de teste, afim de verificar a teoria da ressonância adaptativa.

Após determinarmos o número de variáveis do modelo, com a obtenção dos resultados das redes ArtMap e Fuzzy ArtMap, foi feita a validação dos *clusters*, gerando uma análise comparativa entre as duas técnicas. Com isso na segunda etapa do projeto que seria aplicar as técnicas na base de dados híbrida, resolvemos utilizar apenas as redes fuzzy ArtMap, pois essas se destacaram na 1ª etapa da pesquisa. Os grupos foram analisados quanto às informações de resistência, nesse caso, os indivíduos podem estar em falha terapêutica devido à infecção por vírus resistentes ou não terem sido tratados (*naïve*).

---

## Capítulo 4

### Resultados

---

Para analisar os resultados obtidos foram realizados testes com distintos parâmetros de vigilância com o objetivo de encontrar os agrupamentos que apresentassem os resultados mais significativos.

Para verificar o desempenho da rede utilizamos duas medidas: a sensibilidade e a especificidade. A sensibilidade mede a capacidade da rede em identificar corretamente a resistência entre aqueles que a possuem, ou seja, o quão sensível é o teste. A especificidade mede a capacidade da rede em excluir corretamente aqueles que não possuem a resistência. Assim, na avaliação existem 4 interpretações possíveis para o resultado do teste: duas em que o teste está correto e duas em que está incorreto. O teste está correto quando a rede agrupa como resistente tendo realmente a presença da resistência (resultados verdadeiros positivos), ou negativo na ausência da resistência (resultados verdadeiros negativos). Por outro lado, o teste está incorreto quando ele é positivo na ausência da resistência (falso positivo), ou negativo quando a resistência está presente (falso negativo). Os melhores testes diagnósticos são aqueles com poucos resultados falso-positivos e falso-negativos.

Para analisar os resultados tomamos os seguintes procedimentos:

1. Verificamos os agrupamentos estabelecidos pela rede e selecionamos os pacientes que foram agrupados nos clusters;
2. Usando as informações da base de dados verificamos a característica principal do agrupamento, ou seja, o tipo de terapia predominante;
3. Em sequência de acordo com as características encontrados no item 2, analisamos todos os pacientes na base de dados que se enquadram naquela respectiva terapia, de modo a contrapor com a resposta obtida pela rede;
4. Através do conceito de sensibilidade e a especificidade montamos as tabelas de resultados com as porcentagens de acerto e erro.

## 4.1 - Resultados da Base de Dados RENAGENO

### 4.1.1 – Redes ArtMap – Subtipo B

Antes de validar a capacidade da rede em classificar os pacientes de acordo com o regime terapêutico, observamos a distribuição da amostra separada pelo tipo de resistência, conforme mostra a tabela 4.1.

TABELA 4.1  
Distribuição da Amostra da Base B

BASE	N/RES.	SQV	IDV	NFV	APV	ATV	LPV/r	Total
Treino	37	5	8	17	3	7	15	91
Teste	11	2	4	8	2	3	8	39
Total	48	7	12	25	5	10	23	130

LEGENDA:

N/RES - Não Resistentes

SQV - Resistente ao inibidor Saquinavir

IDV - Resistente ao inibidor Indinavir

NFV - Resistente ao inibidor Nelfinavir

APV - Resistente ao inibidor Amprenavir

ATV - Resistente ao inibidor Atazanavir

LPV/r - Resistente ao inibidor Lopinavir/ritonavir

Na tabela 4.2 temos os agrupamentos mais significativo, obtido com o parâmetro de vigilância 0,5, onde os pacientes que compõem a amostra de teste estão destacados.

TABELA 4.2  
Agrupamento com o Parâmetro de Vigilância 0,5 – Rede ArtMap Subtipo B  
**CONJUNTO DE TREINO E TESTE**

Grupo	Pacientes Selecionados	Característica Principal do Agrupamento
1	26, 35, 40, 41, 42, 44, 50, 51, 53, 54, 79, 80, 83, 88, 110, 113, 120, 121, 122, 123, 124, 128	Resistentes ao Lopinavir/ritonavir
2	1, 2, 3, 4, 5, 6, 7, 9, 12, 19, 20, 21, 22, 24, 27, 28, 39, 40, 45, 46, 47, 56, 57, 66, 67, 68, 78, 85, 86, 87, 89, 90, 94, 95, 97, 100, 102, 105, 109, 111, 112, 114, 116, 117, 118, 129	Não resistentes aos inibidores de protease
3	8, 16, 17, 18, 30, 31, 32, 37, 38, 43, 64, 65, 71, 73, 75, 76, 81, 82, 99, 107, 108, 118, 119	Resistentes ao Nelfinavir
4	10, 11, 25, 48, 55, 91, 92, 130	Resistentes ao Amprenavir
5	13, 14, 23, 59, 84, 93, 101, 103, 104, 106	Resistentes ao Indinavir
6	15, 33, 34, 58, 60, 61, 69, 72, 79, 96, 125, 126, 127	Resistentes ao Atazanavir
7	26, 62, 63, 70, 74, 77, 98, 115	Resistentes ao Saquinavir

As tabelas 4.3, 4.4, 4.5, 4.6, 4.7, 4.8 e 4.9 representam o agrupamento dos resultados da rede ARTMAP, com os dados da base composta por informações do laboratório de virologia molecular da UFRJ, para paciente do subtipo B. Os dados foram agrupados e em sequência foi feita uma análise com a finalidade de verificar a característica daquele grupo, destacando desta forma, a resistência predominante no agrupamento.

Tabela 4.3 - Análise dos Resultados do Agrupamento 1 (resistente ao LPV/r)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao LPV/r	Não Resistentes ao LPV/r
Resistentes ao LPV/r	17 (73,9%)	6 (26,1%)
Não Resistentes ao LPV/r	5 (4,7%)	102 (95,3%)
TOTAL	22	108

Tabela 4.4 - Análise dos Resultados do Agrupamento 2 (não resistentes)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Não Resistentes	Resistentes
Não Resistentes	38 (79,2%)	10 (20,8%)
Resistentes	8 (9,8%)	74 (90,2%)
TOTAL	46	84

Tabela 4.5 - Análise dos Resultados do Agrupamento 3 (resistente ao NFV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao NFV	Não Resistentes ao NFV
Resistentes ao NFV	21 (84,0%)	4 (16,0%)
Não Resistentes ao NFV	2 (1,9%)	103 (98,1%)
TOTAL	23	107

Tabela 4.6 - Análise dos Resultados do Agrupamento 4 (resistente ao APV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao APV	Não Resistentes ao APV
Resistentes ao APV	3 (60,0%)	2 (40,0%)
Não Resistentes ao APV	5 (4,0%)	120 (96,0%)
TOTAL	8	122

Tabela 4.7 - Análise dos Resultados do Agrupamento 5 (resistente ao IDV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao IDV	Não Resistentes ao IDV
Resistentes ao IDV	7 (58,3%)	5 (41,7%)
Não Resistentes ao IDV	3 (2,5%)	115 (97,5%)
TOTAL	10	120

Tabela 4.8 - Análise dos Resultados do Agrupamento 6 (resistente ao ATV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao ATV	Não Resistentes ao ATV
Resistentes ao ATV	9 (90,0%)	1 (10,0%)
Não Resistentes ao ATV	4 (3,3%)	116 (96,7%)
TOTAL	13	117



Tabela 4.9 - Análise dos Resultados do Agrupamento 7 (resistente ao SQV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao SQV	Não Resistentes ao SQV
Resistentes ao SQV	5 (71,4%)	2 (28,6%)
Não Resistentes ao SQV	3 (2,4%)	120 (97,6%)
TOTAL	8	122

#### 4.1.2 – Redes ArtMap – Subtipo C

Antes de validar a capacidade da rede em classificar os pacientes de acordo com o regime terapêutico, observamos a distribuição da amostra separada pelo seu regime, conforme mostra a tabela 4.10

TABELA 4.10  
Distribuição da Amostra da Base C

BASE	N/RES.	SQV	IDV	NFV	APV	ATV	LPV/r	Total
Treino	41	2	4	19	5	5	8	84
Teste	13	1	2	8	2	4	6	36
Total	54	3	6	27	7	9	14	120

LEGENDA:

N/RES - Não Resistentes

SQV - Resistente ao inibidor Saquinavir

IDV - Resistente ao inibidor Indinavir

NFV - Resistente ao inibidor Nelfinavir

APV - Resistente ao inibidor Amprenavir

ATV - Resistente ao inibidor Atazanavir

LPV/r - Resistente ao inibidor Lopinavir/ritonavir

Na tabela 4.11 temos os agrupamentos mais significativo, obtido com o parâmetro de vigilância 0,5, onde os pacientes que compõem a amostra de teste estão destacados.

TABELA 4.11  
Agrupamento com o Parâmetro de Vigilância 0,5 – Rede ArtMap Subtipo C  
**CONJUNTO DE TREINO E TESTE**

Grupo	Pacientes Selecionados	Característica Principal do Agrupamento
1	1, 2, 3, 4, 7, 9, 10, 11, 14, 15, 16, 18, 22, 23, 24, 25, 26, 33, 37, 47, 48, 49, 55, 56, 59, 64, 68, 69, 70, 74, 75, 76, 79, 89, 90, 95, 99, 103, 118	Não resistentes aos inibidores de protease
2	5, 6, 12, 29, 30, 38, 67, 85, 86, 92, 97, 104, 105, 106, 110, 111, 117	Resistentes ao Lopinavir/ritonavir
3	32, 36, 45, 46, 63, 73, 93, 94, 96, 107, 115	Resistentes ao Indinavir
4	17, 19, 20, 21, 27, 28, 31, 34, 43, 44, 50, 51, 52, 65, 66, 98, 100, 101, 102, 109, 112, 114	Resistentes ao Nelfinavir
5	13, 35, 39, 57, 58, 60, 61, 62, 87, 91, 113, 116, 119, 120	Resistentes ao Atazanavir
6	8, 40, 53, 54, 77, 78, 81, 82, 108	Resistentes ao Amprenavir
7	41, 42, 71, 72, 83, 80, 84, 88	Resistentes ao Saquinavir

A seguir encontramos nas tabelas 4.12, 4.13, 4.14, 4.15, 4.16, 4.17 e 4.18 os agrupamentos dos resultados da rede ARTMAP, com os dados da base composta por informações do laboratório de virologia molecular da UFRJ, para paciente do subtipo C. Os dados foram agrupados e em sequência foi feita uma análise com a finalidade de verificar a característica daquele grupo, destacando desta forma, a resistência predominante no agrupamento.

Tabela 4.12 - Análise dos Resultados do Agrupamento 1 (não resistentes)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Não Resistentes	Resistentes
Não Resistentes	34 (63,0%)	20 (37,0%)
Resistentes	5 (7,6%)	61 (92,4%)
TOTAL	39	81

Tabela 4.13 - Análise dos Resultados do Agrupamento 2 (resistente LPV/r)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao LPV/r	Não Resistentes ao LPV/r
Resistentes ao LPV/r	11 (78,6%)	3 (21,4%)
Não Resistentes ao LPV/r	6 (5,7%)	100 (94,3%)
TOTAL	17	103

Tabela 4.14 - Análise dos Resultados do Agrupamento 3 (resistente ao IDV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao IDV	Não Resistentes ao IDV
Resistentes ao IDV	4 (66,7%)	2 (33,3%)
Não Resistentes ao IDV	7 (6,1%)	107 (93,9%)
TOTAL	11	109

Tabela 4.15 - Análise dos Resultados do Agrupamento 4 (resistente ao NFV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao NFV	Não Resistentes ao NFV
Resistentes ao NFV	21 (77,8%)	6 (22,2%)
Não Resistentes ao NFV	1 (1,1%)	92 (98,9%)
TOTAL	22	98

Tabela 4.16 - Análise dos Resultados do Agrupamento 5 (resistente ao ATV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao ATV	Não Resistentes ao ATV
Resistentes ao ATV	7 (77,8%)	2 (22,2%)
Não Resistentes ao ATV	7 (6,3%)	104 (93,7%)
TOTAL	14	106

Tabela 4.17 - Análise dos Resultados do Agrupamento 6 (resistente ao APV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao APV	Não Resistentes ao APV
Resistentes ao APV	4 (57,1%)	3 (42,9%)
Não Resistentes ao APV	5 (4,4%)	108 (95,6%)
TOTAL	9	111

Tabela 4.18 - Análise dos Resultados do Agrupamento 7 (resistente ao SQV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao SQV	Não Resistentes ao SQV
Resistentes ao SQV	2 (66,7%)	1 (33,3%)
Não Resistentes ao SQV	6 (5,1%)	111 (94,9%)
TOTAL	8	112

#### 4.1.3 – Resultados da Rede Fuzzy ArtMap – Subtipo B

Observamos que os resultados da aplicação da Redes ArtMap foram significativos, porém todos os dados usados foram binários, considerando apenas se ocorreu a mutação ou não. Vale destacar que quando ocorre uma mutação temos que levar em consideração, também, para qual foi aminoácido que passou a posição. Por exemplo: quando uma Leucina passa para Alanina temos um grau de relevância totalmente distinto de uma Leucina para uma Prolina. Desta forma com o objetivo de considerar o impacto que o tipo de mutação causaria nos resultados convertemos a base de dados de binário para um conjunto Fuzzy e passamos a utilizar a tecnologia de redes neurais Fuzzy ArtMap, gerando assim novos resultados e criando um comparativo de desempenho das redes.

Na segunda aplicação usamos uma base de dados híbrida, com a junção de dados do Laboratório de Virologia Molecular da UFRJ com a base de dados de Stanford. Criamos a regra de corte de 5% e 7,5% e com as posições preservadas e validamos a resistência através de um programa de Stanford. Após essa validação consideramos apenas pacientes não resistentes, com alto grau de resistência e com potencial médio de resistência.

TABELA 4.19  
Agrupamento com o Parâmetro de Vigilância 0,3 – Rede Fuzzy ArtMap Subtipo B

CONJUNTO DE TREINO E TESTE		
Grupo	Pacientes Selecionados	Característica Principal do Agrupamento
1	1, 2, 3, 4, 5, 6, 7, 9, 10, 19, 20, 24, 27, 28, 39, 40, 45, 46, 47, 56, 57, 67, 68, 70, 77, 84, 85, 87, 89, 91, 92, 94, 95, 97, 102, 105, 109, 116, 117, 118, 129, 130	Não resistentes aos inibidores de protease
2	10, 11, 12, 15, 25, 33, 58, 60, 61, 69, 72, 79, 80, 81, 82, 96, 111, 120, 125, 127	Resistentes ao Atazanavir
3	8, 16, 17, 18, 21, 22, 30, 31, 32, 37, 38, 43, 64, 65, 71, 73, 75, 76, 99, 100, 107, 118, 119	Resistentes ao Nelfinavir
4	13, 14, 23, 55, 48, 59, 90, 93, 101, 103, 104, 106, 126	Resistentes ao Indinavir
5	26, 35, 40, 41, 42, 44, 50, 51, 53, 54, 66, 83, 88, 79, 110, 113, 122, 123, 124, 128	Resistentes ao Lopinavir/ritonavir
6	26, 34, 62, 63, 74, 78, 86, 98, 108, 112, 114, 115	Resistentes ao Saquinavir

As tabelas 4.20, 4.21, 4.22, 4.23, 4.24 e 4.25 expõem os agrupamentos dos resultados da rede Fuzzy ARTMAP, com os dados da base composta por informações do laboratório de virologia molecular da UFRJ, para paciente do subtipo B. Os dados foram agrupados e em sequência foi feita uma análise com a finalidade de verificar a característica daquele grupo, destacando desta forma, a resistência predominante no agrupamento.

Tabela 4.20 - Análise dos Resultados do Agrupamento 1 (não resistente)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Não Resistentes	Resistentes
Não Resistentes	40 (83,3%)	8 (16,7%)
Resistentes	2 (2,4%)	80 (97,6%)
TOTAL	42	88

Tabela 4.21 - Análise dos Resultados do Agrupamento 2 (resistente ao ATV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao ATV	Não Resistentes ao ATV
Resistentes ao ATV	9 (90,0%)	1 (10,0%)
Não Resistentes ao ATV	11 (9,2%)	109 (90,8%)
TOTAL	20	110

Tabela 4.22 - Análise dos Resultados do Agrupamento 3 (resistente ao NFV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao NFV	Não Resistentes ao NFV
Resistentes ao NFV	22 (88,0%)	3 (12,0%)
Não Resistentes ao NFV	1 (1,0%)	104 (99,0%)
TOTAL	23	107

Tabela 4.23 - Análise dos Resultados do Agrupamento 4 (resistente ao IDV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao IDV	Não Resistentes ao IDV
Resistentes ao IDV	9 (75,0%)	3 (25,0%)
Não Resistentes ao IDV	4 (3,4%)	114 (96,6%)
TOTAL	13	117

Tabela 4.24 - Análise dos Resultados do Agrupamento 5 (resistente ao LPV/r)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao LPV/r	Não Resistentes ao LPV/r
Resistentes ao LPV/r	18 (78,3%)	5 (21,7%)
Não Resistentes ao LPV/r	2 (1,9%)	105 (98,1%)
TOTAL	20	110

Tabela 4.25 - Análise dos Resultados do Agrupamento 6 (resistente ao SQV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao SQV	Não Resistentes ao SQV
Resistentes ao SQV	6 (85,7%)	1 (14,3%)
Não Resistentes ao SQV	6 (4,9%)	117 (95,1%)
TOTAL	12	118

#### 4.1.4 – Resultados da Rede Fuzzy ArtMap – Subtipo C

TABELA 4.26

Agrupamento com o Parâmetro de Vigilância 0,4 – Rede Fuzzy ArtMap Subtipo C

CONJUNTO DE TREINO E TESTE		
Grupo	Pacientes Selecionados	Característica Principal do Agrupamento
1	13, 35, 39, 57, 60, 62, 85, 87, 91, 107, 111, 116, 120	Resistentes ao Atazanavir
2	6, 12, 29, 30, 38, 67, 86, 92, 97, 104, 105, 106, 110	Resistentes ao Lopinavir/ritonavir
3	1, 17, 19, 20, 21, 27, 28, 31, 34, 43, 44, 50, 52, 65, 66, 68, 69, 80, 81, 98, 102, 112, 114, 119	Resistentes ao Nelfinavir
4	2, 3, 4, 7, 9, 10, 11, 14, 15, 16, 18, 22, 23, 24, 25, 26, 33, 37, 47, 48, 49, 55, 56, 58, 59, 61, 63, 64, 70, 73, 74, 75, 76, 79, 88, 89, 90, 95, 99, 103, 108, 118	Não resistentes aos inibidores de protease
5	8, 40, 53, 54, 77, 78, 83, 109, 113, 117	Resistentes ao Amprenavir
6	41, 42, 51, 71, 72, 82, 84, 100, 101	Resistentes ao Saquinavir
7	5, 32, 36, 45, 46, 93, 94, 96, 115	Resistentes ao Indinavir

As tabelas 4.27, 4.28, 4.29, 4.30, 4.31, 4.32 e 4.33 expõem os agrupamentos dos resultados da rede Fuzzy ARTMAP, com os dados da base composta por informações do laboratório de virologia molecular da UFRJ, para paciente do subtipo

C. Os dados foram agrupados e em sequência foi feita uma análise com a finalidade de verificar a característica daquele grupo, destacando desta forma, a resistência predominante no agrupamento.

Tabela 4.27 - Análise dos Resultados do Agrupamento 1 (resistente ao ATV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao ATV	Não Resistentes ao ATV
Resistentes ao ATV	7 (77,8%)	2 (22,2%)
Não Resistentes ao ATV	6 (5,4%)	105 (94,6%)
TOTAL	13	107

Tabela 4.28 - Análise dos Resultados do Agrupamento 2 (resistente ao LPV/r)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao LPV/r	Não Resistentes ao LPV/r
Resistentes ao LPV/r	10 (71,4%)	4 (28,6%)
Não Resistentes ao LPV/r	3 (2,8%)	103 (97,2%)
TOTAL	13	107

Tabela 4.29 - Análise dos Resultados do Agrupamento 3 (resistente ao NFV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao NFV	Não Resistentes ao NFV
Resistentes ao NFV	23 (85,2%)	4 (14,8%)
Não Resistentes ao NFV	1 (1,1%)	92 (98,9%)
TOTAL	24	96

Tabela 4.30 - Análise dos Resultados do Agrupamento 4 (não resistentes)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Não Resistentes	Resistentes
Não Resistentes	39 (72,2%)	15 (27,8%)
Resistentes	3 (4,5%)	63 (95,5%)
TOTAL	42	78

Tabela 4.31 - Análise dos Resultados do Agrupamento 5 (resistente ao APV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao APV	Não Resistentes ao APV
Resistentes ao APV	6 (85,7%)	1 (14,3%)
Não Resistentes ao APV	4 (3,5%)	109 (96,5%)
TOTAL	10	110

Tabela 4.32 - Análise dos Resultados do Agrupamento 6 (resistente ao SQV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao SQV	Não Resistentes ao SQV
Resistentes ao SQV	3 (100%)	0 (0,0%)
Não Resistentes ao SQV	6 (5,1%)	111 (94,9%)
TOTAL	9	111

Tabela 4.33 - Análise dos Resultados do Agrupamento 7 (resistente ao IDV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao IDV	Não Resistentes ao IDV
Resistentes ao IDV	3 (50,0%)	3 (50,0%)
Não Resistentes ao IDV	6 (5,3%)	108 (94,7%)
TOTAL	9	111

## 4.2 - Resultados da Base de Dados Híbrida

### 4.2.1 – Redes Fuzzy ArtMap – Subtipo B

Antes de validar a capacidade da rede em classificar os pacientes de acordo com o regime terapêutico, observamos a distribuição da amostra separada pelo seu regime, conforme mostra a tabela 4.34.

TABELA 4.34  
Distribuição da Amostra da Base B - Base Híbrida

BASE	N/RES.	SQV	IDV	NFV	APV	ATV	LPV/r	Total
Treino	18	6	11	32	10	12	16	105
Teste	9	2	4	13	4	6	7	45
Total	27	8	15	45	14	18	23	150

LEGENDA:

N/RES - Não Resistentes

SQV - Resistente ao inibidor Saquinavir

IDV - Resistente ao inibidor Indinavir

NFV - Resistente ao inibidor Nelfinavir

APV - Resistente ao inibidor Amprenavir

ATV - Resistente ao inibidor Atazanavir

LPV/r - Resistente ao inibidor Lopinavir/ritonavir



Na tabela 4.35 temos os agrupamentos mais significativo, obtido com o parâmetro de vigilância 0,4, onde os pacientes que compõem a amostra de teste estão destacados.

TABELA 4.35  
Agrupamento com o Parâmetro de Vigilância 0,4 – Fuzzy ArtMap Subtipo B – Base Híbrida

CONJUNTO DE TREINO E TESTE		
Grupo	Pacientes Selecionados	Característica Principal do Agrupamento
1	1, 3, 8, 10, 12, 15, 16, 24, 28, 31, 37, 40, 45, 46, 54, 56, 74, 80, 87, 89, 91, 108, 110, 115, 118, 121, 123, 134, 141, 143, 145, 150	Não resistentes aos inibidores de protease
2	4, 9, 18, 19, 25, 43, 57, 62, 66, 97, 107, 117, 131, 132, 135, 142, 147	Resistentes ao Indinavir
3	2, 5, 7, 27, 29, 33, 34, 36, 41, 49, 50, 52, 61, 67, 68, 72, 73, 75, 76, 77, 78, 81, 88, 95, 96, 98, 100, 101, 102, 106, 112, 113, 119, 136, 137, 139, 144, 146, 149	Resistentes ao Nelfinavir
4	7, 12, 13, 14, 20, 21, 22, 23, 30, 42, 51, 55, 59, 60, 66, 70, 71, 99, 109, 111, 114, 127, 130, 133, 140	Resistentes ao Lopinavir
5	11, 17, 35, 39, 47, 48, 63, 79, 84, 86, 94, 104, 105, 116, 125, 126, 138, 148	Resistentes ao Atazanavir
6	26, 38, 58, 85	Resistentes ao Saquinavir
7	6, 32, 44, 58, 60, 82, 83, 90, 92, 93, 103, 122, 124, 128, 129	Resistentes ao Amprenavir

Nas tabelas 4.36, 4.37, 4.38, 4.39, 4.40, 4.41 e 4.42 temos os resultados dos agrupamentos dos resultados da rede Fuzzy ARTMAP, com os dados da base híbrida, composta por informações do laboratório de virologia molecular da UFRJ e da base pública de Stanford, para paciente do subtipo B. Os dados foram agrupados e em sequência foi feita uma análise com a finalidade de verificar a característica daquele grupo, destacando desta forma, a resistência predominante no agrupamento.

Tabela 4.36 - Análise dos Resultados do Agrupamento 1 (não resistentes)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Não Resistentes	Resistentes
Não Resistentes	23 (85,2%)	4 (14,8%)
Resistentes	9 (7,3%)	114 (92,7%)
TOTAL	32	118

Tabela 4.37 - Análise dos Resultados do Agrupamento 2 (resistente ao IDV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao IDV	Não Resistentes ao IDV
Resistentes ao IDV	9 (60,0%)	6 (40,0%)
Não Resistentes ao IDV	8 (5,9%)	127 (94,1%)
TOTAL	17	133

Tabela 4.38 - Análise dos Resultados do Agrupamento 3 (resistente ao NFV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao NFV	Não Resistentes ao NFV
Resistentes ao NFV	34 (75,6%)	11 (24,4%)
Não Resistentes ao NFV	5 (4,8%)	100 (95,2%)
TOTAL	39	111

Tabela 4.39 - Análise dos Resultados do Agrupamento 4 (resistente ao LPV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao LPV/r	Não Resistentes ao LPV/r
Resistentes ao LPV/r	19 (82,6%)	4 (17,4%)
Não Resistentes ao LPV/r	6 (4,7%)	121 (95,3%)
TOTAL	25	125

Tabela 4.40 - Análise dos Resultados do Agrupamento 5 (resistente ao ATV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao ATV	Não Resistentes ao ATV
Resistentes ao ATV	14 (77,8%)	4 (22,2%)
Não Resistentes ao ATV	4 (3,0%)	128 (97,0%)
TOTAL	18	132

Tabela 4.41 - Análise dos Resultados do Agrupamento 6 (resistente ao SQV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao SQV	Não Resistentes ao SQV
Resistentes ao SQV	2 (25,0%)	6 (75,0%)
Não Resistentes ao SQV	2 (1,4%)	140 (98,6%)
TOTAL	4	146

Tabela 4.42 - Análise dos Resultados do Agrupamento 7 (resistente ao APV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao APV	Não Resistentes ao APV
Resistentes ao APV	13 (92,9%)	1 (7,1%)
Não Resistentes ao APV	2 (1,5%)	134 (98,5%)
TOTAL	15	135

#### 4.2.2 – Redes Fuzzy ArtMap – Subtipo C

TABELA 4.43  
Distribuição da Amostra da Base C - Base Híbrida

BASE	N/RES.	SQV	IDV	NFV	APV	ATV	LPV/r	Total
Treino	28	2	9	24	10	11	23	106
Teste	11	0	4	9	4	6	9	44
Total	39	2	13	33	14	17	32	150

LEGENDA:

N/RES - Não Resistentes

SQV - Resistente ao inibidor Saquinavir

IDV - Resistente ao inibidor Indinavir

NFV - Resistente ao inibidor Nelfinavir

APV - Resistente ao inibidor Amprenavir

ATV - Resistente ao inibidor Atazanavir

LPV/r - Resistente ao inibidor Lopinavir/ritonavir

Na tabela 4.44 temos os agrupamentos mais significativo, obtido com o parâmetro de vigilância 0,4, onde os pacientes que compõem a amostra de teste estão destacados.

TABELA 4.44  
Agrupamento com o Parâmetro de Vigilância 0,4 – Fuzzy ArtMap Subtipo C – Base Híbrida

CONJUNTO DE TREINO E TESTE		
Grupo	Pacientes Selecionados	Característica Principal do Agrupamento
1	3, 4, 9, 10, 13, 16, 18, 22, 25, 35, 41, 56, 57, 61, 66, 69, 70, 80, 83, 84, 90, 91, 93, 94, <span style="border: 1px solid black; padding: 2px;">106, 107, 110, 132, 133</span>	Resistentes ao Nelfinavir
2	1, 2, 26, 27, 28, 38, 77, 78, <span style="border: 1px solid black; padding: 2px;">111, 113, 116, 121, 130</span>	Resistentes ao Indinavir
3	5, 19, 21, 24, 29, 30, 39, 43, 47, 49, 50, 51, 53, 59, 64, 67, 71, 72, 74, 81, 85, 87, 89, 92, <span style="border: 1px solid black; padding: 2px;">109, 115, 117, 119, 125, 126, 131,</span> <span style="border: 1px solid black; padding: 2px;">138, 139, 143, 144, 145</span>	Não resistentes aos inibidores de protease
4	12, 14, 15, 20, 31, 36, 37, 42, 44, 48, 62, 63, 65, 68, 75, 76, 79, 82, 95, 98, 99, 101, 102, 103, 104, 105 <span style="border: 1px solid black; padding: 2px;">112, 114, 118, 120, 124, 128, 134,</span> <span style="border: 1px solid black; padding: 2px;">136, 140, 141, 150</span>	Resistentes ao Lopinavir
5	8, 23, 34, 45, 52, 54, 55, 58, 73, 96, 97, <span style="border: 1px solid black; padding: 2px;">122, 123, 129, 135, 137, 146</span>	Resistentes ao Atazanavir
6	1, 40, 88, <span style="border: 1px solid black; padding: 2px;">127, 147</span>	Resistentes ao Saquinavir
7	6, 7, 11, 17, 32, 33, 46, 60, 86, <span style="border: 1px solid black; padding: 2px;">108, 142, 148, 149</span>	Resistentes ao Amprenavir

Nas tabelas 4.45, 4.46, 4.47, 4.48, 4.49, 4.50 e 4.51 temos os resultados dos agrupamentos dos resultados da rede Fuzzy ARTMAP, com os dados da base híbrida, composta por informações do laboratório de virologia molecular da UFRJ e da base pública de Stanford, para paciente do subtipo C. Os dados foram agrupados e em sequência foi feita uma análise com a finalidade de verificar a característica daquele grupo, destacando desta forma, a resistência predominante no agrupamento.

Tabela 4.45 - Análise dos Resultados do Agrupamento 1 (resistente ao NFV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao NFV	Não Resistentes ao NFV
Resistentes ao NFV	26 (78,8%)	7 (21,2%)
Não Resistentes ao NFV	3 (2,6%)	114 (97,4%)
TOTAL	29	121

Tabela 4.46 - Análise dos Resultados do Agrupamento 2 (resistente ao IDV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao IDV	Não Resistentes ao IDV
Resistentes ao IDV	9 (69,2%)	4 (30,8%)
Não Resistentes ao IDV	4 (2,9%)	133 (97,1%)
TOTAL	13	137

Tabela 4.47 - Análise dos Resultados do Agrupamento 3 (não resistentes)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Não Resistentes	Resistentes
Não Resistentes	33 (86,8%)	5 (13,2%)
Resistentes	3 (2,7%)	109 (97,3%)
TOTAL	36	114

Tabela 4.48 - Análise dos Resultados do Agrupamento 4 (resistente ao LPV/r)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao LPV/r	Não Resistentes ao LPV/r
Resistentes ao LPV/r	30 (93,8%)	2 (6,2%)
Não Resistentes ao LPV/r	7 (5,9%)	111 (94,1%)
TOTAL	37	113

Tabela 4.49 - Análise dos Resultados do Agrupamento 5 (resistente ao ATV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao ATV	Não Resistentes ao ATV
Resistentes ao ATV	12 (70,6%)	5 (29,4%)
Não Resistentes ao ATV	5 (3,8%)	128 (96,2%)
TOTAL	17	133

Tabela 4.50 - Análise dos Resultados do Agrupamento 7 (resistente ao SQV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao SQV	Não Resistentes ao SQV
Resistentes ao SQV	1 (50,0%)	1 (50,0%)
Não Resistentes ao SQV	4 (2,7%)	144 (97,3%)
TOTAL	5	145

Tabela 4.51 - Análise dos Resultados do Agrupamento 7 (resistente ao APV)

CLASSE	RESPOSTA DA REDE (PORCENTAGEM)	
	Resistentes ao APV	Não Resistentes ao APV
Resistentes ao APV	10 (71,4%)	4 (28,6%)
Não Resistentes ao APV	4 (2,9%)	132 (97,1%)
TOTAL	14	136

---

## Capítulo 5

### Discussão dos Resultados

---

Como relatado nas seções anteriores uma das maiores problemáticas na falha terapêutica é o acúmulo de resistências às drogas antirretrovirais, o que torna o sucesso da terapia da AIDS um problema mundial. A predição a resistência de um antirretroviral possibilita uma alteração no regime terapêutico, antes que essa falha venha causar um aumento na carga viral, possibilitando dessa forma a invasão de doenças oportunas causadas pela baixa imunidade do indivíduo.

Nesse trabalho apresentamos dois métodos baseados na teoria da ressonância adaptativa, que através da seleção de resíduos de aminoácidos (posições de mutação), na sequência da protease do HIV-1, foram capazes de categorizar a resistência aos antirretrovirais. Os primeiros testes foram realizados usando as RNA ArtMap, com uma codificação binária. Afim de verificar a capacidade de classificação desse modelo fizemos testes usando distintos parâmetros de vigilância. A codificação dos resíduos de aminoácidos pela escala de hidrofobicidade possibilitou a extensão do modelo computacional para as RNA Fuzzy ArtMap, levando em consideração as propriedades físico-químicas dos aminoácidos, propriedades estas que são independentes ao Subtipo do HIV-1.

Analisando os diversos experimentos realizados nesse trabalho conclui-se que não houve uma relevância significativa acerca dos resultados obtidos na aplicação da rede neural ArtMap comparada com a rede Fuzzy ArtMap.

Como não foi relevante a diferença dos resultados fica difícil afirmar qual das duas redes teve o melhor desempenho, todavia, pelo foco biológico a aplicação Fuzzy ArtMap é muito mais interessante, pois ela não considera apenas a ocorrência da mutação, mas também, o impacto que a mutação causa.

No tocante as problemáticas dos testes destaco que para ambas as redes, a complexidade foi a escolha do melhor parâmetro de vigilância e mais especificamente para as redes Fuzzy ArtMap a proliferação no número de categorias. Como esse tipo de rede trabalha com intervalo de valores entre 0 e 1, observamos uma dificuldade em juntar ou aproximar pacientes ao grupo que continha sua característica de terapia. Com isso gerar os resultados para Fuzzy ArtMap requereu uma dedicação maior na

escolha dos parâmetros da rede e, em consequência, uma junção de agrupamentos com as mesmas características, diminuindo o excesso de agrupamentos.

Para discutir nossos resultados verificamos a validade preditiva dos mesmos usando os conceitos de sensibilidade (SENS), especificidade (ESP) e acurácia (ACR).

Sensibilidade e especificidade descrevem a proporção do resultado positivo ou negativo em quem, sabidamente está, ou não, resistente. Valores preditivos positivos e negativos contêm informações sobre o poder do teste (sensibilidade e especificidade) e da população a ser examinada (prevalência da resistência), sendo uma medida de maior utilidade clínica, foi descartada em nossa tabela de discussão.

A acurácia foi um fator usado para verificar o grau em que os testes realizados foram capazes de determinar o verdadeiro valor da predição a resistência dos antirretrovirais, ou seja, a acurácia mostra se os resultados obtidos realmente representam a “verdade” ou quanto esses se afastam dela. Como os nossos resultados estão baseados na presença ou ausência de resistências, a acurácia está baseada na **sensibilidade e especificidade**.

## 5.1 – 1ª Etapa – Discussão dos Resultados da Base de Dados RENAGENO - Subtipo B

Com a finalidade de verificar a performance dos dois modelos analisamos os resultados na tabela 5.1.

Tabela 5.1 - Análise dos Resultados da Base de Dados RENAGENO – Subtipo B

TERAPIA	ArtMap			Fuzzy ArtMap		
	SENS	ESP	ACR	SENS	ESP	ACR
Não Resistentes aos Inibidores	83%	88%	86%	95%	91%	92%
Resistentes ao Nelfinavir	91%	96%	95%	96%	97%	97%
Resistentes ao Lopinvair/r	77%	94%	92%	90%	95%	95%
Resistentes ao Amprenavir	38%	98%	95%	Não ocorreu agrupamento		
Resistentes ao Atazanavir	69%	99%	96%	45%	99%	91%
Resistentes ao Indinavir	70%	96%	94%	69%	97%	95%
Resistentes ao Saquinavir	63%	98%	96%	50%	99%	95%



No que se refere aos inibidores de protease NFV e LPV para o Subtipo B, em ambas as redes, verificou-se um feedback promissor. Este fato pode ser justificado pois na regra de corte preservamos posições clássicas de mutações desses dois inibidores (D30, L33, I54, N88, L90), indicando assim, que as posições de mutações selecionadas pelo modelo computacional contribuíram para um melhor desempenho na predição desses inibidores.

Em relação ao inibidor SQV e APV os resultados de sensibilidade foram os mais inferiores, fato que pode ser justificado em função da barreira genética do HIV-1 para os mesmos e, também, na amostra de pacientes experimentados temos um número bem reduzido de resistente a eles, principalmente para o SQV. Além disso com o corte do pré-processamento perdemos posições importantes de resistência para estes casos, como por exemplo a posição I84.

## 5.2 – 2ª Etapa – Discussão dos Resultados da Base de Dados RENAGENO - Subtipo C

Tabela 5.2. - Análise dos Resultados da Base de Dados RENAGENO – Subtipo C

TERAPIA	ArtMap			Fuzzy ArtMap		
	SENS	ESP	ACR	SENS	ESP	ACR
Não Resistentes aos Inibidores	87%	75%	79%	93%	81%	85%
Resistentes ao Nelfinavir	95%	94%	94%	96%	96%	96%
Resistentes ao Lopinavir/ritonavir	65%	97%	93%	77%	96%	94%
Resistentes ao Amprenavir	44%	97%	93%	60%	99%	96%
Resistentes ao Atazanavir	50%	98%	93%	54%	98%	93%
Resistentes ao Indinavir	36%	98%	93%	33%	97%	93%
Resistentes ao Saquinavir	25%	99%	94%	33%	100%	95%

Análogo aos casos do Subtipo B os resultados mais significativos foram para as resistências aos inibidores NFV e LPV/r, fato que pode ser justificada, também, pela preservação de mutações clássicas a esses inibidores

Um diferencial dos resultados para o Subtipo C é que em ambas as redes ocorreu o agrupamento do inibidor APV.

O número de pacientes não resistentes são a maioria nas duas bases de dados para ambos Subtipos e essa categoria foi agrupada de forma eficiente, pelas duas redes, porém novamente, a RNA Fuzzy ArtMap se destacou.

Quando comparamos os resultados obtidos na previsão de resistência para o inibidor de protease NFV, LPV/r e pacientes não resistentes nos Subtipos B e C, verificamos que a acurácia média obtida no Subtipo B na categorização de resistência é superior ao valor obtido no Subtipo C, levando a concluir que o perfil mutacional do Subtipo C é diferente quando comparado ao Subtipo B. Da mesma forma se compararmos os Subtipos de forma independente, focando a análise no tipo de rede neural utilizada, veremos que acurácia da RNA Fuzzy ArtMap é maior, porém em uma pequena escala.

### 5.3 – 3ª Etapa – Discussão dos Resultados da Base de Dados Híbrida

No caso da base de dados híbrida os testes só foram realizados usando as redes neurais Fuzzy ArtMap, pois verificamos na aplicação do modelo anterior que essa rede apresentou melhor performance. Desta forma, nesse tópico, focamos em comparar os resultados pelo Subtipo de virologia e não pelo tipo de rede.

Tabela 5.3 - Análise dos Resultados da Base de Dados Híbrida – RNA Fuzzy ArtMap

TERAPIA	Subtipo B			Subtipo C		
	SENS	ESP	ACR	SENS	ESP	ACR
Não Resistentes aos Inibidores	72%	97%	91%	92%	96%	95%
Resistentes ao Nelfinavir	87%	90%	89%	90%	94%	93%
Resistentes ao Lopinavir/ritonavir	76%	97%	93%	81%	98%	94%
Resistentes ao Amprenavir	87%	99%	98%	71%	97%	95%
Resistentes ao Atazanavir	78%	97%	95%	71%	96%	93%
Resistentes ao Indinavir	53%	95%	91%	69%	97%	95%
Resistentes ao Saquinavir	50%	96%	95%	20%	99%	97%

Os resultados acima correspondem a terceira modelagem aplicada nessa pesquisa, na qual, utilizamos o conjunto Fuzzy para definir os valores de entrada para o treinamento da rede. A aplicação da escala de hidrofobicidade de aminoácidos com seus respectivos pesos moleculares, foi eficiente na codificação dos aminoácidos da

sequência do gene da protease. Esse fato é justificado pois o valor de hidrofobicidade é independente ao Subtipo do HIV-1.

Essa terceira abordagem do trabalho definiu os melhores resultados da pesquisa, o qual da mesma forma que aconteceu com a base de dados RENAGENO, a predição de resistência ao NFV, LPV/r e não resistentes foram os que se destacaram. Todavia os resultados para a resistência de APV e ATV, também foram bem relevantes.

A diferença da média da acurácia dos Subtipos B e C foi bem pequena, porém vale destacar que os melhores resultados foram para o Subtipo C, pois este subtipo possui como característica um índice menor de mutações.

.

---

## Capítulo 6

### Conclusões

---

Nesse trabalho, tratamos o problema na previsão de resistência à terapia antirretroviral nos Subtipos B e C, um problema de difícil solução devido à grande complexidade mutacional do HIV-1.

A contribuição dessa pesquisa consistiu o desenvolvimento de dois modelos computacionais: um usando redes neurais ArtMap e outro usando redes neurais Fuzzy ArtMap. No decorrer da pesquisa como observamos um desempenho maior para as RNA Fuzzy ArtMap, resolvemos testar a tecnologia com duas bases de dados distintas, afim de verificar se a performance dessa rede foi influenciada pelos dados selecionados ou se realmente de acordo, com os pré-processamentos estabelecidos, a rede Fuzzy ArtMap se tonaria uma boa solução para o estudo em questão.

Os modelos computacionais propostos mostraram-se promissores, quanto à utilização de da metodologia de pré-processamento aplicada, assim como, a aplicação das redes neurais ArtMap e Fuzzy ArtMap para predizer à resistência aos inibidores de protease nelfinavir e lopinavir/r e pacientes não resistentes, a partir de informações genótípicas do gene da protease do HIV-1.

Os resultados obtidos pelos modelos não possuem uma diferença relevante de forma a concluir qual dos dois seria mais promissor no foco computacional. Porém como já citamos no decorrer da pesquisa, esse trabalho, é um contribuinte na área biológica, onde podemos afirmar que nesse foco as redes Fuzzy ArtMap são mais relevantes, pois consideram o tipo de mutação ocorrida e o grau de relevância de cada mutação, e como esse nível de relevância pode contribuir ou não na classificação dos pacientes em seus determinados grupos de terapia.

Apesar dos bons resultados obtidos com a aplicação do modelo, há a necessidade de aumentar o número de amostras de pacientes resistentes, principalmente aos inibidores SQV, APV e IDV.

Durante o desenvolvimento da pesquisa tentamos utilizar a base de dados híbrida com 320 pacientes, sem aplicar nenhum tipo de corte nos mesmos, porém surgiram duas problemáticas: a sensibilidade a ordem de treinamento e a proliferação de categorias. Dentre esses dois problemas a proliferação de categorias é um dos

principais para esse tipo de rede. Em motivação a este fato ocorre a necessidade de estudar uma forma para minimizar essa problemática nos trabalhos futuros, visto que, um dos objetivos para os próximos passos da pesquisa é aumentar a amostra de dados de pacientes resistentes.

Com relação a base híbrida o objetivo foi comparar o desempenho da rede Fuzzy ArtMap para os subtipos B e C. Nos conceitos de sensibilidade, especificidade e acuraria, analisando os antirretrovirais com melhores resultados, observamos que os resultados obtidos para o subtipo C foram mais significativos, fato que pode ser considerado, pois o número de mutações é menor do que para o subtipo B.

Mesmo pontuando os resultados como muito bons precisamos fazer alguns ajustes na modelagem e aprimorar o conceito da aplicação das técnicas abordadas. Em virtude disso como trabalhos futuros iremos propor:

- 1) Aumentar a amostra de ambas as bases de dados para os pacientes resistentes;
- 2) Fazer um estudo detalhado nas posições clássicas de mutação de cada inibidor, de forma que não ocorra perda dessas posições com o corte;
- 3) Incluir na modelagem mais dois dados: Contagem de CD4+ e a carga viral, para verificar a existência de correlação entre a contagem desses parâmetros na terapêutica utilizada e os dados do sequenciamento do gene pol, na resistência do vírus HIV.

---

## Referências Bibliográficas

---

ALVES, M. R. **Construção de Base de Regras Fuzzy para Identificação de Malignidades Humanas a partir de Experimentos de Microarranjos de DNA**. 214 p. Tese (Programa de Pós-graduação em Engenharia - Doutor em Engenharia Biomédica). Universidade Federal do Rio de Janeiro, 2006. Rio de Janeiro- RJ.

BARRÉ-SINOUSI, F.; CHERMANN, J. C.; REY, F. *et al.* **Isolation of a Tlymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)**. *Science*, v. 220 n.4599, pp.868–71, 1983.

BEERENWINKEL, N., SHIMIDT, B., WALTER, H. *et al.* **Diversity and complexity of HIV-1 drug resistance: a Bioinformatics approach to predicting phenotype from genotype**. *PNAS* v.99, n.12 (jun), pp. 8271-8276, 2002.

BONET, I.; GARCIA, M. M.; SAYES, Y. *et al.* **Predicting Human imunodeficiency vírus (HIV) drug resistance using recurrent neural networks**. *IWINAC Part I, LNCS* v. 4527, pp. 234–243, 2007.

CARPENTER, G. A.; GROSSBERG, S.; ROSEN, D. B. **Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system**. *Neural Networks*, New York, v.4, n.1, p.759-771, 1991a.

CARPENTER, G. A.; GROSSBERG, S.; REYNOLDS, J. H. **ARTMAP: supervised real-learning and classification of non-stationary data by a self-organizing neural network**. *Neural Network*, New York, v.4, n.5, p.565-588, 1991b.

CARPENTER, G. A.; GROSSBERG, S.; MARKUZON, N.; REYNOLDS, J. H.; ROSEN, D. B. **Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps**. *IEEE Transactions on Neural Networks*, New York, v.3, n.5, p.689-715, 1992.

DEFORCHE, K.; SILANDER, T., CAMACHO, R., *et al.* **Analysis of HIV-1 pol sequences using Bayesian networks: implications for drug resistance.** *Bioinformatics*, v22, p. 2975-2979, 2007.

DRAGHICI, S.; POTTER, R. **Predicting HIV drug resistance with neural Networks,** *Bioinformatics*, v.19, pp.98-107, 2003.

FERNANDES, R.C.S. **Análise de aglomerados das sequências da protease do HIV-1 de indivíduos infectados utilizando mapa auto organizável e K-Médias.** 69 p. Dissertação (Programa de Pós-graduação em Engenharia - Doutor em Engenharia Biomédica). Universidade Federal do Rio de Janeiro, 2010. Rio de Janeiro- RJ.

FERREIRA, A.S. **Resistência primária aos antirretrovirais e mapeamento genético do HIV-1 no Estado do Mato Grosso.** 99 p. Dissertação (Programa de Pós-Graduação em Ciências da Saúde – Mestre em Ciências da Saúde). Universidade Federal de Goiás, 2011. Goiânia - GO.

GALLO, R. C.; SARIN, P. S.; GELMANN, E. P. *et al.* **Isolation of Human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS).** *Science*, v.220, n.4599, pp.865–867, 1983.

GAO, F.; VIDAL, N.; LI, Y., *et al.* **Evidence of Two Distinct Subsubtypes within the HIV-1 Subtype A Radiation,** *AIDS Research and Human Retroviruses*, v. 17(8), pp. 675-688, 2001.

GAO, F.; BAILES, E.; ROBERTSON. **Origin of HIV-1 in the Chimpanzee Pan Troglodytes .** *Nature*, v. 397, pp.436–441, 1999.

GROSSBERG, S. Adaptive pattern classification and universal recoding, I: parallel development and coding of neural feature detectors. *Biological Cybernetics*, New York, v.23, n.3, p. 121-134, 1976.

HEDSTROM L.; LIECHTI G.; GOLDBERG J.B.; GOLLAPALLI D.R. **The antibiotic potential of prokaryotic IMP dehydrogenase inhibitors.** *Curr Med Chem.*18(13):1909-18, 2011.

HEDSTROM L. **Serine Protease Mechanism and Specificity.** Department of Biochemistry, MS 009, Brandeis University, Waltham, Massachusetts 02454, 2002.

KANTOR, R.; KATZENSTEIN, D. **Drug resistance in non-subtype B HIV-1.** *Journal of Clinical Virology*, v. 29, pp. 152-159, 2004.

LEVY, J. A., HOFFMAN, A. D., KRAMER, S. M. *et al.* **Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS.** *Science*, v. 225, n.4664, pp.840–842, 1984.

MINISTÉRIO DA SAÚDE DO BRASIL. **Recomendações para terapia antirretroviral em adultos e adolescentes infectados pelo HIV – 2013/2014.** (Programa Nacional de DST/AIDS), 2014.

MINISTÉRIO DA SAÚDE DST/AIDS. **Boletim Epidemiológico de AIDS/DST 2014.** Disponível em: <http://www.aids.gov.br>, acessado em abril de 2014.

MORGADO, M.G.; GUIMARAES, M.L.; GRIPP, C.B.W.G *et al.* HEC AIDS Clin Res Group. **Molecular epidemiology of HIV in Brazil: High prevalence of B subtype and identification of an HIV-1 subtype D infection in Rio de Janeiro City.** *J of AIDS & Human Retrov*, v. 18, p. 488-494, 1998.

OLIVEROS, M.P.R. **Revolução das Mutações de Resistência aos Inibidores de Protease em Pacientes Infectados pelo HIV-1 Subtipo F.** 129 p. Tese (Programa de Pós-graduação em Ciências – Doutor em Ciências). Universidade de São Paulo, 2009. São Paulo – SP.

OLIVEROS, M.P.R. **Prevalência e Covariação de Mutações Relacionadas aos Inibidores de Protease no Subtipo F do HIV-1.** 111 p. Dissertação (Programa de



Pós-graduação em Ciências – Mestre em Ciências). Universidade de São Paulo, 2005. São Paulo – SP.

PORTAL DA EDUCAÇÃO. **Terapia Antirretroviral**. Disponível em: <http://www.portaleducacao.com.br/farmacia/artigos/7829/terapia-anti-retroviral#ixzz386KI4hT2>, acessado em 01 de junho de 2014

POTTS, K.E.; KALISH, M.L.; LOTT, T. *et al* 1993. **Genetic heterogeneity of the V3 region of the HIV-1 envelope glycoprotein in Brazil**. Brazilian Collaborative AIDS Research Group. AIDS, v. 7, pp. 1191-1197, 1993

RAMBAUT, A., POSADA, D., CRANDALL, A. K. *et al*. **The causes and consequences of HIV evolution**. Reviews Nature, 2004.

SANTOS, P.C.C. **Banco de Dados Inteligente e Ferramentas Associadas de Sequência, Mutações e Resistências aos Antirretrovirais do Virus HIV**. 108 p. Dissertação (Programa de Pós-graduação Interunidades em Biotecnologia – Doutor em biotecnologia). Universidade de São Paulo, 2010. São Paulo – SP.

SANCHES, M.; KRAUCHENCO, S.; MARTINS, N.H. *et al*. 2007. **Structural Characterization of B and non-B Subtypes of HIV-Protease: Insights into the Natural susceptibility to drug resistance**. J. Mol. Biol. v. 369, pp.1029–1040, 2007.

SHAFER, R.W., RHEE, S.Y., PILLAY, D., *et al.*, 2007. **HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance**, *AIDS*, v.21, pp. 215-223.

SHAFER, R. W., 2002, **Genotypic testing for human immunodeficiency virus type 1 drug resistance**, *Clinical Microbiology Reviews*, v.15, pp.247- 277.

SILVA, R. M. **Algoritmo genético e kernel discriminante de Fisher aplicado a identificação de mutações de resistência do HIV-1 aos inibidores antirretrovirais da Protease**. 126 p. Tese (Programa de Pós-graduação em Engenharia Biomédica -

Doutor em Engenharia Biomédica). Universidade Federal do Rio de Janeiro (UFRJ), 2009. Rio de Janeiro - RJ.

SILVEIRA, M.C.G. **Análise de Estabilidade Transitória de Sistemas Elétricos por Redes Neurais ARTMAP Nebulosas Modulares**. 107 P. Tese (Programa de Pós-graduação em Engenharia Elétrica – Doutora em Engenharia Elétrica). Faculdade de Engenharia de Ilha Solteira, 2003. Ilha Solteira – SP.

SING, T., BEERENWINKEL, N. **Mutagenetic tree Fisher kernel improves prediction of HIV drug resistance from viral genotype**, *Advances in Neural Information Processing Systems 19*. MA, MIT, USA, pp. 1-9, 2007.

SOUZA, C.R. **Redes neurais sem-peso aplicadas na categorização de subtipos do HIV-1**. 101 p. Dissertação (Programa de Pós-graduação em Engenharia de Sistemas e Computação - Mestre em Engenharia de Sistemas e Computação). Universidade Federal do Rio De Janeiro (UFRJ), 2010. Rio de Janeiro - RJ.

UNAIDS, 2014. **AIDS epidemic update**. Disponível em: [http://www.unaids.org/en/HIV\\_data/Epidemiology/default.asp](http://www.unaids.org/en/HIV_data/Epidemiology/default.asp), Acessado em 30 de março de 2014.

VELLA S.; PALMISANO L. **Antiviral therapy: state of the HAART**. *antiviral Res* 45: 1-7, 2000.

VANDAMME, A., SONNERBORG, A., AIT-KHALED, M., *et al.*, 2004. **Updated European recommendations for the clinical use of hiv drug resistance testing**, *Antiviral Therapy*, v.9(6), pp.829-848.

WANG, D., LARDER, B. **Enhanced prediction of Lopinavir resistance from genotype by use of artificial neural networks**. *J. Infect. Dis.* 188, 653–660, 2003

WLODAWER, A., GUSTCHINA, A., 2000. **Structural and biochemical studies of retroviral proteases**, *Biochimica et Biophysica Acta*, v. 1477, pp. 16-34.