

UNIVERSIDADE DE LISBOA

Faculdade de Ciências

Departamento de Informática



**ANÁLISE DA PREVALÊNCIA DE HIV EM MULHERES
GRÁVIDAS DE MOÇAMBIQUE (2001-2007)**

Lélio Manuel Rungo Dauane

Mestrado em Tecnologias de Informação Aplicadas às
Ciências Biológicas e Médicas

2010

UNIVERSIDADE DE LISBOA

Faculdade de Ciências

Departamento de Informática



**ANÁLISE DA PREVALÊNCIA DE HIV EM MULHERES
GRÁVIDAS DE MOÇAMBIQUE (2001-2007)**

Lélio Manuel Rungo Dauane

Trabalho orientado pelo Prof. Dr. André Osório Cruz de Azerêdo Falcão

Mestrado em Tecnologias de Informação Aplicadas às
Ciências Biológicas e Médicas

2010

Resumo

A Vigilância Epidemiológica do HIV-SIDA, permite conhecer diferentes aspectos relacionados à distribuição e à disseminação da infecção pelo HIV. Dados obtidos por estes estudos são de fundamental importância para definir programas de prevenção e controle desta doença. Neste trabalho são analisados dados obtidos de 36 Postos Sentinela, distribuídos por todo o país, nos anos de 2001, 2002, 2004 e 2007.

O objectivo deste trabalho é descrever a variação temporal da progressão da epidemia do HIV e estimar a evolução espacial da doença através de uma análise de tendências.

Para proceder a análise foi necessário: filtrar a base de dados, eliminando ruídos aí contidos; combinar dados de fontes diferentes numa única base de dados; seleccionar variáveis de interesse para este trabalho; transformar os dados em formatos apropriados; extrair padrões de interesse; e por fim usar técnicas de visualização, a fim de apresentar o conhecimento extraído.

Os dados mostram uma diversidade regional dos cenários da prevalência do HIV e suas tendências. Enquanto as regiões Norte e Centro apresentam uma tendência estacionária, a região Sul continua registar uma tendência de crescimento da epidemia. As tendências regionais reproduzem-se a nível provincial, embora algumas províncias registem uma aparente diminuição da prevalência. As taxas de prevalências nas zonas transfronteiriças apresentam uma continuidade com as taxas de prevalência em regiões adjacentes de outros países.

Factores regionais e zonas fronteiriças dão a percepção de contribuir para os altos níveis de HIV-SIDA.

O aumento do conhecimento e alteração dos comportamentos individuais e sociais são essenciais para a prevenção do HIV e subversão da actual tendência de agravamento.

Este trabalho desenvolve um modelo de previsão da seroprevalência baseado em redes neuronais artificiais que prevê a redução das taxas de prevalência em todo o país, excepto os cornos das províncias de Niassa e Tete.

Palavras-Chave: HIV, SIDA, Moçambique, Vigilância epidemiológica.

Abstract

The Epidemiological Surveillance of HIV / AIDS, provides information about the distribution and dissemination of HIV infection. Data from these studies are fundamental to design programs for disease prevention and control. We reviewed data from 36 sentinel sites distributed all over the country in the years 2001, 2002, 2004 and 2007.

The purpose of this study is to describe the temporal variation of the progression of the HIV epidemic and to estimate the spatial evolution of the disease through a trends analysis.

To conduct the analysis it was necessary to: filter the database, eliminating noise contained therein; combine data from different sources into one database, select variables of interest for this work, transform the data into appropriate formats, extract patterns of interest, and finally use visualization techniques in order to present the extracted knowledge.

The analyzed data shows a diversity of regional scenarios of HIV prevalence and trends. While the North and Center have a stationary trend, the southern region continues to register an epidemic growth. Regional trends are reproduced at the provincial level, although some provinces registering an apparent decrease in prevalence. The prevalence rates in border areas have a continuity with the prevalence rates in adjacent regions of other countries.

Frontier areas and regional factors give the perception of contributing to the high levels of HIV-AIDS.

The increase of knowledge and the change of individual and social behavior are essential for HIV prevention and subversion of the current trend of deterioration.

In this work was developed a model for the forecast of seroprevalences based on the neuronal networks, which predicts a reduction of the prevalence rates in the country, except the horns of the Niassa and Tete province.

Keywords: HIV, AIDS, Mozambique, Epidemiological surveillance

AGRADECIMENTOS

Este espaço é dedicado àqueles que deram a sua contribuição para que esta dissertação fosse realizada. A todos eles deixo aqui o meu agradecimento sincero.

- Em primeiro lugar agradeço ao Prof. Doutor André Falcão pela forma como orientou o meu trabalho. As notas dominantes da sua orientação foram a utilidade das suas recomendações e a cordialidade com que sempre me recebeu. Estou grato por ambas e também pela liberdade de acção que me permitiu, que foi decisiva para que este trabalho contribuísse para o meu desenvolvimento pessoal.
- Em segundo lugar, agradeço a minha esposa Eulália, as minhas filhas Alicia e Thandywe, pela sua ajuda nos momentos mais críticos, por acreditarem no futuro deste projecto e contribuírem para o meu crescimento profissional e por ser também um exemplo a ser seguido.
- Gostaria ainda de agradecer aos meus pais, Fernando Dauane e Carolina Rungo, pelo amor, carinho, compreensão e carácter com que sempre me educaram e pelo esforço dedicado para que eu pudesse chegar até aqui;
- Aos meus irmãos, pelo carinho, paciência e incentivo.
- Deixo também uma palavra de agradecimento aos professores do DI/FCUL, pela forma como leccionaram o MIBM e por me terem transmitido o interesse por estas matérias. São também dignos de uma nota de apreço os colegas do mestrado pelas discussões e longas conversas, que em muito contribuíram para a realização da parte curricular do mestrado e deste trabalho, em particular, o Luís Filipe por ter feito uma revisão ao trabalho
- Ao Instituto Nacional de Saúde de Moçambique, pela cedência dos dados sem os quais não seria possível realizar esta dissertação.
- Ao meu cunhado Elias Macovela e a minha amiga Clara.
- Aos colegas, amigos moçambicanos e angolanos com quem partilhei o apartamento, um muito obrigado.
- Finalmente, gostaria de deixar um agradecimento muito especial a senhora Domingas Velema, por ter me recebido e acomodado em sua casa quando cheguei a Portugal.

Dedico esta tese

À minha esposa Eulália, as minhas filhas Alicia e Thandywe pelo apoio, força, incentivo, companheirismo e amizade.
Sem elas nada disso seria possível.

Conteúdo

Resumo	5
Abstract	7
Capítulo 1 INTRODUÇÃO	17
1.1 Enquadramento	17
1.2 Motivação	19
1.3 O que é o HIV?	20
1.4 Objectivos	22
1.4.1 Objectivo Geral.....	22
1.4.2 Objectivos Específicos	22
1.5 Metodologia	22
1.6 Organização do documento.....	23
Capítulo 2 VIGILÂNCIA EPIDEMIOLÓGICA EM MOÇAMBIQUE.....	24
2.1 Breve Historial da V.E. do HIV em Moçambique.....	24
2.2 Critérios para selecção dos primeiros Postos Sentinela.....	25
2.3 Selecção de outras Unidades Sanitárias para funcionarem como novos Postos Sentinela.....	26
2.4 Critérios para selecção das participantes	27
Capítulo 3 METODOLOGIA.....	28
3.1 Limpeza de dados	29
3.2 Integração de dados.....	29
3.3 Selecção	30
3.4 Transformação	30
3.5 Extracção.....	30
3.6 Avaliação ou pós processamento	31
3.7 Visualização de resultados	31
3.8 Inferência	31
Capítulo 4 RESULTADOS E DISCUSSÃO.....	33
4.1 Análise da situação actual do HIV em Moçambique.....	33
4.1.1 Taxas de prevalências regionais de HIV	35
4.1.2 Taxas de prevalência provinciais de HIV	37
4.1.3 Taxas de prevalência em diferentes faixas etárias.....	38
4.1.4 Taxas de prevalência de acordo com a ocupação.....	39
4.1.5 Taxas de prevalência do HIV nas zonas rurais e urbanas.....	40

4.1.6	Krigagem.....	42
4.1.7	Erros de krigagem.....	46
4.2	Inferência da prevalência do HIV	48
4.2.1	Redes neuronais artificiais	48
4.2.2	O processamento dos dados	50
4.2.3	Modelação da Rede Neuronal Artificial	51
4.2.4	Validação do modelo.....	51
4.2.5	Previsões das taxas de prevalência	54
	Capítulo 5 CONCLUSÕES	56
	BIBLIOGRAFIA	59
	ANEXOS	63

Lista de Acrónimos

ArcView	Software para Análise Espacial
C.S.	Centro de Saúde
CM	Conselho de Ministros
CNCS	Conselho Nacional de Combate ao Sida
H.G.	Hospital Geral
H.R.	Hospital Rural
HIV	Vírus da Imunodeficiência Humana
INCAM	Inquérito Nacional sobre Causas da Mortalidade em Moçambique
INE	Instituto Nacional de Estatística
INS	Instituto Nacional de Saúde
Interland	Terras do Interior Sem Acesso ao Mar
MISAU	Ministério da Saúde
P.S.	Posto Sentinela
PEN	Plano Estratégico Nacional
RNA	Rede Neuronal Artificial
SIDA	Síndrome de Imuno Deficiência Adquirida
SIG	Sistema de Informação geográfico
U.S.	Unidade Sanitária
V.E.	Vigilância Epidemiológica.

Lista de Figuras

Figura 1. Esquema de fases da descoberta de informação.....	28
Figura 2. Esquema da base de dados do INS.....	29
Figura3. Localização espacial dos Postos Sentinela.....	33
Figura 4. Prevalências regionais em Moçambique versus anos.....	35
Figura 5. Prevalências provinciais em Moçambique versus anos.....	37
Figura 6. Taxas de prevalências na faixa etária de 15 – 49 anos.....	38
Figura 7. Taxas de prevalências versus ocupação.....	39
Figura 8. Prevalências em zonas urbanas e rurais.....	40
Figura 9. Prevalências em zonas urbanas.....	41
Figura 10. Prevalências em zonas rurais.....	41
Figura 11. Evolução espacial das prevalências em Moçambique (2001- 2002).....	44
Figura 12. Evolução espacial das prevalências em Moçambique (2004- 2007).....	45
Figura 13. Prevalências em Moçambique e em países vizinhos.....	46
Figura 14. Erros da krigagem das prevalências em Moçambique (2001-2002).....	47
Figura 15. Erros da krigagem das prevalências em Moçambique (2002-2007).....	47
Figura 16. Exemplo de uma RNA com quatro camadas.....	49
Figura 17. Correlação entre valores previstos e reais.....	52
Figura 18. Relação entre o erro e o valor real.....	53
Figura 19. Histograma de resíduos.....	53
Figura 20. Variação do erro ao longo do tempo.....	53
Figura 21. Comparação entre prevalências de 2007 e as previsões para 2011.....	54
Figura 22. Previsões das prevalências do HIV-SIDA em 2013 e 2015.....	55

Lista de Tabelas, Fórmulas e Anexos

Tabela 1. Tabela explicativa dos atributos da tabela da modelação da RNA.....	32
Tabela 2. Prevalências do HIV nos Postos Sentinela no período de estudo.....	34
Tabela 3. Validação dos resultados com base no erro.....	52
Fórmula 1. Fórmula para o cálculo da taxa de prevalência.....	30
Fórmula 2. Fórmula para modelação da RNA.....	32
Fórmula 3. Fórmula para o calculo do erro do modelo da RNA.....	32
Anexo . Código R do treino da rede neuronal artificial e previsões	63

Capítulo 1 INTRODUÇÃO

1.1 Enquadramento

Com o presente trabalho **“Análise da Prevalência de HIV-SIDA em Mulheres Grávidas de Moçambique 2001-2007”** pretende-se dar a conhecer o perfil epidemiológico do HIV-SIDA, que tem causado altas taxas de mortalidade na população moçambicana.

Moçambique é um país que se localiza na costa oriental de África, e divide-se em 11 províncias, incluindo a cidade de Maputo que tem estatuto de província. Contudo, é muito frequente, mesmo a nível oficial, referir-se a Moçambique como tendo apenas 10 províncias, devido a não ser feita a separação entre a cidade e a província de Maputo.

Tendo como capital a cidade de Maputo, tem a Sul as províncias de Maputo, Gaza e Inhambane, ao Centro as províncias de Manica, Sofala, Tete e Zambézia e a Norte as províncias de Nampula, Niassa e Cabo Delgado.

Com uma área aproximada de 799.380 Km², faz fronteira a Norte com a República da Tanzânia, a Noroeste com o Malawi e a Zâmbia, a Oeste com o Zimbabwe e a República da África do Sul, e a Sul com a Suazilândia e ainda a África do Sul. O clima em Moçambique é de uma maneira geral tropical e húmido.

A epidemia do HIV-SIDA, representa um fenómeno global, dinâmico e instável, cuja forma de ocorrência nas diferentes regiões do mundo depende, entre outros determinantes, do comportamento humano individual e colectivo. O SIDA destaca-se entre as enfermidades infecciosas emergentes pela grande magnitude e extensão dos danos causados às populações. Desde a sua origem, cada uma de suas características e repercussões tem sido exaustivamente discutida pela comunidade científica e pela sociedade em geral [2].

A vigilância do HIV é recomendada pela Organização Mundial da Saúde (OMS), como sendo fundamental para colecta de dados para retratar uma situação mais recente da infecção [25].

Os dados obtidos de programas de vigilância do HIV são colhidos por meio de estudos transversais repetidos, periodicamente, numa determinada população. Estes estudos permitem investigar como a infecção pelo HIV se apresenta num determinado subgrupo populacional, num determinado espaço de tempo. Ao conjunto de estudos transversais periódicos dá-se o nome de "Vigilância por Postos-Sentinela". Resumidamente denominado de "Projeto-Sentinela", refere-se à colecta sistemática de dados para estabelecer as tendências espaço-temporais da infecção por HIV em subgrupos populacionais específicos, denominados, por sua vez, de "Populações Sentinela" [25].

Um dos objectivos do Projecto-Sentinela é o de estimar a prevalência de infecção pelo HIV na população-base da qual a amostra foi seleccionada. Aplicada ao denominador da população total, fornece o número total de infecções prevalentes. Mulheres gestantes são consideradas especialmente importantes para este tipo de investigação, porque elas são, geralmente, representativas de todas as mulheres em idade reprodutiva [5].

Estudos populacionais em países vizinhos mostraram que numa epidemia madura como a de Moçambique, as mulheres tendem a infectar-se mais cedo que os homens [2].

Embora o volume de dados disponíveis em Moçambique tenha crescido exponencialmente, continua a haver a necessidade de fortalecer a capacidade de análise desses dados. Além disso é necessário criar capacidade para centralizar e analisar as diversas fontes de dados, facilitando assim, o acesso aos dados e sua interpretação.

A resposta nacional face aos índices de HIV em Moçambique é recente. O CNCS foi criado em 2000 e o PEN II foi elaborado em 2003, fixando objectivos e estratégias para o período de 2005 a 2009. As acções definidas como prioritárias foram: prevenção; assistência jurídica, redução do estigma e discriminação, tratamento, mitigação, investigação e coordenação da resposta nacional. O PEN II assume a participação da sociedade civil na sua operacionalização e aponta a necessidade de respostas diversificadas de acordo com os diferentes cenários da epidemia no país. Isto exige das organizações a adequação das suas propostas às linhas de prioridade do PEN II e a interacção estreita com actores locais que influenciam os costumes e práticas nas comunidades [7].

Em Março de 2010, o Conselho de Ministros aprovou o PEN III, com um horizonte temporal de quatro anos (2010 – 2014), com o objectivo de agir de acordo com as evidências demonstradas em estudos e evidências baseadas em boas práticas versus evidências baseadas na pesquisa operacional [6].

Além de estimativas e as projecções, este trabalho também é de grande importância para entender o impacto futuro da epidemia e a resposta nacional necessária.

1.2 Motivação

Os estudos epidemiológicos que permitem conhecer diferentes aspectos relacionados à distribuição e à disseminação da infecção pelo HIV são de fundamental importância para subsidiar e saber como actuar nos programas de prevenção e controle do HIV-SIDA.

O INCAM, lançado pelo INE, aponta a malária como a principal causa de morte, no País, seguida do HIV/SIDA, com 29 e 27% respectivamente [15].

O HIV/SIDA é a principal causa de mortalidade entre pessoas dos 15 aos 49 anos, atingindo o pico no grupo dos 25 aos 49 anos.

O HIV/SIDA é o principal responsável pela mortalidade nas províncias do Sul, nomeadamente a de Gaza, Maputo província e Cidade, especialmente em zonas urbanas.

O perfil da epidemia do HIV em Moçambique é predominantemente heterossexual, estando mais disseminada entre as mulheres. Dentre os países da África subsahariana, a prevalência, embora alta, é menor do que a encontrada na África do Sul e maior do que a da Tanzânia, países com quem Moçambique faz fronteira. Há grande heterogeneidade entre as regiões do país, sendo possível encontrar um perfil mais próximo ao da África do Sul, por exemplo, nas províncias do Sul, onde a prevalência de infecção entre mulheres grávidas é alta. Mais próxima da Tanzânia, nas províncias do Norte, onde a percentagem da população muçulmana é maior e a seroprevalência entre gestantes é a mais baixa do país [8].

Com uma população predominantemente rural, atingida por frequentes desastres naturais, alta taxa de desemprego, baixa qualificação da mão-de-obra, e estando o país envolvido no processo de reconstrução e retomada de crescimento após a prolongada guerra civil, Moçambique apresenta preocupantes indicadores de saúde: a esperança de vida saudável ao nascer oscila em torno dos 37 anos; a mortalidade materna está estimada em torno de 1.000 por 100.000 nascidos vivos; e a mortalidade infantil atinge a taxa de 143 por 1.000 [15].

Apenas 12% das mulheres fazem uso de contraceptivos e menos da metade dos partos é realizada por profissionais treinados. O percentual de profissionais de saúde actuando no sistema público é de 3 por 100.000 habitantes, havendo, ainda, dificuldade de acesso a unidades de saúde em função das distâncias, da falta de estradas e de meios de transporte e do número reduzido de serviços. Os dados também apontam que a epidemia ainda está em expansão no sul do país [31].

A pluralidade de etnias, idiomas locais e religiões existente no país é um obstáculo para que mensagens e acções educativas, elaboradas a partir da perspectiva “ocidental” ou apenas na língua oficial, o Português, atinjam um número suficiente de pessoas.

Portanto, este documento poderá também servir de um instrumento orientador de políticas de redução das taxas de prevalências do HIV-SIDA em Moçambique.

1.3 O que é o HIV?

O vírus da imunodeficiência humana (HIV) é o agente causador da SIDA (Síndrome de Imunodeficiência Adquirida) sendo um vírus linfotrópico com afinidade preferencial para os linfócitos T CD4+ (responsáveis, em parte, pelo controlo do sistema imunológico).

De maneira análoga a outras viroses, o HIV é um parasita que se replica dentro das células hospedeiras [35], sendo que o tipo mais comum do vírus é conhecido como HIV-1 existindo outro tipo, o chamado HIV-2 que é, geralmente menos virulento, produzindo no entanto os mesmos efeitos registados para o HIV-1.

Foi isolado pela primeira vez em 1983 pelos pesquisadores Robert Gallo, nos EUA, e Luc Montagnier, na França. Em 1986, um comité internacional recomendou o termo HIV (vírus da imunodeficiência humana) para denominá-lo, reconhecendo-o como capaz de infectar seres humanos [18].

O caso mais antigo da infecção pelo HIV registado foi observado num marinheiro em Manchester que morreu com uma patologia semelhante à da SIDA em 1959. No entanto, a autenticidade deste facto não está confirmada. Além disso, análises genéticas de sequências de amostras obtidas no período entre 1971 a 1976, de membros de uma família Norueguesa infectados antes de 1971, mostraram que as

viroses foram adquiridas a partir de grupos estranhos ao local referido ou seja, uma estirpe de vírus existente, principalmente, na África Ocidental [35].

Embora não se conheça a origem do vírus, sabe-se que existe semelhança com a família de retro vírus que infecta os macacos verdes africanos, que vivem na África sub-Sahariana, chamada de vírus da imunodeficiência símia (SIV). Esta hipótese de introdução do SIV em humanos foi proposta por um antropólogo que estudou a tribo Igjiwi oriunda do Zaire. Este observou que, em rituais religiosos, o homem sacrificava o animal, fazendo a ingestão do seu sangue. Assim, o vírus SIV pode ter sido transmitido ao homem, sofrido mutação e atacado a espécie humana. Por isso, supõe-se que o HIV tenha origem no continente africano [18]. Além disto, a transferência de espécies cruzadas de HIV-1 e HIV-2 de chimpanzés e “*mangabeys*” para humanos pode ter provocado uma alteração da virulência, isto porque se sabe que alguns hospedeiros símios não desenvolvem SIDA. Esta alteração de virulência pode ser melhor detectada pela resposta do hospedeiro à infecção do que pelas propriedades do vírus. Uma possível explicação para este facto, é a de que a infecção do HIV activa o sistema imunitário humano em seu benefício, promovendo um conjunto de células tolerantes ao progresso da infecção e à apoptose (morte programada das células). A eventual diminuição da quantidade de linfócitos T-CD4⁺ denuncia o poder que o vírus tem contra o sistema imunitário [32].

Alguns estudos demonstraram que existe uma elevada taxa de variabilidade genética do vírus, tendo-se identificado, até agora, pelo menos doze subtipos genéticos, designados de A a J, pertencentes ao grande grupo do HIV-1 (estirpe dominante na epidemia global). Para muitas das regiões do genoma do HIV-1, os sub-tipos B e D estão mais relacionados entre si do que com qualquer outro sub-tipo do mesmo grupo. Contudo, análises filogenéticas das regiões env mostraram uma associação não usual entre viroses do sub-tipo B e F [32].

1.4 Objectivos

Neste trabalho gostaria de repensar a forma de abordar as prevalências, além das prevalências dos Postos Sentinela contribuir com a visão espacial da epidemia com base na interpolação de dados.

1.4.1 Objectivo Geral

- Análise da prevalência do HIV em mulheres gestantes de Moçambique

1.4.2 Objectivos Específicos

- Descrição temporal da progressão da doença em Moçambique
- Estimacão da evoluçao espacial da doença através de uma análise de tendências

1.5 Metodologia

Este projecto recorre às tecnologias de informacão e procura responder simultaneamente às necessidades de extracção, manipulacão e visualizacão de dados. SãO usadas metodologias de análise baseadas em aprendizagem automática, em dados de quatro censos, 2001, 2002, 2004 e 2007, referentes as rondas de V.E. do HIV em Moçambique.

Portanto a metodologia focalizar-se-á nas seguintes etapas:

Limpeza de dados: etapa onde sãO eliminados ruídos e dados inconsistentes.

Integracão de dados: etapa onde diferentes fontes de dados podem ser combinadas produzindo um único repositório de dados

Seleccão: etapa onde foram seleccionadas as variáveis que interessavam ao projecto

Transformacão: etapa onde os dados foram transformados em formatos apropriados para aplicacão de algum algoritmo.

Extracção: etapa essencial no processo, consistindo na aplicação de técnicas a fim de se extrair os padrões de interesse.

Avaliação ou pós processamento: etapa onde foram identificados os padrões interessantes de acordo com os objectivos do projecto

Visualização de resultados: etapa onde foram utilizadas técnicas de representação de conhecimento a fim de apresentar o conhecimento extraído.

Inferência: etapa onde foram utilizadas algumas técnicas de data mining para fazer previsões para os próximos anos.

1.6 Organização do documento

Este documento está organizado da seguinte forma:

Capítulo 1. Introdução, na qual se enquadra e define o problema e objectivos genéricos da dissertação.

Capítulo 2. Descreve a V.E. do HIV em Moçambique e os critérios usados para a selecção dos primeiros Postos Sentinela.

Capítulo 3. Descreve a metodologia utilizada ao longo do trabalho.

Capítulo 4. Apresenta e discute os resultados alcançados no projecto

Capítulo 5. São expostas as principais conclusões do trabalho e futuras direcções.

Capítulo 2 VIGILÂNCIA EPIDEMIOLÓGICA EM MOÇAMBIQUE

2.1 Breve Historial da V.E. do HIV em Moçambique

Em Moçambique, as prevalências do HIV-SIDA obtêm-se através de um processo permanente de V.E. em mulheres grávidas. As rondas de V.E. servem para estimar a prevalência de HIV e Moçambique, como alguns outros países, serve-se do mesmo sistema para fazer as suas estimações [23].

Os dados são recolhidos em unidades sanitárias que funcionam como Postos Sentinela para a V.E.

Em 1988 entrou em funcionamento na cidade de Maputo o Posto Sentinela C.S. José Macamo, em 1994 começaram a operar mais dois postos, no C.S. nº3 na cidade de Tete e no C.S. Eduardo Mondlane na cidade de Chimoio. Com a inclusão de mais um posto sentinela em 1996, o C.S. Ponta Gea na cidade da Beira, o número de postos passou para quatro, até ao ano de 1999. Até aqui não abrangia a região norte e as zonas rurais.

Em 2000, o número de Postos Sentinela aumentou de 4 para 20, sendo 11 urbanos e 9 rurais, dos quais 7 foram localizados na região Norte, 8 na região Centro e 5 na região Sul.

Em 2001 passou-se de 20 a 36 Postos Sentinela, numero que se mantém até ao presente.

2.2 Critérios para selecção dos primeiros Postos Sentinela

Moçambique possui 128 distritos administrativos e era impossível o estabelecimento de 1 Posto Sentinela em cada distrito. Deste modo, fez-se corresponder distritos similares a cada posto sentinela com as mesmas características.

Os critérios de similaridade tomaram em conta os seguintes aspectos:

Similaridade nos **movimentos populacionais** (ex: quantidade de tráfego de transportes comerciais, volume de comerciantes no sector informal que se deslocam, quantidade de trabalhadores migratórios nas minas e na agricultura, etc.);

Facilidades de **comunicação** entre esses distritos (vias de acesso, transporte, rádio, língua, acesso aos países do *interland*);

Serviços de Saúde (ex: Acesso aos serviços de saúde, Frequência das DTS, dos principais síndromas de DTS diagnosticados nesses distritos).

Similar frequência percentual da **existência de grupos de alto risco de infecção ao HIV** (ex: trabalhadores de sexo, camionistas de longo curso, trabalhadores migrantes, jovens fora da escola, entre outros);

Similaridades **de culturas tradicionais** (ex: ritos de iniciação sexual, tabús, práticas culturais em relação ao género e sexualidade, língua);

Similar **situação económica** (agricultura de subsistência, comércio informal, emprego e desemprego, pobreza segundo os indicadores de carência);

Nível de escolaridade (ex: Taxa de analfabetismo, Taxa de crianças em escola primária, Taxa de jovens em escola secundária)

2.3 Selecção de outras Unidades Sanitárias para funcionarem como novos Postos Sentinela

Este trabalho teve em conta a localização geográfica das províncias nos corredores comerciais, do seguinte modo:

- Para os corredores comerciais do Limpopo e do Maputo, na região Sul, as províncias: Cidade de Maputo, Maputo Província e Gaza;
- Para o corredor comercial da Beira na região Centro, as províncias: Sofala, Manica e Tete;
- Para o corredor comercial de Nacala, na região Norte, as províncias: Nampula e Niassa.

Assim, e de acordo com a província a que pertence um determinado corredor comercial, se observaram as localidades, vilas e cidades onde ainda não funcionava um Posto Sentinela, e de acordo com as características das Unidades Sanitárias (existência de laboratório clínico com pessoal e de consulta pré-natal com Enfermeira de SMI em funcionamento), assim como outras características de movimentos populacionais e comércio, tais como:

- Existência nessa localidade, vila ou cidade de paragem para camionistas de longo curso, ou
- de zonas fronteiriças com outros países, ou
- frequência elevada de chegada de turistas, ou
- perspectivas de futuros grandes investimentos que empregariam numerosos jovens;

Para as outras províncias onde a estrada dos corredores comerciais anteriormente mencionados não tem transcurso pelas mesmas, foram analisados outros factores, tais como:

Região Sul:

Para a província de Inhambane houve necessidade de se pôr em funcionamento um Posto Sentinela na zona Rural, assim como em analisar as potencialidades de outras Unidades Sanitárias noutros distritos, tendo talvez em conta a maior frequência de mineiros que trabalham na África do Sul e maior movimentação turística;

Região Centro :

Para a província da Zambézia, analisou-se a futura movimentação comercial pela construção de novas rodovias, para as províncias vizinhas e para o Malawi.

Região Norte :

Para a província de Cabo Delgado, analisou-se os distritos onde passará o futuro corredor comercial proveniente da Tanzânia, assim como localidades, vilas e cidades com futuros grandes investimentos que terá como consequência o emprego de muitos jovens.

2.4 Critérios para selecção das participantes

- Ter idade compreendida entre os 15 – 49 anos.
- Possuir gravidez confirmada
- Frequentar o Posto Sentinela pela primeira vez, durante a gravidez actual
- Residir dentro da área de cobertura do Posto sentinela
- Ter participado na palestra pré natal em que os detalhes da ronda de V.E. eram apresentados.
- Ter consentido oralmente para participar na ronda

Capítulo 3 METODOLOGIA

Neste capítulo será apresentada a metodologia utilizada para a realização deste trabalho, iniciando-se com a apresentação do esquema das etapas da descoberta de informação (fig.1).

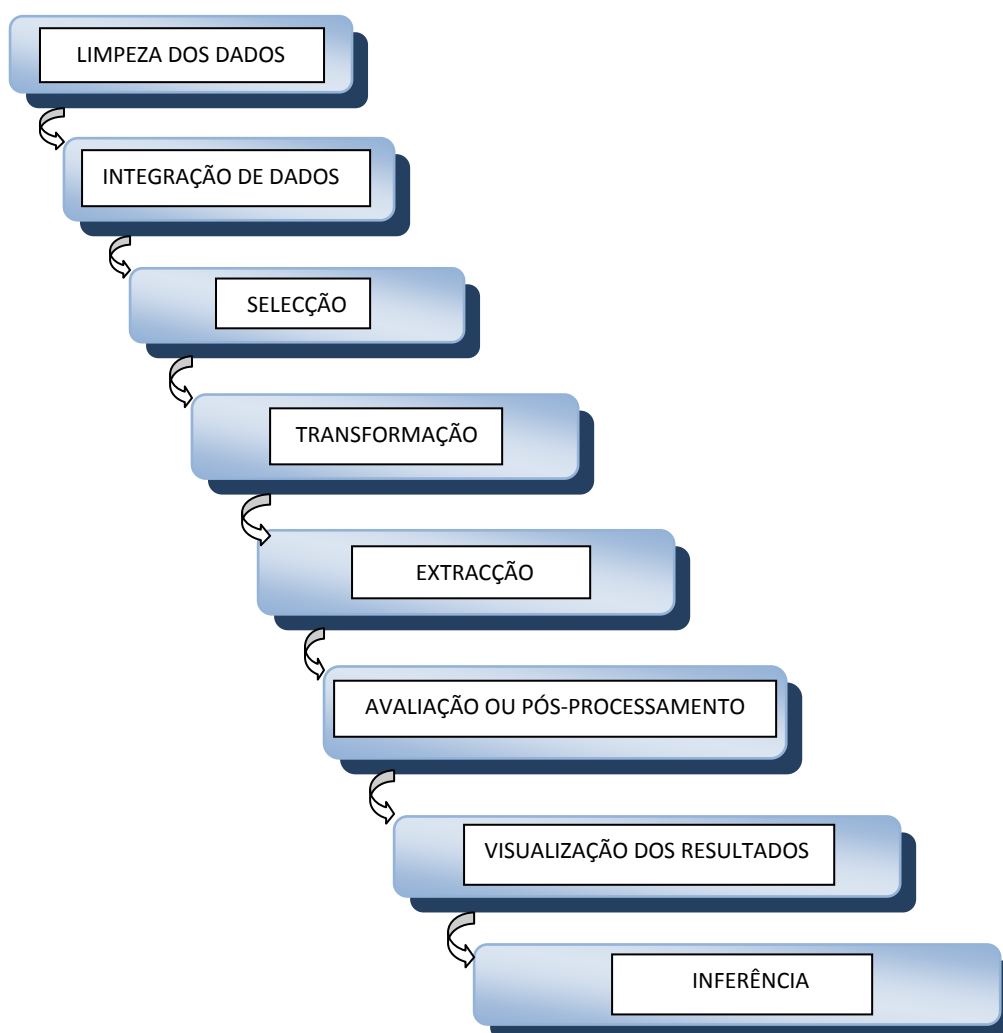


Fig.1. Esquema de fases da descoberta de informação

3.1 Limpeza de dados

Os dados usados neste trabalho foram fornecidos pelo INS, no formato sav (SPSS), como mostra a figura 3. Nesta etapa foram eliminados os ruídos e dados inconsistentes na variável *repro* (0;1). Segundo o Relatório sobre a Revisão dos Dados de V.E do HIV – Ronda 2007, de forma a garantir a fiabilidade dos dados é necessário filtrar a base de dados nesta variável, sendo o valor 1 correspondente aos dados validados. Consequentemente, foram validados 42707 dados de um total de 43108. Os ruídos foram causados por má digitação ou dados em falta.

The image shows two screenshots of the SPSS Data Editor interface. The top screenshot displays a data table with columns: provn, provs, pronome, regio, codn, cods, codnome, ps, Idade, Gesta, Para, and inm. The bottom screenshot displays a data table with columns: hv, proven, reside, ocugrav, ocugrav otr, ocuparc, ocuparc otr, ano, repro, and several var columns.

Fig.2. Esquema da base de dados do INS.

3.2 Integração de dados

Nesta etapa usou-se o Google Earth para recolher e agregar as coordenadas geográficas dos pontos amostrados. A partir do site Map Library [37], foi também obtido o mapa de Moçambique no formato *shapefile*, formato que permite a leitura no software ARCVIEW.

3.3 Selecção

Foram seleccionadas as variáveis que interessavam ao projecto:

provs, região, codnome, idade, hiv, ano, ocugrav e repro.

provs – nome da província

regiao – zona Sul/Centro/Norte

Idade – idade da gestante

codnome – nome do Posto Sentinela

ocugrav – ocupação da grávida

ano – ano relativo a VE

HIV 0: significa que o resultado do teste é negativo; **1:** significa resultado positivo.

repro - variável de validação de dados (0 e 1), usando os valores da variável repro 1.

3.4 Transformação

O resultado desta etapa foi a transformação da base de dados em formato *x/s* (Excel) e CSV, e a transformação dos dados das coordenadas geográficas em shapefiles, usando a ferramenta **CSV to Shapefile Converter** para posterior uso no ARCVIEW.

3.5 Extracção

A taxa de prevalência reflecte as condições de saúde de uma população e foi calculada dividindo o número de mulheres seropositivas pelo total de mulheres testadas, de acordo com a fórmula 1:

$$(1) \quad Prev = \frac{\text{N}^{\circ} \text{ de Pessoas Infectadas (HIV positivas)}}{\text{População Total (HIV negativa+positiva)}}$$

Para facilitar a análise de dados recorreu-se ao uso de uma tabela dinâmica (*Pivot table*), do software Excel, que é uma tabela interactiva que resume uma grande quantidade de dados rapidamente, ou os combina de tabelas diferentes. É possível girar suas linhas e colunas para ver resumos diferentes dos dados de origem, filtrar os dados exibindo páginas diferentes ou exibir os detalhes das áreas de interesse.

3.6 Avaliação ou pós processamento

Depois do processamento dos dados identificaram-se algumas tendências das taxas de prevalências, tanto a nível de províncias, regiões e nacional e o comportamento das prevalências nas diferentes faixas etárias e profissões.

3.7 Visualização de resultados

Os avanços computacionais e o aprimoramento nas técnicas de mapeamento, permitem uma avaliação cada vez mais precisa da qualidade dos atributos mapeados. Permitem ainda, detectar erros a eles associados, introduzidos durante a determinação do modelo de representação espacial a ser utilizado, através de interpolações de dados.

Para a visualização de dados produziram-se gráficos no Excel e mapas no ARCVIEW .

3.8 Inferência

Neste subcapítulo espera-se chegar ao modelo ideal das prevalências para os próximos anos, com base nas prevalências já estabelecidas.

Para os cálculos das previsões, foi necessário transformar os ficheiros das imagens de interpolações dos anos 2001, 2002, 2004 e 2007 em imagens *raster*. Destas imagens, foram retirados valores pixel a pixel, obtendo-se um ficheiro com 69864 pontos estruturados em 213 colunas e 328 linhas. Todos os pontos que se encontravam fora do mapa de Moçambique foram considerados nulos.

Com os dados do ficheiro, derivou-se outro ficheiro de diferenças de tempos (anos), de acordo com a sua posição na coluna e linha, contendo 160770 linhas e 6 colunas. O

objectivo deste ficheiro, era encontrar valores das taxas de prevalência nos anos não amostrados, para determinar a forma da curva de forma a prever o modelo.

A seguinte função descreve o modelo em que se procura encontrar uma prevalência P2 no tempo T2, a partir dos dados da prevalência P1 no tempo T1.

$$(2) \quad P2 = f(P1, T1, T2)$$

Tabela.1. Tabela explicativa dos atributos do ficheiro

Atributo	Explicação
r	Posição do valor do pixel na linha
c	Posição do valor do pixel na coluna
T1	Tempo 1 (ano)
T2	Tempo 2 (ano)
Prev1	Valor da taxa de prevalência 1 num determinado ano
Prev2	Valor da taxa de prevalência 1 num determinado ano

O software R (versão 2.10.1) foi usado para a modelação da rede neuronal. Para tal, foi usado o pacote **nnet**, que contém as funções necessárias para a obtenção de modelos de redes neuronais e suas previsões.

O modelo da rede neuronal obteve-se através do pacote **nnet** do software **r**. Após a testagem do modelo, recorreu-se ao ficheiro das diferenças e substituíram-se as variáveis **t1** pelo último ano (2007) e o **t2**, pelo ano que se pretendia fazer a previsão. O *script* da modelação da rede neuronal e suas previsões para os próximos anos encontra-se em anexo.

Finalmente, foram feitas as previsões para os anos de 2011, 2013 e 2015, houve necessidade de se fazer o processo inverso até a visualização dos mapas.

O desempenho da RNA após o treino, foi avaliado em relação à classificação dos pixéis pertencentes ao conjunto de teste. Para isso calculou-se o erro (ver fórmula 3), a relação valores previstos/valores verdadeiros, a relação erro/valores verdadeiros e o histograma dos resíduos.

$$(3) \quad \text{Erro} = \frac{\Sigma(P-V)^2}{nV}$$

Onde :

P = valor previsto

V = valor verdadeiro

nV = total dos valores verdadeiros

Capítulo 4 RESULTADOS E DISCUSSÃO

4.1 Análise da situação actual do HIV em Moçambique

A análise e discussão de resultados são feitas com base em mapas, tabelas e gráficos. A fig.3, apresenta a distribuição e localização dos Postos Sentinela que fizeram parte da rede de V.E, nas rondas entre os anos de 2001 a 2007.

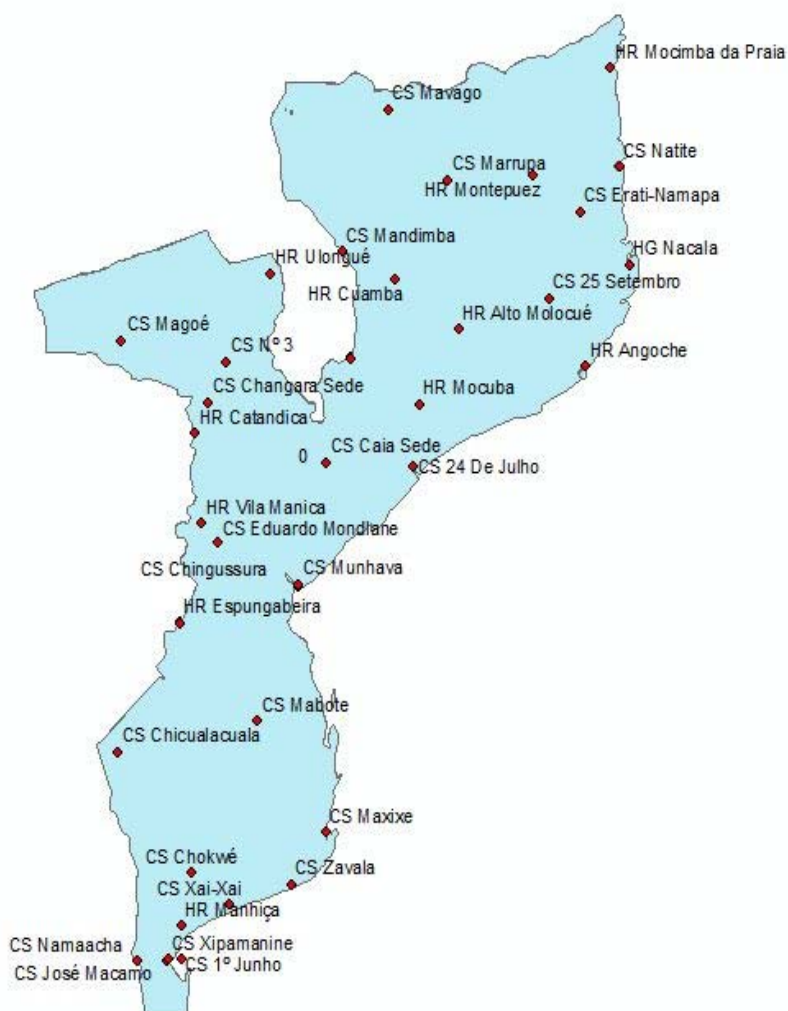


Fig.3 . Localização espacial dos Postos Sentinela

A tabela 2, mostra a evolução das prevalências nos Postos Sentinela. Os valores a vermelho são as prevalências mais altas e os valores a verde as mais baixas, do ano a que dizem respeito. Os C.S de Chicualacuala em Gaza e de Marrupa em Niassa, só entraram em funcionamento a partir de 2004. Os C.S. da Munhava e de Xipamanine nas cidades da Beira e Maputo, respectivamente, deixaram de funcionar a partir de 2004 por serem redundantes, estavam muito próximos a outros dois Postos Sentinela.

Os P.S. a laranja localizam-se em zonas urbanas e os a branco em zonas rurais.

Tabela 2. Prevalências do HIV nos Postos Sentinela no período de estudo

Postos Sentinela	Província	2001	2002	2004	2007
01 - CS José Macamo	Maputo Cidade	14,00	18,00	18,98	24,67
02 - HR Manhiça	Maputo Província	15,00	14,67	23,64	28,52
03 - CS Xai-Xai	Gaza	19,67	23,67	27,67	34,92
04 - CS Chokwé	Gaza	19,33	22,00	22,56	26,00
05 - CS Maxixe	Inhambane	6,00	9,67	13,33	9,70
06 - CS 24 De Julho	Zambézia	22,00	25,00	33,33	35,45
07 - HR Mocuba	Zambézia	14,67	14,00	18,52	15,15
08 - HR Milange	Zambézia	18,33	14,00	22,14	16,34
09 - CS Ponta Gêa	Sofala	27,33	35,67	34,56	28,14
10 - HR Catandica	Manica	10,00	16,72	18,98	9,36
11 - CS Eduardo Mondlane	Manica	23,67	24,33	25,51	25,49
12 - CS Nº 3	Tete	17,33	21,74	25,84	18,73
13 - CS Changara Sede	Tete	20,00	13,33	19,72	9,55
14 - CS Mandimba	Niassa	6,33	14,33	10,44	5,41
15 - HR Cuamba	Niassa	5,67	10,33	16,79	11,33
16 - CS 25 Setembro	Nampula	11,00	11,67	10,67	10,34
17 - HG Nacala	Nampula	6,67	8,00	8,16	7,43
18 - HR Angoche	Nampula	5,00	7,00	5,67	5,03
19 - CS Natite	Cabo Delgado	9,67	11,33	10,33	14,95
20 - HR Montepuez	Cabo Delgado	3,33	4,67	10,42	7,32
21 - C.S. Xipamanine	Maputo Cidade	17,33	16,00	---	---
22 - CS 1º Junho	Maputo Cidade	15,33	17,67	22,97	18,82
23 - CS Namaacha	Maputo Província	16,67	21,00	18,73	21,69
24 - CS Zavala	Inhambane	5,67	4,67	7,33	11,04
25 - CS Mabote	Inhambane	15,05	12,71	8,47	15,75
26 - HR Alto Molocué	Zambézia	5,02	6,67	14,00	7,14
27 - CS Chingussura	Sofala	21,67	29,67	29,05	29,87
28 - C.S. Munhava	Sofala	27,00	27,00	---	---
29 - CS Caia Sede	Sofala	7,67	12,00	18,77	15,38
30 - HR Vila Manica	Manica	14,67	21,00	15,67	16,33
31 - HR Espungabera	Manica	9,03	10,33	12,71	8,65
32 - HR Ulógue	Tete	11,33	10,96	13,47	6,00
33 - CS Mágoe	Tete	14,41	12,33	12,16	12,26
34 - CS Mavago	Niassa	2,00	3,67	4,05	2,67
35 - CS Erati-Namapa	Nampula	7,67	7,02	11,04	7,94
36 - HR Mocimboa da Praia	Cabo Delgado	5,67	8,67	4,67	7,33
37 - CS Chicualacuala	Gaza	---	---	21,00	14,08
38 - CS Marrupa	Niassa	---	---	2,72	8,64

4.1.1 Taxas de prevalências regionais de HIV

O gráfico da fig.4 mostra a evolução da prevalência de casos de HIV, nas três regiões do país, nos anos 2001, 2002, 2004 e 2007.

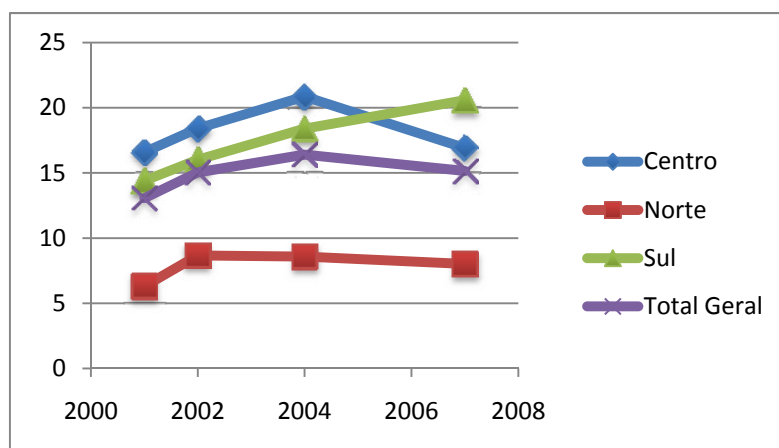


Fig.4. Prevalências nas três regiões(%), versus anos.

É de realçar que a prevalência nacional (Total Geral) tende a estabilizar-se, o mesmo se poderá dizer da região Norte do país.

As taxas de prevalência elevadas no centro do país são atribuídas a vários factores, inclusive os cerca de dois milhões de refugiados que voltaram para Moçambique vindo de países vizinhos (Malawi e Zâmbia), após o Acordo Geral de Paz em 1992, países esses onde a taxa de incidência é alta. Outro factor apontado é a mobilidade da população ao longo dos corredores de transporte que ligam Moçambique, nomeadamente o porto de Beira, ao Zimbabwe e Malawi.

A ligeira descida da taxa de prevalência na zona Centro, dá indicações de que a epidemia encontra-se na sua fase de *plateau*.

Entende-se por *plateau* numa epidemia a fase em que a doença atinge uma estabilidade ou regista uma ligeira descida, visto que, todas epidemias começam por uma fase de crescimento lento, seguido por um crescimento rápido e no fim estabilizam, ou tem uma ligeira redução, com ou sem alguma intervenção. No Uganda aconteceu um facto semelhante em que as taxas de prevalência do HIV mantiveram-se durante muito tempo altas e estáveis.

Uma das principais vias rodoviárias do Sul do país é o Corredor de Maputo. Estradas ligam as Províncias do Sul do país ao Zimbabwe, a África do Sul e à Suazilândia.

Mercadorias, mineiros, camionistas e os seus ajudantes, pequenos comerciantes e turistas transitam regularmente pela região. Segundo os dados da figura 4, esta área está a tornar-se no maior epicentro da epidemia de HIV em Moçambique.

Supõe-se que os factores de disseminação do vírus neste corredor são o baixo nível de escolaridade, a pobreza aguda, o elevado desemprego e os comportamentos sexuais de risco. Assim, a população trabalhadora distante da família, as profissionais do sexo e os seus parceiros são considerados grupos de risco.

Existem várias ONG's a trabalhar no Corredor de Maputo, com o objectivo de promover a prevenção. Estas defendem a fidelidade, o início tardio das relações sexuais, a redução do número de parceiros e o uso do preservativo para conter a propagação do vírus.

Estudos epidemiológicos efectuados revelam uma redução de transmissão do HIV da mulher para o homem nos casos em que este é circuncidado [4]. Há evidências substanciais de que a circuncisão masculina protege não apenas contra a aquisição do HIV, mas também contra o carcinoma peniano, infecções urinárias e infecções sexualmente transmissíveis ulceradas [24].

A prática de circuncisão masculina, em Moçambique, é variável de uma província para outra e até a nível de distrito.

A comunidade que mais adere à prática de circuncisão masculina no país é muçulmana, que representa aproximadamente 18 por cento da população, de um universo de aproximadamente 20 milhões de habitantes em todo o país. Na região Norte onde a população é maioritariamente muçulmana, as prevalências continuam as mais baixas do país.

4.1.2 Taxas de prevalência provinciais de HIV

As taxas de prevalência nas províncias reflectem-se nas regiões (fig.5). Por excepção na zona Sul a província de Inhambane apresenta uma taxa baixa comparando com Gaza, Maputo Cidade e Província, visto que a mesma não faz fronteira com algum país do *Interland* e os naturais desta província também tem o hábito de fazer a circuncisão masculina.

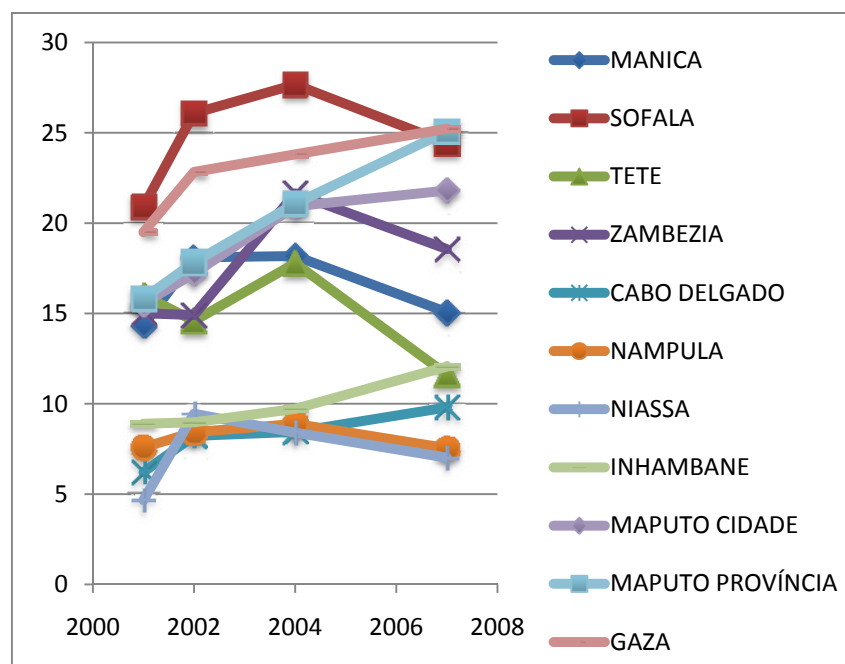


Fig 5. Prevalências nas províncias (%), versus anos.

Moçambique situa-se numa região onde estão os países com as taxas de prevalência de HIV mais elevadas do mundo. Tais são os casos do Botswana, África do Sul e Zimbabwe. Nestes países, a epidemia começou antes que em Moçambique, e está a causar uma inversão dos níveis de crescimento económico e esperança de vida [14].

Na ultima ronda de V.E do HIV, as taxas de prevalências desceram nas províncias de Tete e Manica, apresentando uma curva de seroprevalência que reflecte a situação das zonas fronteiriças. Estes efeitos poderão ser resultado do aumento da mortalidade verificada, efeitos da educação e influenciados pela curta estadia nos outros países e/ou esforços de prevenção que estão sendo implementados [6].

A abertura de corredores de desenvolvimento, que visam facilitar a circulação de pessoas e mercadorias entre os países da região, a emigração de trabalhadores moçambicanos para a África do Sul que regressam infectados e a transmissão interna parecem ser os factores responsáveis pela expansão da epidemia em Moçambique.

Os dados da fig.5, espelham cenários diferentes nas três regiões do país, na região Sul a epidemia demonstra níveis crescentes de prevalência, onde todas as taxas observadas situam-se acima de 20%, com excepção da Província de Inhambane, que registou a prevalência mais baixa. Nas quatro províncias da região Centro (Sofala, Zambézia, Manica e Tete), o cenário observado é de estabilização. Na região Norte (Niassa, Cabo delgado e Nampula) as taxas de prevalência continuam a ser menores em comparação ao resto do país, mostrando também uma estabilização.

Segundo estudos citados em [6] nas áreas com altos níveis de circuncisão, tais como as províncias nortenhas de Cabo Delgado, Nampula e Niassa, há um efeito protector aparente da circuncisão masculina ao nível da população, adicionando-se ao facto dos valores da religião muçulmana serem bastante rígidos.

4.1.3 Taxas de prevalência em diferentes faixas etárias

A prevalência descrita para diferentes grupos etários e por definição, a prevalência de adultos refere-se a prevalência de indivíduos na faixa etária dos 15 aos 49 anos (fig.6)

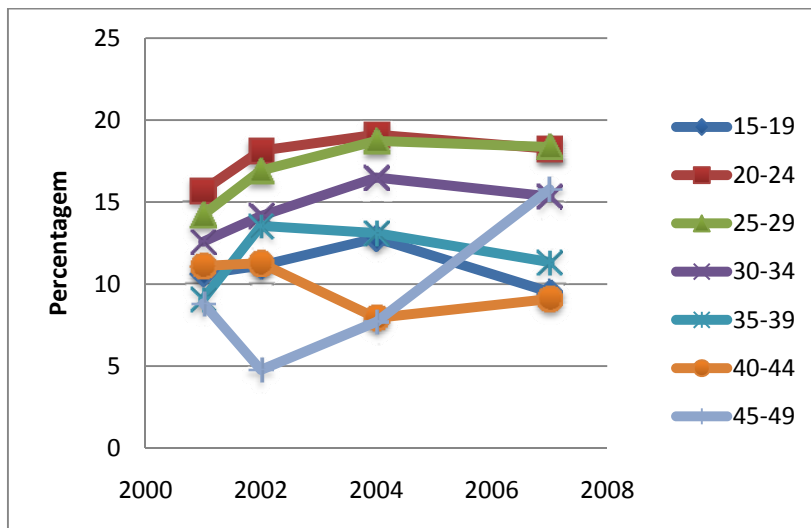


Fig.6. taxa de prevalências de HIV na faixa etária de 15 – 49 anos

As prevalências são mais altas nas idades compreendidas entre os 20-29 anos, idades mais activas no sexo. Embora a taxa de prevalência esteja a descer, não deixa de ser preocupante que as prevalências entre os jovens de 15 – 19 anos estejam nos níveis observados.

A taxa de prevalência na faixa etária entre os 45 e 49 anos, parece ter um crescimento assustador, mas tal, poderá dever-se ao facto da amostra nesta faixa etária ser muito pequena e como tal ser pouco representativa.

Segundo [34] mulheres e homens incorporam nas suas práticas os valores do poder masculino como certos e justos. Os modelos de educação e a influência religiosa determinam que a jovem não esteja preparada, quando adulta, para negociar com o parceiro o exercício da sexualidade seja na reprodução como no prazer. Nesta situação, as mulheres estão, por um lado, sujeitas à contaminação de doenças sexualmente transmissíveis, sem que tenham a possibilidade de se precaverem, não tendo direito a escolha sobre o seu corpo.

4.1.4 Taxas de prevalência de acordo com a ocupação

Na base de dados, os dados sobre a proveniência, tempo de residência e ocupação da gestante, só estão disponíveis a partir de 2004. Portanto, só existem dados de 2004 e 2007 (fig.7).

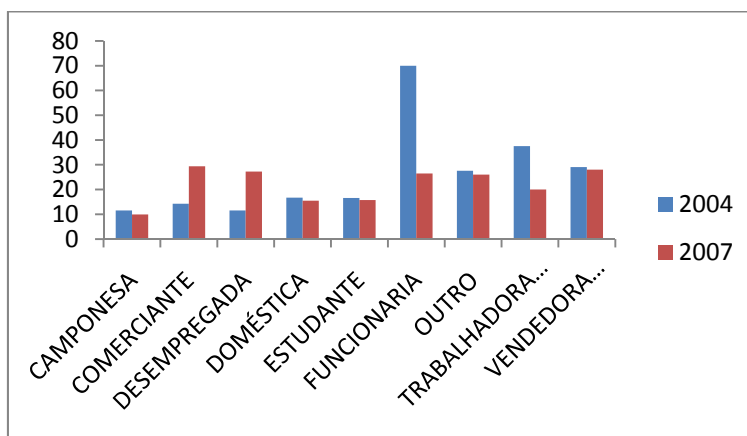


Fig.7. Prevalências do HIV(%), versus ocupação

É de realçar que as enfermeiras tentaram codificar *ocupação* segundo uma lista predefinida pelo INS, mas quando não conseguiram encaixar a resposta da gestante, preencheram com OUTRO.

Os principais canais através dos quais a epidemia de HIV afecta o desenvolvimento sócio económico passam pela afectação da força de trabalho e pelo nível e distribuição de poupanças. A epidemia tem impacto primário na população em idade de trabalho,

onde se concentra a incidência de enfermidade e a mortalidade relacionadas com o HIV. Assim, pessoas com importantes papéis socioeconómicos (tanto homens como mulheres) deixam de dar sua plena contribuição para o desenvolvimento. Os efeitos, naturalmente, não se limitam a simples cálculo de perdas de mão-de-obra, mas têm implicações muito mais profundas para a estrutura de famílias, a sobrevivência de comunidades e problemas a longo prazo de sustentação da capacidade produtiva [13].

O efeito directo vem da perda de pessoas com formação, experiência e educação adequadas, que têm a função de manter o fluxo de mão-de-obra recém-formada.

4.1.5 Taxas de prevalência do HIV nas zonas rurais e urbanas

As taxas de prevalência do HIV são mais altas em zonas urbanas que em zonas rurais (fig.8).

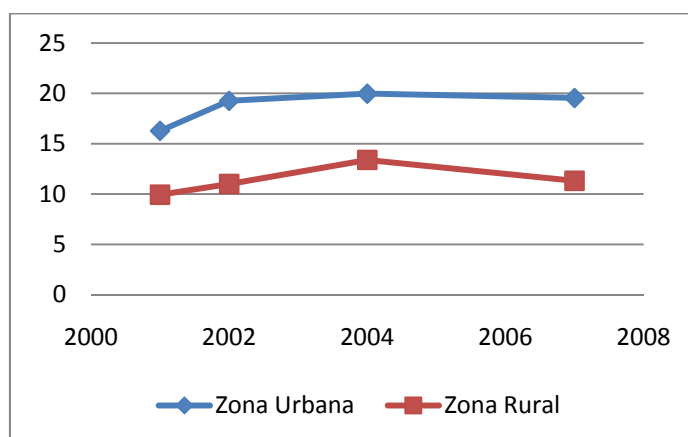


Fig. 8. Prevalências do HIV em zonas rurais/urbanas

As zonas urbanas tem melhores condições socioeconómicas comparativamente as zonas rurais. Consequentemente, essas condições são um atractivo tanto de indivíduos nacionais e estrangeiros para esses locais.

O estudo “As características socioeconómicas dos distritos com postos de vigilância epidemiológica do HIV/SIDA e sua influência na taxa de prevalência do HIV/SIDA em Moçambique” [1], concluiu que os distritos com melhores condições socioeconómicas e de bem-estar tendem a exibir taxas de prevalência do HIV mais elevadas do que os

distritos mais pobres. Este resultado é consistente com o de outros estudos realizados na África sub-Sahariana [16].

As figuras 9 e 10 mostram as taxas de prevalência de HIV nas zonas urbanas e nas zonas rurais, respectivamente.

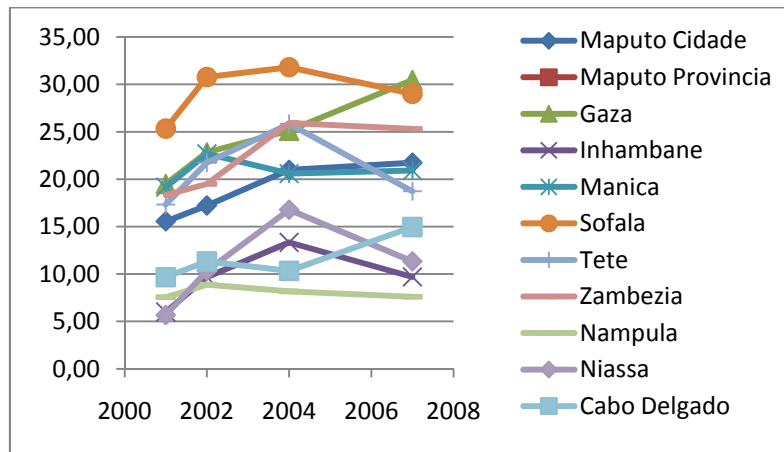


Fig.9 Prevalências do HIV em zonas urbanas

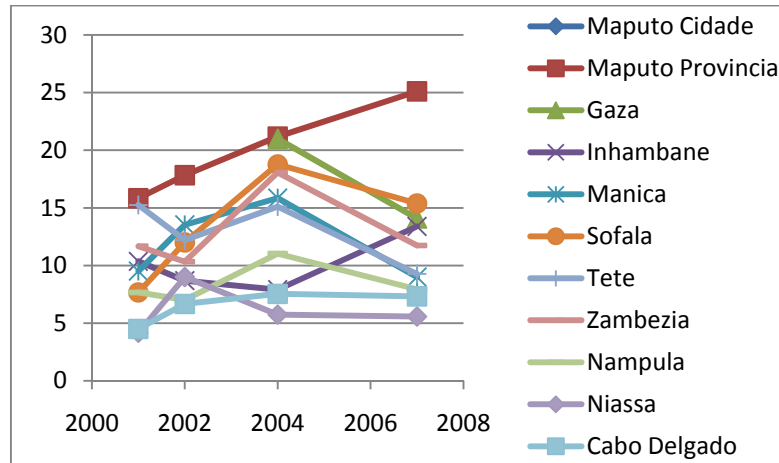


Fig.10 Prevalências do HIV em zonas rurais

[4], estratificou os dados de prevalência em seis partes, por área geográfica e residência, nomeadamente:

Norte Urbano: A prevalência do HIV é relativamente baixa e estável.

Norte Rural: A epidemia é caracterizada pela mais baixa prevalência no país mas desde 2001 as taxas cresceram. Existem elevadas taxas de circuncisão, mas a cobertura de prevenção é fraca.

Centro Urbano: A tendência da seroprevalência é preocupante, continuando alta e estável.

Centro Rural: A prevalência do HIV é relativamente baixa e em redução nos últimos anos.

Sul Urbano: A seroprevalência é alta e em crescimento, o que é motivo de grande preocupação.

Sul Rural: A seroprevalência é relativamente alta, particularmente nas zonas rurais, e regista-se um crescimento na prevalência nos últimos anos.

A confluência de níveis elevados de comportamentos de alto risco, que incluem parceiros concorrentes múltiplos, parceiros ocasionais, alta troca de parceiros, grande movimento de pessoas devido a actividades económicas e interacção com principais grupos de maior risco, cria uma tendência que produz uma “hiper-endemia” em grande parte do sul e zonas urbanas de Moçambique [4].

4.1.6 Krigagem

A estatística espacial calcula estimativas dentro de um contexto regido por um fenómeno natural, com distribuição no espaço e supõe que os valores das variáveis, consideradas como regionalizadas, sejam espacialmente correlacionados. O principal objectivo é modelar variações espaciais contínuas assumindo uma estrutura de correlação espacial da variável analisada [11]. Devido a essa característica tem sido grande a sua aplicação principalmente para efectuar estimativas e/ou simulações de variáveis em locais não amostrados.

[20], aponta que os mapas de isovalores, que mostram a variabilidade dos dados, são resultados cada vez mais comuns que se espera dos SIG's, assim como as estimativas dos dados de pontos não amostrados, por meio de valores em pontos amostrados. O autor coloca também que nestas duas situações, os problemas de interpolação surgem, tornando necessário o uso de metodologias específicas, e as soluções deveriam vir com os erros associados às estimativas.

A interpolação é uma técnica utilizada para a estimativa do valor de um atributo em locais não amostrados, a partir de pontos amostrados na mesma área ou região. A interpolação espacial converte dados de observações pontuais em campos contínuos, produzindo padrões espaciais que podem ser comparados com outras entidades espaciais contínuas. O raciocínio que está na base da interpolação é que, em média, os valores do atributo tendem a ser similares em locais mais próximos do que em locais mais afastados. Esse conceito também fundamenta a base das relações espaciais entre fenómenos geográficos, utilizando a correlação espacial como meio de diferença dos atributos estimados [3].

Os métodos de interpolação mais comuns dos SIG's em geral pertencem a duas categorias: globais e locais, sendo os globais mais utilizados em superfícies de tendência, e os locais podem ser polinómios de baixa ordem, funções *spline*, poliedros, triangulação e médias móveis ponderadas. Porém, estes métodos não fornecem os erros associados às estimativas. Somente o método da krigagem o faz por meio de um "modelo contínuo de variação espacial". Os modelos de atributos quantitativos indeterminados no campo baseiam-se nas médias e nos erros médios aleatórios. As médias são armazenadas nos mapas criados e estes erros definidos por meio do cálculo dos desvios-padrão. Mas geralmente não é suficiente apenas o desvio-padrão para a identificação dos erros, mas também a inclusão das auto correlações, as correlações espaciais dos atributos [20]. Por isso a técnica da krigagem se torna importante.

Segundo [3], os métodos de krigagem dependem de modelos matemáticos e estatísticos, assim como da noção de auto correlação. Na estatística clássica, assume-se que as observações são independentes, ou seja, não há correlação entre as observações. Na geoestatística, a informação dos locais espaciais permite o cálculo das distâncias entre as observações e modelar a auto correlação como uma função da distância.

A krigagem universal faz um tipo de regressão com as coordenadas espaciais como sendo as variáveis explicativas, mas ao invés de assumir que os erros são independentes, eles são modelados para serem auto correlacionados [3].

A Krigagem usada para a predição não requer que os dados tenham distribuição normal. Entretanto, a normalidade é necessária para se obter mapas de quantis e de probabilidade na krigagem ordinária, simples e universal. Considerando apenas a predição criada por médias ponderadas, a krigagem é considerada o melhor estimador não viciado. No caso de uma distribuição normal dos dados, é o melhor estimador entre todos os estimadores não viciados, não apenas aqueles com médias ponderadas. Ela depende também do pressuposto de que os erros aleatórios são estacionários de

segunda ordem, ou seja, têm média zero e a covariância entre dois erros aleatórios depende apenas de sua distância e direção que os separa, não de sua posição [11].

Uma das desvantagens das abordagens pontuais, é que elas possuem fronteiras locais, sendo difícil a percepção do que está a acontecer com o objecto de estudo, neste caso, o país como um todo.

A possibilidade de colocar os dados num mapa permite-nos perceber certas tendências espaciais, que em forma de tabela ou gráfico não seriam de tão fácil percepção.

Para estimação da evolução espacial do HIV-SIDA em Moçambique, foi necessário recorrer a técnicas de interpolação de dados (krigagem).

A krigagem pode ser a ferramenta ideal para se utilizar, pois apresenta a evolução espacial da doença ao longo do tempo (figs.11 e 12)

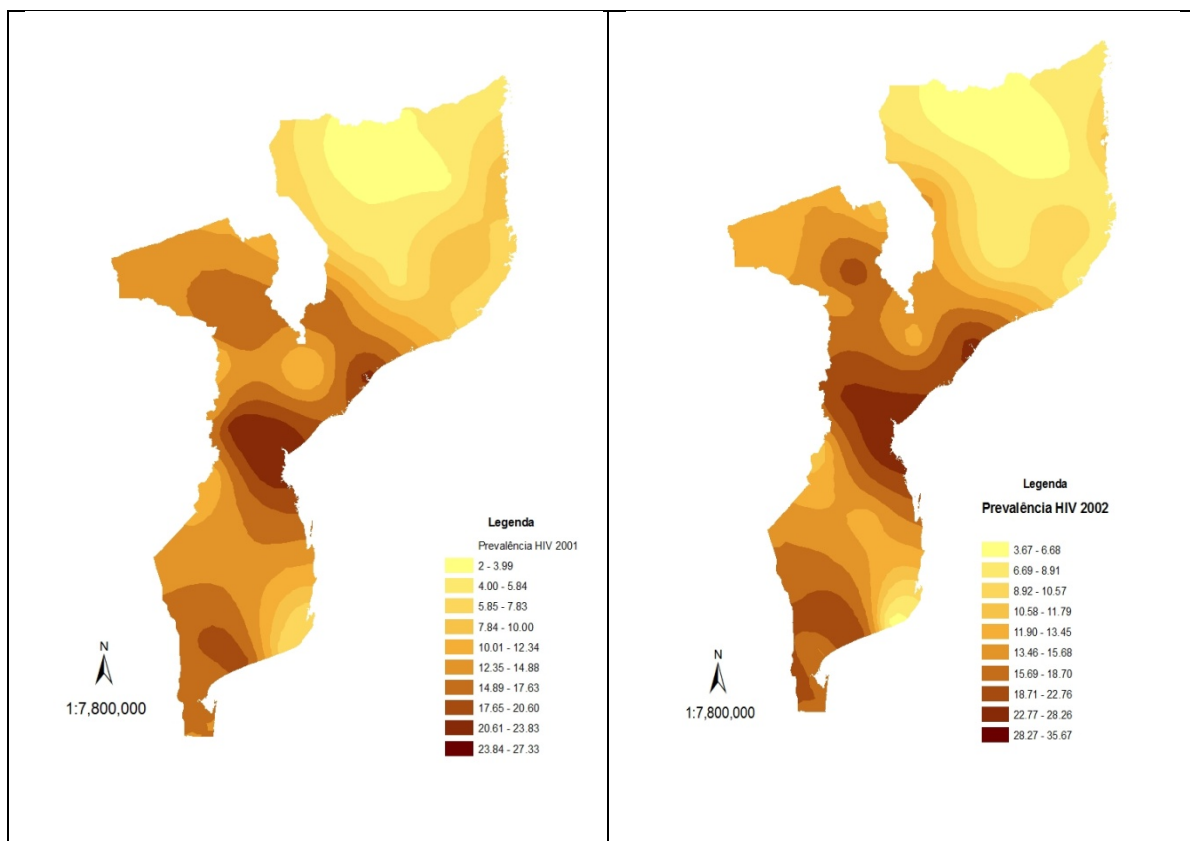


Fig.11 Mapas mostrando a evolução espacial das prevalências do HIV-SIDA em Moçambique (2001-2002)

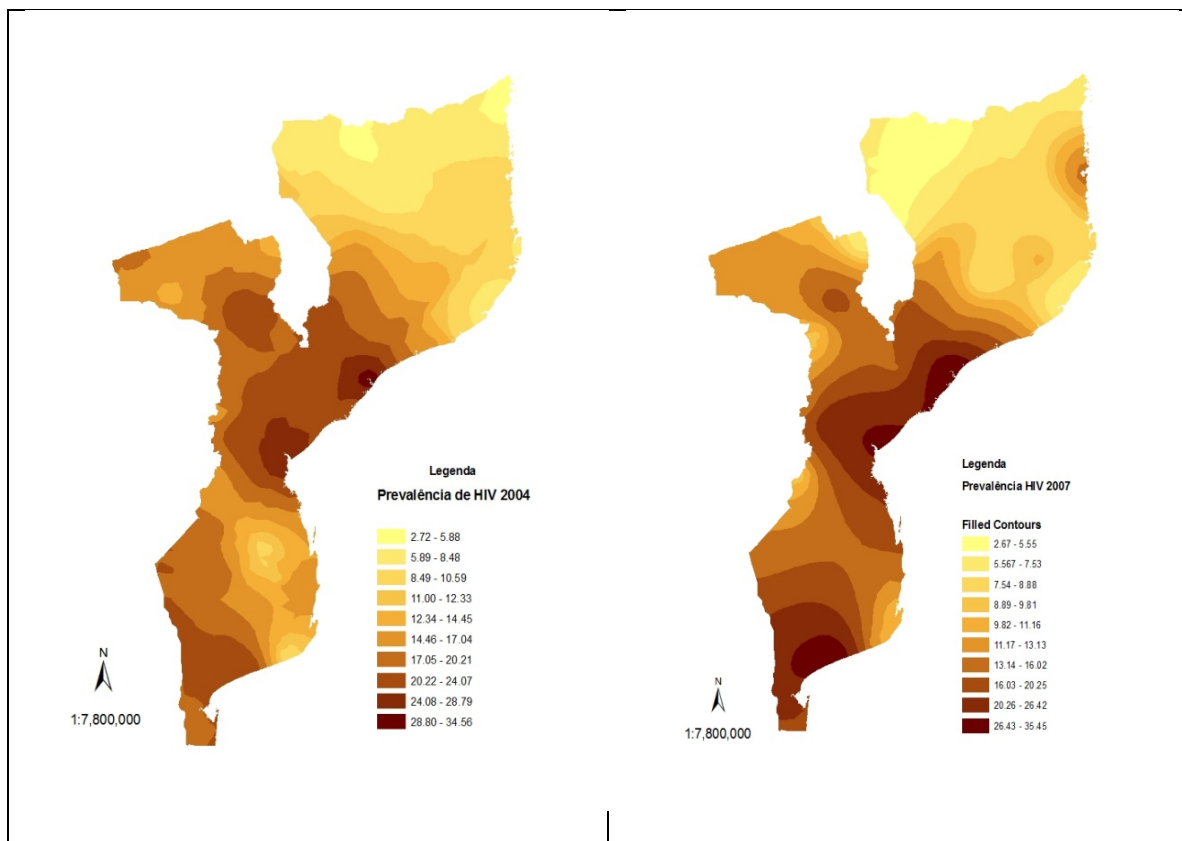


Fig.12 Mapas mostrando a evolução espacial das prevalências do HIV-SIDA em Moçambique (2004-2007)

Uma análise comparativa entre as taxas de prevalência da doença entre os anos 2001 e 2007, na base da informação fornecida pelos mapas é ilustrativa quer dos níveis de alastramento do SIDA, quer da regionalização das maiores taxas de prevalência em áreas de migração acentuada e de grande mobilidade populacional, como são as regiões Sul e centro do país.

Segundo [6], as províncias com ligações socioculturais e económicas contínuas, parecem ter tendências semelhantes da epidemia HIV às dos seus vizinhos internacionais. As tendências da epidemia do HIV destas zonas em Moçambique, reflectem as que ocorrem nos países vizinhos devido à porosidade das fronteiras, ligações culturais e linguísticas e/ou rotas de transporte (fig. 13).

O mapa da fig. 13, ilustra as semelhanças na sero prevalência nas zonas fronteiriças da Tanzânia, Malawi, Zâmbia, Zimbabwe, África do Sul e Suazilândia desde 2004.

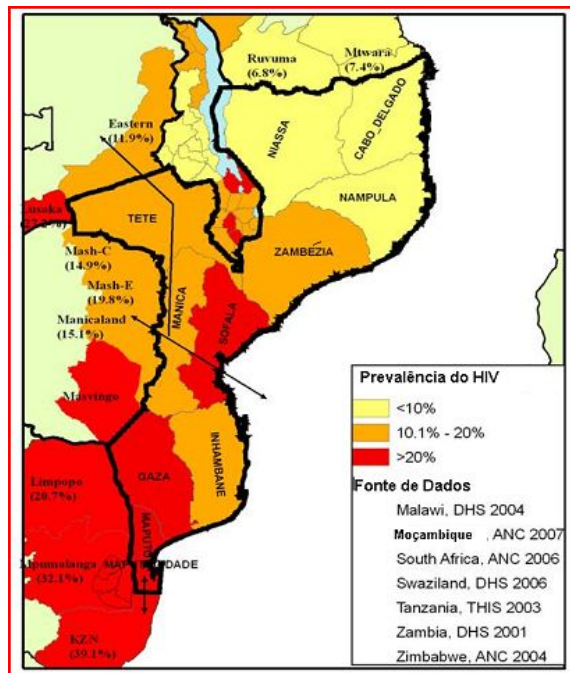


Fig.13. Prevalência do HIV em Moçambique e países vizinhos. Fonte: [3]

- Maputo - Cidade e Província de Maputo são semelhantes às zonas fronteiriças da África do Sul.
- A província de Gaza é semelhante às zonas fronteiriças da África do Sul e Zimbabwe.
- Manica e Sofala são semelhantes às zonas fronteiriças do Zimbabwe (Corredor da Beira).
- Tete é semelhante às zonas fronteiriças da Zâmbia.
- Niassa e Cabo Delgado são semelhantes às zonas fronteiriças da Tanzânia.
- Tete e Zambézia são semelhantes às zonas fronteiriças do Malawi associadas às rotas de transporte.

4.1.7 Erros de krigagem

As técnicas geoestatísticas de estimativa são superiores por permitirem o cálculo do erro associado, chamado variância de krigagem. Este, pode ser mapeado e oferecer informações úteis sobre a confiabilidade dos valores interpolados.

Pelos mapas apresentados, nota-se claramente que o erro associado a interpolação é mínimo nos pontos amostrados, sendo maiores nas áreas não amostradas (fig.14).

Estes resultados foram obtidos, ajustando-se os respectivos variogramas para se obter o menor erro e consequentemente a maior dependência espacial possível.

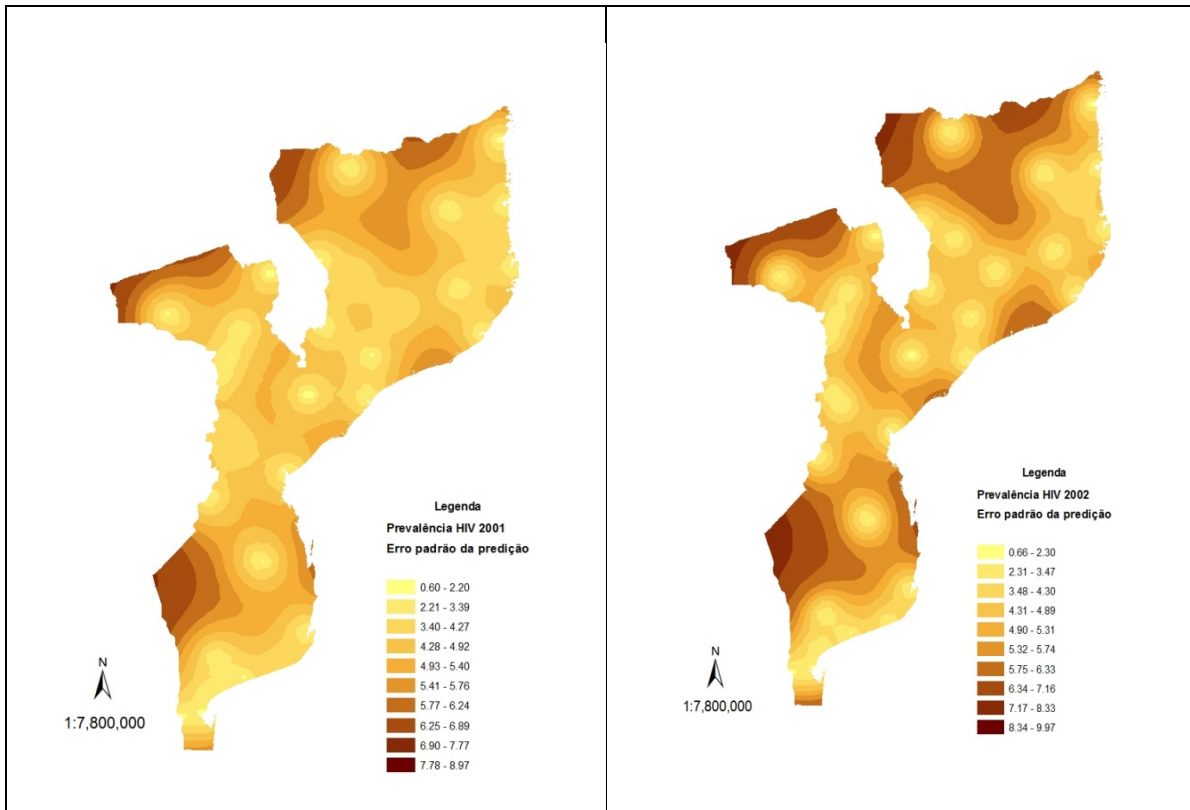


Fig.14 Erros da krigagem dos anos 2001 e 2002

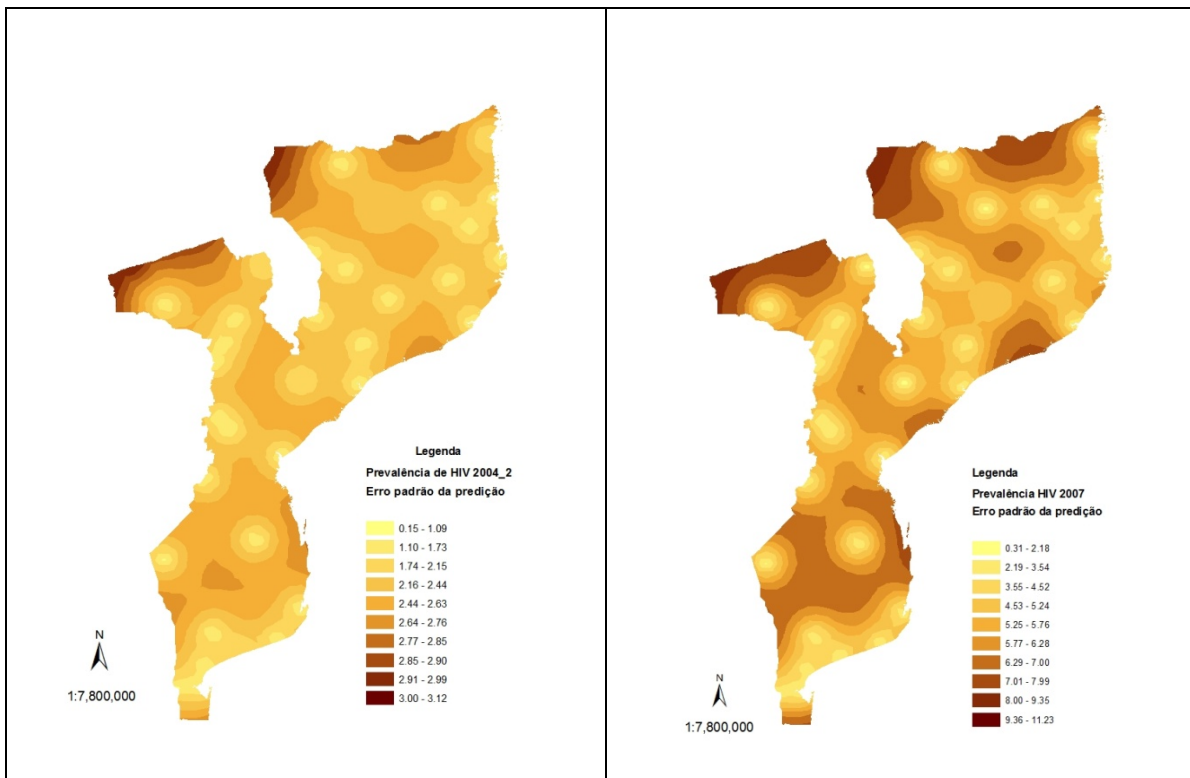


Fig.15. Erros da krigagem dos anos 2004 e 2007

Segundo [11], a natureza estrutural de um conjunto de dados é definida a partir da comparação de valores tomados simultaneamente em dois pontos, segundo uma determinada direcção. A função variograma é uma medida da variância das diferenças nos valores da variável regionalizada entre pontos separados por uma certa distância. Pontos mais próximos, por estarem correlacionados, terão essa variância pequena, aumentando à medida que os pontos se distanciam.

4.2 Inferência da prevalência do HIV

4.2.1 Redes neuronais artificiais

Ao se analisar o desenvolvimento da sociedade, especialmente nas últimas décadas, verifica-se que é cada vez maior o número de informações geradas.

Muitas dessas informações estão armazenadas em grandes bases de dados que, se devidamente tratadas em termos analíticos, podem gerar resultados de significativa relevância para a proposição, elaboração e/ou avaliação dos mais diversos processos e políticas, que cercam os vários sectores sociais (produção, serviços, actividades públicas entre outros).

A partir da definição dos objectivos que se espera atingir a partir da análise de certos dados, é possível escolher, dentre diversas opções, a ferramenta de *data mining* que seja a mais adequada.

Quando se analisam questões relativas a previsões de comportamento futuros de certas variáveis, uma das ferramentas que pode ser utilizada, de maneira a assegurar a obtenção de resultados satisfatórios, relaciona-se com as RNA's [17].

O estudo das RNA's tem alta relevância para os pesquisadores de Inteligência Artificial, sendo objecto de crescentes pesquisas e servindo como valiosa ferramenta para diversas aplicações, como a sua aplicação nesta dissertação, para a previsão de taxas de prevalência de HIV-SIDA.

As RNA's foram desenvolvidas numa tentativa de criar um modelo computacional, que simule a estrutura e funcionamento do cérebro humano a partir de estudos neurobiológicos, buscando a chamada Inteligência Natural [26].

A RNA é uma implementação de um algoritmo construído, tomando como base o funcionamento do cérebro humano e caracteriza-se por possuir grande número de elementos de processamento bem simples, inspirados no funcionamento de um neurónio e grande número de conexões entre estes elementos de processamento. Cada conexão tem um peso associado, este peso representa o nível de interação ou acoplamento entre elementos de processamento e se a sua natureza é excitatória ou inibitória [17].

Um modelo neuronal consiste numa rede de células autónomas, dotadas de capacidade de processamento. As células são ligadas por conexões, cada uma com um peso associado, que corresponde à influência da célula no processamento do sinal de saída. Pesos positivos correspondem a factores de reforço do sinal de entrada e pesos negativos correspondem a factores de inibição. Os modelos, geralmente, apresentam um conjunto de células de entrada, por onde são passadas as informações para a rede, um conjunto de células de saída, que apresentam os sinais de saída da rede e um conjunto de células intermediárias [22].

Segundo [10], as RNA's possuem várias características vantajosas que estimulam as pesquisas, como o controle altamente paralelo e distribuído; a capacidade de aprendizagem através da observação de um conjunto de exemplos, sendo estes com ou sem informação de respostas desejadas a cada estímulo; robustez, uma queda de desempenho gradual na eventualidade de mal funcionamento, ou até mesmo inutilização de parte da rede; manutenção do desempenho na presença de ruído e capacidade de lidar com dados incompletos.

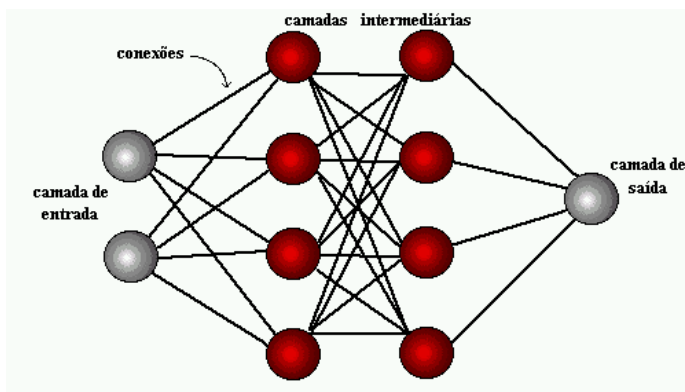


Fig16. Exemplo de RNA com 4 camadas

O processo de aprendizagem das RNA's é realizado quando ocorrem várias modificações significantes nas sinapses dos neurónios. Essas mudanças ocorrem de acordo com a activação dos neurónios. Se determinadas conexões são mais usadas, estas são reforçadas enquanto as demais são enfraquecidas. É por isso que quando uma RNA é implantada para uma determinada aplicação, é necessário um tempo para que esta seja treinada [12].

Existem, basicamente, 3 tipos de aprendizagem nas RNA's:

Supervisionada: neste tipo, a rede recebe um conjunto de entradas padronizadas e seus correspondentes padrões de saída, onde ocorrem ajustes nos pesos sinápticos até que o erro entre os padrões de saída gerados pela rede tenha um valor desejado;

Não-supervisionada: neste tipo, a rede trabalha os dados de forma a determinar algumas propriedades dos conjuntos de dados. A partir destas propriedades é constituída a aprendizagem;

Híbrida: neste tipo ocorre uma "mistura" dos tipos supervisionada e não-supervisionada. Assim, uma camada pode trabalhar com um tipo enquanto outra camada trabalha com o outro tipo.

4.2.2 O processamento dos dados

O software escolhido para lidar com as RNA's foi o R que é ao mesmo tempo uma linguagem de programação e um ambiente para computação estatística e gráfica [36].

A escolha do R, tem ainda outros motivos relacionados com o facto de ser um software gratuito e de código aberto. Estas características, em conjunto com as suas reconhecidas qualidades, fazem dele uma ferramenta quase ideal para aprender a programar dentro dum contexto de análise de dados e sistemas de apoio `a decisão [27].

Apesar do seu carácter gratuito, é uma ferramenta muito potente com boas capacidades ao nível da programação e um conjunto bastante vasto de pacotes que acrescentam bastantes potencialidades a já poderosa versão base (S PLUS) do mesmo [30].

O objectivo deste subcapítulo é estimar as taxas de prevalência para os próximos anos a partir de dados de prevalências já conhecidos.

Para tal, testaram-se um conjunto de metodologias para a otimização do modelo.

Testaram-se modelos de regressão lineares e não lineares, como a diferença de médias, mas os mesmos não convergiam.

4.2.3 Modelação da Rede Neuronal Artificial

Ao implementarmos uma rede neuronal, normalmente o conjunto de dados é separado em dois conjuntos: conjunto de treino e conjunto de teste.

O conjunto de treino é utilizado para o treino da rede e ajuste dos parâmetros da rede, devendo conter um número estatisticamente significativo de casos em estudo, de modo a constituir uma amostra representativa do problema que se pretende estudar [26].

O conjunto de teste é utilizado para verificar a capacidade de generalização da rede sob condições reais de utilização. Os dados do conjunto de teste não devem ser usados para ajuste dos parâmetros da rede [26].

A habilidade de generalização da rede se refere a seu desempenho ao classificar padrões do conjunto de teste. Deficiências na capacidade de generalização da rede podem ser atribuídas ao problema de sobre ajuste (overfitting). Esse problema ocorre quando, após um certo período de treino, a rede se especializa no conjunto de treino e perde a capacidade de generalização. Diz-se então que a rede memorizou os padrões de treino, gravando suas peculiaridades e ruídos, proporcionando perdas na capacidade de generalização quando essa é utilizada para classificar os padrões pertencentes ao conjunto de teste [26].

A rede neuronal baseia-se nos dados a ela exibidos para extrair o modelo desejado. Portanto, a fase de treino deve ser rigorosa e verdadeira, a fim de serem evitados modelos hipotéticos.

4.2.4 Validação do modelo

Como teste de validação dos resultados, ou seja, para avaliar o poder de previsão da RNA, realizaram se previsões dos dados já conhecidos. Esses valores previstos, foram

então comparados com os valores reais, possibilitando, assim, calcular os erros da previsão, que correspondiam a diferença entre os valores previstos e os reais.

Dentre os principais resultados obtidos neste trabalho, quatro merecem ser destacados:

a. os melhores resultados para a previsão mediante o uso das RNAs, foram os obtidos com o treino da rede realizado com o algoritmo sendo 10 o número ideal de neurónios na camada intermediária, 10000 iterações e um declínio de 0.001 (Tabela.3). Estes parâmetros afectam o desempenho da RNA, devendo ser cuidadosamente escolhidos.

Tabela 3. validação dos resultados com base nos erros

Número d iterações	Neurónios na c. Intermédia	Declínio	Erro
100	2	0.001	0.000677
10000	2	0.001	0.000486
5000	5	0.001	0.000276
10000	8	0.001	0.000191
10000	6	0.001	0.000221
10000	10	0.001	0.000163
1500	9	0.0001	0.000223

b. Existe uma boa correlação entre os valores previstos e os valores reais. Para melhor representar os resultados foi representada uma linha que mostra uma forte correlação entre as duas variáveis.

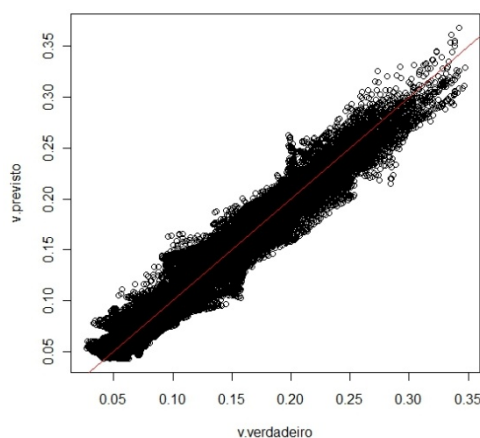


Fig.17. Correlação entre os valores previstos e os valores reais

c. A relação entre o erro e o valor real, mostra que os valores reais concentram-se numa faixa de erro mínima.

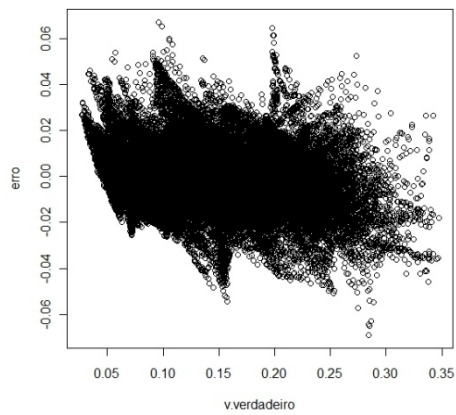


Fig.18 . Relação entre o erro e o valor real

d. Os resíduos apresentam uma distribuição normal, o que testa a variável quanto a fiabilidade do modelo da RNA.

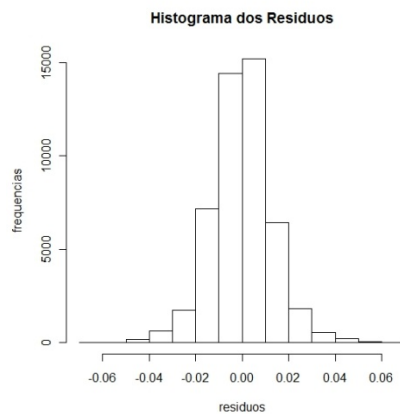


Fig.19.histograma dos residuos

e. Apresenta-se a figura 20, de forma a ver que para intervalos maiores há tendencialmente um erro maior.

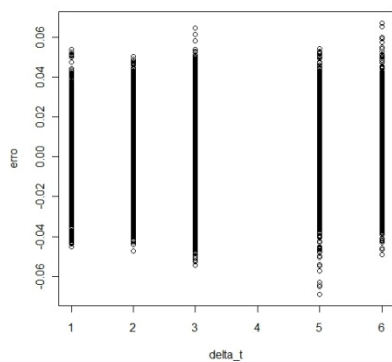


Fig. 20 variação do erro ao longo do tempo(

4.2.5 Previsões das taxas de prevalência

As figuras 21 e 22 mostram as projecções das taxas de prevalência do HIV para os anos 2011, 2013 e 2015, respectivamente, comparando com as taxas de prevalência do ano de 2007.

É de salientar que a escolha dos anos 2011, 2013 e 2015, deve-se a política da vigilância epidemiológica do HIV-SIDA em Moçambique que é feita de dois em dois anos.

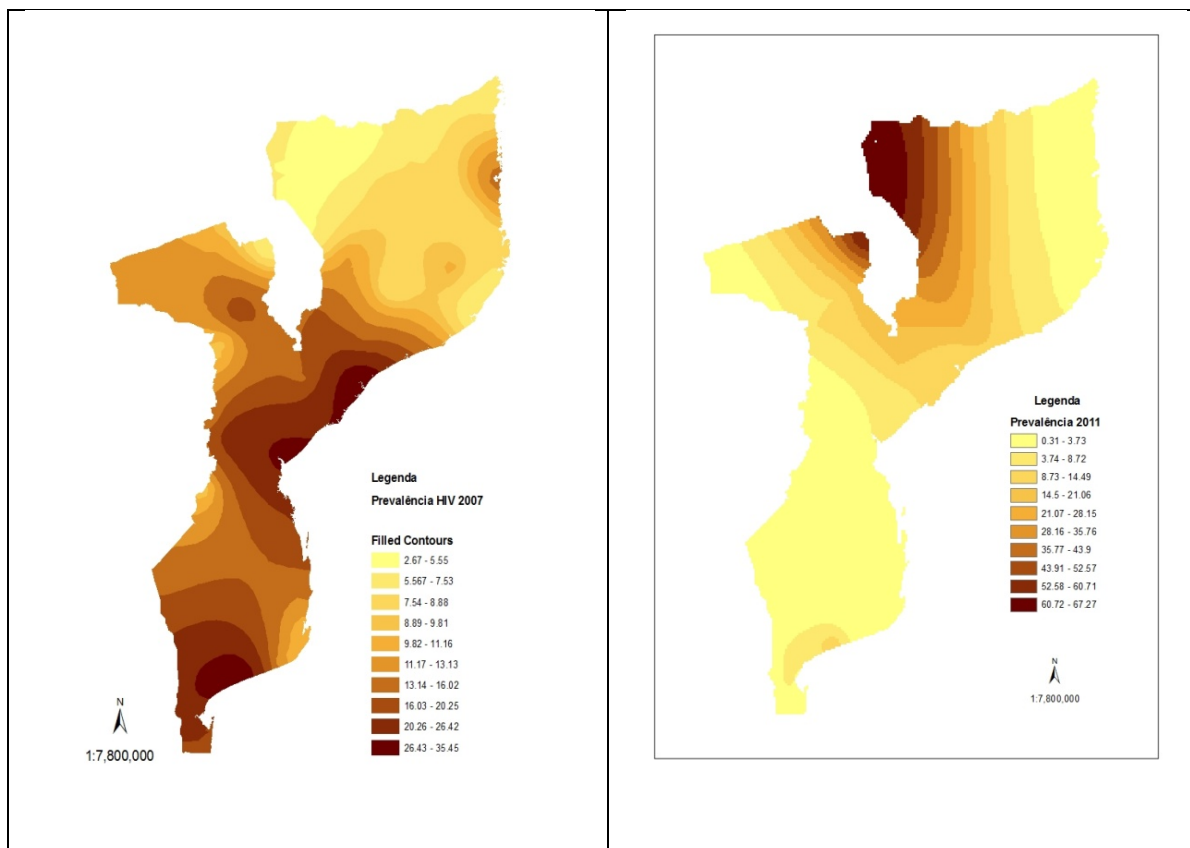


Fig.21 Comparação entre as prevalências do HIV-SIDA em 2007 e as previsões para os anos 2011.

Um facto curioso nas projecções, indicam uma subida acentuada das taxas de prevalência no corno da província de Niassa e Tete, contrariando a tendência das prevalências a nível nacional que é de redução.

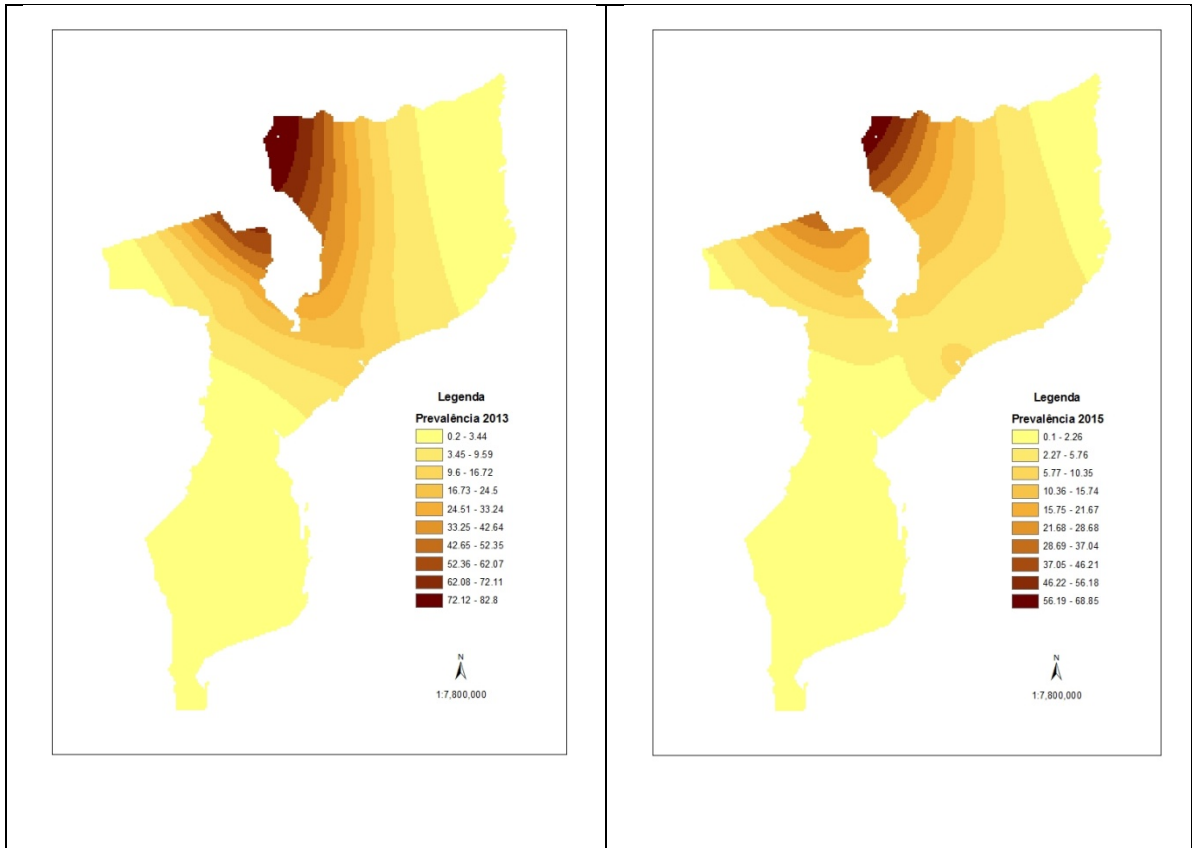


Fig.22 Previsões das prevalências do HIV-SIDA em 2013 e 2015.

Capítulo 5 CONCLUSÕES

Moçambique é um país pobre que enfrenta problemas de raízes históricas, história do colonialismo, guerra civil e desastres naturais. Os problemas agravam-se com a epidemia do HIV-SIDA, que já está a prejudicar o desenvolvimento do país.

O aumento do conhecimento e alteração dos comportamentos individuais e sociais são essenciais para a prevenção do HIV e subversão da actual tendência de agravamento.

A taxa de prevalência de HIV a nível nacional foi estimada em 16%, sendo alta, parece estar a atingir a fase *plateau*.

A zona Sul possui as maiores taxas de prevalência de HIV-SIDA, com tendência a subir.

Supõe-se que a epidemia atingiu a fase *plateau* na zona Centro do país.

A zona Norte continua com as taxas de prevalência mais baixas do país e parece estar numa fase estacionária.

As províncias com maior número de infectados são Sofala, Gaza, Maputo Cidade e Província, registando valores acima dos 20%.

Factores regionais e zonas fronteiriças dão a percepção de contribuir para os altos níveis de HIV-SIDA, províncias como Maputo Cidade e Província, Gaza, Manica e Tete, são grandes corredores de transporte.

As projecções feitas para os próximos anos indicam uma descida das taxas de seroprevalência a nível nacional, mas também indicam um aumento das taxas para as regiões dos cornos das províncias de Niassa e Tete.

A interpolação de dados é importante para estimar a evolução espacial da doença, análise usada para a previsão de dados nos locais não amostrados. É interessante visualizar alguns padrões das taxas de prevalência em províncias que fazem fronteira com os países vizinhos, porque parece haver uma continuidade das taxas de prevalência através da fronteira.

O baixo nível social das mulheres, a sua dependência económica e social em relação aos homens e a falta de controlo da sua sexualidade cria barreiras para a mulher na

sua habilidade de negociar sexo seguro, o que faz com que este grupo seja o grupo com maior risco de infecção.

Com o aumento de taxa de mortalidade devido a sida, a tendência será de aumentar o número de lares liderados por crianças, o que vai criar condições para aumentar a vulnerabilidade destes grupos.

Esta questão é hoje uma das principais ameaças ao desenvolvimento de Moçambique, com reflexos evidentes na produtividade da força de trabalho e em indicadores como a esperança média de vida da população. O número de órfãos devido à doença é igualmente preocupante, representando previsivelmente um esforço adicional em termos dos serviços de saúde e de assistência social

Apesar de o governo estar consciente da gravidade da epidemia, são necessárias acções urgentes para estancar o HIV-SIDA.

A rede neuronal artificial é uma técnica robusta que tem sido utilizadas como modelo de classificação e previsão no campo da epidemiologia.

Este trabalho demonstrou que o processo de descoberta de conhecimento em bases de dados e o uso de técnicas de data mining são de extrema importância porque podem auxiliar no processo de tomada de decisões.

Finalmente, referir que este estudo baseia-se em análise de dados e o futuro dirá se está correcto. Ainda precisa de ser validado com os dados das taxas de prevalências das próximas rondas de Vigilâncias Epidemiológicas do HIV-SIDA.

BIBLIOGRAFIA

1. ARNALDO, C. & A. Francisco. As características socioeconómicas dos distritos com postos de vigilância epidemiológica do HIV/SIDA e sua influência na taxa de prevalência do HIV/SIDA em Moçambique. Centro de Estudos de População. Faculdade de Letras. Universidade Eduardo Mondlane. Maputo. 2004.
2. BRITO, A. M.; Castilho E. A.; Szwarcwald. C. L. AIDS and HIV infection in Brazil: a multifaceted epidemic. Revista da Sociedade Brasileira de Medicina Tropical. 34(2): 207-217. mar-abr. 2000.
3. CÂMARA, G.; MEDEIROS, J. S. Princípios básicos em geoprocessamento. In: ASSAD, E. D.; SANO, E. E. (Ed.). Sistemas de informações geográficas: aplicações na agricultura. 2.ed. ver. ampl. Brasília, DF: Embrapa-SPI: Embrapa-CPAC. pp.3-11. 1998.
4. CNCS. Relatório do Projecto de Triangulação de Moçambique: síntese dos dados sobre as tendências das epidemias Nacional e locais e a análise de cobertura e intensidade dos esforços de prevenção - Processo, Principais Constatações e Recomendações. Maputo: Conselho Nacional de Combate ao SIDA. 2009.
5. COLEMBUNDERS R. L. & HEYWARD W. L. Surveillance of AIDS and HIV infection: opportunities and challenges. Health Policy. 15:1-11. 1990.
6. CONSELHO de MINISTROS. Plano Estratégico de Combate ao Sida 2010-2014. Maputo. Moçambique. 2010.
7. CONSELHO de MINISTROS. Plano Estratégico de Combate ao Sida. Maputo. Moçambique. 2004.
8. CONSELHO NACIONAL DE COMBATE AO SIDA, CNCS. Plano Estratégico Nacional de Combate ao HIV/SIDA, Partes I, II, III, Maputo. Moçambique. 2004.

9. CRESSIE, N. Statistics for spatial data. New York: Wiley. 1993.
10. DAHMER, Alessandra. Segmentação de imagens Ecocardiográfica Utilizando Redes Neurais e Medidas de Textura. Porto Alegre: CPGCC. 1998.
11. DIGGLE, P. J., RIBEIRO JÚNIOR, P. J. Model Based Geostatistics. In: SINAPE, 14., 2000, Caxambu. Anais... São Paulo: Associação Brasileira de Estatística. 2000.
12. HAYKIN, S. Neural networks: a comprehensive foundation. IEEE Press, New York. 1994.
13. ILO/AIDS, MOÇAMBIQUE: O impacto do HIV/AIDS em recursos humanos. Março de 2004. Acedido em 16.08.2010. http://www.ilo.org/public/portugue/region/eurpro/lisbon/pdf/moz_ilo_aids.pdf
14. INSTITUTO NACIONAL DE ESTATISTICA. Impacto Demográfico do HIV-SIDA Em Moçambique. Actualização. – Ronda de V.E. 2007. Maputo. Moçambique.
15. INSTITUTO NACIONAL DE ESTATISTICA. Inquérito Nacional Sobre Causas de Mortalidade 2007/2008. Relatório Preliminar. Maputo. Moçambique. 2009.
16. KIRUNGA, Christine T. & James P.M. Ntozi. Socio-economic determinants of HIV serostatus: a study of Rakai District, Uganda. *Health Transition Review*7 (Supplement): 175-188. 1997.
17. KOSKO, B. Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Prentice-Hall. 1992.
18. KUBY, J. Immunology, 5nd Edition, W.H.Freeman and company, New York. 2003.
19. LAMAS, C. La Geodemografía Y la investigación de medios. IV Seminario de AEDEMO sobre Medios Impresos, Radio y Publicidad Exterior. Bilbao, Noviembre de 1994.
20. LOURENÇO, R.W. Comparação entre métodos de interpolação para Sistemas de Informações Geográficas. Dissertação de mestrado elaborada junto ao curso de Pós-Graduação em Geociências – área de concentração em

Geociências e Meio Ambiente do Instituto de Geociências e Ciências Exatas da Universidade Estadual Paulista, Campus de Rio Claro. Rio Claro. SP. 1998.

21. McCUTCHAN, F.E. The Molecular Epidemiology of Infectious Diseases (ed. Thompson, R.C.A.), Hodder Arnold, Londres, 143-167.2000.
22. MENDONÇA, E. A. - HYCONES III: Sistema de Apoio à UTI Cardiológica. Curso de Pós-Graduação em Medicina - Instituto de Cardiologia do Rio Grande do Sul e Fundação Universitária de Cardiologia. Porto Alegre.1996.
23. MISAU PNC ITS/HIV-SIDA. Relatório sobre a Revisão dos Dados de V.E. do HIV-Ronda 2007. Maputo: Ministério da Saúde. Moçambique. 2008.
24. MOSES S, Bailey RC, Ronald AR. Male circumcision: assessment of health benefits and risks. *Sex Transm Infect* 1998;74:368-73.
25. ORGANIZAÇÃO MUNDIAL DA SAUDE. Módulo de Treinamento: Vigilância da Infecção pelo HIV. Programa Mundial de Controle da Aids, Genebra. Tradução pela Coordenação Nacional de DST e Aids, Ministério da Saúde, Brasília. 1993.
26. OSORIO, F. S. Um Estudo sobre Reconhecimento Visual de Caracteres Através de Redes Neurais. Porto Alegre: CPGCC. 1991.
27. R DEVELOPMENT CORE TEAM. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 2006.
28. RODDICK, J. F. e SPILLIOPOULOU, M.,A bibliography of temporal, spatial and spatio-temporal data mining research, SIGKDD Explorations. Vol. 1, No. 1, (in press). 1999.
29. TOBLER, W. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46 (2). pp.234-240. 1970.
30. TORGO, L. Programação, Análise de Dados e Sistemas de Apoio à Decisão usando o R. 2003.
31. UNAIDS, Country Profile 2007. Disponível em: http://www.unaids.org/en/Regions_Countries/Countries/mozambique.asp. Acesso em 28 ago. 2010.

32. WEISS, R.A. Gulliver's travels in HIVland, *Nature*, 410, 963-967. 2001
33. WHO. Carrying out HIV Sentinel Surveillance. World Health Organization, New Delhi India. 1994
34. WLSA, Feminização do Sida em Moçambique: Estudo Piloto. Maputo. Moçambique. 2003.
35. ZHU, T., Korber, B.T., Nahmias, A. J., Hooper, E., Sharp, P. M., Ho, D.D. An African HIV-1 sequence from 1959 and implications for origin of the epidemic, *Nature*, vol 391, 594-597. 2003.
36. Página do software R. acedida em Setembro de 2010. <http://cran.r-project.org/>
37. Página da MAP LIBRARY, acedida em Março de 2010. <http://www.maplibrary.org/stacks/Africa/Mozambique/index.php>

ANEXOS

O R possui uma quantidade variada de métodos, que podem ser usados para obter modelos para problemas de classificação. Para tal, apresenta um package **nnet**, que contém as funções necessárias para a obtenção de redes neuronais. **nnet** e **predict**, são as funções mais importantes do package.

nnet(x, y, size, decay, maxit, trace=T)

x: uma matriz ou dataframe com as variáveis de entrada na rede

y: uma matriz ou dataframe com as variáveis de saída da rede

size: número de nodos na camada intermédia. Se zero a rede não contém camada intermédia

decay: constante de decaimento dos pesos. Serve para evitar o overfitting da rede pois a cada iteração reduz os valores dos pesos da rede.

maxit: número máximo de iterações da rede. 100 é o valor por defeito

trace: como se fosse um interruptor para o rastreamento da aoptimização, por defeito é True.

Script do R para modelação da Rede Neuronal Artificial e previsão das prevalências para os anos de 2011, 2013 e 2015

```
# carregamento dos dados
```

```
dfs <- read.table('c:\\dfs.csv',header=T,sep=";",dec=".")
```

```
# carregamento da biblioteca nnet
```

```
library(nnet)
```

```
# verificação de valores existentes na rede
```

```

n<-length(rownames(dfs))
# escolha de 70% dos dados, para o treino.
trains<-sample(n, n*.7)
# normalização dos dados ( fazer com que assumam valores entre zero e um)
cols<-dfs$c/max(dfs$c)
rows<-dfs$r/max(dfs$r)
t1<-(dfs$t1-2000)/20
t2<-(dfs$t2-2000)/20
prev1<-dfs$prev1/100
prev2<-dfs$prev2/100
# organização dos dados num data frame
dfs2<-data.frame(rows, cols, t1,t2, prev1,prev2)
# definição dos dados de treino e de teste
dat_tr<-dfs2[trains,]
dat_te<-dfs2[-trains,]
# treino da rede neuronal
nn<-nnet(dat_tr[,1:5], dat_tr[,6], size=2, decay=0.001, maxit=10000, trace=T)
#previsão de dados
preds<-predict(nn, dat_te[,1:5])
# equação da soma das diferenças dos quadrados para a determinação do erro
dif2<-(preds-dat_te[,6])^2
sum(dif2)/

```

```
# projecção para 2011
proj11 <- read.csv2('c:\\proj11.csv')
cols<-proj11$col/max(proj11$col)
rows<-proj11$row/max(proj11$row)
t1<-(proj11$t1-2000)/20
t2<-(proj11$t2-2000)/20
prev1<-proj11$prev2007/100
proj11<-data.frame(rows, cols, t1,t2, prev1)
preds11<-predict(nn, proj11[,1:5])
```

```
# projecção para 2013
proj13 <- read.csv2('c:\\proj13.csv')
cols<-proj13$col/max(proj13$col)
rows<-proj13$row/max(proj13$row)
t1<-(proj13$t1-2000)/20
t2<-(proj13$t2-2000)/20
prev1<-proj13$prev2007/100
proj13<-data.frame(rows, cols, t1,t2, prev1)
preds13<-predict(nn, proj13[,1:5])
```

```
# projecção para 2015
proj15 <- read.csv2('c:\\proj15.csv')
cols<-proj15$col/max(proj15$col)
rows<-proj15$row/max(proj15$row)
t1<-(proj15$t1-2000)/20
t2<-(proj15$t2-2000)/20
```

```
prev1<-proj15$prev2007/100
proj15<-data.frame(rows, cols, t1,t2, prev1)
preds15<-predict(nn, proj15[,1:5])

# valores reais das prevalencias
preds11<-preds11*100
preds13<-preds13*100
preds15<-preds15*100

# ###Transformação dos dados no R em ficheiro csv
write.csv(preds11, file="preds11.csv")
write.csv(preds13, file="preds13.csv")
write.csv(preds15, file="preds15.csv")
```