

UNIVERSIDADE CANDIDO MENDES
PROGRAMA DE PÓS GRADUAÇÃO (MESTRADO) EM PESQUISA
OPERACIONAL E INTELIGÊNCIA COMPUTACIONAL

UBIRATAN ROBERTE CARDOSO PASSOS

COMPUTAÇÃO EVOLUTIVA E APRENDIZADO DE MÁQUINA
APLICADOS AO APOIO DO DIAGNÓSTICO DA CARDIOPATIA
ISQUÊMICA

CAMPOS DOS GOYTACAZES – RJ
Novembro de 2014
UNIVERSIDADE CANDIDO MENDES

PROGRAMA DE PÓS GRADUAÇÃO (MESTRADO) EM PESQUISA
OPERACIONAL E INTELIGÊNCIA COMPUTACIONAL

Ubiratan Roberte Cardoso Passos

**COMPUTAÇÃO EVOLUTIVA E APRENDIZADO DE MÁQUINA
APLICADOS AO APOIO DO DIAGNÓSTICO DA CARDIOPATIA
ISQUÊMICA**

Dissertação apresentada ao Programa de Pós-Graduação em Pesquisa Operacional e Inteligência Computacional, da Universidade Candido Mendes – Campos/RJ, para obtenção do grau de MESTRE EM PESQUISA OPERACIONAL E INTELIGÊNCIA COMPUTACIONAL.

Orientador: Prof. Dr. Ítalo de Oliveira Matias

CAMPOS DOS GOYTACAZES – RJ

Novembro de 2014

UBIRATAN ROBERTE CARDOSO PASSOS

**COMPUTAÇÃO EVOLUTIVA E APRENDIZADO DE MÁQUINA
APLICADOS AO APOIO DO DIAGNÓSTICO DA CARDIOPATIA
ISQUÊMICA**

Dissertação apresentada ao Programa de Pós-Graduação em Pesquisa Operacional e Inteligência Computacional, da Universidade Candido Mendes – Campos/RJ, para obtenção do grau de MESTRE EM PESQUISA OPERACIONAL E INTELIGÊNCIA COMPUTACIONAL.

Aprovado em ____ de dezembro de 2014

BANCA EXAMINADORA

Orientador: Prof. Dr. Ítalo de Oliveira Matias
UCAM – Universidade Candido Mendes

Examinador: Prof. Dr. Dalessandro Soares Vianna
UFF – Universidade Federal Fluminense

Examinador: Prof^ª. Dr^ª. Marcilene de Fátima Dianin Vianna
UFF – Universidade Federal Fluminense

CAMPOS DOS GOYTACAZES – RJ

Novembro de 2014

À minha família, que sempre me apoiaram e comigo estiveram em cada caminhada de minha vida, mostrando-me que longe podemos ir com amor e fé.

À minha amada esposa, que comigo caminhou por toda esta jornada, me fortaleceu, por mim orou e muito suportou para que eu hoje pudesse estar aqui.

AGRADECIMENTOS

À Deus, pelo dom da vida e sabedoria.

Ao prof. Dr. Ítalo de Oliveira Matias, que acreditou em mim e me incentivou dedicando seu tempo a me orientar e me levar na busca do saber, ao Professor Dr. Dalessandro Viana, Eduardo Shimoda, Aldo Shimoia e demais professores e colegas de curso.

Aos funcionários da UCAM, pelo apoio e bom trabalho realizado auxiliando-me sempre que necessário.

Aos meus amigos, Nivaldo Antônio da Silva, Marlito Pereira, Marcelo Albuquerque Schuster, Vinícius Machado, Flávio Targa Veríssimo, Hélio Carlos, Tiago Costa, Antônio Carlos Ribeiro e Jocimar Fernandes e todos os demais que sempre me incentivaram a continuar.

À Sandro Moretti, que gentilmente cedeu parte de sua pesquisa para que eu pudesse analisar.

“A ciência sem a religião é manca, a religião sem a ciência é cega.”

Albert Einstein

LISTA DE FIGURAS

Figura 1. Áreas relacionadas a IA.....	17
Figura 2. Estrutura Básica de um Sistema Especialista.....	18
Figura 3. Esquema de um sistema e uma máquina de aprendizado.	20
Figura 4. Aterosclerose.....	24
Figura 5. Ciclo de Vida de um Sistema RBC.....	28
Figura 6. Exemplo de codificação binária de um indivíduo.....	34
Figura 7. Método da Roleta.....	35
Figura 8. Mutação.....	36
Figura 9. Modelo de McCulloch e Pitts	38
Figura 10. Modelo MLP de McCulloch e Pitts.....	39
Figura 11. Fluxograma – Elaboração do Projeto.....	44
Figura 12. Caso de Uso – Visão Geral do Projeto.....	49
Figura 13. Caso de Uso – Consultas.....	50
Figura 14. Caso de Uso – Inteligência Computacional.....	50
Figura 15. Fluxograma – Algoritmo Genético.....	51
Figura 16. Algoritmo – População Inicial.....	52
Figura 17. Algoritmo – Função de Aptidão.....	53
Figura 18. Algoritmo – Função de Aptidão.....	53
Figura 19. Fluxograma – Conversão de Imagens.....	54
Figura 20. Imagem Padrão Sinal de ECG.....	54
Figura 21. Imagem Padrão Sinal de ECG - Monocromática.....	54
Figura 22. Imagem Padrão Sinal de ECG (Monocromática) – Busca por Padrões.....	55
Figura 23. Matriz binária sinal de EGC – Busca por Padrões.....	55
Figura 24. Algoritmo de Ativação da RNA.....	56
Figura 25. Representação numérica dos valores dos atributos.....	61
Figura 26. Comparação entre ferramentas e método tradicional - Variância.....	66
Figura 27. ANOVA e Teste de Tukey.....	67

LISTA DE GRÁFICOS

Gráfico 1. Erro médio do treinamento – Dados Clínicos.	62
Gráfico 2. Erro médio do treinamento – Sinais de ECG..	62
Gráfico 3. Erro médio do treinamento – Sobreposição RNA – Dados clínicos x RNA Imagens ECG.	63
Gráfico 4. Evolução do Algoritmo – Acertos máximos e médios.....	64
Gráfico 5. Análise de Convergência. Fonte:.....	64
Gráfico 6. Comparação de desempenho sensibilidade e especificidade.....	69

LISTA DE TABELAS

Tabela 1. Risco de morte, por 100 mil habitantes por doença isquêmica cardíaca (DIC) e doença cerebrovasculares (DCV); em homens (H) e em mulheres (M)	23
Tabela 2. Medidas de distância mais comuns utilizadas para cálculos de similaridade.....	31
Tabela 3. Comparação entre ferramentas e método tradicional - Variância.....	68

LISTA DE QUADROS

Quadro 1. Atributos considerados nos testes de similaridade do algoritmo.	58
Quadro 2. Pesos atribuídos por especialistas.	58
Quadro 3. Pesos atribuídos por especialistas.	59
Quadro 4. Configuração e resultados do processo de aprendizado das RNA's.	63
Quadro 5. Pesos atribuídos pelo AG.	65
Quadro 6. Comparação – Médias dos Pesos atribuídos pelos Especialistas e pelo AG.	66

LISTA DE SIGLAS

AVC	Acidente Vascular Cerebral
AVC M	Acidente Vascular Cerebral nas Mulheres
AVC H	Acidente Vascular Cerebral nos Homens
AM	Aprendizado de Máquina
AD	Árvores de Decisão
CE	Computação Evolutiva
CSCW	Computer-Supported Cooperative Work
DICOM	Digital Imaging Communications in Medicine
DIC	Doença Isquêmica Cardíaca
DIC M	Doença Isquêmica Cardíaca nas Mulheres
DIC H	Doença Isquêmica Cardíaca nos Homens
DCV	Doenças Cardiovasculares
DCNT	Doenças Crônicas Não Transmissíveis
ECG	Eletrocardiograma
FN	Falso Negativo
FP	Falso Positivo
FANN	Fast Artificial Neural Network
IA	Inteligência Artificial
IC	Inteligência Computacional
OCR	Optical Character Recognition
OMS	Organização Mundial de Saúde
PACS	Picture Archiving and Communication System
RBC	Raciocínio Baseado em Casos
RAM	Random Access Memory
RNA	Redes Neurais Artificiais
RNA MLP	Redes Neurais Artificiais Multilayer Perceptron
RL	Regressão Logística
SAD	Sistema de Apoio à Decisão
SSD	Sistema de Suporte à Decisão
SE	Sistema Especialista
TCP/IP	Transmission Control Protocol / Internet Protocol
UTI	Unidade de Terapia Intensiva
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

LISTA DE EQUAÇÕES

Equação (a)	Euclidiana
Equação (b)	Euclidiana Ponderada
Equação (c)	Manhattan
Equação (d)	Hamming
Equação (e)	Minkowsky Ponderada
Equação (f)	Tchebyshev
Equação (g)	Probabilidade de Seleção Roleta
Equação (h)	Função Discreta Contínua
Equação (i)	Conversão de imagem para monocromática
Equação (j)	Conversão tons de cinza
Equação (k)	Limiarização

SUMÁRIO

1. INTRODUÇÃO	10
1.1. Justificativa.....	10
1.2. Objetivos.....	11
1.3. Formulação do Problema.....	11
1.4. Organização do Trabalho.....	12
2. FUNDAMENTAÇÃO TEÓRICA	13
2.1. Trabalhos Relacionados.....	13
2.1.1. Trabalhos relacionados à utilização de inteligência computacional aplicada ao diagnóstico médico.....	13
2.1.2. Trabalhos relacionados à utilização de inteligência computacional aplicada ao diagnóstico de doenças cardíacas.	14
2.1.3. Comentários Finais	16
2.2. Inteligência Computacional.....	16
2.3. Sistemas Especialistas	17
2.4. Aprendizado de Máquina.....	19
2.4.1. Aprendizado Supervisionado.....	20
2.4.2. Aprendizado Não-Supervisionado.....	21
2.4.3. Aprendizado Semi-Supervisionado.....	21
2.5. Gestão do Conhecimento.....	21
2.6. Conhecimento Médico.....	22
2.7. Doenças Cardiovasculares.....	23
2.7.1. Cardiopatia Isquêmica	24
2.7.2. Eletrocardiograma.....	25
2.7.3. Raciocínio Baseado em Casos	27
2.7.3.1. Raciocínio Baseado em Casos no Diagnóstico Médico.....	29
2.7.3.2. Cálculo de Similaridade	30
2.7.4. Computação Evolutiva	32
2.7.4.1. Algoritmos Genéticos.....	32
2.7.4.2. Representação.....	33
2.7.4.3. Avaliação.....	34
2.7.4.4. Seleção	34
2.7.4.5. Cruzamento	35

2.7.4.6.	Mutação.....	36
2.7.5.	Redes Neurais Artificiais	36
2.7.5.1.	MLP – Multi-Layer Perceptron – Perceptron Multicamadas.....	38
2.7.6.	Computação Gráfica.....	40
2.7.6.1.	Imagem Digital.....	40
3.	METODOLOGIA	43
3.1.	Classificação da Pesquisa	43
3.2.	Suposições e Hipóteses.....	43
3.3.	Fluxo de Trabalho.....	44
3.4.	Coleta e Análise de Dados.....	44
3.5.	Delimitação do Escopo.....	45
3.6.	Resultados Esperados	45
3.7.	Técnicas, recursos e configurações	46
4.	PROPOSTA DO CARDIOCARE.....	48
4.1.	Arquitetura.....	49
4.2.	Algoritmos	51
4.2.1.	Algoritmos Genéticos (RBC).....	51
4.2.2.	Computação Gráfica / RNA.....	54
5.	RESULTADOS.....	57
5.1.	Parametrização e Atributos.....	57
5.2.	Resultados.....	61
5.2.1.	Resultados RNA	61
5.2.2.	Resultados RBC (AG).....	63
5.2.3.	Validação	66
6.	CONSIDERAÇÕES FINAIS	70
6.1.	Conclusões.....	70
6.2.	Trabalhos Futuros	71
7.	REFERÊNCIAS	72

RESUMO

COMPUTAÇÃO EVOLUTIVA E APRENDIZADO DE MÁQUINA APLICADOS AO APOIO DO DIAGNÓSTICO DA CARDIOPATIA ISQUÊMICA

As enfermidades cardíacas são atualmente uma das maiores causas de morte por doenças não transmissíveis no país. A grande dificuldade em se realizar um diagnóstico preciso e ágil é um dos fatores que contribuem para esta realidade, o que se agrava devido a falta de recursos e profissionais em algumas localidades. Surge então neste cenário como possível solução diversos avanços tecnológicos que permitem auxiliar os profissionais médicos no processo de diagnóstico. Baseado nesta afirmativa é proposto neste trabalho uma ferramenta híbrida, que combina diversas técnicas de inteligência computacional (Redes Neurais Artificiais, Algoritmos Genéticos, Raciocínio Baseado em Casos e Computação Gráfica), capaz de através de análises de históricos clínicos e imagens de sinais de eletrocardiogramas, indicar com até 97,01% de acerto o diagnóstico positivo ou negativo para cardiopatia isquêmica. O trabalho em questão foi desenvolvido em parceria com profissionais médicos especialistas, tendo sido também utilizada uma base de dados contendo dados clínicos de aproximadamente 1020 pacientes, com seus respectivos diagnósticos. Foi utilizada também uma base de imagens contendo 92 imagens de sinais de eletrocardiogramas, para submissão a análise da Rede Neural Artificial desenvolvida neste projeto. Ao final dos testes a ferramenta apresentou taxas de acerto expressivas sendo também avaliada pela comunidade médica que apoiou o projeto, recebendo boa aceitação dos mesmos, cumprindo desta forma seu objetivo.

PALAVRAS-CHAVE: Algoritmos Genéticos, Redes Neurais Artificiais, Raciocínio Baseado em Casos, Cardiopatia Isquêmica, Suporte à Decisão Médica.

ABSTRACT

EVOLUTIONARY COMPUTING AND LEARNING MACHINE APPLIED TO SUPPORT THE DIAGNOSIS OF ISCHEMIC HEART DISEASE

Cardiac diseases are currently a leading cause of death from non-communicable diseases in the country. The difficulty in performing an accurate and responsive diagnosis is one of the factors contributing to this situation, which is exacerbated by a lack of resources and personnel in some locations. Then emerges as a possible solution in this senary many technological advances that enable assist medical professionals in the diagnosis process. Based on this statement is proposed in this paper a hybrid tool that combines various computational intelligence techniques (Artificial Neural Networks, Genetic Algorithms, Case-Based Reasoning and Computer Graphics), capable of through analysis of clinical signs and historical images of electrocardiograms, indicate up to 97.01% accuracy the positive or negative diagnosis of ischemic heart disease. The work in question was developed in partnership with medical specialists and was also utilized a base of satisfaction data clinical data from approximately 1,020 patients, with their respective diagnoses. A database of images containing 92 images of signs of electrocardiograms, for submission to the Artificial Neural Network analysis developed in this project was also used. At the end of the tool tests showed significant hit rates also being evaluated by the medical community who supported the project, receiving good acceptance thereof, thus fulfilling their goal.

KEYWORDS: Genetic Algorithms, Neural Networks, Case-Based Reasoning, Ischemic Heart Disease, Medical Decision Support.

1. INTRODUÇÃO

1.1. Justificativa

O aprendizado humano dá-se principalmente por experiências. No âmbito da IA, as técnicas de RBC são capazes de imitar essa característica humana na tentativa de resolver problemas, realizando buscas em uma base de conhecimento na qual casos similares estão armazenados. Esta busca de casos similares é um dos problemas centrais do RBC.

O trabalho em questão refere-se a uma pesquisa motivada pela necessidade de resolver problemas concretos, tendo como finalidade prática investigar o problema de atribuição de pesos para os diversos atributos analisados pelo sistema RBC no domínio estudado. O sucesso dos sistemas RBC dependem fundamentalmente da seleção dos casos mais similares, sendo necessário desta forma buscar formas de se otimizar esta definição de pesos. O AM de máquina preocupa-se com como se construir sistemas computacionais que melhorem seu desempenho em determinadas tarefas através de experiências, dentre os métodos de AM pode-se destacar os AG's. AG's são métodos de otimização e busca inspirados nos mecanismos de evolução dos seres vivos, onde otimização consiste em tentar várias soluções e utilizar a informação obtida neste processo de forma a encontrar soluções cada vez melhores (CARVALHO e LACERDA, 1999).

A complexidade e importância do diagnóstico médico, fazendo com que erros possam levar a resultados catastróficos elevam a importância do tema aqui discutido (GOMES, 2011). Gomes (2011) esclarece ainda que não é possível afirmar que todo diagnóstico é realizado sem

que haja equívocos, uma vez que esta é uma atividade predominantemente intuitiva. Soma-se a tudo o fato de que, em diversas localidades existe carência profissional de diversas especialidades, e um médico, por mais que seja capacitado não é capaz de atender a todas as especialidades. Estes fatores criam um ambiente propício à aplicação e propagação de ferramentas inteligentes que possam de alguma forma reduzir a complexidade da atividade de diagnósticos, trazendo desta forma uma série de vantagens para a sociedade em geral (SETIAWAN, VENKATACHALAM e HANI, 2009).

1.2. Objetivos

O principal objetivo deste trabalho é propor o desenvolvimento de uma ferramenta híbrida, onde seja possível registrar as experiências dos especialistas médicos, sobretudo àquelas relacionadas ao domínio estudado (cardiopatia isquêmica), permitindo posteriormente que o sistema utilize deste conhecimento para soluções de problemas futuros através da aplicação das técnicas de inteligência artificial (Reconhecimento Baseado em Casos, Algoritmos Genéticos e Redes Neurais Artificiais).

Mais especificamente, a solução aqui proposta conta com os princípios das tecnologias cima citadas de modo a:

- Permitir a inserção de informações relacionadas a consultas médicas, exames e diagnósticos, permitindo desta forma o crescimento da base de conhecimento do sistema;
- Buscar, recuperar, adaptar e solucionar novos casos através de operadores genéticos e neurais, quando não forem encontradas soluções específicas na base de conhecimentos;
- Facilitar a identificação e remoção de casos considerados desnecessários na base de conhecimentos, contribuindo para a melhoria da ferramenta;
- Permitir a inserção de novas soluções adaptadas, baseadas nos conhecimentos anteriores, enriquecendo o conhecimento da ferramenta proposta.

1.3. Formulação do Problema

Este projeto segue as diretrizes do projeto IACVIRTUAL (ALMEIDA, 2006). Seu propósito é definir o grau de correlação entre cada um dos atributos/sintomas analisados com o

diagnóstico final. Atualmente o processo de diagnóstico é feito com pouco ou sem qualquer suporte originado de ferramentas inteligentes, sendo este um processo totalmente subjetivo, mesmo quando estes profissionais lançam mão de exames mais conclusivos como sinais de eletrocardiograma, cuja interpretação pode variar de profissional para profissional.

Deve-se ressaltar que o projeto CardioCare não busca a substituição do profissional médico especialista, mas sim servir como sistema de suporte ao processo decisório. Assim, a ideia é manter a atribuição de pesos pelo especialista, mas apresentar uma forma de suporte a esta decisão com o cálculo automático dos pesos das características.

1.4. Organização do Trabalho

O trabalho em aqui apresentado está dividido em 6 capítulos, conforme descrito abaixo:

- a) Capítulo 1, **Introdução**. Breve explicação do assunto abordado, justificando o e apresentando os objetivos do trabalho;
- b) Capítulo 2, **Inteligência Computacional**. Apresentação de uma visão geral da área de inteligência computacional, com histórico das principais tecnologias aplicadas neste trabalho;
- c) Capítulo 3, **Metodologia**. Descreve de forma clara o processo de elaboração deste trabalho, as fontes de dados, as principais referências, e detalhes dos processos;
- d) Capítulo 4, **Proposta do CardioCare**. Apresentação da arquitetura, proposta e propósito da ferramenta que deu origem a este trabalho;
- e) Capítulo 5, **Resultados**. Demonstração dos resultados dos testes e treinamentos realizados com a ferramenta. Compostos de dados estatísticos, informações sobre os processos de simulação, resultados comparativos e agrupados por cada uma das técnicas aplicadas e utilizadas pela ferramenta;
 - Capítulo 6, **Conclusões e Trabalhos Futuros**. Apresentação das conclusões às que se pode alcançar com o desenvolvimento deste trabalho, bem como apresentação de futuras propostas para o mesmo (tanto em termos de implementação quanto aplicação).

2. FUNDAMENTAÇÃO TEÓRICA

2.1. Trabalhos Relacionados

Motivados principalmente pela implantação de novas tecnologias (principalmente as que envolvem modelagens ou simulações de ambientes reais e desenvolvimento de aplicações médico-hospitalares, voltadas a otimizar os processos da área de saúde) as pesquisas na área médica direcionam-se para o melhoramento dos serviços prestados, sendo portanto, um instrumento que procura otimizar o atendimento e minimizar os riscos à saúde dos pacientes.

Na literatura é possível encontrar vários exemplos de utilização de tais recursos na área da saúde, sendo alguns deles apresentados abaixo.

2.1.1. Trabalhos relacionados à utilização de inteligência computacional aplicada ao diagnóstico médico.

Atualmente existem muitas soluções destinadas ao suporte ao trabalho do especialista médico, como por exemplo o projeto HealthNet (BARBOSA *et al.*, 2001)(BARBOSA *et al.*, 2003), que propõe a realização de telediagnóstico e de segunda opinião a partir de uma rede metropolitana de computadores em Recife/PE, tendo o suporte de dois hospitais locais de grande porte. Na referida proposta, os hospitais responsáveis pelo suporte apoiam outros menores no Estado com serviços de telediagnóstico (GLAUCIA, 2012).

Outra experiência, que objetiva projetar e implementar, através de videoconferência, Mejia *et al.* (2003) propõe projetar uma rede de intercâmbio de informações relacionadas aos

pacientes em tempo real. O projeto prevê a construção de uma rede de Telemedicina, na qual as informações podem ser estáticas ou dinâmicas, análogas ou digitais, com dados, imagens (Padrão DICOM¹), sinais biométricos e voz.

Cita-se também o projeto *Co-Surgeon*, sistema capaz de combinar simulação cirúrgica 3D² com tecnologia *Computer-Supported Cooperative Work* (CSCW) para viabilizar, através de acesso remoto de múltiplos usuários em uma rede TCP/IP, a manipulação de modelos anatômicos 3D e simulação de cirurgias compartilhando a mesma visão (GLAUCIA, 2012).

Barros e Wangenheim (2000) apresentaram um trabalho que consiste no acesso a um banco de imagens, em conformidade ao padrão DICOM, num ambiente em que equipes médicas podem examinar uma mesma imagem e elaborarem conjuntamente um laudo através de conferências de voz. Hsu *et al.* (2008) propõe, baseado na combinação de CSCW e *Picture Archiving and Communication System* (PACS), um sistema de teleconsulta visando apoio ao diagnóstico remoto. O conceito de comunidade virtual de internet também é adaptado para fornecer uma forma de conexão rápida e conveniente para os participantes.

Araújo (2008) desenvolveu o *Health-Image*, que é um ambiente computacional voltado para a transmissão e manipulação de imagens médicas, visando o apoio ao diagnóstico através da Web. Em comum todos os trabalhos apresentam soluções que fazem uso de tecnologias computacionais para suporte à atividade do diagnóstico médico.

2.1.2. Trabalhos relacionados à utilização de inteligência computacional aplicada ao diagnóstico de doenças cardíacas.

No tocante à implementação de sistemas inteligentes de apoio a decisão médica, pode-se citar o projeto *Lepidus* (UNIFESP/USP). Trata-se de um sistema inteligente de apoio à decisão médica que fornece uma lista de hipóteses candidatas, quando lhe é informado um conjunto de atributos (sintomas e sinais). O projeto é baseado em RNA's para processar sintomas e sinais e fornecer respostas, a taxa média de acerto deste sistema é de 84% (SILVA *et al.*, 2002).

¹ DICOM, abreviação de *Digital Imaging Communications in Medicine* (ou Comunicação de imagens digitais em medicina), é o conjunto de normas para tratamento, armazenamento e transmissão de informação médica (imagens médicas) num formato eletrônico, estruturando um protocolo.

² Imagens 3D são imagens de duas dimensões elaboradas de forma a proporcionarem a ilusão de terem três dimensões.

Jara *et al.* (2009) utilizaram inteligência artificial para detectar e prever doenças miocárdicas, tal solução utiliza sinais vitais para, por um lado realizar a detecção de sintomas através de um sistema de regras e, por outro lado, efetuar a previsão da doença através de algoritmos de cronobiologia. A proposta em questão utilizou uma base de conhecimento para representar informações humanas gerais, detalhes de coração e parâmetros de eletrocardiograma (ECG) baseados em ontologias e sistemas de regra para detectar anomalias e padrões de informações do eletrocardiograma.

Ainda no diagnóstico cardíaco, Mo e Xu (2010) propuseram a utilização de SVM com base na função de *Kernel* híbrido e Leite *et al.* (2010) utilizaram redes neurais competitivas na classificação de arritmias cardíacas.

Antonio (2008) aplicou RNA MLP (*Multilayer Perceptron*) com *backpropagation* na identificação de áreas hiperdensas e/ou hipodensas anormais em exames de Tomografia Computadorizada de Crânio, em conjunto com o histórico clínico do paciente, para auxílio no diagnóstico médico emergencial, a partir da detecção da ocorrência do Acidente Vascular Cerebral (AVC) hemorrágico ou isquêmico. Leite (2011) apresentam um modelo *fuzzy* para pré-diagnóstico de AVC aplicado ao monitoramento de pacientes internados em Unidade de Terapia Intensiva (UTI).

Almeida (2006) propõem em seu trabalho o desenvolvimento de uma aplicação baseada em algoritmos genéticos e raciocínio baseado em casos que, através da análise de uma série de atributos do histórico clínico do paciente, determina qual a relação de cada um destes com o diagnóstico final, sendo ao o algoritmo capaz de definir a probabilidade de este paciente apresentar diagnóstico positivo ou não para a doença obstrutiva coronariana, obtendo taxas de até 90% de acertos.

Muthukaruppan e Er (2012) apresentaram uma ferramenta híbrida baseada em lógica *fuzzy*, preparada para auxílio ao diagnóstico das doenças do trato coronariano capazes de através de análise de uma série de atributos fornecer ao especialista diagnósticos com até 93,27 de precisão. Ainda nesta mesma linha de pesquisa, Khatibi e Montazer (2010) apresentaram, também baseada em lógica *fuzzy*, uma ferramenta capaz de, através de uma séries de inferências e análises de vários atributos indicar com precisão de até 91,58% os riscos de um pacientes apresentar alguma doença do trato coronariano.

Por fim, Başçiftçi e Incekara (2012) apresentam uma ferramenta, também baseada em lógica *fuzzy*, entretanto, com suporte a web, capaz de realizar análises de informações do histórico do paciente e com base neste oferecer resultados com até 86,1% de acertos para doenças do trato cardíaco.

2.1.3. Comentários Finais

Considerando os trabalhos apresentados, nota-se que muitas das propostas apresentadas realizam diagnósticos com apoio da inteligência computacional. Entretanto poucas apresentam soluções híbridas, principalmente no que tange a aplicação de algoritmos genéticos combinados com as técnicas de raciocínio baseado em casos e redes neurais artificiais. Apesar de em bom número existe ainda alguma carência de sistemas gratuitos de qualidade aptos a realizar diagnósticos através do histórico do paciente, oferecendo diagnóstico diferenciado através da análise de sinais cardíacos.

2.2. Inteligência Computacional

Sobre IA, pode-se afirmar que não é possível defini-la formalmente, uma vez que para isto seria necessário definir primeiramente a própria inteligência. Entretanto é possível encontrar algumas propostas de definições operacionais, tais como: “uma máquina inteligente se ela é capaz de solucionar uma classe de problemas que requerem inteligência para serem solucionados por seres humanos” (MCCARTHY e HAYES, 1969); “Inteligência Artificial é a parte da ciência da computação que compreende o projeto de sistemas computacionais que exibam características associadas, quando presentes no comportamento humano, à inteligência” (BARR e FEIGENHAUM, 1981); ou ainda “Inteligência Artificial é o estudo das faculdades mentais através do uso de modelos computacionais” (CHARNIAK e MCDERMOTT, 1985). Alguns outros autores preferem não propor uma definição para o termo concentrando-se em seus objetivos, a saber: “tornar os computadores mais úteis e compreender os princípios que tornam a inteligência possível” (WINSTON, 1985)

Atualmente pode-se identificar duas principais linhas de pesquisa para a construção de sistemas inteligentes: a linha **conexionista** e a linha **simbólica** (MINSKY e PAPERT, 1969).

IA é uma vasta área de conhecimento que entre outras engloba as seguintes técnicas: Redes Neurais Artificiais (RNA's), Lógica Fuzzy, Sistemas Especialistas (SE), Sistemas de Suporte a Decisão (SSD), Computação Evolutiva (CE) e Aprendizado de Máquina (AM) (SETIAWAN, VENKATACHALAM e HANI, 2009), baseadas nas mais diversas áreas do conhecimento. Na Figura 1 pode-se observar as tecnologias relacionadas a IA, e quais as áreas onde ela é aplicada.

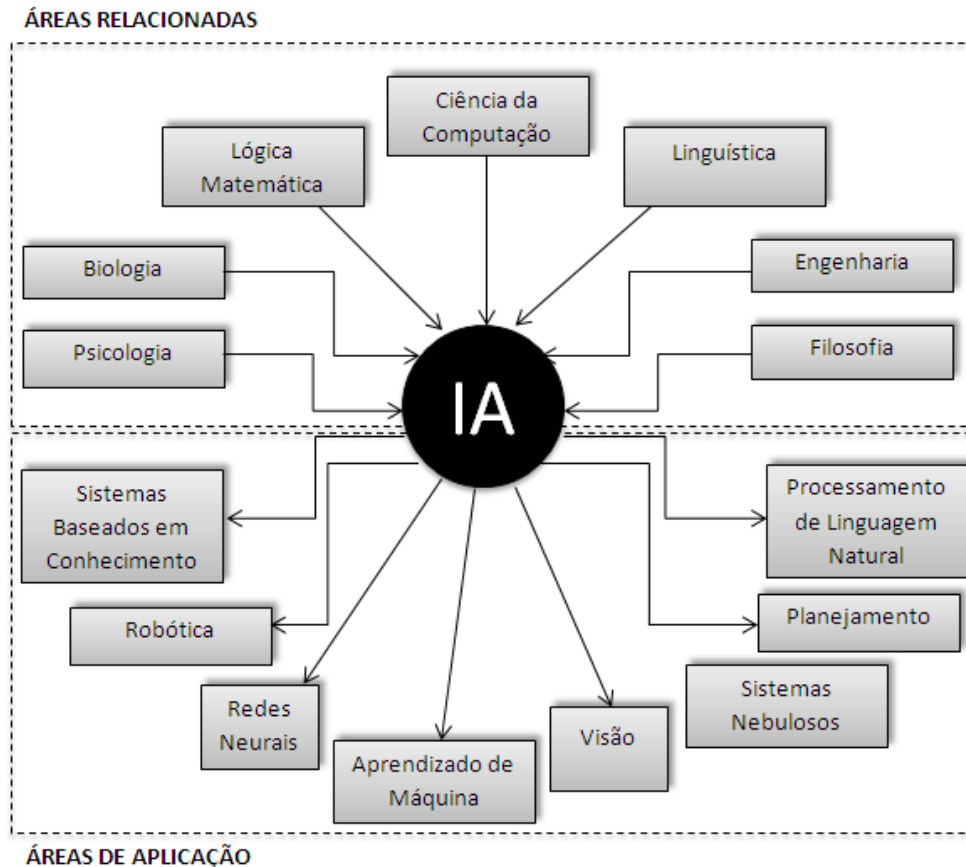


Figura 1. Áreas relacionadas a IA. Fonte: (SETIAWAN, VENKATACHALAM e HANI, 2009).

Logo, neste trabalho foram aplicadas as seguintes tecnologias: Sistemas Especialistas, Aprendizado de Máquina, Computação Evolutiva, Algoritmos Genéticos e Redes Neurais Artificiais além de Raciocínio Baseado em Casos, formando assim uma ferramenta híbrida de aplicação de métodos heurísticos para auxiliar no diagnóstico da cardiopatia isquêmica.

2.3. Sistemas Especialistas

Sistemas especialistas é um termo proveniente dos sistemas conhecidos como sistemas de produção, nos quais pares de expressões constituam um conjunto de condições e ações. Emil Leon Post, introduziu a ideia de sistemas de produção ainda em 1936 propondo sistemas hoje conhecidos como sistemas de Post. Estes sistemas são constituídos de um conjunto de regras para especificação sintática de transformações realizadas sobre uma cadeia de caracteres, representadas por um método geral para processamento de dados.

Estes sistemas apresentam um conjunto expressivo de benefícios se comparados a outros desenvolvidos com base em técnicas tradicionais. Isto se deve ao fato de serem

compostos por algum tipo de inteligência e conhecimento. Dentre as vantagens destes sistemas pode-se destacar:

- a) São capazes de estender as facilidades de tomada de decisão para muitas pessoas. Desta forma, o conhecimento dos especialistas pode ser distribuído de forma que possa ser utilizado por um grande número de pessoas;
- b) Podem melhorar a produtividade e desempenho dos seus usuários devido a sua capacidade de prover conhecimento que, em condições normais, demandariam certa quantidade de tempo para ser disseminado e utilizado no processo de tomada de decisão;
- c) Auxiliam na redução do grau de dependência que as organizações mantêm quando expostas a situações críticas ou inevitáveis, como por exemplo, a falta de um especialista;
- d) Mostram-se adequados também para serem utilizados em treinamentos em grupo, de uma forma prática e agradável, servindo inclusive como instrumento de coleta de informações pós-treinamento.

A estrutura básica de um sistema especialista é composta de três elementos fundamentais: base de conhecimento, motor de inferência e interface com o usuário como pode ser observado na Figura 2.

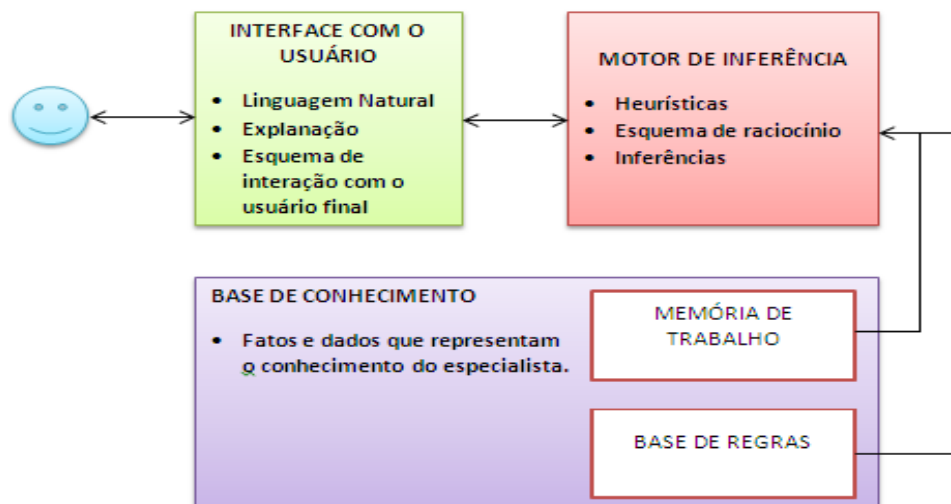


Figura 2. Estrutura Básica de um Sistema Especialista. Fonte: Adaptado de (NILSON, 1982).

Rich (1983) descreve o **motor de inferência** como sendo primordial para a existência de um sistema especialista, trata-se de seu núcleo. Sua função consiste em intermediar as relações entre os fatos e regras e heurísticas que compõem a base de conhecimento.

A tradicional **base de conhecimento** referente ao armazenamento de dados estáticos aqui é substituído por uma base de regras, fatos e também heurísticas que correspondem ao

conhecimento dos especialistas de um determinado domínio. Esta base de fatos e regras realiza interações com o usuário e com o motor de inferência permitindo desta forma a identificação dos problemas a serem resolvidas, as possíveis soluções além do processo de raciocínio e inferências que geram as conclusões sobre o mesmo (CHECLAND, 1981).

Aluri (1988) afirma que a **interface com o usuário** é a etapa que requer maior dedicação de tempo com projeto e implementação. Heurísticas são procedimentos informais, logo um problema submetido a um sistema especialista é endereçado por estratégias de busca, assim sendo estes sistemas sempre retém elementos de memória que lhes permitem o encaixe e encadeamento com estas estratégias. Logo é necessário que a interface seja bastante flexível conduzindo a um processo de navegação intuitiva durante a interação do usuário com o sistema, tornando-a assim eficiente. A interface **flexível** com o usuário permite que este descreva o problema ou os objetivos que deseja alcançar, além de permitir que usuário e sistema adotem um modelo estruturado de consultas (ALURI, 1988).

2.4. Aprendizado de Máquina

A partir da evolução dos computadores os aspectos técnicos ganharam força, tendo sido realizadas pesquisas relevantes nessa área, demonstrando que máquinas podem mostrar um nível significativo de aprendizado mesmo as fronteiras dessa habilidade não estando claramente definidas (PRATI, 2006).

Aprendizado é qualquer mudança em um sistema que melhore o seu desempenho a partir da segunda vez em que este repete a mesma tarefa ou outra tarefa similar. O aprendizado de máquina (AM) é uma parte da Inteligência Artificial (IA) responsável pelo desenvolvimento de teorias computacionais focadas na criação do conhecimento artificial. Aprendizado de Máquina (Machine Learning, em inglês), é a área da IA que objetiva o desenvolvimento de técnicas computacionais sobre processo de aprendizado.

Barone (2003) propõe a seguinte definição de aprendizado de máquina: “Aprendizado é o processo pelo qual um sistema inteligente é adaptado através do estímulo do ambiente no qual está inserido”, sendo o tipo de aprendizado determinado pela maneira a qual esta adaptação é realizada. A forma como os estímulos do ambiente se relacionam com os ajustes realizados na Máquina de Aprendizado é que definem os diversos métodos de aprendizado. Basicamente pode-se considerar os seguintes métodos de aprendizado: Supervisionado, Não-Supervisionado e Semi-Supervisionado, e em Indutivos e Transdutivos. Silva (2008) define o esquema de apresenta o esquema de aprendizado da forma como é apresentada na Figura 3.

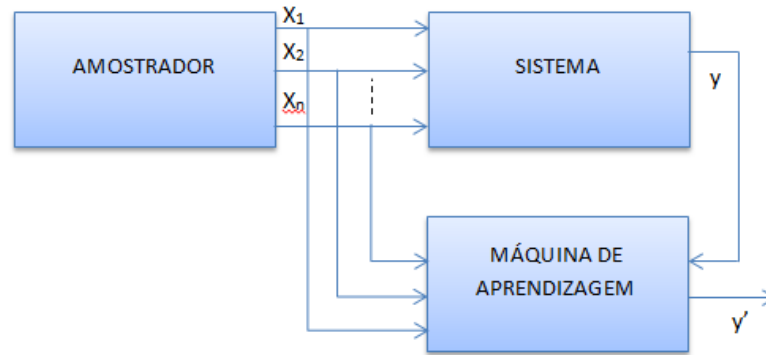


Figura 3. Esquema de um sistema e uma máquina de aprendizado. Fonte: Adaptado de (Silva, 2008)

Nota-se pela figura acima que o sistema obtém as informações das observações realizadas a partir de x_i , sendo as saídas dadas por y_i , e as respostas da máquina de aprendizado buscam emular o comportamento do sistema dado por y' . O Aprendizado de Máquina apresenta a vantagem de apresentar boas soluções para problemas onde não se pode encontrar soluções exatas, esta técnica reduz a necessidade de inserção de conhecimento de um especialista para que seja possível encontrar a solução do problema além de permitir o encontro de regras e relacionamentos implícitos nos dados, mas dificilmente observáveis por especialistas humanos. O aprendizado dos sistemas podem ser classificados basicamente como: **indutivo**, podendo ser **não supervisionado** e **supervisionado**, e dentro dos estudos de aprendizado supervisionado, classificam-se como **classificação** e **regressão** Russell & Norving (2004).

Para uma determinada classe de problemas o aprendizado indutivo pode se apresentar como uma abordagem complexa ou proibitiva, levando até mesmo a comportamentos indesejados. Para estes casos utiliza-se a **inferência transdutiva** que consiste numa abordagem alternativa que indica “pular” o passo de indução de regras, realizando a predição de um fenômeno e relacionando-o diretamente o mesmo com os padrões anteriormente observados.

2.4.1. Aprendizado Supervisionado

Neste tipo de aprendizado, o objetivo é induzir conceitos a partir de exemplos pré-classificados, ou seja, exemplos rotulados como uma classe conhecida, possuindo estas classes valores discretos. O problema em questão pode ser classificado como de categorização sendo os valores destas classes contínuos, então o problema é categorizado como de regressão. Bigus (1996) relata que o aprendizado supervisionado é utilizado quando se dispõe de um banco de dados com tantas perguntas quanto respostas, utilizando este para treinamentos de redes neurais

na obtenção de classificações, funções de aproximação ou modelagem e previsão baseadas no tempo.

2.4.2. Aprendizado Não-Supervisionado

Para casos onde existe uma incerteza sobre a saída esperada é necessário utilização de métodos probabilísticos, permitindo assim simular uma experiência não vivida. Para estes procedimentos é amplamente difundida a utilização da aprendizagem **bayesiana** ou **redes bayesianas**, sendo os dados definidos como evidências e hipóteses como teorias probabilísticas. Para realização destes cálculos utiliza-se a regra proposta por Russell & Norving (2004). Um dos modelos mais comuns de aprendizado não supervisionado é o **aprendizado por reforço**. Este tipo de aprendizado se baseia em dados de um ambiente totalmente observável. O objetivo deste tipo de aprendizado é descobrir o quanto uma política é boa, ou seja, descobrir sua utilidade.

2.4.3. Aprendizado Semi-Supervisionado

Neste tipo de aprendizado uma parte dos dados que é utilizada no treinamento é classificada, enquanto outra é constituída de dados não rotulados. Este paradigma mostra-se bastante útil quando os casos apresentam uma amostragem limitada do conjunto de treinamento e não fornece informação suficiente para utilizar a indução como regra geral. Assim sendo um conjunto de teste é utilizado como fonte extra de informação para resolução do problema proposto (SILVA, 2008).

Esta abordagem tem se mostrado bastante útil em problemas onde o espaço de amostragem é demasiadamente grande tornando inviável a geração de uma amostra estatisticamente representativa, ou até mesmo em casos em que o classificador mostra-se demasiadamente custoso.

2.5. Gestão do Conhecimento

Gestão do conhecimento pode ser entendida como uma coleção de processos que governa a criação, disseminação e utilização do conhecimento para atingir plenamente os objetivos de uma organização.

Entretando, Moresi (2001) define gestão do conhecimento como sendo conjunto de atividades que tem por objetivo desenvolver e controlar todo tipo de conhecimento em uma

organização, visando sobre tudo seus objetivos futuros Nonaka e Takeuchi (1997) dividem o conhecimento de duas formas:

- a) **Conhecimento explícito:** que pode ser articulado na linguagem formal, principalmente em afirmações gramaticais, expressões matemáticas, especificações manuais entre outros.
- b) **Conhecimento tácito:** mais difícil de articular na linguagem formal, trata-se do conhecimento incorporado à experiência individual, envolvendo fatores intangíveis como, por exemplo: crenças pessoais, perspectivas e sistemas de valor.

Assim sendo, pode-se afirmar que a gestão do conhecimento tem como atividade fundamental mapear dinamicamente conhecimentos, experiências e práticas, registrando e assegurando assim que tal conhecimento esteja sempre disponível para os mais diversos processos de tomada de decisão (ALMEIDA, 2006).

2.6. Conhecimento Médico

Constantes avanços científicos têm causado mudanças significativas na medicina. Assim como grande parte das áreas de conhecimento atuais a profissão de médico requer constante processo de atualização, e a geração de conhecimento, sobretudo por parte destes profissionais é hoje um importante elemento de sucesso, mas também um grande desafio profissional (SILVA, 2005). RBC, na visão de Devenport e Prusak (1998), é uma forma de se combinar a computação com o poder da narrativa na codificação do conhecimento. A tecnologia fica a incumbência de extrair o conhecimento de uma ou várias narrativas, denominados casos, sobre um determinado domínio, no caso do trabalho em questão, conhecimento médico. No entanto, diferente dos sistemas especialistas tradicionais que exigem uma série de regras bem estruturadas e sem duplicidades, esta técnica permite que as estruturas de casos possam refletir um pensamento.

O processo de diagnóstico médico é realizado através de uma série de inferências sobre a natureza do problema. Estas inferências consistem em observações realizadas sobre dados retirados das queixas, sinais, sintomas, testes, exames e varias outras manipulações referentes ao paciente, alguns destas observações são de conhecimento clínico, fisiológico, bioquímico, anatômico e patológicos, sobre casos semelhantes e sua biologia subjacente, a experiência médica em realizar diagnósticos semelhantes, senso comum e intuição (KASSIRER e GORRY, 1978).

Os profissionais de saúde utilizam amplamente suas experiências passadas no processo de reunir e interpretar informações. Estas experiências são essenciais, pois reduzem a necessidade de se ter um número considerável de questões desnecessárias, de ordenar testes de diagnósticos supérfluos e de tornar a tarefa de informação manejável e eficiente (KASSIRER e GORRY, 1978).

Este conhecimento pode ser transferido em sua íntegra para bases de dados, conhecidas também como bases de conhecimento, sendo esta a “memória” dos sistemas RBCs, a partir da qual estes sistemas serão capazes de analisar e encontrar as melhores soluções para os problemas aos quais for submetidos, desde que estes tenham alguma relação com o referido domínio.

2.7. Doenças Cardiovasculares

Segundo Gomes (2011), 29,4% de todas as mortes registradas durante um ano no Brasil são causadas por doenças cardiovasculares, isto equivale a afirmar que, anualmente mais de 308 mil pessoas vêm a óbito principalmente por acidente cardiovascular cerebral (AVC) e infartos. Através de pesquisas realizadas, inclusive em bases de dados do Ministério da Saúde (GOMES, 2011) afirma ainda que das vítimas, 60% são homens com idade média de 56 anos. Estes índices fazem com o que o Brasil esteja entre os 10 países com maior número de mortes causadas por doenças cardiovasculares.

A principal característica das doenças cardiovasculares é a presença da aterosclerose, acúmulo de placas de gorduras nas artérias ao longo dos anos que impede a passagem do sangue. A Tabela 1 apresenta a adaptação do resultado de uma pesquisa realizada por Mansur e Favarato (2012), que mostra a taxa de mortalidade por doenças cardiovasculares no Brasil, sendo assim especificadas as legendas: doença isquêmica cardíaca (DIC) e doença cerebrovasculares (DCV); em homens (H) e em mulheres (M).

Tabela 1. Risco de morte, por 100 mil habitantes por doença isquêmica cardíaca (DIC) e doença cerebrovasculares (DCV); em homens (H) e em mulheres (M)

Ano	DIC H	DIC M	AVC H	AVC M
1990	194,87	120,28	206,09	160,17
1991	194,80	114,86	193,48	148,59
1992	178,55	109,50	192,96	148,41
1993	188,62	116,04	207,45	160,73
1994	186,74	116,07	202,70	158,71
1995	183,86	119,13	196,27	155,17
1996	175,95	110,06	173,60	137,07
1997	173,04	109,31	175,23	137,04

1998	176,97	110,42	177,90	136,85
1999	177,63	111,00	175,18	136,83
2000	162,12	96,99	156,68	119,82
2001	162,15	97,16	159,11	120,30
2002	164,00	99,47	158,35	121,38
2003	167,12	99,55	160,96	121,62
2004	171,92	103,35	162,08	124,13
2005	164,95	98,49	155,09	121,35
2006	174,48	103,40	165,42	128,58
2007	151,49	87,77	139,64	106,51
2008	152,67	87,83	138,70	105,40
2009	149,36	84,82	134,09	102,27
variação (%)	-23,35	-29,50	-34,90	-36,15

Fonte: Adaptado de (MANSUR E FAVARATO, 2012).

Pela tabela percebe-se uma sensível redução no risco de morte para ambos os sexos ao longo de 19 anos, o que pode ser explicado devido as ações de prevenção e evolução nos processos de diagnósticos precoce de tais enfermidades. Contudo, como já mencionado anteriormente, estas continuam sendo uma das principais causas de morte por doenças crônicas não transmissíveis, não somente no Brasil, mas em todo o mundo.

2.7.1. Cardiopatia Isquêmica

De acordo com o Brats (2014), “a doença arterial coronariana é resultante do processo de estreitamento ou oclusão das artérias coronarianas por aterosclerose, uma doença que afeta o revestimento endotelial das grandes e médias artérias”. Este estreitamento dificulta a circulação sanguínea e distribuição de oxigênio pelo corpo. Basta tão somente não receber oxigênio para que uma região entre em colapso, como pode ser observado na Figura 4.

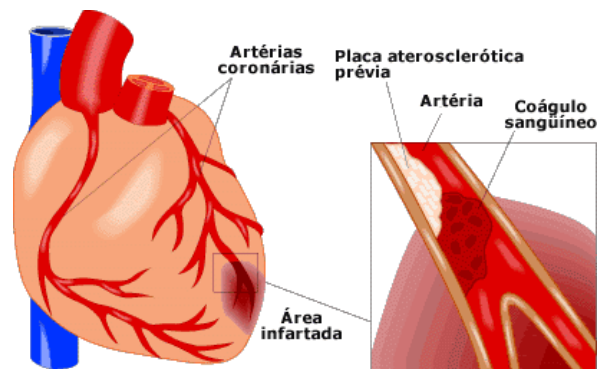


Figura 4. Aterosclerose. Fone: Adaptado de (GOMES, 2011)

A aterosclerose pode ocorrer por origem genética ou comportamental (sendo este o principal motivo de sua ocorrência). Obesidade, sedentarismo, tabagismo, hipertensão, colesterol alto e consumo excessivo de álcool constituem as principais razões para a ocorrência de entupimentos das artérias. Os comportamentos acima citados foram apontados pelo estudo

AFIRMAR (Fatores de Risco Associados com o Infarto do Miocárdio no Brasil), realizado pelo Instituto Dante Pazzanese de Cardiologia, tendo sido entrevistados 3.550 pacientes de 51 cidades brasileiras entre os anos de 1997 e 2000, o maior já realizado no País.

Neste tipo de enfermidade em 90% dos casos a obstrução da coronariana é causada pela formação de placa ateromatosa. Estas placas podem produzir uma grave redução do lume da artéria coronariana numa proporção superior a 50% do diâmetro desta (ou superior a 75% da área transversal do lume), o que pode levar a uma importante redução do fluxo sanguíneo causando assim um desequilíbrio na distribuição da oferta e consumo de oxigênio (isquemia miocárdica). Com o passar do tempo a má distribuição do oxigênio para os tecidos pode causar lesões irreversíveis (GOMES, 2011).

Esta é uma enfermidade que pode evoluir de forma assintomática, insidiosa, durante décadas, vindo a se manifestar clinicamente das mais diversas formas (BRATS, 2014). Segundo o médico cardiologista Ubiratan Rosa Passos, especialista pela Sociedade Brasileira de Cardiologia (SBC), “esta é uma enfermidade cujo diagnóstico pode oscilar de fácil à difícil diagnóstico”, o qual também fez a seguinte afirmação:

“[...] temos que considerar as queixas clínicas, quando sintomáticas, essas doenças sugerem exames bem específicos que podem confirmar ou excluir o diagnóstico da DAC. Contudo, quando assintomáticas, esses exames passam a ser questionados, devido a seu alto custo, por se tratar de um método invasivo, e seus efeitos nocivos ao paciente. Em casos como este é importante pesquisar o histórico do paciente, como fatores de risco, como fumo, diabetes, sedentarismo entre outros [...]” (PASSOS, 2013)³.

Esta afirmação reforça a importância dos métodos e ferramentas que auxiliam os profissionais médicos no diagnóstico precoce e correto desta enfermidade evitando assim o agravamento da mesma. Atualmente um dos exames mais utilizados para diagnóstico da esta enfermidade é o eletrocardiograma, sendo a coronarioscopia aplicada somente para confirmação ou afastamento do diagnóstico por se tratar de um método caro e invasivo (CHORRO, CIVERA e MERINO, 2007).

2.7.2. Eletrocardiograma

³ Ubiratan Rosa Passos – Médico Cardiologista especialista pela Sociedade Brasileira de Cardiologia – SCMC – Santa Casa de Misericórdia de Castelo – Governo do Estado do Espírito Santo.

O eletrocardiograma (ECG) é a representação gráfica obtida através dos potenciais elétricos originados do coração quando eletrodos são instalados na superfície do organismo humano. Os eletrodos são ligados aos membros e à parede torácica sendo seus sinais amplificados e registrados através do eletrocardiógrafo. Deve-se no entanto esclarecer que o ECG registra apenas diferenças de potencial instantâneas entre os eletrodos (CHAVES e MOREIRA, 2001).

Os métodos conhecidos de investigação diagnóstica da Doença Isquêmica Cardíaca (DIC) são em sua grande maioria dispendiosos, invasivos e de difícil aplicação, sendo assim, além da anamnese e exame clínico o ECG pode ser largamente empregado devido a suas características. Trata-se de um exame relativamente simples e de baixo custo que proporciona dados sobre o aspecto do miocárdio, permitindo desta forma ao profissional especialista diagnosticar manifestações da DIC, tais como infarto do miocárdio, isquemia e a hipertrofia cardíaca e até mesmo detectar o risco de futuros eventos cardíacos ou de mortalidade por doenças relacionadas. Estudos demonstram que pequenas alterações nos sinais de um exame de ECG podem ser entendidas como “preditoras” de manifestações clínicas de enfermidades cardíacas (ROSE, et al., 1982) (HAMPTON, 1984) (KANNEL, et al. 1987) (KNUSTSEN, et al. 1988) (MENOTTI e SECCARECCIA, 1997)

Chaves e Moreira (2001) afirmam ainda que, apesar de suas limitações o ECG é o exame auxiliar mais utilizado no diagnóstico de doenças cardíacas. Isto se dá pelo fato de este ser um exame não invasivo, barato e muito versátil.

Entretanto a interpretação de um eletrocardiograma é totalmente subjetiva podendo variar de um especialista para o outro, sendo que, a omissão é uma das principais causas de erros em sua interpretação, quando estes ocorrem. Logo a adoção de uma abordagem sistemática quando se observam os traços de um ECG torna-se necessária. Para análise de um sinal de ECG os seguintes pontos devem ser cuidadosamente observados.

1. padronização (calibração) e características técnicas (incluindo a colocação dos eléctrodos e artefatos);
2. frequência cardíaca;
3. ritmo;
4. intervalo PR;
5. intervalo QRS;
6. onda P;
7. voltagem do complexo QRS;
8. eixo eléctrico médio do complexo QRS;

9. progressão da onda R nas derivações pré-cordiais;
10. onda T;
11. onda U.

Estudar e entender estes parâmetros torna-se importante para configuração do sistema aqui proposto.

2.7.3. Raciocínio Baseado em Casos

Em seu cotidiano o ser humano se depara com inúmeras situações, algumas desconhecidas e outras não, mas em geral, a maioria necessita de um processo decisório para sua resolução. Na medida em que essas experiências são vividas aumenta-se também a quantidade de decisões tomadas corretamente, assim como a quantidade de conhecimentos acumulado. Todos os casos (decisões certas ou erradas) são influenciados pela quantidade, tipo e relevância das experiências ou situações vividas ao longo do tempo.

Raciocínio Baseado em Casos (RBC) é uma poderosa técnica utilizada para solução automatizada de problemas. Esta técnica pode ser aplicada de forma ampla e direta em um amplo aspecto de tarefas, todas relacionadas à Inteligência Artificial (IA). Basicamente RBC propõe-se a resolver um problema lembrando uma situação similar anterior e, então, reutilizando informações e conhecimento de referida situação.

O primeiro sistema RBC foi desenvolvido em 1983 por Janet Kolodner, o referido sistema ficou conhecido como (Cyrus) e era baseado em um modelo de memória dinâmica servindo posteriormente de base para outros sistemas RBC. Modelos como este têm forte influência sobre os sistemas RBC uma vez que o conhecimento sobre o domínio pode melhorar a qualidade do raciocínio, encurtar o caminho de busca da solução ou mesmo preencher lacunas do espaço do problema que os casos naturalmente não cobririam.

Pode-se então dizer que a Técnica de Raciocínio Baseado em Casos (RBC) baseia-se na busca de soluções para problemas atuais em experiências passadas, utilizando para isto um método inspirado em uma das principais características humanas, a lembrança.

Rezende (2003) afirma que RBC são sistemas que utilizam conhecimentos repassados explicitamente para resolver problemas através da manipulação do conhecimento e informações de maneira inteligente, auxiliando desta forma na resolução de problemas que requerem uma gama considerável de conhecimento humano

Estes sistemas estão habilitados para funcionarem como modelos cognitivos permitindo assim que os mesmos entendam aspectos do comportamento e pensamento dos seres humanos.

Outra característica dos sistemas RBC é que os mesmos possuem um conhecimento específico, focado em exemplos concretos de casos, totalmente orientados a metas e soluções, diferente do que ocorre em outras técnicas de IA.

Um sistema de RBC deve ser capaz de questionar o usuário usando linguagem de fácil entendimento, desenvolver linha de raciocínio a partir dessas informações e do conhecimento nele inserido, explicar seu raciocínio devendo ser capaz de interpretar o processo e apresentá-lo de forma compreensível. (REZENDE, 2003). Sobre sistemas RBC, o padrão mais aceito para seus processos é o Ciclo de RBC, como é apresentado na Figura 5.

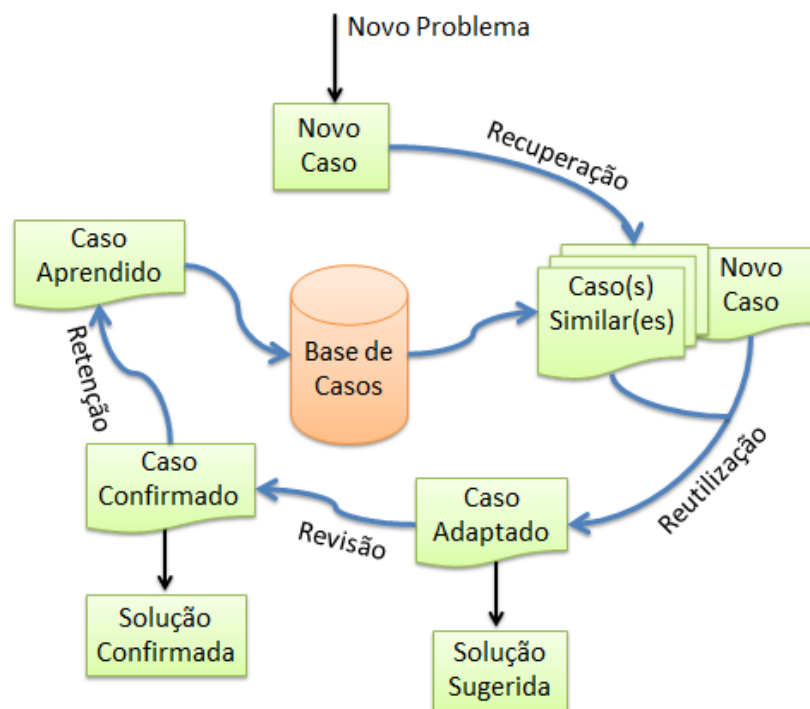


Figura 5. Ciclo de Vida de um Sistema RBC. Fonte: Adaptado de (REZENDE, 2003).

Rezende (2003) cita o ciclo do RBC em quatro etapas de execução, comumente conhecidas como 4R's (Recuperação, Reuso, Revisão e Retenção), explicadas abaixo:

a) **Recuperação:** a partir da apresentação ao sistema de um novo problema é feita a recuperação na base de casos daquele mais parecido com o problema em questão. Isto é feito a partir da identificação das características mais significantes em comum entre os casos;

b) **Reuso:** a partir do caso recuperado é feita a reutilização da solução associada àquele caso. Geralmente a solução do caso recuperado é transferida ao novo problema diretamente como sua solução;

c) **Revisão:** é feita quando a solução não pode ser aplicada diretamente ao novo problema. O sistema avalia as diferenças entre os problemas (o novo e o recuperado), quais as

partes do caso recuperado são semelhantes ao novo caso e podem ser transferidas adaptando assim a solução do caso recuperado da base à solução do novo caso;

d) **Retenção:** é o processo de armazenar o novo caso e sua respectiva solução para futuras recuperações. O sistema irá decidir qual informação armazenar e de que forma.

Resumidamente pode-se então explicar assim o processo: a partir de um novo problema de entrada, é recuperado um conjunto de casos similares candidatos a resolver o problema atual. A partir destes, o caso candidato a solução é adaptada ao problema atual, validado (etapa que pode ocorrer com participação do usuário ou ser automática) e por fim esse novo caso é aprendido e armazenado na base de casos para uma posterior utilização.

2.7.3.1. Raciocínio Baseado em Casos no Diagnóstico Médico

A área médica tem despertado especial interesse nos pesquisadores e desenvolvedores de sistemas de suporte à decisão devido à suas características e importância para o bem social, dentre as várias técnicas aplicadas, os sistemas RBC vem sido largamente aplicados devido as suas características. Sistemas de Suporte a Decisão (SSD), aplicados à medicina, é todo software que tem por objetivo auxiliar o profissional médico na solução de um determinado problema. Vale lembrar que todo SAD tem como característica a busca de informações em meios externos, combinando essas informações a dados já conhecidos, esses sistemas oferecem grande auxílio à medicina. Conforme a Organização Mundial da Saúde (OMS, 2014), a quantidade de informações referentes à área da saúde dobra a cada três anos, surgindo novos métodos de diagnósticos e terapêutica, novos princípios químicos, inovações da área da biologia molecular e da genética, entre outros avanços medicinais, o que torna crescente o interesse pelo estudo destes sistemas.

A elaboração de um diagnóstico com grandes chances de sucesso para eventos agudos da cardiopatia isquêmica tem como principal base os seguintes elementos (RABELO, 1993): (i) história clínico do paciente; (ii) exame físico; (iii) eletrocardiograma. Desta forma, com as ferramentas corretas, havendo disponibilidade de recursos e um profissional com algum conhecimento sobre a enfermidade. Neste cenário, mesmo um posto de saúde pequeno, em regiões afastadas pode realizar tal procedimento com grandes chances de sucesso, reduzindo assim as possibilidades de óbito de um determinado paciente. Entretanto, é comum que, mesmo estes postos disponibilizando tais recursos, não disponham de profissionais capazes de realizar o pré-diagnóstico corretamente devido ao grau de subjetividade do processo, além das inúmeras indicações diagnósticas para um grupo semelhante de sintomas.

Neste contexto, um sistema inteligente, capaz de auxiliar os profissionais médicos na realização de diagnósticos específicos (cardiopatia isquêmica, por exemplo) como é o caso de um sistema baseado nas técnicas de RBC, cujo “conhecimento” é baseado em experiências pode ser de grande validade, pois, uma vez que o ambiente estiver parametrizado e povoado com um grupo específico de informações iniciais (dados clínicos), o sistema será capaz de sugerir um diagnóstico provável através das várias experiências que este mantém acumulada (ALMEIDA, 2006) (SILVA, 2008).

2.7.3.2. Cálculo de Similaridade

Um dos processos mais importantes de um sistema RBC, de acordo com Almeida (2006), é a recuperação de casos que sejam úteis na resolução de um dado problema. Esta visão dá-se pelo fato de ser este processo o responsável por encontrar casos na base de conhecimento que forneçam a melhor solução para a situação a ser resolvida, além de ser pré-requisito para os demais processos. Almeida (2006) divide este processo em quatro tarefas: identificação de características, busca, identificação de compatibilidade e seleção.

A recuperação de um caso esta fortemente ligada a forma como estes casos estão representados na base de dados, dos índices utilizados (quando aplicados) e do algoritmo de busca e cálculo de similaridade. O desempenho na busca também é um fator muito importante para alguns domínios, principalmente quando se tem uma base com dezenas de milhares de casos para serem analisados, podendo ser entendido com o um problema de busca, mas com uma importante diferença: não se pode esperar encontrar nesta base de conhecimento casos exatamente iguais aos procurados, mesmo por que os algoritmos de busca em bancos de dados procuram valores exatos em relação à consulta. Sistemas RBC baseiam-se em semelhanças e não necessariamente em exatidão. Assim sendo, espera-se que algoritmos específicos sejam utilizados e que seja levada em conta a representação dos casos, tipos de atributos que compõem os índices, objetivo do raciocinados e determinada operação de busca, a precisão da solução e a necessidade de buscas sequenciais ou paralelas (SILVA, 2008).

Os métodos de busca podem interferir nas medidas de similaridade de forma que, a similaridade possa ser calculada automaticamente ou ser uma tarefa posterior à busca, desta forma é possível refinar os casos fornecidos pela recuperação. Existem basicamente duas abordagens de busca, baseada no cálculo da distância entre os casos ou na sua representação (LIAO, ZHANG e MOUNT, 1998). A primeira abordagem leva considera os casos mais similares, estes são escolhidos com base na distância entre eles, de acordo com a avaliação de

uma medida de similaridade. A segunda abordagem relaciona a busca à representação, ou indexação dos casos de acordo com sua organização procurando assim o mais similar entre eles.

Bergmann (2002) apresenta vários métodos de recuperação de casos, tais como: **força bruta, indução, indução guiada pelo conhecimento, SQL ou recuperação de padrões e recuperação validada** (BERGMANN, 2002), (SIMOUDIS, 1992).

Bergmann (2002) sugere ainda um método conhecido como indexação, contemplando a indução, as redes de recuperação de casos e a recuperação separação e diminuição, sendo o último destinado para situações complexas de similaridade. Assim como em outros métodos relacionados a IA, no RBC existe a ideia de que problemas semelhantes possuem soluções semelhantes, sendo que muitas das medidas de similaridade são construídas segundo esta condição. Muitas das vezes as similaridades são encontradas nas características mais profundas, extraídas das características superficiais, possibilitando assim o encontro de uma solução satisfatória para o problema (SIMOUDIS, 1992) (BOMER, 1994).

A escolha da medida de similaridade esta diretamente ligada ao domínio e ao objetivo do sistema. Quando o domínio e estruturas são complexas, pode ser necessária a intervenção humana nos últimos estágios para selecionar os melhores casos. As medidas de similaridade mais utilizadas são àquelas ligadas à distância e as demais aprendizado de máquina (LIAO, ZHANG e MOUNT, 1998) (PAL e SIMON, 2004), são apresentadas na Tabela 2 algumas medidas de distância que podem ser utilizadas para medidas de similaridade.

Tabela 2. Medidas de distância mais comuns utilizadas para cálculos de similaridade.

Equação (a)	Euclidiana	$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$	Muito utilizada para cálculos de valores reais. É a medida de distância mais usada.
Equação (b)	Euclidiana Ponderada	$d(x, y) = \sqrt{\sum_{i=1}^n W_i^2 (x_i - y_i)^2}$	Cada atributo possui um peso, indicando a sua relevância na composição do caso.
Equação (c)	Manhattan	$d(x, y) = \sum_i^n x_i - y_i $	Também conhecida como <i>City-block</i> . Um exemplo simples de uso é a distância entre peças de xadrez, onde o caminho a ser percorrido não é uma linha reta.
Equação (d)	Hamming	$d = \sum_{i=1}^n \delta_i, \quad \text{onde}$	Hamming foi concebida para tratar valores binários e é muito usada

		$\delta_i = \begin{cases} 1 & \text{se } x_i \neq y_i \\ 0 & \text{em outros casos} \end{cases}$	para calcular similaridade de atributos simbólicos.
Equação (e)	Minkowsky Ponderada	$d(x, y) = \left(\sum_{i=1}^n W_i x_i - y_i \right)^{\frac{1}{k}}$	É uma generalização da distância euclidiana e Manhattan, quando $k=2$ e $k=1$, respectivamente.
Equação (f)	Tchebyshev	$d(x, y) = \max\{ x_1 - y_1 , \dots, x_n - y_n \}$	É um caso especial onde o valor de k na distância de Minkowski.

Fonte: Adaptado de (PAL e SIMON, 2004)

As medidas das distâncias podem não ser usadas diretamente para medir similaridade, mas, sim, como um estágio anterior ao cálculo da similaridade propriamente dita.

2.7.4. Computação Evolutiva

Trata-se de um ramo da Inteligência Artificial que propõe um paradigma para resolução de problemas baseado na Seleção Natural (Darwin, 1859). Inspirado no evolucionismo Darwiniano, como os Algoritmos Genéticos (AG's) é uma das áreas da IA com maior inspiração biológica (ALVES, 2009).

Métodos evolucionários são baseados em populações. Nestes casos, ao invés de trabalhar com uma única solução, caminhando com esta pelo espaço de soluções buscando aquela que será ótima, é utilizada uma população de soluções, que evoluem baseadas em algum princípio definido na busca pela solução ótima.

Os métodos evolucionários baseiam-se em heurísticas as quais buscam no espaço de soluções àquela que mais se adequa, seguindo os mais variados processos biológicos. Os algoritmos genéticos, que serão abordados na seção 2.4.4.1, são inspirados na *Teoria da Evolução das Espécies*, de Charles Darwin (Darwin, 1859), para a evolução da população de soluções.

2.7.4.1. Algoritmos Genéticos

Com base na *Teoria Evolucionária da Espécies*, de Charles Darwin (Darwin, 1859), foi proposto em 1975, por Holland os chamados *Algoritmos Genéticos* (HOLLAND, 1975). Estes algoritmos propõem a utilização de uma população de indivíduos, ficando a pressão ambiental responsável por estimular a seleção natural (*survival of the fittest*) ou seja, sobrevivência do

mais apto, desta forma, o desempenho geral da população cresce com o passar das gerações. Este mecanismo é facilmente entendido como um processo de otimização, pois, dada uma função objetivo a ser maximizada ou minimizada, uma população de possíveis soluções é criada de forma aleatória. Ainda utilizando esta função objetivo como medida de grau de adequação das soluções testadas, cria-se uma tendência em se escolher os melhores indivíduos para compor a nova geração de soluções, através de processos de cruzamento e mutação principalmente. Estes procedimentos podem ser repetidos até que as condições estabelecidas para a convergência sejam atingidas, esperando assim encontrar um ótimo global para o problema, entre o conjunto de soluções encontradas.

Algoritmos genéticos podem ser expressos de uma forma básica através do pseudocódigo representado abaixo.

```

gerar população aleatoriamente
avaliar os indivíduos
ranquear os indivíduos
repetir
    selecionar melhores indivíduos
    cruzar indivíduos pais escolhidos
    mutar indivíduos selecionados
    avaliar os indivíduos
até que condição de parada seja satisfeita

```

Desta forma, o processo evolucionário faz com que as soluções tendam a iniciar o processo disperso, caminhando na direção dos ótimos com o passar das gerações.

Toda heurística, segundo as quais os métodos evolucionários utilizados nos algoritmos genéticos são baseados, foram inspiradas em diversos processos biológicos, dentre os quais pode-se destacar: *Ant Colony Optimization* (DORIGO, MANIEZZO e COLOMI, 1996) e *Particle Swarm Optimization* (KENNEDY e EBERHART, 1995), baseados nos princípios de *Swarm Intelligence*. Este métodos assim como o *Algoritmo Clonal* (CASTRO, 2000), é uma versão modificada do AG com supressão do operador de cruzamento e os conhecidos *Sistemas Imunológicos Artificiais*, que modelam o sistema imunológico humano para otimização (DASGUPTA, 1998).

2.7.4.2. Representação

A primeira atividade a se realizar quando da definição de um AG é a forma de representação dos indivíduos. Determina-se neste ponto qual a relação entre o universo dos problemas onde ocorrerá a evolução das soluções. Aos elementos que compõem o conjunto de soluções dá-se o nome de *fenótipos*, ao passo que, as codificações que podem levar estes à

solução dos problemas são denominados *cromossomos* (comumente também chamados de *genótipos*) (OLIVEIRA & XEXÉO, 2011). Desta forma a codificação consiste em mapeamento entre os fenótipos e os cromossomos que representarão as possíveis soluções. Este conjunto de soluções é frequentemente representado por cadeias de valores binários assim como pode ser observado pela Figura 6.

	0	1	2	3	4	5	6	7
Binária	0000	0001	0100	0111	1000	1001	1110	1111

Figura 6. Exemplo de codificação binária de um indivíduo. Fonte: Adaptado de (OLIVEIRA E XEXÉO, 2011)

Logo, cada valor gerado pelo algoritmos é convertido em um valor binário, permitindo assim maior facilidade na manipulação dos dados analisados pelo AG.

2.7.4.3. Avaliação

Cronologicamente o próximo passo a ser executado é a definição da função de desempenho, também conhecida como *fitness*. O papel desta função é guiar o processo evolutivo do algoritmo. Entretanto é comum que durante o processo de mapeamento direto entre os espaços de soluções e de busca ocorram convergências prematuras, principalmente nos casos onde existe uma grande diferença entre os valores de desempenho e os indivíduos da população (PACHECO, 1999).

2.7.4.4. Seleção

Depois de realizada a criação e avaliação da população inicial e serem atribuídos seus desempenhos, chega-se o momento de selecionar os indivíduos que farão parte da próxima geração de possíveis soluções. A escolha destes indivíduos deve seguir o princípio da seleção probabilística: logo, os indivíduos de maior desempenho terão maior probabilidade de serem selecionados, entretanto, permite ainda que indivíduos de desempenho inferior também sejam selecionados, garantindo desta forma a diversidade genética das novas populações, um fator primordial para o sucesso de qualquer processo evolutivo.

Método da Roleta: Neste método cada um dos indivíduos tem a probabilidade de ser selecionado para a próxima geração de acordo com seu desempenho relativo ao da população, como demonstrado na equação abaixo.

Equação (g) – Probabilidade Seleção Roleta

$$P_{sel} = \frac{f_i}{\sum f}$$

Logo o processo de seleção pode ser apresentado como uma roleta, onde as probabilidades de cada indivíduo ser selecionado podem ser observadas na área correspondente a cada um, como demonstrado na Figura 7.

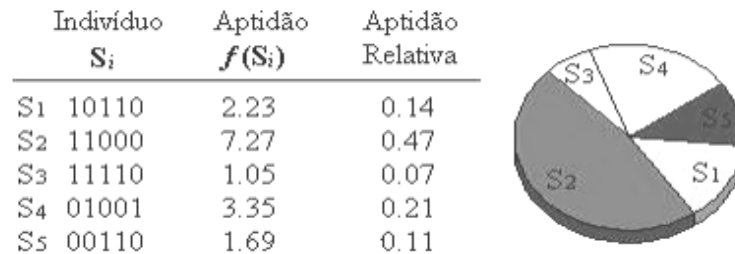


Figura 7. Método da Roleta. Fonte: Adaptada de (OLIVEIRA E XEXÉO, 2011).

Cada um dos indivíduos da população é representado na roleta de uma forma proporcional à sua aptidão, assim sendo, indivíduos com alta aptidão passam a ocupar uma parcela maior na roleta, ao passo que, indivíduos de menor aptidão ocupam uma fatia menor. Tal roleta será “girada” n_{pop} vezes, sendo n_{pop} o número de indivíduos da população, sendo os indivíduos escolhidos de acordo com a “posição onde parar a roleta”. Desta forma, os indivíduos com maior *fitness* serão aqueles que têm a maior probabilidade de serem sorteados, sendo assim escolhidos para formar a nova geração (OLIVEIRA & XEXÉO, 2011).

```

definir l
repetir
  selecionar individuo_1
  selecionar individuo_2
  r = valor_sorteado entre 0 e 1
  se r < l então
    selecionar o melhor entre os 2 indivíduos
  senão
    selecionar o pior entre os 2 indivíduos
  fimse
enquanto npop não terminar

```

Elitismo: Busca garantir que melhor indivíduo seja preservado na próxima geração, impedindo desta forma que, uma solução ótima seja perdidas quando, quando aplicadas algumas das outras operações as quais são pertinentes a algoritmos genéticos.

2.7.4.5. Cruzamento

Segundo Eiben e Smith (2003), o cruzamento é o principal dos operadores dos algoritmos genéticos. Através deles, dois indivíduos podem ter seus genótipos recombinados,

formando assim um ou dois novos indivíduos. Cruzamentos envolvendo mais de dois indivíduos são perfeitamente possíveis, ainda que este tipo de operação não tenha qualquer embasamento biológico. No entanto, cruzamentos como estes raramente são utilizados, ainda que alguns escritores afirmam que apresentem grandes efeitos positivos no processo evolutivo (EIBEN e SMITH, 2003).

A lógica do cruzamento é descrita da seguinte forma: dados dois ou mais indivíduos que possuam as características desejadas, o cruzamento entre estes indivíduos combina essas características. Trata-se de um procedimento utilizado a milênios pelo ser humano, quando o mesmo cruza animais ou plantas de características superiores, tentando desta forma obter variações cada vez melhores. Os métodos de cruzamento mais utilizados são **cruzamento com um ponto de corte, cruzamento com n pontos de corte, cruzamento uniforme, cruzamento por variável e o cruzamento por variável entre vários indivíduos.**

2.7.4.6. Mutação

Trata-se de um operador que atua sobre somente um genótipo de cada vez, gerando desta forma um novo indivíduo levemente modificado. De acordo com Eiben e Smith (2003), o principal objetivo da mutação nos algoritmos genéticos é manter a diversidade genética da população, assim como o papel do cruzamento é levar o algoritmo a um novo ponto no espaço de buscas. A mutação procura garantir que estes espaços esteja conectados. As propriedades pertinentes a mutação, garantem que dado um tempo suficiente, o AG atingirá um ótimo global do problema.. A Figura 8. ilustra o processo.

Antes da Mutação: 1 1 1 0 0
 Depois da Mutação: 1 1 0 0 0

Figura 8. Mutação. Fonte: Adaptada de (SILVA, 2008).

Logo, para se realizar este procedimento, basta tão somente decidir, qual o valor da taxa de probabilidade de mutação p_m , será aplicada para que cada gene sofra ou não a mutação. De uma forma geral, os valores mais indicados na literatura variam entre 0.01 e 0.03 (SILVA, 2008)

2.7.5. Redes Neurais Artificiais

Das diversas técnicas utilizadas para proporcionar às máquinas a capacidade de aprendizado, as que mais se destacam são às baseadas em sistemas naturais, seja eles biológicos ou físicos, dentro os quais pode-se destacar a computação evolutiva, apresentada anteriormente, Sistemas Nebulosos, Swarm Intelilenge e as Redes Neurais Artificiais, que será discutida nesta seção.

RNA's podem ser aplicadas nas mais diversas áreas da ciência, assim também como da vida humana, dentre as quais, pode-se levantar como mais comuns as seguintes:

- a) Análise de imagens;
- b) Classificação de dados;
- c) Diagnóstico médico;
- d) Identificação de fraude de cartão de crédito;
- e) Processamento de sinais e eliminação de ruídos;
- f) Reconhecimento e síntese de fala;
- g) Reconhecimento óptico de caracteres (OCR); e
- h) Robótica.

RNA's são basicamente modelos matemáticos semelhantes a estruturas neurais biológicas com capacidade computacional adquirida por meio de aprendizagem e generalização. Estas redes buscam assemelhar-se com o sistema nervoso dos seres vivos, simulando sua capacidade de processamento de informações. Tem-se então uma metáfora da maneira como o cérebro humano processa as informações utilizadas em computação.

Os neurônios biológicos são como circuito de entradas binárias combinadas por uma soma ponderada (de pesos) produzindo desta forma uma entrada efetiva, este seria o funcionamento de um neurônio biológico. Constituindo-se tais neurônios de:

- a) Entradas (dendritos) - onde são aplicados os sinais;
- b) Pesos (sinapses) - onde fica retido o conhecimento;
- c) Função soma - somatório da relação dos sinais de entrada e dos pesos sinápticos;
- d) Função de ativação - função que dependendo do valor do somatório irá ou não ativar a saída
- e) dependendo do *threshold* (sigmoidal, escada, Gaussian);
- f) Saída (axônio) - saída de interface.

Sendo assim, pode-se generalizar que um neurônio tem a seguinte funcionalidade: 1) Sinais são apresentados a entrada; 2) É realizada a multiplicação dos sinais com os pesos sinápticos; 3) É realizado a soma ponderado dos sinais; 4) Se o nível for maior que um limite (*threshold*), então a saída será ativada (SILVA, 2008), vide Figura 9.

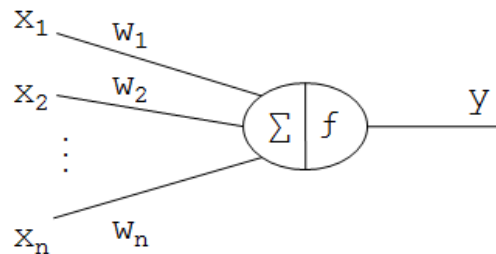


Figura 9. Modelo de McCulloch e Pitts. Fonte: (SILVA, 2008).

Uma RNA pode ser considerada um grafo orientado, onde neurônios artificiais são nodos ligados através de conexões diretas com pesos para neurônios de saída e de entrada. Uma RNA é dividida em três camadas, entrada, oculta e saída. RNA's podem ser classificadas quanto a sua estrutura, seu comportamento e algoritmo que utiliza, que é em muitos casos determinado pelo seu comportamento, sendo neste trabalho aplicada a *Multi-Layer Perceptron* com algoritmos *backpropagation* de aprendizado supervisionado.

O modelo proposto por McCulloch e Pitts apresenta grande semelhança com um modelo biológico. As entradas x_1 a x_n representam os dendritos, ou seja, prolongamentos dos neurônios naturais utilizados para receberem estímulos nervosos. Os pesos w_1 a w_n representam as conexões sinápticas entre os neurônios, pontos responsáveis por enfatizar os estímulos vindos de alguns neurônios. A saída y representa no modelo o axônio, prolongamento do neurônio natural, responsável pela condução do estímulo nervoso a outros neurônios. Por fim, a função de decisão $f(\sim x)$ tem a mesma função de decisão de disparar (ou não) os neurônios naturais.

Contudo, em 1969, a publicação do livro *Perceptrons*, por Marvin Minsk e Seymour Papert (1969), ficou provado que por si o *Perceptron* simples era capaz de solucionar somente problemas linearmente separáveis, incapaz por exemplo de solucionar o problemas OU-EXCLUSIVO. Esta constatação reduziu consideravelmente o interesse nas Redes Neurais Artificiais – RNA's. Sendo então seu desenvolvimento retornado, entre outros motivos, devido ao desenvolvimento do método da retropropagação (*Backpropagation*) de erros e a introdução das Redes Perceptron Multicamadas (Multi-Layer Perceptron – MLP).

2.7.5.1. MLP – Multi-Layer Perceptron – Perceptron Multicamadas

Sendo um dos princípios das RNA's a simulação da essência dos sistemas neurais biológicos, é necessário para que isto ocorra a criação de neurônios artificiais definidos como:

I. Ele recebe um número de entradas (que a partir de dados originais, ou a partir da saída de outros neurônios na rede neural). Cada entrada vem através de uma ligação que tem uma força (ou peso); estes pesos correspondem a eficácia sináptica num neurônio biológico. Cada neurônio tem também um valor de limiar individual. A soma ponderada das entradas é formada e subtraída do limiar, para compor a ativação do neurônio (também conhecido como o potencial pós-sináptico, ou PSP, do neurônio);

II. O sinal de ativação é transmitido através de uma função de ativação (também conhecido como uma função de transferência), para produzir a saída do neurônio.

Assim sendo, o algoritmo propõe o uso da função degrau (ou seja, a saída fornecida pelo neurônio específico é 0, se o estímulo de entrada for menor do que zero, e 1 se o estímulo de entrada for maior ou igual a 0), logo o neurônio em questão age como um próprio neurônio biológico, como descrito anteriormente. O esquema proposto possibilita o surgimento de pesos negativos, o que pode gerar sinapses inibitórias ao invés de excitatórias sobre o neurônio, assim como também acontece no cérebro biológico. Uma representação esquemática de rede MLP pode ser observada na Figura 10.

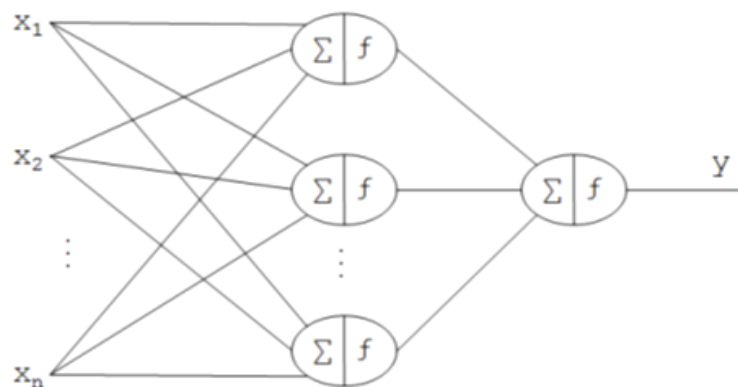


Figura 10. Modelo MLP de McCulloch e Pitts. Fonte: (SILVA, 2008)

As camadas escondidas das Redes MLP são responsáveis por mapear o espaço de entrada no chamado *Espaço de Características*, onde problemas não lineares podem ser transformados em lineares e, então, facilmente resolvidos. O processo de aprendizado de uma rede MLP consiste em se encontrar os pesos correspondentes às ligações entre os neurônios de forma a se obter o comportamento desejado.

2.7.6. Computação Gráfica

Dentro da ciência da computação, a computação gráfica destaca-se como área responsável pela geração e manipulação de imagens em geral, através de sistemas computadorizados (Viera e Filho, 1999). Geralmente relacionada a jogos digitais, além de possuírem um conjunto de técnicas aplicáveis a qualquer ramo da informática, tal como a geração de interfaces gráficas com usuário, artes digitais, softwares de uso geral, etc., em muitos trabalhos, como neste, esta técnica será aplicada na preparação de imagens correspondentes a derivações de sinais impressos de eletrocardiogramas, permitindo que as mesmas sejam analisadas por um sistema de inteligência artificial.

2.7.6.1. Imagem Digital

O correto entendimento de o que realmente são imagens digitais, depende fortemente do bom entendimento do que são funções contínuas e discretas, e suas diferenças. Funções contínuas não podem ser perfeitamente reproduzidas em meio computacionais, devido a suas limitações físicas e tecnológicas, que determinam os sistemas computacionais com recursos limitados. Tais funções equivalem às funções reais, de maneira tal que, uma função define uma regra onde para cada variável independente x , é possível encontrar o valor da variável dependente y ou $f(x)$ (GONZALEZ & WOODS, 1992), segundo demonstrado na equação abaixo.

Equação (h) – Função Contínua

$$f: R \rightarrow R$$

$$x \rightarrow f(x)$$

Para possibilitar a representação de uma função contínua em ambientes computacionais, é preciso uma função discreta equivalente, ou seja, uma função cujo domínio e contradomínio sejam números naturais, isto é, um conjunto finito de valores, ao contrário das funções contínuas que podem possuir infinitos valores em seus domínios e contradomínios (GONZALEZ & WOODS, 1992).

Imagens digitais podem ser representadas pela atribuição discreta de uma característica de tons/cor, brilho e transparência para cada ponto da função discreta, representada, ou através

da atribuição de um nível de cinza, por exemplo, para imagens monocromáticas, como representado na equação que segue.

Equação (i) – Conversão de tons de cinza

$$f(x, y) = \text{cor no ponto } (x, y); \text{ para } x \in [0, X] \text{ e } y \in [0, Y]$$

, onde x e y são números reais no intervalo 0 a X e 0 a Y

Assim sendo, uma imagem digital monocromática será o resultado da atribuição dos níveis de cinza para cada uma das coordenadas (x, y) de uma função discreta. A intensidade da imagem de saída é definida como nível de cinza (L), para cada um dos pontos (x, y) , estando esta intensidade dentro dos limites de intervalo, como descrito na equação representada a seguir.

Equação (j) – Definição dos tons de cinza.

$$L_{min} \leq L \leq L_{max}, \text{ sendo } L_{min} \text{ e } L_{max} \text{ valores finitos e positivos.}$$

Criar imagens digitais a partir de uma imagem real é um processo composto de duas etapas, que juntas permitirão a representação digital da informação que foi capturada pelo dispositivo. Estas etapas são a aquisição e a digitalização (GONZALEZ & WOODS, 1992).

Aquisição: primeiro momento da criação de uma imagem digital a partir de uma imagem real e **digitalização**, resultado do processo anterior deve ser transformado em um formato manipulável para o sistema computacional.

Dadas as características de algumas imagens, em alguns casos é necessário um processo de limiarização, ou seja, detecção e separação das informações contidas em uma determinada imagem digital, permitindo posteriormente a criação de uma imagem binária constituída de apenas 0's e 1's. Contudo para que isto se torne possível é necessário definir o nível de cinza N , de uma forma que, qualquer valor inferior ou igual a este seja considerado o objeto e todo valor superior a este seja considerado o fundo, fazendo desta forma que ao final do processo tenha-se somente dois tipos de informação. Em notações matemáticas a limiarização é uma operação onde uma entrada $f(x,y)$ de nível cinza N gera uma saída $g(x,y)$ denominada imagem limiarizada, representado na equação a seguir. (GONZALEZ & WOODS, 1992).

Equação (k) - Limiarização

$$g(x, y) = 1 \text{ se } f(x, y) \geq T$$

$$g(x, y) = 0 \text{ se } f(x, y) < T$$

Assim sendo pode-se definir os pixels de valor igual a 1 como sendo objeto, e os pixels com valor igual a 0 como sendo o fundo, T é então um nível de cinza predefinido. Quando se trabalha com imagens que possuem somente duas informações esta técnica mostra-se bastante útil, entretanto, se a imagem possuir uma série de informações, a melhor representação da mesma não seria uma imagem binária (GONZALEZ & WOODS, 1992), assim sendo, dadas as

características das imagens analisadas neste trabalho, **Afinamento**, **Thinning** ou **Esqueletização** foram aplicadas com o propósito de preparar as imagens para utilização na ferramenta. Estes conceitos serão apresentados de forma sucinta para que não haja fuga do escopo deste trabalho.

3. METODOLOGIA

3.1. Classificação da Pesquisa

Analisando do ponto de vista do problema, esta pode ser classificada como uma pesquisa quantitativa, isto pelo fato de os resultados desta permitirem a classificação e análise científica, uma vez que é realizada sobre dados referentes ao diagnóstico da cardiopatia isquêmica. Quanto aos meios, esta pesquisa classifica-se como sendo de laboratório, pois a mesma trata simulações realizadas em computador.

Sob a ótica da natureza e dos fins, esta pesquisa classifica-se como metodológica aplicada (VERGARA, 2000) (MORESI, 2003). Isto pelo fato de o estudo está associado a caminhos, formas, maneiras e procedimentos para se atingir um determinado fim. Também por ser fundamentalmente motivada pela necessidade de se resolver problemas concretos, tendo, portanto, finalidade prática, ao contrário da pesquisa pura, movida basicamente pela curiosidade intelectual do pesquisador e sobretudo situado no nível de especulação.

3.2. Suposições e Hipóteses

Esta pesquisa pretende testar a seguinte hipótese:

A melhor solução determinada pelo AG, cujos pesos para a função de similaridade (dada pela distância euclidiana ponderada) global do RBC, cuja acuidade auxiliará o especialista

médico no diagnóstico da cardiopatia isquêmica. Testa também a hipótese de que através do uso da IC, é possível determinar maior taxa de acertos os reais diagnósticos para esta enfermidade através da análise de sinais de eletrocardiogramas, que serão analisados pela própria ferramenta.

3.3. Fluxo de Trabalho

Segundo Malta e Merhy (2004), o fluxograma permite instrumentalizar o trabalhador no seu próprio processo de trabalho. Assim sendo, para que o projeto fosse desenvolvido de forma sistemática e organizada foi criado e respeitado um fluxo simplificado de trabalho, como pode ser observado na Figura 11.

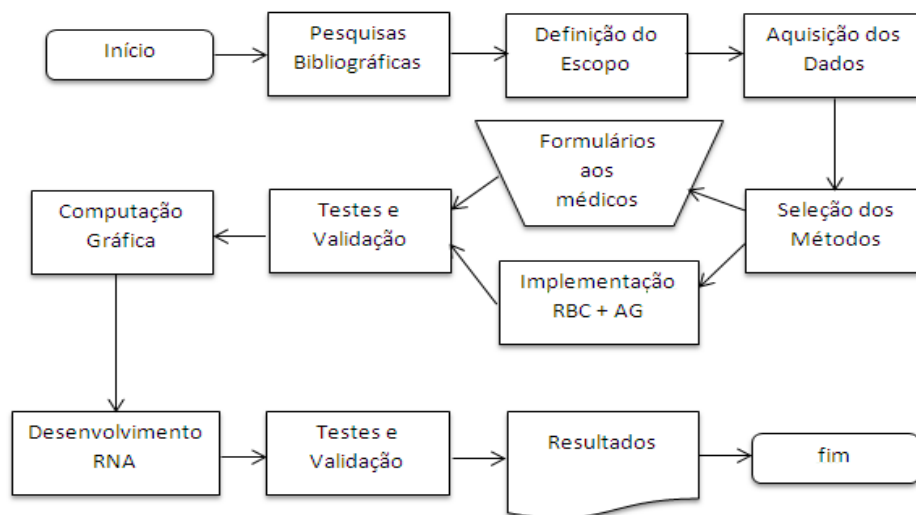


Figura 11. Fluxograma – Elaboração do Projeto. Fonte: O autor.

O fluxograma apresentado na imagem acima descreve de forma resumida o ciclo de vida da proposta em questão. Logo é possível entender cada etapa do processo, desde a concepção da ideia, até disponibilização da ferramenta para testes em campo.

3.4. Coleta e Análise de Dados

Para confecção deste trabalho foi utilizada uma base de dados contendo aproximadamente 997 registros, dos quais 592 apresentavam diagnósticos positivos para a cardiopatia isquêmica, gentilmente cedido por (ALMEIDA, 2006). Utilizou-se também imagens extraídas do bando de imagens de sinais de eletrocardiogramas, adquiridas através do web site da *European Union e European Society of Cardiology*, base esta que contém 92

registros de imagens de eletrocardiogramas distribuídos livremente, sendo cada um resultado de 10 minutos de sinais monitorados devidamente anotados por cardiologistas.

Por fim, para seleção dos atributos a serem considerados como os mais relevantes e definição primária da relevância de cada um, foi realizada a combinação dos trabalhos de (ALMEIDA, 2006) (DETRANO, 1989) e (SETIAWAN, VENKATACHALAM e HANI, 2009), de onde foram levantados os atributos para confecção de um questionário enviado para vários especialistas médicos para validação dos atributos e levantamento primário dos pesos.

3.5. Delimitação do Escopo

O projeto em questão apresentará uma interface gráfica com usuário amigável, pois destina-se também a coordenar as atividades gerais de uma clínica médica. Entretanto o escopo principal do projeto é oferecer suporte ao diagnóstico da cardiopatia isquêmica, através da análise e atribuição de pesos. Após esta etapa o sistema deve ser capaz de indicar o percentual de probabilidade de um determinado indivíduo apresentar diagnóstico positivo ou não para esta enfermidade.

Através do módulo gerenciado pelo cérebro artificial a ferramenta deve ser capaz de, oferecer valores preditivos que confirmem ou revoguem o diagnóstico anterior permitindo inclusive comparações através da análise de imagens contendo sinais de eletrocardiogramas.

Apesar de possibilitar a adaptação e treinamento para outras enfermidades esta ferramenta não esta habilitada e nem se propõe a diagnosticar qualquer outra enfermidade que não seja a cardiopatia isquêmica.

3.6. Resultados Esperados

Os resultados obtidos por cada um dos métodos passaram por diversas análises estatísticas, os quais permitiram quantificar e qualificar os mesmos, dentre os diversos testes realizados pode-se citar:

- a) Especificidade: capacidade de um teste discriminar dentro de um número de suspeitos, àqueles que estão efetivamente doentes;
- b) Sensibilidade: capacidade que o mesmo teste tem de ser negativo, para indivíduos que efetivamente não estão doentes;
- c) Prevalência: número de casos efetivamente levados a teste dentro da população; e

- d) Acurácia: a capacidade de um teste realmente identificar corretamente as pessoas com uma determinada doença e excluir àquelas que realmente estão saudáveis.

Testes como análise de variância (ANOVA⁴) e Tukey⁵ também foram realizados permitindo comparar o desempenho da aplicação aqui proposta com outras disponíveis na literatura.

Espera-se com este trabalho os seguintes resultados:

- i) Determinação de um AG de qualidade para cálculo dos pesos das funções de similaridade aplicadas ao módulo RBC;
- ii) Implementação de um módulo baseado em RNA para análise de imagens de sinais de eletrocardiogramas para segunda opinião na análise dos diagnósticos;
- iii) Desenvolvimento de uma ferramenta gratuita eficaz no que tange o auxílio na atividade de diagnóstico médico;
- iv) Validação da ferramenta proposta através da utilização de casos reais da área de cardiologia contendo diagnósticos da Cardiopatia Isquêmica;
- v) Comparação, com auxílio de especialista médico e métodos estatísticas, dos resultados, das melhores soluções e outras ferramentas com o mesmo propósito.

3.7. Técnicas, recursos e configurações

Este trabalho lançou mão das seguintes técnicas referentes à inteligência computacional: Algoritmos Genéticos, Raciocínio Baseado em Casos, Computação Gráfica e Redes Neurais Artificiais.

O algoritmo genético proposto, por ter sido implementado paralelamente a distribuição dos formulários exigiu a capacidade de ter suas configurações referentes ao número de atributos e valores máximos e mínimos dos atributos configuráveis durante a execução do mesmo, permitindo assim adaptação para novos casos.

Foram também implementadas as seguintes funções para funcionamento do algoritmo: seleção por roleta (com ranqueamento), seleção por elitismo, cruzamento de um e vários pontos e **mutação**. A validação da solução deu-se principalmente através da função de cálculo de distância euclidiana ponderada, apresentada na Tabela 2, tendo sido a taxa de similaridade entre

⁴ ANOVA – Análise de Variância.

⁵ Tukey – Teste de Tukey – Diferença Significante.

os casos ajustada para 80%, indicando que somente os casos com no mínimo 80% de similaridades deveriam ser levados a testes, conforme será explicado no capítulo 5.

O projeto também contempla aplicação de Redes Neurais Artificiais, o que permite análise dos casos sob uma ótica diferenciada, permitindo inclusive que, através da aplicação de técnicas de computação gráfica imagens e sinais de eletrocardiogramas possam ser analisados pelo cérebro artificial, fornecendo desta forma ao profissional especialista, melhores condições para o processo decisório.

O projeto exigia também, para que pudesse ser continuado uma taxa mínima de sucesso nas predições de diagnósticos realizadas pelo módulo de RBC, esta taxa foi fixada em 85%, por ser também a taxa média de sucesso nos diagnósticos realizados pelos grupos de especialistas (ALMEIDA, 2006).

Com pode ser observado na Figura 11 mostrada acima, foi desenvolvido também um cérebro artificial para o projeto, permitindo assim a análise dos atributos pode uma nova ótica, através de métodos diferentes dos aplicados em algoritmos genéticos. Além de realizar análise de dados clínicos, assim como ocorre com o módulo onde foram aplicadas técnicas de RBC + AG, o cérebro artificial permite a análise de sinais de eletrocardiogramas através de imagens. Para tornar isto possível foi desenvolvido no projeto um pequeno módulo de computação gráfica, encarregado de recuperar as imagens dos sinais de eletrocardiogramas, preparar a imagem e convertê-las para um conjunto binário que servirá de entrada para a RNA.

O tratamento de todas as imagens foi realizado através de um algoritmo inspirado no algoritmo de *OTSU*, permitindo desta forma criar as entradas necessárias para o treinamento da RNA.

A RNA utilizada para reconhecimento da enfermidade tratada neste trabalho foi desenvolvida usando-se como base a biblioteca FANN (*Fast Artificial Neural Network*), distribuída gratuitamente nas mais diversas linguagens de programação capazes de implementar inúmeras estruturas de RNA's.

4. PROPOSTA DO CARDIOCARE

Como explanado anteriormente a ferramenta em questão foi concebida através da aplicação de vários conceitos relacionados à inteligência computacional, possuindo esta um módulo para cálculo de similaridade entre casos, permitindo desta forma a aplicação das técnicas de raciocínio baseado em casos, aplicação de algoritmos genéticos para definição dos pesos de cada um dos atributos verificados. É aplicada também redes neurais artificiais, MLP (*multilayer perceptron*) com *backpropagation*, para, não somente analisar os mesmos atributos utilizados no módulo de RBC, mas também para análise de imagens de sinais de eletrocardiogramas.

A solução proposta oferece ao usuário não somente suporte ao processo de tomada de decisão, mas também apoio operacional às atividades de uma clinica ou consultório médico. O ambiente utilizado para o desenvolvimento desta ferramenta foi:

- Softwares:
 - Sistema Operacional Windows 8.1;
 - Plataforma de Desenvolvimento, Visual C# Express;
 - Linguagem de Programação C#;
- Hardware:
 - Notebook;
 - Processador Intel Core i3 – Segunda Geração;
 - 6 GB Memória RAM;

Toda tecnologia utilizada para desenvolvimento da ferramenta em questão pode ser adquirida por um custo relativamente baixo pelas organizações promotoras da saúde coletiva pública e/ou particular.

4.1. Arquitetura

Apesar de suas características a solução proposta não apresenta uma arquitetura complexa, mostrando-se assim de fácil utilização de estudo, permitindo evoluções futuras. Basicamente a arquitetura da ferramenta procura atender a um número muito pequeno de requisitos. Sua arquitetura básica pode ser observada no diagrama de casos de uso representado na Figura 12.

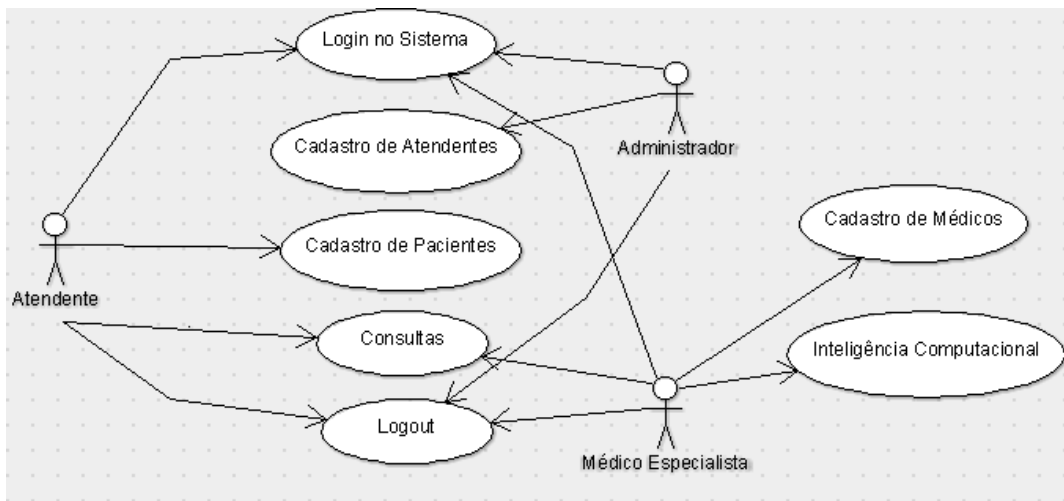


Figura 12. Caso de Uso – Visão Geral do Projeto. Fonte: O autor.

Pode-se ver pela imagem a interação não somente do especialista médico, mas também de outros atendentes ao sistema. Isto se dá pelo fato de o sistema ser desenvolvido também para atendimentos ambulatoriais e permissão de triagem pré-consulta.

Ao se analisar mais profundamente alguns módulos do sistema, mais especificamente os módulos **consultas** e **inteligência computacional**, tem-se os seguintes casos de uso observados nas Figuras 13 e 14.

O módulo de consulta aberto tanto aos médicos quanto aos atendentes permite inicializar uma consulta, realizar o procedimento de triagem e acompanhar a evolução do paciente na medida em que este comparece às consultas.

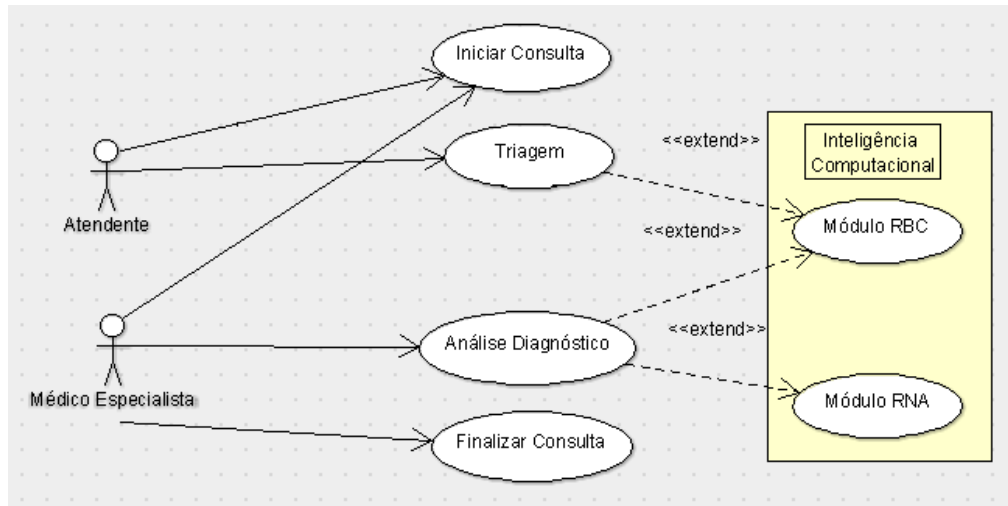


Figura 13. Caso de Uso – Consultas. Fonte: O autor.

Sempre que necessário o módulo de consultas pode acessar informações/métodos pertinentes ao módulo de inteligência computacional. Para isto é necessário que o operador indique ao sistema sua intenção. Uma vez iniciado este módulo se encarregará de observar a todos os atributos disponíveis para então realizar sua previsão de diagnóstico.

Através do módulo de inteligência artificial o usuário poderá configurar não somente os tipos de operações a serem realizadas pelo algoritmo genético, como também poderá configurar algumas das opções de aprendizado do módulo de redes neurais artificiais, bem como escolher qual a melhor medida de similaridade para o módulo RBC. Entretanto estas opções estarão disponíveis somente para o **Operador Administrador** do sistema.

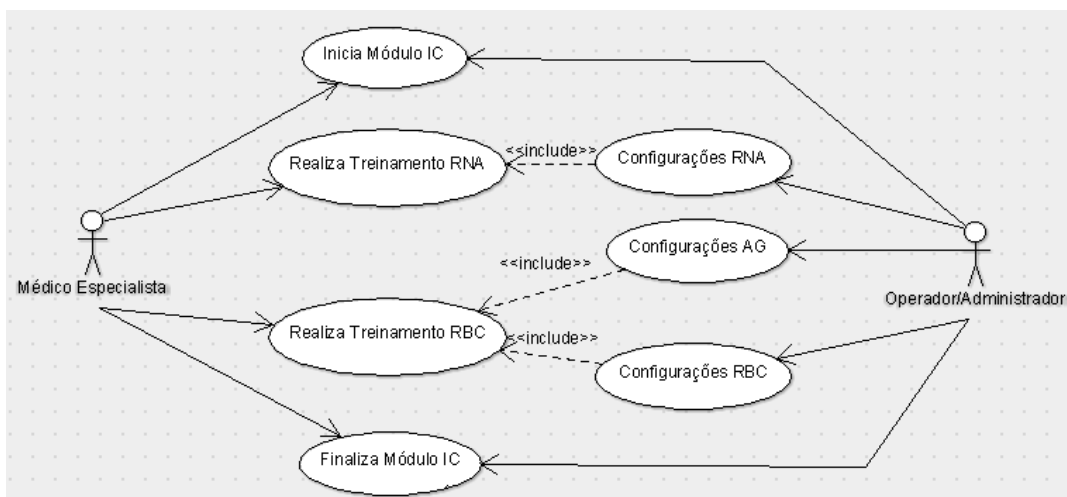


Figura 14. Caso de Uso – Inteligência Computacional. Fonte: O autor.

Logo, sempre que necessário ou de tempos em tempos novo treinamento podem ser realizados no sistema, compondo assim novos conhecimentos para o memo. Tal treinamento pode ser realizado tanto pelo médico especialista quanto pelo operador administrador, sendo

este o único que pode interferir nas configurações dos módulos específicos de inteligência computacional.

4.2. Algoritmos

A ferramenta proposta aplica basicamente os seguintes algoritmos chave para seu correto funcionamento: Algoritmos Genéticos (para módulo de raciocínio baseado em casos), algoritmos de computação gráfica (baseado no algoritmo de OTSU) e algoritmo de Redes Neurais Artificiais – MPL – *Backpropogation*, utilizado a partir da importação da biblioteca *FANN – Fast Artificial Neural Network*.

4.2.1. Algoritmos Genéticos (RBC)

O ciclo básico do algoritmo genético proposto no trabalho, bem como sua associação às técnicas de raciocínio baseado em casos pode ser observado de forma básica através do fluxograma apresentado na Figura 15.

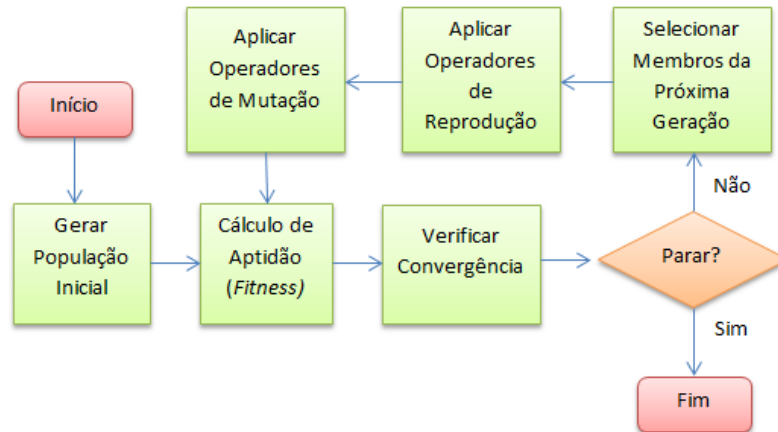


Figura 15. Fluxograma – Algoritmo Genético. Fonte: O autor.

Ao iniciar o módulo de RBC é ativado o algoritmo genético para cálculo de similaridade do mesmo.

Então segundo as configurações já estabelecidas é gerada uma população inicial, a partir da qual é feito um cálculo de similaridade. A partir destes resultados são realizados algoritmos de seleção **elitismo** e **roleta (com ou sem ranqueamento)**. Posteriormente são realizados o **cruzamentos (de um ou vários pontos)** e **mutação**.

As características do algoritmo proposto podem ser observados abaixo:

- Os genes do cromossomo são representados por números binários;

- A população inicial será gerada aleatoriamente;
- O número de indivíduos de cada população bem como a quantidade de gerações é definida de forma flexível;

- O critério de parada do algoritmo também é definido de forma flexível;

Os trechos do algoritmo acima descrito podem ser observado abaixo. A Figura 16. contém o trecho de código que realiza a geração da população inicial de forma aleatória.

```

for (int x = 0; x < nInd; x++)
{
    Random rand = new Random();
    retorno = "";
    l_genes = new List<Genes>();
    for (int j = 0; j < natrid; j++)
    {
        l_alelos = new List<Cromossomos>();

        for (int i = 0; i < 4; i++)
        {
            int fator = DateTime.Now.Millisecond;
            alelo = new Cromossomos();

            fator = fator * ((DateTime.Now.Millisecond * DateTime.Now.Millisecond) * (i * DateTime.Now.Millisecond * j) + 1);
            double valRandom = (double)rand.NextDouble();

            if (valRandom > 0.50)
            {
                alelo.alelo = '1';
                l_alelos.Add(alelo);
                retorno = retorno + "1";
            }
            else
            {
                alelo.alelo = '0';
                l_alelos.Add(alelo);
                retorno = retorno + "0";
            }
            Thread.Sleep(9);
        }
        Genes gene = new GenesControle().criaGene(l_alelos);
        l_genes.Add(gene);
    }
}

```

Figura 16. Algoritmo – População Inicial. Fonte: O autor.

Como explicado anteriormente a população inicial é gerada aleatoriamente. Buscando garantir o mínimo de possibilidade de repetição entre os valores sorteados, foi criado um fator de multiplicação para ser aplicado à função *rand*⁶ do C#⁷.

Para os testes realizados o cálculo de aptidão ou *fitness* foi realizado com base na função de distância euclidiana ponderada. O trecho do algoritmo que realiza este cálculo pode ser observado na Figura 17.

⁶ Função que gera valores numéricos aleatoriamente (é possível definir limites máximos e mínimos)

⁷ Linguagem de programação criada e distribuída pela Microsoft Corp.

```

if ((atrib_c >= 0) && (atrib_t >= 0))
{
    for (int x = 0; x < vetInd.gene[liY].cromossomo.Count; x++)
    {
        cromosoma = cromosoma + vetInd.gene[liY].cromossomo[x].alelo;
    }
    distancia = atrib_c - atrib_t;
    distancia = (double)Math.Pow(distancia, 2);
    geneReal = converteReal(0,1,4,cromosoma);
    distancia = (double)((double)geneReal * (double)distancia);
    //distancia = (double)Math.Pow(distancia, 2);
    sim = sim + (double)Math.Sqrt(distancia);
}
else
{
    for (int x = 0; x < vetInd.gene[liY].cromossomo.Count; x++)
    {
        cromosoma = cromosoma + vetInd.gene[liY].cromossomo[x].alelo;
    }
    distancia = 1;
    distancia = (double)Math.Pow(distancia, 2);
    geneReal = (double)converteReal(0, 1, 4, cromosoma);
    distancia = (double)((double)geneReal * (double)distancia);
    sim = sim + (double)Math.Sqrt(distancia);
}
}

```

Figura 17. Algoritmo – Função de Aptidão. Fonte: O autor.

Sendo que, para cada atributo é realizada um cálculo local de similaridade, havendo posteriormente um cálculo global no qual é aplicado a mesma função de similaridade.

Sabe-se que o processo acima citado é realizado para cada indivíduo da população corrente. No trabalho em questão o algoritmo foi configurado para realizar testes com todos os registros armazenados no banco de dados, assim como demonstrado no fluxograma apresentado na Figura 18.

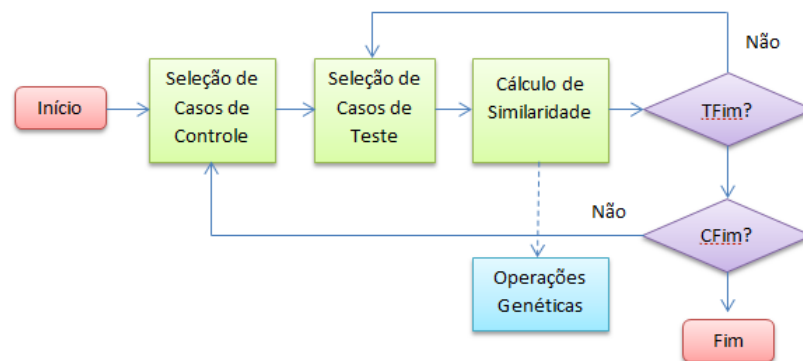


Figura 18. Algoritmo – Função de Aptidão. Fonte: O autor.

A cada teste de aptidão o algoritmo entra em um novo processo repetitivo, para realizar as verificações com cada um dos valores (pesos) encontrados pelo algoritmo genético, para cada atributo (sintoma) e posteriormente para o paciente. Este processo foi referenciado na Figura 15., que apresenta o fluxograma do mesmo. Por serem demasiadamente extensos, estes algoritmos poderão ser analisados nos documentos anexos a este trabalho.

4.2.2. Computação Gráfica / RNA

Buscando oferecer um segundo diagnóstico ou mesmo a possibilidade de realizar indicações de diagnóstico baseado em imagens de sinais de eletrocardiogramas, foi implementado o módulo RNA. Entretanto, para que este módulo viesse a funcionar corretamente foi necessário converter as imagens (escaneadas ou recuperadas diretamente em forma digital) em matrizes binárias, permitindo assim gerar as entradas necessárias para a RNA. O fluxograma apresentado na Figura 19. representa de forma clara o processo descrito acima.

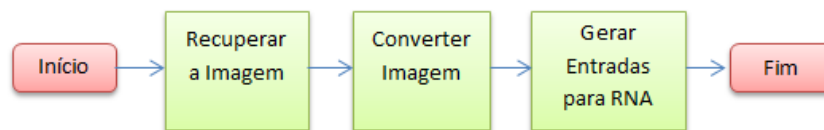


Figura 19. Fluxograma – Conversão de Imagens. Fonte: O autor.

Para que a solução seja capaz de realizar uma indicação de diagnóstico por análise de imagens, o usuário deve antes realizar o treinamento do sistema através da recuperação da imagem que, depois de submetida aos algoritmos adequados (que poderão ser observados nos apêndices) gerarão as entradas necessárias para o treinamento da RNA.

O procedimento pode ser melhor compreendido analisando a sequencia de figuras abaixo, a iniciar pela Figura 20.

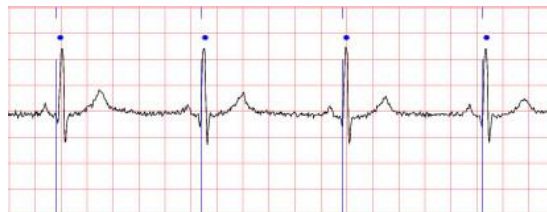


Figura 20. Imagem Padrão Sinal de ECG. Fonte: O autor.

A figura acima simula uma das possíveis entradas para o módulo de treinamento e análise de imagens de sinais de ECG, contudo, na forma em que se encontra esta imagem não é capaz de fornecer uma entrada aceitável para o módulo de RNA, sendo necessário tratamentos, como podem ser observados nas Figuras 21, 22 e 23.

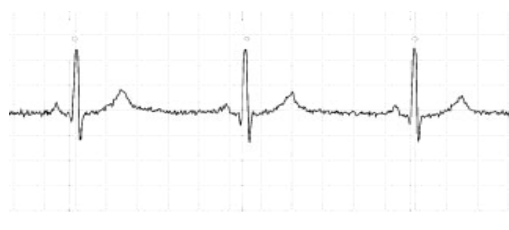


Figura 21. Imagem Padrão Sinal de ECG - Monocromática. Fonte: O autor.

A conversão da imagem em um padrão monocromático é o primeiro passo da preparação da mesma para geração das entradas para a RNA. Uma vez convertida, o sistema analisará somente 2 cores buscando nesta sequência de cores padrões para separação de um trecho que se apresenta repetidamente na imagem, o que pode ser observado na Figura 22.

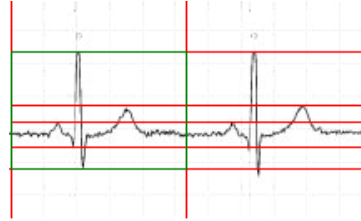


Figura 22. Imagem Padrão Sinal de ECG (Monocromática) – Busca por Padrões. Fonte: O autor.

As linhas vermelhas observadas indicamos pontos de máximo e mínimo local, tanto horizontal quanto verticalmente, indicando assim ao sistema onde existem padrões na imagem analisada. O retângulo verde indica a área de repetição, sendo esta separada e convertida em uma matriz binária, como indicado na Figura 23.

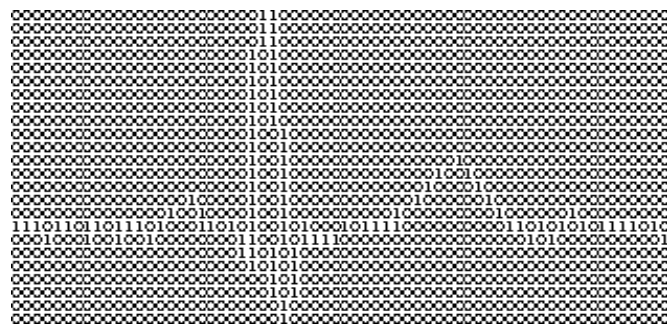


Figura 23. Matriz binária sinal de ECG – Busca por Padrões. Fonte: O autor.

Uma vez construída a matriz binária correspondente à imagem, é possível então gerar o arquivo de entrada para a RNA, composto da seguinte forma: **linha com dados de entrada, linha com saída esperada**, e assim sucessivamente, até serem consideradas todas as imagens separadas para o treinamento. Gerado o arquivo o mesmo é submetido como entrada para o algoritmo apresentado na Figura 24 (lembrando que, o algoritmo esta escrito na linguagem C, sendo este adicionado como referência externa no projeto CardioCare escrito em C#).

```
struct fann *ann = fann_create_standard(num_layers, num_input,
    num_neurons_hidden, num_output);

fann_set_activation_function_hidden(ann, FANN_SIGMOID_SYMMETRIC);
fann_set_activation_function_output(ann, FANN_SIGMOID_SYMMETRIC);

fann_train_on_file(ann, "image_bin.data", max_epochs,
    epochs_between_reports, desired_error);

fann_save(ann, "img_training.net");

fann_destroy(ann);
```

Figura 24. Algoritmo de Ativação da RNA. Fonte: O autor.

O algoritmo acima representado recebe como entrada o arquivo contendo as informações sobre as imagens para o treinamento (bits de entrada e bit de saída). Sendo os bits de saída representados por 0 (ausência da enfermidade) e 1 (presença da enfermidade), podendo a RNA retornar qualquer valor entre estas faixas indicando o grau de indicação. Uma vez realizado o treinamento, a função retorna o arquivo de configuração da rede para futura utilização.

5. RESULTADOS

Para validação do algoritmo proposto e comprovação de sua eficiência e eficácia, o mesmo foi exposto a um exaustivo processo de treinamento e experimento (utilização assistida). Após treinado o algoritmo foi submetido a várias simulações, com informações de prontuários previamente retirados da mesma base de dados utilizadas para testes, contudo, os prontuários utilizados não participaram da etapa de treinamento. Os resultados obtidos nos testes permitiram sua submissão a várias análises o que tornou possível avaliar seu desempenho e aplicabilidade.

5.1. Parametrização e Atributos

Os atributos utilizados pelo algoritmo como termos de comparação foram criados através de orientações provenientes de especialistas e análise de estudos bibliográficos (ALMEIDA, 2006), (SONI, 2011) e (DETRANO, 1989) levando aos atributos representados no Quadro 1.

Atributo	Descrição	Ordem
Idade	informada em anos e então classificada em um das 18 faixas previstas	1
Sexo	M ou F	2
Pressão Sistólica	informada em mmHg e então classificada em uma das 9 faixas previstas	3
Pressão Diastólica	informada em mmHg e então classificada em uma das 9 faixas previstas	4
Frequência Cardíaca	informada em bpm e então classificada em uma das 8 faixas previstas	5
Intensidade da Dor	um código representando uma das 3 faixas previstas	6
Tempo de início do último episódio de dor	um código representando uma das 4 faixas previstas	7
Tempo de início dos Sintomas	um código representando uma das 8 faixas previstas	8
Antecedentes	informação da existência ou não de 12 tipos de antecedentes	9
Característica da dor	informação da existência ou não de 7 características específicas da dor	10
Alterações no Eletrocardiograma	informação da existência ou não de 10 tipos de alterações no ECG	11
Local da dor	informação da existência ou não de dor em 6 locais do corpo	12

Outros sintomas	informação da existência ou não de 6 outros sintomas	13
Sinais de ausculta	informação da existência ou não de 4 características observadas na ausculta	14
Sinais observados	informação da existência ou não de 5 sinais observados no exame	15
Fatores de risco	informação da existência ou não de 7 fatores considerados de risco	16
Sinais de pulso	informação da existência ou não de 4 características nos sinais de pulso	17

Quadro 1. Atributos considerados nos testes de similaridade do algoritmo. Fonte: Especialistas e adaptação de (ALMEIDA, 2006), (SONI, 2011) e (DETRANO, 1989)

Para cada atributo foi definida uma faixa mínima e máxima para atribuição de pesos, o modelo aqui proposto seguiu o sugerido por (ALMEIDA, 2006). A faixa de valores sugerida e utilizada neste projeto foi de [0..1], sendo **zero** o menor peso, representando assim menor interferência no resultado final e **um** o maior peso, ou seja, maior interferência no resultado final.

Para definição inicial dos pesos, oferecendo assim parâmetros para comparação dos resultados obtidos pelo sistema, utilizou-se os valores atribuídos pelo **especialista 01**⁸ e **especialista 02**⁹. Os pesos atribuídos pelos especialistas são apresentados no Quadro 2.

Atributo	Especialista 01	Especialista 02
Idade	0,6	0,7
Sexo	0,3	0,3
Pressão Sistólica	0,4	0,4
Pressão Diastólica	0,6	0,5
Frequência Cardíaca	0,7	0,7
Intensidade da Dor	0,7	0,8
Tempo de início do último episódio de dor	0,6	0,6
Tempo de início dos Sintomas	0,5	0,6
Antecedentes	1,0	0,8
Característica da dor	0,6	0,8
Alterações no Eletrocardiograma	0,8	1,0
Local da dor	0,7	0,8
Outros sintomas	1,0	0,9
Sinais de ausculta	0,5	0,6
Sinais observados	0,7	0,8
Fatores de risco	0,6	0,8
Sinais de pulso	0,4	0,4

Quadro 2. Pesos atribuídos por especialistas. Fonte: Especialistas e (ALMEIDA, 2006)

Cada especialista consultado, segundo seus conhecimentos, atribuiu valores a cada um dos atributos, de acordo com o que julgam ser mais relevante na determinação do diagnóstico da cardiopatia isquêmica. Outros 15 formulários foram distribuídos para outros especialistas médicos na cidade de Cachoeiro de Itapemirim e cidades vizinhas, obtendo 7 respostas. Os resultados foram adicionados aos dois que já constavam no projeto, dando origem ao Quadro

⁸ Dr. Hervaldo Sampaio de Carvalho, médico cardiologista e professor da Universidade de Brasília.

⁹ Dr. Ubiratan Rosa Passos, médico cardiologista especialista pela Sociedade Brasileira de Cardiologia.

3, sendo também calculada a média dos pesos atribuídos pelos especialistas, sendo a média utilizada para comparação com os pesos posteriormente indicados pelo A.G.

Atributo	Esp. 01	Esp. 02	Esp. 03	Esp. 04	Esp. 05	Esp. 06	Esp. 07	Esp. 08	Esp. 09	Média
Idade	0,6	0,7	0,4	0,3	0,5	0,1	0,4	0,3	0,6	0,43
Sexo	0,3	0,3	0,4	0,4	0,3	0,4	0,4	0,6	1,0	0,45
Pressão Sistólica	0,4	0,4	0,2	0,1	0,3	0,5	0,3	0,3	0,5	0,33
Pressão Diastólica	0,6	0,5	0,5	0,8	0,1	0,3	0,2	0,6	0,3	0,43
Frequência Cardíaca	0,7	0,7	0,6	0,5	0,8	0,5	0,3	0,5	0,2	0,53
Intensidade da Dor	0,7	0,8	0,4	0,1	0,4	0,6	0,1	0,2	0,1	0,37
Tempo de início do último episódio de dor	0,6	0,6	0,2	0,3	0,9	0,5	0,2	0,4	1,0	0,52
Tempo de início dos Sintomas	0,5	0,6	0,3	0,1	0,2	0,5	0,5	0,4	0,3	0,37
Antecedentes	1,0	0,8	1	0,9	1,0	0,8	0,8	0,5	0,9	0,85
Característica da dor	0,6	0,8	0,2	0,5	0,7	0,7	0,8	0,6	0,8	0,63
Alterações no Eletrocardiograma	0,8	1,0	0,9	0,3	0,2	0,2	0,8	0,5	0,8	0,61
Local da dor	0,7	0,8	0,8	0,6	0,9	0,8	0,8	0,5	1,0	0,76
Outros sintomas	1,0	0,9	0,8	0,9	0,6	0,8	0,9	0,8	0,5	0,80
Sinais de ausculta	0,5	0,6	0,3	0,7	0,8	0,5	0,5	0,7	0,1	0,52
Sinais observados	0,7	0,8	0,7	0,4	0,8	0,4	0,3	0,3	0,8	0,57
Fatores de risco	0,6	0,8	0,2	0,4	0,7	0,6	0,5	0,5	0,8	0,56
Sinais de pulso	0,4	0,4	0,2	0,8	0,1	0,2	0,1	0,5	0,1	0,31

Quadro 3. Pesos atribuídos por especialistas. Fonte: Especialistas e (ALMEIDA, 2006)

Obtidas as informações iniciais para avaliação do algoritmo, definiu-se as configurações dos algoritmos (A.G. e RNA), mesmo sendo possível alterar estes valores durante a execução do programa. Foram então definidos os seguintes valores para execução, treinamento e testes dos mesmos:

- a) **Algoritmos Genéticos:** Taxa de Cruzamento – $T_c = 0,70$; Taxa de Mutação – $T_m = 0,03$, valores estes indicados na bibliografia estudada, como pode ser observado em (ALMEIDA, 2006). O número de indivíduos de cada geração, bem como a quantidade de gerações foi definida empiricamente, ficando configurado como 30 o número de indivíduos e 20 as gerações.

- b) **Redes Neurais Artificiais:** Taxa de aprendizado – foram treinadas com as seguintes taxas: 0.3, 0.7 e 0.9. Camada Oculta – foram configuradas da seguinte forma: 5, 15 e 25 neurônios. Função de Ativação: Sigmoide. Tais valores foram definidos de forma empírica.

Para que a ferramenta pudesse interpretar as informações fornecidas pelo sistema no tocante ao histórico clínico do paciente, os atributos foram convertidos em valores numéricos, como pode ser observado na Figura 25.

<u>DADOS CLÍNICOS</u>
<u>Idade</u>
<u>Sexo</u> (0 = Masculino; 1 = Feminino)
<u>Pressão Sistólica</u>
<u>Pressão Diastólica</u>
<u>Frequência Cardíaca</u>
<u>Intensidade da Dor</u> (0 = Dor leve; 1 = Dor Moderada; 2 = Dor Severa)
<u>Tempo de início do último episódio de Dor</u> (0 = Até uma hora; 1 = Até duas horas; 2 = Até três horas; 3 = Até quatro horas; 4 = Até cinco horas; 5 = Até seis horas; 6 = Entre seis e doze horas; 7 = Acima de doze horas)
<u>Tempo de início dos Sintomas</u> (0 = Sem dor; 1 = Entre 30 minutos e 12 horas; 2 = Entre 12 e 48 horas; 3 = Acima de 48 horas)
<u>Antecedentes</u> (0= Intervenção coronária prévia (mais de 6 meses); 1 = Intervenção coronária prévia (últimos 6 meses); 2 = Cardiopatia Isquêmica na família (Idade<55); 3 = Dor/desconforto precordial há menos de 2 semanas; 4 = Episódios repetidos de dor nos últimos 2 meses; 6 = Evento coronário prévio; 7 = Histórico de doença cardíaca não coronariana; 8 = Histórico de doença ulcerosa gastro-duodenal; 9 = Histórico de dor que melhora com ansiolítico; 10 = Histórico esofagite/disp. gasosa/cólica biliar; 11 = Marcapasso; 12 = Prolapso da válvula mitral)
<u>Característica da dor</u> (0 = Depende da posição do paciente; 1 = Desencadeada com os esforços ou emoções; 3 = Em aperto ou opressão; 4 = Em peso ou queimor; 5 = Em pontada; 6 = Espontânea em repouso; 7 = Ventilatório dependente)
<u>Alterações no Eletrocardiograma</u> (0 = Marcapasso; 1 = BRE; 2 = SVE; 3 = Segmento ST supradesnivelado; 4 = Segmento ST infradesnivelado; 5 = Onda T Apiculada; 6 = Onda T Baixa Probabilidade; 7 = Onda T Probabilidade Intermediária; 8 = Onda T Alta Probabilidade; 9 = Complexo QRS com padrão Q ou QS;
<u>Local da Dor</u> (0 = Braço esquerdo e/ou direito; 1 = Dorso; 2 = Hemitórax direito; 3 = Mandíbula; 4 Região epigástrica; 5 = Região precordial ou médio-esternal)
<u>Outros Sintomas</u> (0 = Dispnéia aos esforços; 1 = Dispnéia em repouso; 2 = Febre; 3 = Náuseas e vômitos; 4 = Palpitações associadas; 5 = Sudorese)
<u>Sinais de Ausculta</u> (0 = Atrito pericárdico; 1 = Batimento cardíaco com ritmo de galope; 2 = Crepitos de base; 3 = Nenhuma anormalidade)
<u>Sinais Observados</u> (0 = Dor a palpação no local referido; 1 = Facies de sofrimento agudo; 2 = Palidez; 3 = Pele viscosa; 4 = Vômitos)

Fatores de Risco (0 = Diabete; 1 = Dislipidemia; 2 = Hipertensão arterial; 3 = Obesidade; 4 = Pós-menopausa; 5 = Stress; 6 = Tabagismo)
Sinais de Pulso (0 = Arritmico; 1 = Filiforme; 2 = Não Existe Diferença de pulso MI_MS; 3 = Sim Existe Diferença de pulso MI_MS)

Figura 25. Representação numérica dos valores dos atributos. Fonte: O autor.

Convertendo os valores de texto para valores numéricos, o sistema torna-se capaz de analisar cada um dos atributos segundo as funções de cada método. Apesar de terem sido atribuídos empiricamente, a alteração destes valores não interfere nos cálculos e resultado final apresentado pelos métodos.

Para o treinamento baseado em dados clínicos foram utilizados 500 casos. Após o treinamento os módulos foram submetidos a testes com 134 casos que não foram utilizados durante o processo de treinamento.

O treinamento e simulação para análise de imagens de sinais de eletrocardiogramas foi realizado somente para o módulo de RNA, foram adquiridas 92 imagens, sendo destas 60 utilizadas para o treinamento e 32 para simulação.

Os resultados do treinamento e das simulações podem ser observados no tópico 5.2. Resultados.

5.2. Resultados

Uma vez que a solução apresentada se trata de uma ferramenta híbrida que combina tecnologias e módulos diferentes, cada um apresentando resultados separados (que podem por opção do usuário serem agrupados e comparados para geração de um resultado único), nos testes e simulações abordou-se o conceito de prova e contraprova.

Aplicou-se então o mesmo teste no módulo de RBC e RNA, sendo os resultados divididos por módulo, buscando assim melhor legibilidade e entendimento para o leitor. Logo os resultados ficaram assim organizados. Tópico 5.2.1. Resultados RNA; Tópico 5.2.2. Resultados RBC (AG); Tópico 5.2.3. Validação.

5.2.1. Resultados RNA

Após a realização dos treinamentos nos quais foram utilizados as especificações já citadas, foi possível observar a evolução da RNA na medida em que novos casos eram sendo submetidos ao teste. Nas simulações pode-se observar uma curva de erro médio semelhante a

apresentada no Gráfico 1, que corresponde ao resultado do processo de treinamento da RNA, para diagnóstico da cardiopatia isquêmica baseada nos dados clínicos, ou seja, sem a realização de qualquer análise nas imagens dos sinais de eletrocardiograma.



Gráfico 1. Erro médio do treinamento – Dados Clínicos. Fonte: O Autor.

Nota-se pelo gráfico que, a medida em que o treinamento avança ocorre também a redução da taxa de erro médio, chegando o sistema a um ponto de mínima convergência entre as épocas 4000 e 4500, não sendo necessário assim prosseguir com o treinamento.

Ao ser submetida ao treinamento para análise de imagens de sinais de eletrocardiogramas, a RNA apresentou o seguinte gráfico de curva do erro médio, vide Gráfico 2.

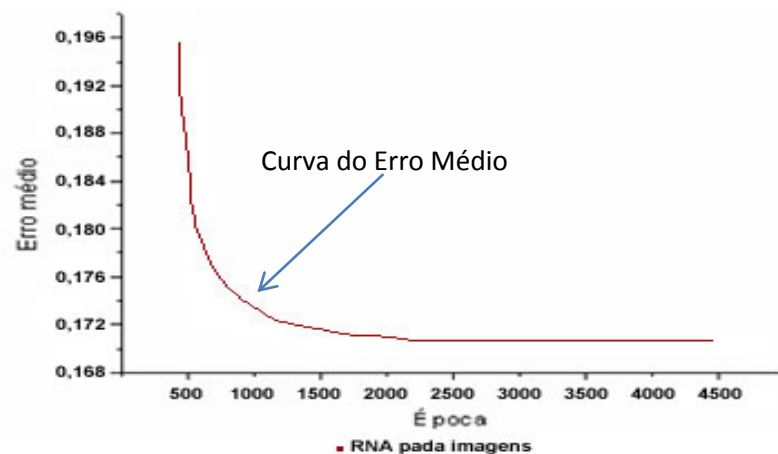


Gráfico 2. Erro médio do treinamento – Sinais de ECG. Fonte: O Autor.

No treinamento para reconhecimento de padrões e indicação de diagnóstico a partir da análise de sinais de eletrocardiogramas, a rede apresenta estagnação na evolução da taxa de acertos ainda antes da 2500ª época, podendo ai ser encerrado seu treinamento.

Sobrepondo então a evolução da RNA para os dois treinamentos, obtêm-se o Gráfico 3.

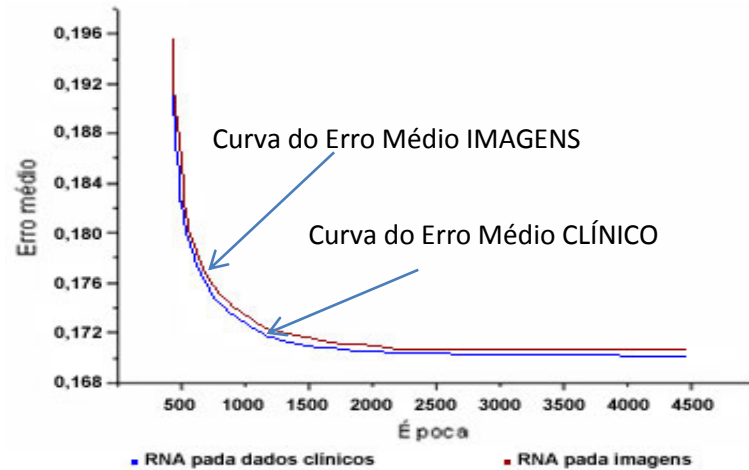


Gráfico 3. Erro médio do treinamento – Sobreposição RNA – Dados clínicos x RNA Imagens ECG. Fonte: O Autor.

É possível então notar que, apesar de tênue, existe uma superioridade nas taxas de acertos da RNA treinada para dados clínicos sobre a RNA treinada para análise de imagens de sinais de EGC's, o que pode por um lado ser explicado, por exemplo, devido à necessidade de tratamento e conversão das imagens que pode gerar alguma perda de informação.

As informações obtidas durante o processo de treinamento e simulação do módulo RNA forneceu subsídios para definição do seguinte quadro de resultados, vide Quadro 4.

Nº Neurônios	Épocas	Tx. Aprendizado	Função de Ativação	Treino %	Teste%	Fonde Dados
5	3700	0.3	Sigmoide	81,432	81,687	Clinico
5	3700	0.3	Sigmoide	81,323	80,942	Imagem
10	2709	0.7	Sigmoide	80,987	80,980	Clinico
10	2342	0.7	Sigmoide	79,438	78,983	Imagem
15	1876	0.9	Sigmoide	87,049	87,001	Clinico
15	1765	0.9	Sigmoide	85,093	84,984	Imagem
20	3200	0.3	Sigmoide	76,094	75,093	Clinico
20	3100	0.3	Sigmoide	89,091	88,984	Imagem
25	2903	0.7	Sigmoide	83,239	82,364	Clinico
25	2876	0.7	Sigmoide	82,264	82,153	Imagem

Quadro 4. Configuração e resultados do processo de aprendizado das RNA's. Fonte: O Autor.

Confirmando assim, como pode ser observado pelo quadro acima, a capacidade alcançada pela RNA para diagnosticar a enfermidade estudada a partir da análise de dados clínicos do histórico do paciente e/ou através da simples análise dos sinais do ECG.

5.2.2. Resultados RBC (AG)

Assim como realizado no módulo de RNA, o módulo de RBC por AG foi submetido a vários procedimentos de testes e simulações. Os dados e configurações utilizados para os

processos foram discutidos anteriormente, bem como os métodos aplicados. Após realizados os procedimentos o algoritmo retornou o seguinte gráfico de evolução, vide Gráfico 4.

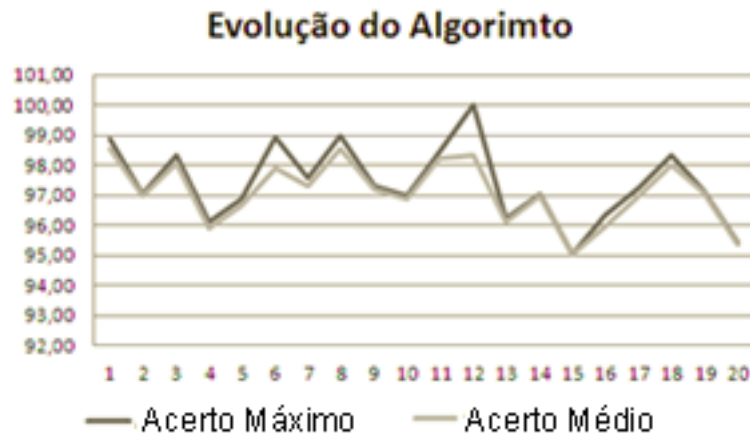


Gráfico 4. Evolução do Algoritmo – Acertos máximos e médios. Fonte: O Autor.

O gráfico acima demonstra a evolução das soluções para cada uma das gerações. A linha mais escura representa o melhor *fitness* de cada geração, enquanto as linhas mais claras representam o *fitness* médio da geração. Nota-se pelo gráfico que o pico do algoritmo ocorre na 12ª geração, após isto o mesmo apresenta variações com tendência negativa, como se pode observar no Gráfico 5. de análise de convergência.

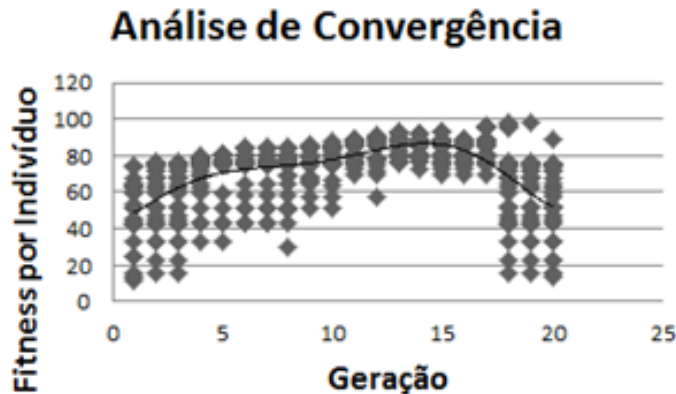


Gráfico 5. Análise de Convergência. Fonte: O Autor.

O gráfico acima evidencia a tendência negativa dos resultados futuros, entretanto não se pode afirmar com total segurança que, a partir da 20ª geração não venha surgir soluções com melhores *fitness*. Entretanto a tendência negativa indicada pela curva de convergência indica uma situação contrária, o que justifica a interrupção dos testes.

Ao final dos testes o algoritmo alcançou picos de acertos de até 99%, entretanto o valor aceito para validação dos resultados foi a taxa média de acertos, que alcançou valores de até 97,01%.

Realizados treinamentos, todos os valores (pesos) gerados por cada um dos indivíduos em cada uma das gerações foram registrados em banco de dados para futuras análises, foram também separados os 9 melhores indivíduos/soluções para comparação com os dados informados pelo grupo de especialistas médicos, resultando desta forma no quadro 5.

Atributo	Ind. 01	Ind. 02	Ind. 03	Ind. 04	Ind. 05	Ind. 06	Ind. 07	Ind. 08	Ind. 09	Média
Idade	0,80	0,80	0,80	0,80	0,80	0,27	0,80	0,80	0,80	0,74
Sexo	0,20	0,20	0,20	0,20	0,20	0,47	0,20	0,47	0,20	0,26
Pressão Sistólica	0,53	0,53	0,60	0,53	0,53	0,53	0,53	0,53	0,53	0,53
Pressão Diastólica	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47
Frequência Cardíaca	0,80	0,80	0,80	0,80	0,80	0,53	0,80	0,80	0,80	0,47
Intensidade da Dor	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,54
Tempo de início do último episódio de dor	0,53	0,53	0,80	0,80	0,53	0,67	0,53	0,00	0,53	0,47
Tempo de início dos Sintomas	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,53
Antecedentes	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,60	0,53	0,47
Característica da dor	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,57
Alterações no Eletrocardiograma	0,53	0,67	0,53	0,53	0,80	0,53	0,53	0,53	0,53	0,47
Local da dor	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,47	0,56
Outros sintomas	0,53	0,53	0,53	0,53	0,53	0,80	0,53	0,53	0,53	0,46
Sinais de ausculta	0,47	0,47	0,40	0,47	0,47	0,47	0,47	0,47	0,47	0,46
Sinais observados	0,60	0,53	0,53	0,53	0,53	0,53	0,53	0,80	0,53	0,56
Fatores de risco	0,47	0,47	0,40	1,00	0,47	0,47	0,47	0,47	0,33	0,50
Sinais de pulso	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,67	0,53	0,53

Quadro 5. Pesos atribuídos pelo AG. Fonte: O Autor.

O quadro acima, como já descrito apresenta valores dos 9 melhores indivíduos encontrados pela solução. Estes indivíduos foram selecionados para que se pudesse calcular as médias de cada atributo com o mesmo número de amostras obtidas através dos pesos atribuídos pelos especialistas. O Quadro 6. apresenta a comparação entre os resultados.

Atributo	Média Especialistas	Média A.G.	Varição
Idade	0,43	0,74	0,31
Sexo	0,45	0,26	0,19
Pressão Sistólica	0,33	0,53	0,20
Pressão Diastólica	0,43	0,47	0,04
Frequência Cardíaca	0,53	0,47	0,06
Intensidade da Dor	0,37	0,54	0,17
Tempo de início do último episódio de dor	0,52	0,47	0,05
Tempo de início dos Sintomas	0,37	0,53	0,17
Antecedentes	0,85	0,47	0,38
Característica da dor	0,63	0,57	0,06
Alterações no Eletrocardiograma	0,61	0,47	0,14
Local da dor	0,76	0,56	0,20
Outros sintomas	0,80	0,46	0,34
Sinais de ausculta	0,52	0,46	0,06
Sinais observados	0,57	0,56	0,01
Fatores de risco	0,56	0,50	0,06

Sinais de pulso	0,31	0,53	0,22
------------------------	------	------	------

Quadro 6. Comparação – Médias dos Pesos atribuídos pelos Especialistas e pelo AG. Fonte: O Autor.

Identificado os pontos de divergência, é possível realizar um levantamento mais preciso dos atributos que merecem maior atenção dos especialistas permitindo inclusive uma discussão com os mesmos para aprimoramento da ferramenta e de seu trabalho.

5.2.3. Validação

Depois de analisados os resultados dos pesos aplicados para cada atributo analisado, foi realizada a comparação da ferramenta em questão, aqui denominada (**AP04**), com outras aplicações de finalidades semelhantes, além do próprio método tradicional.

As comparações que buscavam comprovar a equivalência de qualidade da aplicação apresentada em relação a outros métodos disponíveis na literatura levaram em consideração três ferramentas propostas periódicos internacionais, a saber: (MUTHUKARUPPAN e ER, 2012), apresentando em seu trabalho uma ferramenta híbrida baseada em lógica fuzzy com taxas de até 93,27% de acertos (**AP01**), (KHATIBI e MONTAZER, 2010), também desenvolvendo uma ferramenta híbrida baseada em lógica fuzzy obtiveram taxas de acerto na ordem de 91,58% (**AP02**) e (BAŞÇIFTÇI e INCEKARA, 2012), propuseram uma aplicação online capaz de atingir até 86,1 % de acuidade (**AP03**). A percentagem de acerto dos métodos tradicionais (**TRAD**) foram calculadas com base na base de dados utilizada para o treinamento da aplicação aqui proposta, uma vez que esta base continha não somente os diagnósticos reais, mas também aqueles indicados pelo profissional médico. Na Figura 26. a seguir, pode-se observar os resultados do processo comparativo, onde foram aplicados métodos estatísticos de cálculo de variância (HAIR e JOSEPH, 2007), preparando assim para aplicação de outros métodos que classificam a ferramenta em relação as demais com a quais foi comparada.

	AP01	AP02	AP03	AP04	TRAD
1	89,18	88,37	83,6	96,34	86,38
2	90,53	89,43	85,5	97,85	85,99
3	89,73	90,05	85,3	97,98	85,73
4	92,02	90,87	84,2	95,54	86,21
5	91,98	90,63	86	97,63	87,35
6	93,23	90,58	86,00	96,89	85,33
7	92,15	91,21	86	97,77	87,07
8	93,27	91,47	83,3	96,98	86,98
9	93,01	91,58	84,3	94,99	86,31
10	93,26	91,32	86	97,41	87,02

(a) amostragem

RESUMO				
Grupo	Contagem	Soma	Média	Variância
AP01	10	918,36	91,836	2,306004444
AP02	10	905,51	90,551	1,033587778
AP03	10	850,19	85,019	1,140921111
AP04	10	969,38	96,938	1,046906667
TRAD	10	864,37	86,437	0,429445556

(b) resumo

Figura 26. Comparação entre ferramentas e método tradicional - Variância. Fonte: O Autor.

Nota-se na Figura 26 que, de cada ferramenta/método foram retirados 10 amostras de forma aleatória, sendo as taxas de acerto global de cada um deles apresentado no quadro (a) são apresentadas as amostras retiradas para os métodos comparativos, e aplicados métodos estatísticos para análise de variância obteve-se o resumo apresentado na tabela (b).

Posteriormente submeteu-se os dados a análises de variância (ANOVA) buscando desta forma confirmar ou rejeitar a hipótese de igualdade entre as médias resultando nas tabelas apresentadas na Figura 27.

ANOVA							TUKEY		
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico	Amos	Média	Classif.
Entre grupos	891.9368	4	222.98	187.16571	2.0058E-27	2.5787392	AP04	96.94	A
Dentro dos grupos	53.61179	45	1.1914				AP01	91.84	B
							AP02	90.55	B
Total	945.5486	49					TRAD	86.44	C
							AP03	85.02	C

Figura 27. ANOVA e Teste de Tukey. Fonte: O Autor.

Além de ter sido testado quanto a sua aptidão global em indicar corretamente diagnósticos positivos ou negativos para a enfermidade estudada. Uma vez que o sistema foi calibrado e os pesos de cada um dos atributos distribuídos conforme indicado nos quadros anteriores, submeteu-se o mesmo a novas avaliações, buscando agora medir sua taxa de **especificidade**, **sensibilidade**, **prevalência** e **acurácia**, cuja importância fora outrora evidenciada por Flores (2005), quando em se tratando de análises clínicas e laboratoriais, devendo sempre ferramentas, métodos ou demais operações que façam menção ao diagnóstico médico ou clínico demonstrar os resultados destes testes.

Para análise destes testes, o sistema foi submetido a novas simulações, sendo desta vez selecionados aleatoriamente e utilizados 500 registros retirados da própria base de dados utilizada para treinamento do algoritmo. Tais registros foram selecionados para composição de uma base de simulações, que se encarrega de realizar as entradas para o sistema simulando sua utilização, registrando todos os resultados necessários para realização dos testes citados, gerando após as devidas análises os resultados apresentados na Tabela 3 apresentada abaixo.

Tabela 3. Comparação entre ferramentas e método tradicional - Variância.

Legenda	Valores	Observações
Total de Simulações	500	
Indicações Positivas	323	
Verdadeiros Positivos (VP)	289	
Falsos Positivos (FP)	13	
Indicações Negativas	177	
Verdadeiros Negativos (VN)	162	
Falsos Negativos (FN)	15	
Sensibilidade	95,065%	$VP / (VP + FN)$
Especificidade	92,571%	$VN / (FP + VN)$
Prevalência	63,465	$(VP + FN) / (VP + FN + VN + FP)$
Acurácia	96,868	$(VP + VN) / (VP + FN + VN + FP)$

Fonte: O Autor.

Ainda seguindo os conceitos apresentados por Flores (2005), pode-se afirmar que os testes realizados permitem: definir a real capacidade apresentada pela ferramenta em apresentar resultados positivos, para pacientes que realmente apresentam a cardiopatia isquêmica (**sensibilidade**); a **especificidade**, refere-se a medição da capacidade da ferramenta em indicar e diagnosticar um verdadeiro negativo, ou seja, indicar que um indivíduo está sadio quando este realmente não apresenta a cardiopatia isquêmica; a quantidade de casos testados dentro de todos selecionados para as simulações é dado pela **prevalência**; por fim a veracidade do teste, ou seja o quanto o teste se aproxima ou se afasta da “verdade” é dado pela **acurácia**.

Todos os resultados obtidos pela ferramenta, no que se refere aos testes acima citados, foram confrontados também com resultados obtidos através de diferentes métodos e ferramentas também destinadas a este fim (ALMEIDA, 2006), os quais cita-se: **Análise Linear de Fisher (Fisher)**, **Árvores de Decisão (AD)**, **Regressão Logística (RL)** e **Redes Neurais Artificiais (RNA)**, além do método aqui proposto, **Algoritmos Genéticos com Reconhecimento Baseado em Casos (AG+RBC)**, sendo os resultado dos mesmos disponibilizados no Gráfico 6.

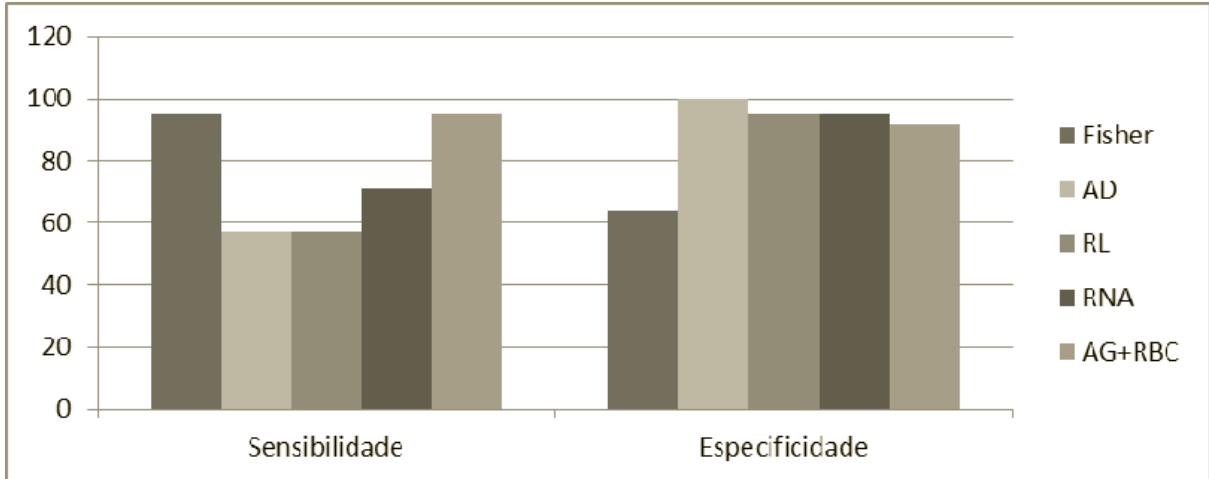


Gráfico 6. Comparação de desempenho sensibilidade e especificidade. Fonte: Adaptado de Almeida (2006)

Pode-se, pelo Gráfico 6 notar que o método aqui proposto apresenta taxas de sensibilidade bem próximas às apresentadas pelo modelo linear de Fisher, que, dos métodos em questão foi o que apresentou a melhor taxa. Quanto a especificidade as taxas ficaram bem próximas dos modelos de Árvores de Decisão, Regressão Logística e Redes Neurais Artificiais, que foram os métodos com melhores resultados, demonstrando assim que a ferramenta aqui proposta apresenta resultados que se enquadram no grupo das soluções (encontradas na bibliografia a que se teve acesso) que apresentam os mais expressivos resultados.

6. CONSIDERAÇÕES FINAIS

6.1. Conclusões

O principal objetivo do trabalho é a apresentação de uma ferramenta de suporte a decisão específica para o diagnóstico da cardiopatia isquêmica, com possibilidades de adaptação e aprendizado para que seja utilizada no diagnóstico de outras enfermidades.

Esta ferramenta é apresentada como uma opção às soluções existentes, na qual é aplicada uma proposta híbrida, unindo características da computação evolutiva (Algoritmos Genéticos) e aprendizado de máquina (Raciocínio Baseado em Casos e Redes Neurais Artificiais). Desta forma, além de predizer diagnósticos através da análise de dados clínicos o sistema também se tornou capaz de analisar imagens de sinais de eletrocardiogramas, oferecendo assim aos profissionais médicos mais informações para suporte à decisão.

Dividia a análise da proposta em módulos, aquele em que foi aplicada computação evolutiva e aquele em que foi aplicado RNA, pode-se concluir que: os resultados apresentados nos experimentos realizados (considerando somente o módulo de computação evolutiva) mostra que a aplicação proposta apresenta taxas de acertos sensivelmente superiores às demais aplicações com as quais foi comparada, mostrando-se ainda mais eficiente quando comparadas aos métodos tradicionais de diagnóstico.

O módulo de RNA apresenta taxas menores, podendo se enquadrar nas mesmas taxas de acertos dos métodos tradicionais, entretanto, deve-se considerar que, o módulo de RNA apresenta resultados inferiores quando utilizado somente análise das imagens de sinais de ECG's como entradas, excluindo-se os demais achados das análises clínicas, levando-se a crer que, a ferramenta possui grande capacidade de classificação e predição diagnóstica mesmo que com número reduzido de informações, bastando para tal que essas informações tenham alto grau de relevância, o que também é computado pela ferramenta.

Dados os resultados das análises estatísticas realizadas para comprovação dos bons resultados apresentados pela aplicação proposta pode-se perceber as grandes vantagens obtidas através de pequenas adaptações realizadas nos algoritmos e a seleção de uma função de similaridade adequada ao problema em questão, levando a aplicação a resultados expressivos na indicação da possibilidade da presença da doença obstrutiva coronariana nos casos analisados, bem como boas taxas de especificidade e sensibilidade. Desta forma, baseado nos resultados demonstrados no capítulo 5, a ferramenta aqui proposta enquadra-se em níveis de qualidade e capacidade de acertos observados nas principais ferramentas de mesma finalidade observadas na literatura.

6.2. Trabalhos Futuros

Como trabalhos futuros sugerimos primeiramente a revisão do algoritmo proposto e definição de um sistema inteligente que seja capaz de ao longo do tempo armazenar e manter em sua base de dados principal somente as melhores soluções para um dado problema, evitando assim o crescimento desordenado dessa base de conhecimentos que pode trazer ao sistema, ao longo do tempo uma considerável queda em seu desempenho, não melhorando significativamente os resultados apresentados, uma vez que esta base estará composta não somente de bons, mas também maus resultados.

Propõe-se também a formação de outras bases de conhecimento e adaptação do algoritmo realização de treinamentos em outras patologias de difíceis diagnósticos, confirmando desta forma o poder adaptativo da solução aqui proposta. Por fim a criação de um sistema inteligente de seleção de casos para teste baseado em princípios estatísticos que impeça o algoritmo de realizar comparações desnecessárias sem, no entanto comprometer sua qualidade.

Finalizando com a adaptação da ferramenta para que a mesma seja projetada para ser uma solução embarcada em sistemas de leituras de sinais cardíacos, e mesmo possibilitando sua utilização em dispositivos móveis alertando a seus usuários sobre a necessidade de consultarem seus médicos cardiologistas.

7. REFERÊNCIAS

ALMEIDA, S. M. Algoritmo Genético para Determinação de Relevância de Sintomas no Diagnóstico de Cardiopatia Isquêmica, Brasília, 2006.

ALVES, W. J. B. Reconhecimento de Pessoas por meio de Algoritmo Genético Aplicado em Medidas das Proporções Áureas da Face Humana. **FPeO/Fafibe**, 2009.

ARCELLI, C.; BAJA, G. S. A one-pass two-operation processo to detect the skeletal pixels on the 4-distant transform. **IEEE transactions on patterns analysis and machine intelligence**, p. 411,414, 1989.

ARCELLI, C.; CORDELLA, L. P.; LEVIALDI, S. From local maxima to connected skeletons. **IEEE transaction os pattern analysis and machine intelligence**, p. 134-143, 1981.

BARBOSA, A. K. P. et al. HealthNet: um Sistema Integrado de Telediagnóstico e Segunda Opinião Médica. **Boletim Bimestral Sobre Tecnologia de Redes.**, 2001.

BARBOSA, A. K. P.; NOVAES, M. A.; VASCONCELOS, A. M. L. A Web Application to Support Telemedicine Services in Brazil. **Proceedings of AMIA Sumposium, Washington - USA**, p. 56-60, 2003.

BARONE, D. **Sociedades Artificiais: A Nova Fronteira da Inteligência das Máquinas**. Porto Alegre: Bookman, 2003.

BARR, A.; FEIGENHAUM, E. A. **The Handbook of Artificial Intelligence**. California: William Kaufmann inc - Volume I-II, 1981.

BARROS, E. M.; WANGENHEIM, A. V. Sala de Laudo Virtual: Um Ambiente de Teleradiologia para Diagnóstico Cooperativo via Internet. **TELEMED 2000**, 2000.

BAŞÇIFTÇI, F.; INCEKARA, H. Design of Web-based Fuzzy Input Expert System for the analysis of serology laboratory tests. **Journal Of Medical System**, p. 2187-2191, 2012.

BERGMANN, R. **Experience Management: Foundations, Development Methodology, and Internet-Based Applications**. 1ª Edição. Berlin: Springer, 2002.

BIGUS, J. P. **Data Mining With Neural Networks: Solving Business Problems From Application Development to Decision Support.** [S.l.]: McGraw-Hill, 1996.

BOMER, K. Structural Similarity as Guidance in Case-Based Design. **Topics in Case-Based Reasoning**, p. 197-208, 1994.

CARVALHO, A. C. P. L. F.; LACERDA, E. G. Introdução aos Algoritmos Genéticos. **Anais do XIX Congresso Nacional da Sociedade Brasileira de Computação**, 15 jul. 1999. Vol. 2. p. 51-126.

CASTRO, L. N. D. The Clonal Selection Algorithm With Engineering Applications. **Workshop on Artificial Immune Systems**, p. 36-37, 2000.

CHARNIAK, E.; MCDERMOTT, D. **Introduction to Artificial Intelligence.** MA, USA: Addison-Wesley Publishing Company, 1985.

CHAVES, P. C.; MOREIRA, A. L. **Eletrocardiograma - Texto de Apoio.** Porto: Univerdidade do Porto, 2001.

CHECLAND, P. B. **Systems Thinking, Systems Practice.** Chichester: John Wiley, 1981.

CHORRO, F. J.; CIVERA, R. G.; MERINO, V. L. **Cardiologia Clínica.** Valência: Universitat de Valência, 2007.

DASGUPTA, D. **Artificial Immune Systems and Their Applications.** New York: Secaucus, 1998.

DAVENPORT, T. H.; PRUSAK, L. **Conhecimento Empresarial: Como as Organizações Gerenciam o seu Capital Intelectual.** Rio de Janeiro: Campus, 1998.

DETRANO, R. E. A. International application of a new probability algorithm for the diagnosis of coronary artery disease. **American Journal of Cardiology**, p. 304-310, 1989.

DORIGO, M. V.; MANIEZZO; COLOMI, A. The Ant System: Optimization by a Colony of Cooperating Agents. **IEEE Transactions on Systems, Mans and Cybernetics Part B: Cybernetics**, p. 29-41, 1996.

EIBEN, A. E.; SMITH, J. E. **Introduction Evolutionary Computing.** [S.l.]: SpringerVerlag, 2003.

FLORES, R. E. **A medicina baseada em evidências e o diagnóstico laboratorial.** NewsLab. São Paulo - SP, v. 73: [s.n.]. 73 2005. p. 92-103.

GLAUCIA, R. M. A. S. Método Fuzzy para Auxílio ao Diagnóstico de Câncer de Mama em Ambiente Inteligente de Telediagnóstico Colaborativo para Apoio à Tomada de Decisão, Natal, 2012.

GOMES, R. C. Ministério da Saúde. **Portal Brasil**, p. Gomes, R. C. Doenças Cardiovasculares Causam quase 30% das Mortes no País. No Brasil, homens são os mais afetados pelas doenças cardiovasculares, como infarto e AVCs. Ministério da Saúde, 2011. Disponível em <<http://www.brasil.gov.br/saude/2011/09/doenca>, 06 set. 2011. Disponível em: <<http://www.brasil.gov.br/saude/2011/09/doencas-cardiovasculares-causam-quase-30-das-mortes-no-pais>>. Acesso em: 22 abril 2014.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing.** [S.l.]: Addison-Wesley Publishing Company, Inc, 1992.

HAIR, G.; JOSEPH, F. **Análise Multivariada de Dados**. [S.l.]: Bookman, 2007.

HAMPTON, J. R. **The Importance of Minor Abnormalities in the Resting Eletrocardiogram**. Suíça: Euro Heart, 1984.

HOLLAND, J. H. **Adaptation in Natural and Artificial Systems**. Michigan: University of Michigan, 1975.

HSU, Y. et al. Design and Implementation of Teleconsultation System of Instant Treatment. **Bioinformatics and Biomedical Engineering - ICBEE 2008. The 2nd International Conference on. Shangay, China.**, p. 1355-1358, 2008.

JARA, A. J. et al. **An Antology and Rule Based Intelligent Information System to Detect and Predicted Myocardiol Diseases**. Information Technology and Applications in Biomedicine. ITAB 2009, 9th International Conference on.. Lamarca, Chipre: [s.n.]. 2009. p. 1-6.

KANNEL, W. B. et al. Nonespecific Eletrocardiographic Adnormality as a Predictor of Coronary Heart Disease. **The Framingham Study am Heart**, p. 113-370, 1987.

KASSIRER, J. P.; GORRY, G. A. Problem Solving: a Behavioral Analysis. **Analns of Internal Medicine**, 1978. Vol. 89, n.1 p. 245-255.

KENNEDY, J.; EBERHART, R. Particle Swarm Optimization. **In Proceedings of the IEEE International Conference on Neural Network**, p. 1942-1948, 1995.

KHATIBI, V.; MONTAZER, G. A. A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment. **Expert Systems with Applications**, p. 8536-8542, 2010.

KNUSTSEN, R. et al. The predictive value of resting electrocardiograms for 12-year incidence of coronary heart disease in the Honolulu Heart Program. **J Clin Epidemiol**, p. 293-302, 1988.

LEITE, C., R. M. et al. **Classification of Cardiac Arrhythmias Using Competitive Networks**. Proceeding of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society - EMBC2010. Buenos Aires, Argentina: [s.n.]. 2010.

LEYMARIE, F.; LEVINE, M. D. Simulating the rassfire transform using an active contour model. **IEEE transactions os pattern analysis and machine intelligence**, p. 56-75, 1992.

LIAO, T. W.; ZHANG, Z.; MOUNT, C. R. Similarity Measures for Retrieval in Case-Based Reasoning Systems. **Applied Artificial Intelligence**, p. 267-288, 1998.

MACKAY, D. J. C. **Information Theory. Inference and Learning Algorithms**. Cambridge : Cambridge University Press, 2002.

MALTA, D. C.; MERHY, E. E. Avaliação do Projeto Vida e do Acolhimento no Sistema Único de Saúde de Belo Horizonte. **REME Rev Min Enfermagem**, p. 259-267, 2004.

MANSUR, A. D. P.; FAVARATO, D. Mortalidade por doenças cardiovasculares no Brasil e na região metropolitana de São Paulo. **Arquivos Brasileiros de Cardiologia**, 05 set. 2012. 755-761.

MCCARTHY, J.; HAYES, P. J. Some philosophical problems from the standpoint of artificial intelligence. In: MICHIE, D.; MALTZER, B. **Machine Intelligence**. Edinburgh: Edinburgh University Press, 1969. p. 463-502.

MCCORDUCK, P. **Machines Who Think**. San Francisco: Freeman, 1979.

MEJIA, S. et al. Antioquia's Telemedicine Network: Informatic and Telecommunication Technologies to Service of Health. **Engineering in Medicine and Biology Society. Proceedings of the 25th Annual International Conference of IEEE**, p. 3649-3652, 2003.

MENOTTI, A.; SECCARECCIA, F. Eletrocardiographic Minnesota code Findings Predicting Short Term Mortality in Asymptomatic Subjects. **The Italian RIFLE Pooling Project. G Ital Cardiol**, p. 27:40-9, 1997.

MINSKY, M. L.; PAPERT, S. A. **Perceptrons: An Introduction to Computational Geometry**. Cambridge, MA, USA: MIT Press, 1969.

MO, Y.; XU, S. **Application of SVM Based on Hybrid Kernel Function in Heart Disease Diagnoses**. Intelligent Computing and Cognitive Informatics (ICICCI). Malaysia: [s.n.]. 2010. p. 462-465.

MORESI, A. D. Gestão da Informação e do Conhecimento. In: TARAPANOFF, K. **Inteligência Organizacional e Competitiva**. Brasília: Universidade de Brasília, 2001.

MORESI, E. A. D. **Manual de Metodologia da Pesquisa**. Brasília: Universidade Católica de Brasília, 2003.

MUTHUKARUPPAN, S.; ER, M. J. A Hybrid Particle Swarm Optimization Based Fuzzy Expert System for the Diagnosis of Coronary Artery Disease. **Expert System Applications**, p. 11657-11665, 2012.

NILSON, N. S. **Principles of Artificial Intelligence**. Berlin: Springer Verlag, 1982.

NONAKA, I. M.; TAKEUCHI, H. **Criação de Conhecimento na Empresa**. Rio de Janeiro: Campus, 1997.

OMS. World Health Statistics. **World Health Organization**, 22 mar 2014. Disponível em: <http://apps.who.int/iris/bitstream/10665/112738/1/9789240692671_eng.pdf?ua=1>. Acesso em: 23 jun 2014.

PACHECO, M. A. C. **Algoritmos Genéticos: Princípios e Aplicações**. [S.l.]: Departamento de Engenharia Elétrica. Pontifícia Universidade Católica do Rio de Janeiro, 1999.

PAL, S. K.; SIMON, C. K. S. **Foundations of Soft Case-Based Reasoning. 1ª Ed.** [S.l.]: John Wiley & Sons, 2004.

PRATI, R. C. Novas Abordagens em Aprendizado de Máquina para Geração de Regras, Classes Desbalanceadas e Ordenação de Casos, São Carlos, 2006.

RABELO, J. R. . E. A. Um Sistema Especialista para Diagnóstico de Cardiopatias Isquêmicas. **Informédica**, p. 5-11, 1993.

REZENDE, S. O. **Sistemas Inteligentes: Fundamentos e Aplicações**. Barueri: Manole, 2003.

- ROSE, G. A. et al. **Metodos de encuesta sobre enfermedades cardiovasculares**. Genebra: Organizacion Mundial de la Salud, 1982.
- RUSSELL, S.; NORVING, P. **Inteligência Artificial 2º ed**. Rio de Janeiro: Elsevier, 2004.
- SCHANK, R. C. **Dynamic Memory: A Theory of Learning in Computers and People**. New York: Cambridge University Press, 1982.
- SETIAWAN, N. A.; VENKATACHALAM, P. A.; HANI, A. F. M. Diagnosis of Coronary Artery Disease Using Artificial Intelligence Based Decision Support System. **Proceedings of The Internacional Confernce of Man-Machine Systems (ICoMMS)**, 2009.
- SILVA, R. et al. Lapidus On Line: Sistema de Apoio à Decisão Médica na Internet. **Anaos do VIII Congresso Brasileiro de Informática em Saúde - CBIS'02**, 2002.
- SILVA, R. P. Modelo de Apoio ao Diagnóstico Médico Aplicado ao Raciocínio Baseado em Casos, Brasília, 2005.
- SIMOUDIS, E. Using Case-Based Retrieval For Customer Technical Support. **IEEE Expert**, p. 7-13, 1992.
- SONI, J. E. A. Intelligent and Effective Heart Disease Prediction System Using Weighted Associative Classifiers. **International Journal on Computer Science And Engineering (IJCSE)**, 2011.
- VERGARA,. **Projetos e relatórios de pesquisas em administração**. São Paulo: Atlas, 2000.
- WINSTON, P. H. **Artificial Intelligence (2nd Edition)**. MA, USA: Addison-Wesley Publishing Company, 1985.