



III Simpósio Internacional de Gestão de Projetos (III SINGEP)
II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)
ISSN:2317-8302

APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL NO AUXÍLIO AO DIAGNÓSTICO DO RISCO DE DESENVOLVER CÂNCER COLORRETAL

ELENICE GONÇALVES SOUSA CONCEIÇÃO
UNINOVE – Universidade Nove de Julho

EDQUEL BUENO PRADO FARIAS
UNINOVE – Universidade Nove de Julho

RENATO JOSÉ SASSI
UNINOVE – Universidade Nove de Julho



APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL NO AUXÍLIO AO DIAGNÓSTICO DO RISCO DE DESENVOLVER CÂNCER COLORRETAL

Resumo

O Câncer Colorretal é uma neoplasia que se origina na camada superficial do revestimento intestinal e com o tempo vai atingindo as camadas mais profundas. Os sinais e sintomas relacionados a esse tipo de câncer são variados e pouco específicos. Em muitos momentos não é possível dizer se esses sintomas significam ou não que um paciente tenha risco de desenvolver a doença, o que pode retardar a descoberta e o correto encaminhamento para o tratamento. Técnicas de Inteligência Artificial como: a Teoria dos *Rough Sets* (RS) (que é um modelo matemático usado na redução de atributos), a Rede Neural Artificial (RNA) (que é um modelo matemático inspirado no funcionamento do cérebro humano) e o Sistema Especialista (SE) (capaz de emitir decisão tal qual um especialista humano de determinada área do conhecimento) podem auxiliar no diagnóstico. O objetivo deste trabalho foi aplicar técnicas de Inteligência Artificial no auxílio ao diagnóstico do risco de desenvolver câncer colorretal. Os resultados levaram a concluir que a aplicação de técnicas de Inteligência Artificial pode auxiliar no diagnóstico do risco de desenvolver câncer colorretal.

Palavras-chave: *Inteligência Artificial, Câncer Colorretal, Teoria dos Rough Sets, Sistemas de Apoio à Decisão, Sistema Especialista.*

Abstract

The colorectal cancer is a neoplasia that originates on the surface layer of the intestinal lining and, with the time, reaches the deeper layers. The signs and symptoms related to this cancer are varied and unspecific. In many situations it is not possible to say if these symptoms mean that a patient is at risk for developing the disease or not, which can delay the discovery and correct referral for treatment. Artificial Intelligence techniques such as: the theory of Rough Sets (RS) (which is a mathematical model used in the reduction of attributes); the Artificial Neural Network (ANN) (which is a mathematical model inspired by the workings of the human brain) and the Expert System (ES) (which is able to issue a decision as a human expert in a particular field of knowledge) can help in the diagnosis. The aim of this study was to apply Artificial Intelligence techniques to help to diagnose the risk of developing colorectal cancer. The results led to the conclusion that the application of Artificial Intelligence techniques can help in diagnosing the risk of developing colorectal cancer.

Keywords: *Artificial Intelligence, Colorectal Cancer, Rough Sets Theory, Decision Support Systems, Expert System.*



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

1 Introdução

Albertin e Moura (2004) afirmam que a Tecnologia da Informação (TI) tem sido considerada como um dos componentes mais importantes do ambiente empresarial atual, sendo que as organizações brasileiras têm utilizado ampla e intensamente esta tecnologia, tanto em nível estratégico como operacional.

Segundo Carneiro (2009) a TI na Medicina tem como objetivo fornecer informação para o médico, hospital, laboratório ou clínica, de modo a melhorar a qualidade do serviço aumentando a competitividade do negócio.

Um Sistema de Informação hospitalar é definido por Johanston (1993) como um sistema computadorizado que, instalado em um ambiente hospitalar, objetiva registrar informações sobre os pacientes de tal forma que possam ser compartilhadas por todos os setores do hospital que delas necessitem.

Sistemas de Suporte a Decisão Clínica (SSDC) são definidos como aplicações informáticas que auxiliam médicos e profissionais de saúde na tomada de decisão, por meio de acesso oportuno a conhecimento médico armazenado eletronicamente, de forma a melhorar as suas práticas médicas. (SARTIPI et al, 2011).

Sistemas de suporte a decisão servem como uma segunda opinião ao clínico no momento do diagnóstico. A utilização dessa tecnologia ajuda na prevenção do diagnóstico errado e no correto encaminhamento de pacientes que apresentam risco de alguma enfermidade, principalmente quando não fica muito clara ao clínico a situação do paciente somente pela análise de seus sintomas.

O futuro do diagnóstico médico está se movendo em novas áreas, utilizando técnicas como a mineração de dados e a Inteligência Artificial (IA). Essas técnicas estão sendo estudadas, melhoradas e desenvolvidas. Esta área de pesquisa é de grande importância, pois causa mudanças diretas na vida das pessoas. Ao melhorar a precisão e o tempo com que o câncer é diagnosticado torna-se mais fácil eliminá-lo antes que ele se agrave e atinja um estágio em que pouco pode ser feito.

Na busca para minimizar dúvidas durante o diagnóstico médico surgem os Sistemas de Apoio a Decisão, que tem como objetivo reduzir erros e incertezas no momento da tomada de decisão.

Aproveitando os recursos da informática, os sistemas automatizados de apoio à decisão possuem a capacidade de incorporar e representar uma enorme quantidade de informação médica e de codificar estratégias de seleção e decisão que levem a respostas úteis para o processo de tomada de decisão de um profissional de saúde. (PISA et al, 2004)

Diversas técnicas da Inteligência Artificial podem ser utilizadas, como a Teoria dos Rough Sets, as Redes Neurais Artificiais (RNAs) e os Sistemas Especialistas (SEs), onde os problemas são resolvidos da mesma maneira que um especialista humano resolveria.

Este trabalho tem como objetivo utilizar as técnicas de Inteligência Artificial para desenvolver uma solução que provê uma confiável fonte de consulta aos clínicos no momento da tomada de decisão relacionada ao câncer colorretal.



2 Referencial Teórico

2.1 Câncer

O câncer é definido como um tumor maligno, mas não é uma doença única e sim um conjunto de mais de 200 patologias, caracterizado pelo crescimento descontrolado de células anormais (malignas) e como consequência ocorre a invasão de órgãos e tecidos adjacentes envolvidos, podendo se disseminar para outras regiões do corpo, dando origem à tumores em outros locais. Essa disseminação é chamada de metástase. (ABRALE, 2014)

O câncer é a principal causa de morte nos países economicamente desenvolvidos e a segunda principal causa de morte em países em desenvolvimento. (WORLD HEALTH ORGANIZATION, 2008) A incidência do câncer está aumentando nos países em desenvolvimento economicamente como resultado do crescimento e envelhecimento da população e também, cada vez mais, com a adoção de estilos de vida associados ao câncer, incluindo tabagismo, sedentarismo e dietas “ocidentalizadas”. (JEMAL, 2011).

Todos os dias novas pessoas são diagnosticadas com algum tipo de câncer, muitas já tardiamente, o que reduz em muito as chances de sucesso nos tratamentos e aumenta significativamente as chances de morte do paciente. O diagnóstico precoce é essencial para que o tratamento se inicie o mais breve possível, aumentando assim as chances de sucesso e recuperação do paciente.

2.1.1 Câncer colorretal

O câncer colorretal é uma neoplasia que se origina de qualquer porção do colón, reto ou canal anal. A doença começa na camada superficial do revestimento intestinal e com o tempo vai atingindo as camadas mais profundas. (HOSPITAL DO CÂNCER DE BARRETOS, 2014).

O câncer do intestino grosso (câncer colorretal) é um dos tipos com maior incidência em todo o mundo, principalmente nas regiões mais desenvolvidas. No Brasil é o segundo tumor mais incidente em mulheres e o terceiro em homens (exceto os casos de câncer de pele não melanoma). São estimados cerca de 32.600 novos casos em 2014. (A.C. CAMARGO, 2014).

Os sinais e sintomas relacionados ao câncer colorretal são variados e pouco específicos. Os mais comuns são: presença de sangue ou muco nas fezes, alteração do hábito intestinal e dores abdominais ou pélvicas. A anemia costuma estar associada a tumores do cólon direito e lesões no reto podem causar puxos e tenesmos. Durante a avaliação clínica a coleta de história familiar detalhada é de grande importância para não se negligenciar o diagnóstico de uma síndrome hereditária. (ROSSI et al, 2010).

De acordo com o INCA (2013) o câncer é uma patologia com localizações e aspectos clínico-patológicos múltiplos e não possui sintomas ou sinais patognomônicos, podendo ser detectado em vários estágios de evolução histopatológica e clínica. Destes fatos resulta, em grande parte, a dificuldade do seu diagnóstico e a afirmativa de que a suspeita de câncer pode surgir diante dos sintomas os mais variados possíveis. O paciente, ao procurar um médico, não sabe ainda a natureza da sua doença e, assim, não procura diretamente um especialista. Setenta por cento dos diagnósticos de câncer são feitos por médicos não-cancerologistas, o que evidencia a importância destes profissionais no controle da doença.

O médico chega a uma suposição diagnóstica através de várias etapas, durante as quais deve proceder a uma análise cuidadosa, com base principalmente em seu conhecimento do



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

caso e da patologia, olhando sempre o paciente como um todo, não se restringindo ao sistema-alvo da sua especialidade. (INCA, 2013)

O SSDC oferece aos médicos, funcionários, pacientes ou a outros indivíduos conhecimento e informações específicas da pessoa, filtrada de forma inteligente ou apresentados em momentos apropriados, para melhorar a saúde e assistência médica. Isto engloba uma variedade de ferramentas e intervenções como alertas e lembretes computadorizados, orientações clínicas, aparelhos de ordem, painéis e relatórios de dados de pacientes, modelos de documentação, apoio ao diagnóstico e ferramentas de fluxo de trabalho clínico. SDC tem sido eficaz na melhoria dos resultados em algumas instituições de saúde e locais de prática médica por tornar o conhecimento médico necessário prontamente disponível para os usuários do conhecimento. (OSHEROFF et al, 2006, tradução nossa)

As ferramentas de suporte a decisão clínica não substituem a opinião e conhecimento de um especialista, nem dá menos importância ao papel do médico. Essas ferramentas suportam a decisão criando uma situação mais confiável onde os especialistas podem confirmar seus diagnósticos ou dúvidas a respeito de um paciente de maneira mais precisa.

2.2 Inteligência Artificial

Para Russell e Norvig (1995) os principais objetivos da IA são desenvolver métodos e sistemas para resolver problemas, geralmente resolvido pela atividade intelectual dos seres humanos, por exemplo, o reconhecimento da imagem língua, processamento da fala, planejamento e previsão, sistemas de informação, reforçando assim de computador, e para desenvolver modelos que simulam os organismos vivos e o cérebro humano, em especial, melhorando assim a nossa compreensão de como o cérebro humano funciona.

2.2.1 Teoria dos *Rough Sets*

A Teoria dos Rough Sets (RS) foi proposta por Zdzislaw Pawlak (PAWLAK, 1982) em 1982 como um novo modelo matemático para representação do conhecimento e tratamento de incerteza, tendo sido usada, posteriormente, para o desenvolvimento de técnicas para classificação aproximada em aprendizado de máquina. Devido a essas características, tem-se utilizado essa teoria em IA, especialmente nas áreas de aquisição de conhecimento, raciocínio indutivo e descoberta de conhecimento em base de dados.

Os conceitos de RS têm se mostrado muito úteis quando aplicados a problemas do tipo: redução de atributos, descoberta de dependência entre atributos e na descoberta de padrões entre os dados (PAWLAK, 1996).

A redução de atributos em RS é feita através dos chamados Redutos (RED), que são subconjuntos de atributos capazes de representar o conhecimento da base de dados com todos os seus atributos iniciais (PAWLAK, 1982). Este procedimento de eliminação de atributos irrelevantes é uma das características da Teoria.

2.2.2 Mapas Auto-Organizáveis

As Redes Neurais Artificiais (RNAs) são uma classe especial de sistemas modelados seguindo analogia com o funcionamento do cérebro humano, sendo formadas por neurônios artificiais conectados de maneira similar aos neurônios do cérebro humano (GOEBEL E GRUENWALD, 1999).



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

Uma de suas principais vantagens é sua variedade de aplicação, mas em compensação, os seus dados de entrada são difíceis de serem formados e os modelos produzidos são difíceis de entender (HARRISON, 1998).

De acordo com Farias (2012) os principais modelos de RNAs são: Modelos de Hopfield, Bam e ART, Modelo RBF, Modelos Recorrentes (Perceptron) e Modelo de Kohonen (SOM).

Um Mapa Auto-Organizável (Self-Organizing Maps, ou SOM) é uma arquitetura de rede neural artificial, com aprendizado não supervisionado, baseada em um mapa de neurônios, cujos pesos são adaptados para verificar padrões semelhantes em relação a um conjunto de treinamento (KOHONEN, 2001). Sua principal característica é o mapeamento ordenado dos padrões de entrada de elevada dimensão em reticulados de neurônios de saída com dimensão menor, comumente duas, o que facilita a visualização dos dados.

Para uma dada base de dados com N amostras com d atributos cada, em que d determina a dimensão dos padrões de entrada, ocorrerá um mapeamento desses padrões para um reticulado de neurônios de saída arranjados em 2D.

A rede SOM é uma arquitetura de rede neural artificial, estruturada em duas camadas, entrada e saída. Os neurônios da camada de saída são comumente dispostos em um mapa de duas dimensões, com uma dada relação de vizinhança.

A Figura 1 ilustra essa arquitetura, com d atributos na camada de entrada e um conjunto de unidades u (neurônios) arranjados na forma de um mapa em 2D na camada de saída. Cada u é caracterizado por sua posição x e y no mapa, que é representado por $u=[u_x, u_y]$, respectivamente, que resulta em um vetor 2D igual a $u=[u_x, u_y]$. Cada u tem associado um vetor protótipo $m_u = [m_{1u}, m_{2u}, \dots, m_{du}]$, sendo d a dimensão do protótipo, a mesma do padrão de entrada.

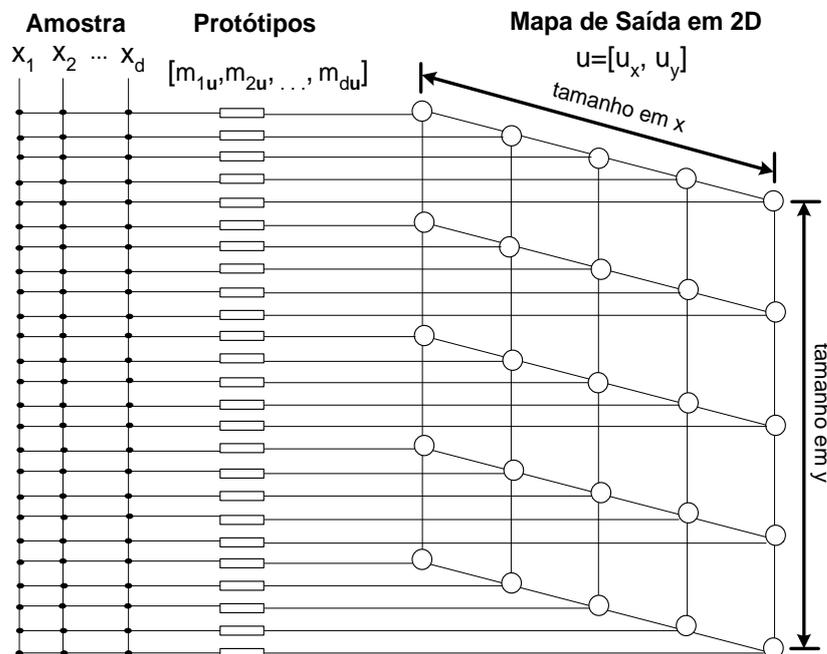


Figura 1 - Ilustração de uma rede SOM (2D). Fonte Kohonen (1982)

O algoritmo de aprendizado da rede SOM é realizado em um processo iterativo, onde, no primeiro passo, $t=0$, inicializa o vetor protótipo (m) randomicamente. Porém, a inicialização do m , pode ser feita de outras maneiras (KOHONEN, 2001).

O algoritmo de treinamento da rede SOM é também chamado de competitivo. Em cada passo do processo (ou época), uma amostra x é randomicamente escolhida do conjunto



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

de treinamento. A distância, geralmente euclidiana, entre x e todos os vetores protótipos m é calculada. A unidade com menor distância, chamada de BMU (best-matching unit) é o u com protótipo m mais próximo a x , conforme a Equação 1:

$$\|x - mbmu\| = \arg \min_u \|x - mu\| \quad (1)$$

A seguir, os vetores protótipos são atualizados. O BMU e sua vizinhança topológica são movidos para próximos a x , como se fosse um “arrasto”, que ilustra a atualização do neurônio vencedor (BMU) e sua vizinhança em direção a x . Os círculos em preto e cinza correspondem às situações antes e depois da atualização, respectivamente. As linhas mostram a relação da vizinhança.

A regra para a atualização dos vetores protótipos da unidade u é dada pela Equação 2:

$$m_i(t+1) = m_i(t) + \alpha(t) h_{bi}(t) [x - m_u(t)] \quad (2)$$

Onde t é o número de épocas, $\alpha(t)$ é a taxa de aprendizado e $h_{bi}(t)$ é o kernel da vizinhança centrado no neurônio vencedor. O kernel pode ser Gaussiano, como na Equação 3:

$$h_{bi}(t) = e^{-\frac{\|r_b - r_i\|^2}{2\sigma^2(t)}} \quad (3)$$

Onde r_b e r_i são as posições do neurônio vencedor b e do neurônio i no mapa da rede SOM e $\sigma(t)$ é o raio da vizinhança. Conforme a distância (b,i) aumenta e o número de épocas (t) aumenta, $h_{bi} \rightarrow 0$. A taxa de aprendizado $\alpha(t)$ e o raio da vizinhança $\sigma(t)$ diminuem monotonicamente com o tempo.

Devido às características da rede SOM de capacidade de quantização vetorial e de projeção vetorial, ele também pode ser utilizado na análise dos dados (KASKI & KOHONEN, 1996), (CURRY, 2003). A quantização vetorial é feita com a projeção de N amostras de entrada para m protótipos, que representam todo o conjunto de dados original.

A partir dos protótipos, realizam-se a formação de clusters e visualização das amostras em duas dimensões (JIN, SHUM, LEUNG et al; 2004).

O primeiro passo da rede SOM na análise de dados é a redução do conjunto de registros para os protótipos, os quais são utilizados no agrupamento ou na visualização dos dados.

A motivação para o uso dos protótipos é que a complexidade computacional do passo subsequente, por exemplo, o agrupamento de dados, é reduzido.

Quando utilizado na tarefa de clusterização, a rede SOM, em seu processamento das amostras, descarta ruídos com média zero e efeito de amostras discrepantes (outliers) na geração dos vetores protótipos (VESANTO e ALHONIEMI, 2000). Com os vetores protótipos, outros algoritmos de clusterização podem ser utilizados, como a própria rede SOM ou o K-médias. Um estudo comparativo entre a rede SOM e K-médias é apresentado por (ULTSCH, 1995).

2.2.3 Sistemas Especialistas

Sistemas Especialistas (SEs) são programas de computador que procuram atingir soluções de determinados problemas do mesmo modo que se supõe que os especialistas humanos resolvam se estiverem sob as mesmas condições. Apesar das limitações das



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

máquinas, é possível, atualmente, a construção de Sistemas Especialistas com alto grau de desempenho, dependendo da complexidade de sua estrutura e do grau de abrangência desejado (LIA1997).

Segundo RIBEIRO (1987) um Sistema Especialista (SE) é aquele que é projetado e desenvolvido para atender a uma aplicação determinada e limitada do conhecimento humano. É capaz de emitir uma decisão, com apoio em conhecimento justificado, a partir de uma base de informações, tal qual um especialista de determinada área do conhecimento humano.

Para GOLDSCHMIDT (2010) são geralmente desenvolvidos para atender a uma aplicação determinada e limitada do conhecimento humano. São também, capazes de emitir uma decisão e flexíveis para incorporação de novos conhecimentos para melhorar seu raciocínio. Utilizam conhecimento justificado e bases de informações, tal qual um especialista humano de determinada área do conhecimento.

Segundo WEISS (1988) os SEs possuem a seguinte arquitetura:

- a) base de conhecimentos representa a informação (fatos e regras) que um especialista utiliza;
- b) editor de bases é o meio pelo qual a Shell permite a implementação das bases desejadas;
- c) máquina de inferência o motor de inferência é o responsável pela ação repetitiva de buscar, analisar e gerar novos conhecimentos;
- d) banco de dados global são as evidências apontadas pelo usuário do Sistema Especialista durante uma consulta.



3 Materiais e Métodos

A metodologia de pesquisa adotada foi definida como bibliográfica e experimental. A pesquisa bibliográfica é elaborada a partir de material já publicado, constituído principalmente de livros, artigos de periódicos e atualmente com material disponibilizado na Internet (GIL, 2002).

A base de dados utilizada foi fornecida por um hospital público inglês no formato de planilhas Excel e o pré-processamento foi realizado utilizando o Office EXCEL da empresa Microsoft. Foram eliminados os elementos inconsistentes e atributos que não continham informações relevantes ao processo de aquisição de conhecimento. Alguns registros foram substituídos. A redução dos atributos foi realizada utilizando a opinião de um especialista e também o software Rosetta, que faz análise de dados no âmbito da teoria dos RS.

A rede SOM foi implementada utilizando o Software Somine da empresa Viscovery, que é um aplicativo para mineração de dados exploratória, análise de cluster visual, segmentação e classificação baseada em Mapas Auto-Organizáveis entre outras funções.

3.1 Pré-processamento e redução de atributos

Atributos com mais de 60% de dados em branco/nulos e registros com diagnóstico ‘desconhecido’ foram excluídos, pois somente criariam ‘ruído’ nos resultados e seriam irrelevantes para o propósito deste trabalho. Alguns dos atributos eram muito relevantes ao trabalho, mas continham certa quantidade de registros nulos/em branco. Para resolver o problema, foi calculada a média dos valores dos atributos e os espaços em branco/nulos foram preenchidos com esse valor.

Para definir quais seriam os atributos mais relevantes para o propósito do trabalho duas técnicas foram utilizadas: conhecimento de um especialista (Médico Residente se especializando na área de Gastro-Cirurgia) e o RS.

- Especialista

Verificando todos os atributos recebidos originalmente, o especialista analisou, com base em seu conhecimento e experiência, quais dos atributos seriam os mais relevantes na identificação do câncer colorretal. Após essa análise os atributos foram reduzidos de 54 para 28.

- Teoria dos *Rough Sets*

A ferramenta Rosetta foi utilizada para realizar a redução dos atributos. A técnica utilizada foi SAVGeneticReducer (Genetic Algorithm).

3.2 Rede SOM

Para a criação do Data Mart a partir da interface do Somine foi feita a importação da planilha com os dados já reduzidos. Para esta tarefa os dados foram normalizados no Excel antes da exportação, onde todos os registros contendo ‘Sim’ foram transformados em ‘1’ e os registros contendo ‘Não’ foram transformados em ‘2’.

Para a criação do modelo foi utilizado todo o conjunto de dados (100%) e todos os atributos foram assinalados com a mesma prioridade (peso 1).



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

Os parâmetros do mapa Kohonen foram os seguintes: número de nós 1000 (number of nodes), formato de mapa automático (automatic map format), relação 75 de 100 (RATIO), treinamento normal (training Schedule), tensão 0.5 (tension).

3.3 O Sistema Especialista

O Sistema Especialista para auxiliar na identificação de câncer colorretal foi implementado utilizando a ferramenta ExSinta que utiliza regras de produção para modelar o conhecimento humano.

O objetivo do ExSinta é simplificar ao máximo as etapas de criação de um SE completo. Para tanto, já oferece uma máquina de inferência básica, fundamentada no encadeamento para trás (backward chaining) (LIA1999).

A base de conhecimento foi gerada a partir da base de dados fornecida por um hospital público inglês. Utilizando as informações de correlação e os dados extraídos dos relatórios da rede SOM juntamente com as regras criadas algoritmos de decisão foi possível obter todas as informações necessárias para alimentar o SE.

O primeiro passo foi inserir as variáveis na base de conhecimento do ExSinta, trabalho totalmente manual. Cada variável precisa possuir um valor, que pode ser numérico, uni valorado ou multivalorado. Foi criado também uma variável objetivo, que é literalmente o objetivo do Sistema Especialista. No caso o objetivo é a identificação do risco para câncer colorretal, então a variável objetivo foi definida como 'Possui risco para câncer Colorretal', que é uni valorada, podendo ser 'sim' ou 'não'.



4 Análise dos resultados

- Rough Sets

Para a redução de atributos os dados foram submetidos a ferramenta Rosetta, que utiliza a teoria dos Rough Sets. Após a utilização do algoritmo genético SAVGeneticReducer (Genetic Algorithm) um conjunto compacto dos atributos mais relevantes foi obtido, contendo 19 registros.

Os outros atributos foram excluídos da base de dados, mantendo-se somente os que foram selecionados pelo Rough Sets.

- Rede SOM

Após a base de dados normalizada e consistente ser importada para a ferramenta Somine, um mapa foi gerado pela rede SOM, contendo nove clusters. Foram criados mapas para cada um dos atributos exibindo a relação de cada atributo com os clusters criados. Na Figura 2 podemos observar que alguns dos atributos definem e ocupam completamente um cluster.

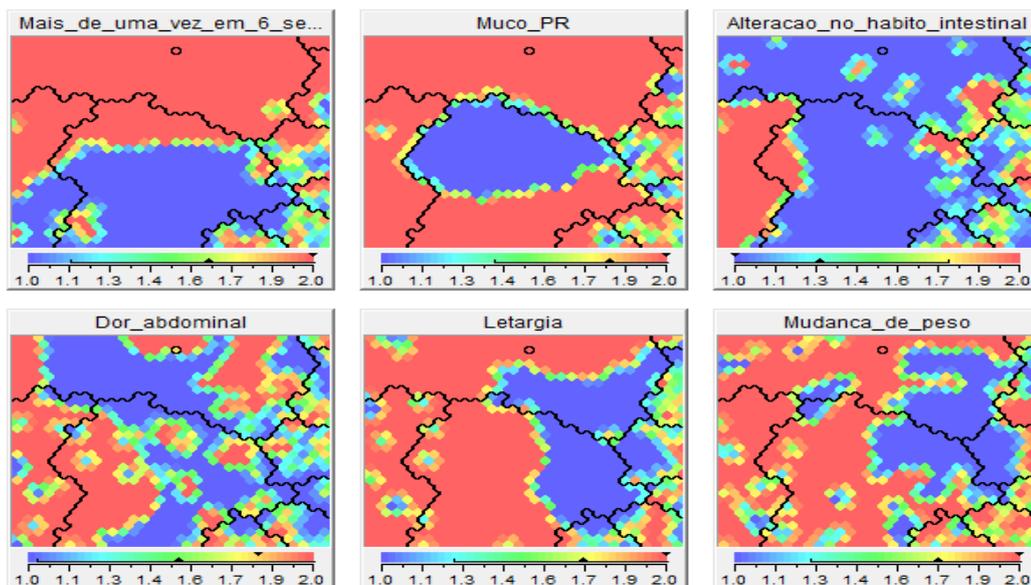


Figura 2 - Clusters gerados pela rede SOM (atributos)

A partir das informações fornecidas pela rede SOM é possível observar também a correlação dos elementos, sendo valores mais próximos de 1 mais relevantes, e valores mais distantes de 1 menos relevantes para o conjunto de dados. O Somine disponibiliza também estatísticas sobre as variáveis como valores faltantes, repetição de valores, relevância e descrição detalhada dos atributos.

Ao final de todo o processo a ferramenta gera um relatório detalhado contendo informações sobre os atributos, parâmetros utilizados, atributos priorizados, segmentação, entre outros que são muito importantes para a implementação do SE.

- Sistema Especialista



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

Após análise do relatório gerado pelo SOMine um total de 152 regras foram geradas pelo ExSinta. Cada regra possui um fator de confiança que mostra o quão confiável é o resultado numa escala de entre 10 e 100.

Ao iniciar o SE a tela de boas-vindas é exibida, mostrando ao usuário o objetivo do programa. Sequencialmente o usuário é direcionado as perguntas referentes à identificação do risco do câncer. Na Figura 3 observa-se um exemplo de pergunta feita pelo sistema.

Identificando o câncer colorretal

Qual o valor de Níveis de HB ?
(Entre um valor numérico)

Valor:

Figura 3 – Exemplo de pergunta feita ao usuário no SE

O usuário irá responder as perguntas existentes até que as premissas do SE tenham sido alcançadas. Neste momento o SE mostra o resultado ao usuário, conforme Figura 4:

Valor	CNF (%)
Não	100

Fechar

Ajuda

Figura 4 – Exemplo de resultado referente a consulta realizada no SE

Na tela de resultados (Figura 5) é possível acessar a aba 'O sistema' onde as regras do SE são exibidas mostrando ao usuário como a decisão foi tomada.

O sistema especialista

Fechar

Ajuda

REGRA 1
SE Níveis de HB < 12,05
E Sangramento Per Rectum = Sim
E Muco Per Rectum = Sim
ENTÃO Possui risco de cancer colorretal = Não CNF 100%

REGRA 2
SE Níveis de HB < 12,05
E Sangramento Per Rectum = Sim
E Muco Per Rectum = Sim
E Níveis de HB >= 9,35
E Ex-Fumante = Não
E Urgencia ao defecar = Não
ENTÃO Possui risco de cancer colorretal = Sim CNF 60%

Figura 5 – Exemplo da aba 'O sistema' mostrando as regras de decisão do SE

4.1 Discussão dos resultados



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

O SE foi testado e os resultados validados. Obteve-se êxito na realização dos testes e os resultados foram de acordo com o esperado.

Foi possível obter um bom nível de acurácia nas regras geradas e todas as possibilidades, considerando-se a base de dados utilizada, foram cobertas. Um total de 152 regras foram implementadas ao SE.

O SE foi testado com perguntas contendo os mais variados graus de confiança para garantir que o resultado final fosse acurado e correspondente aos dados imputados pelo usuário.

Com a implementação do SE foi possível:

- Identificar pacientes com baixo nível de risco para câncer colorretal;
- Identificar pacientes com alto nível de risco para câncer colorretal;
- Identificar quais são os atributos mais e menos importantes a serem considerados no diagnóstico do para câncer colorretal;
- Identificar as melhores técnicas da Inteligência Artificial e algoritmos para a identificação do câncer colorretal;
- Obter resultados mais precisos utilizando somente atributos que são relevantes ao diagnóstico.



5 Conclusão

Neste trabalho foi possível estudar a aplicação de técnicas da Inteligência Artificial como a Teoria dos Rough Sets, a Rede SOM e o Sistema Especialista aplicados no auxílio ao diagnóstico do câncer colorretal de modo a prover uma fonte confiável de apoio aos profissionais da saúde.

A realização deste trabalho permitiu um melhor entendimento de técnicas da IA, tais como Rede SOM e Rough Sets, e suas aplicações, bem como a geração de algoritmos de decisão, que foram essenciais na implementação do SE.

A opinião do especialista teve papel muito importante na seleção dos atributos e na geração das regras de decisão, pois, baseado em sua experiência, somente dados relevantes foram mantidos, fazendo com que o SE fosse mais preciso e acurado.

O SE é o produto final desta pesquisa, e seu desenvolvimento foi facilitado devido ao uso da ferramenta ExSinta, que é amigável e reduziu o tempo de implementação do SE.

O objetivo desta pesquisa foi alcançado ao demonstrar que as técnicas da IA podem auxiliar no diagnóstico do câncer colorretal. Com a implementação do SE a identificação da doença ocorre utilizando-se somente os sintomas relevantes e torna o processo mais prático, eliminando incertezas.

O sistema informa uma porcentagem sobre o quanto aquele resultado é verdadeiro para determinado paciente, baseado nas perguntas respondidas. Assim o clínico pode analisar cada caso e, junto com sua experiência, fazer o diagnóstico mais preciso.

A utilização do SE traz como benefícios a redução do erro médico durante o diagnóstico do câncer colorretal e o melhor aproveitamento do tempo gasto com o paciente, aumentando a qualidade e eficácia.

Este trabalho promove a aproximação das áreas, integrando a medicina e a informática, unindo os conhecimentos para o benefício das pessoas.

O curto tempo foi a maior limitação na realização deste trabalho, pois não foi possível comparar a eficácia de mais algoritmos de decisão para a geração das regras e realizar uma sessão de testes mais intensiva com o SE.

A base de dados utilizada também foi uma limitação, pois é formada por informações oriundas de questionário realizado pessoalmente em papel, contendo alguns dados inconsistentes que tiveram que ser pré-processados.

A continuidade da pesquisa se dará no uso de Raciocínio Baseado em Casos (RBC), técnica que resolve novos problemas a partir de soluções utilizadas para resolver problemas anteriores. Esta técnica permite maior conhecimento no sistema, pois o aprendizado ocorre com a experiência, e novos casos sempre são adicionados e adaptados para serem usados no futuro.



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

6 Referências

ABRALE. O que é câncer. 2012. Disponível em: <<http://www.abrale.org.br/pagina/o-que-e-cancer>> Acesso em 30 jun. 2014

A.C. CAMARGO. Colorretal / Intestino Grosso. Disponível em: <<http://www.accamargo.org.br/tudo-sobre-o-cancer/colorretal--intestino-grosso/12/>> Acesso em 30 jun. 2014

ALBERTIN, A. L.; MOURA, R. M. de. Tecnologia de informação. São Paulo, Atlas, 2004.

CARNEIRO, L. R.; Tecnologia da Informação e a Medicina, 2009. Disponível em <<http://timedicina.blogspot.com.br/2009/06/tecnologia-da-informacao-e-medicina.html>> acessado em 28 10 2011 as 17:44hs

CURRY, B. In: The Kohonen Self-organizing Map as an Alternative to Cluster Analysis: An Application to Direct Marketing. In: International Journal of Market Research, JunE, 2003.

FARIAS, E. B. P. Implementação de um sistema especialista no Service Desk de um Hospital Público. 2012. 90 f. TCC (Graduação) - Curso de Sistemas de Informação, Universidade Nove de Julho, São Paulo, 2012.

GIL, A. C. Como Elaborar Projetos de Pesquisa. Atlas, Brasil 1987.

GOLDSCHMIDT, R. R. Inteligência Computacional, Rio de Janeiro: IST-Rio, 2010. 143 p.

GOEBEL, M., GRUENWALD, L.A. Survey of data mining and knowledge discovery software tools. In: SIGKDD Explorations, Junho 1999.

HARRISON, T.H. Intranet data warehouse. Editora Berkeley, 1998.

HOSPITAL DO CÂNCER DE BARRETOS. Câncer Colorretal (Cólón, Reto e Intestino Grosso). Disponível em: <<http://www.hcancerbarretos.com.br/cancer-colorretal>>. Acesso em: 30 jun. 2014.

INCA. Colorretal: Diagnóstico. Instituto Nacional de Câncer. Disponível em: <<http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/colorretal/diagnostico1>>. Acesso em: 06 jun. 2014.

JEMAL, A.; BRAY, F.; CENTER, M.M.; FERLAY, J.; WARD, E.; FORMAN, D. Global Cancer Statistics. CA: A Cancer Journal for clinicians. [Estados Unidos], p. 69-90. mar. 2011.

JIN, H.; SHUM, W. H.; LEUNG, K. S.; WONG, M. L. In: Expanding self-organizing map for data visualization and cluster analysis. Information Sciences, 163, 2004.

JOHANSTON, H. Sistemas de Informação Hospitalar: presente e futuro. 1993. Revista Informédica 1993;1(2):5-9. Disponível em: <<http://www.informaticamedica.org.br/informed/halley.htm>> Acesso em: 22 abr 2010.



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

KASKI, S.; KOHONEN, T. In: Exploratory Data Analysis by the Self-organizing Map: Structures of Welfare and Poverty in the World. In: Proceedings of the third International Conference on Neural Networks in the Capital Markets. World Scientific, Singapore, 1996.

KOHONEN, T. Self-Organizing Maps. Springer Series in Information Sciences, Vol. 30, third edition. Springer, Berlin, Heidelberg, New York, 2001.

LIA, Laboratório de Inteligência Artificial. Uma ferramenta para criação de sistemas especialistas. Endereço eletrônico: <<http://www.lia.ufc.br>>

LIA. Expert Sinta. (1996) “Expert SINTA – Uma ferramenta visual para criação de sistemas especialistas –Versão 1.1– Manual do Usuário”. – Laboratório de Inteligência Artificial /LIA – UFC, CE.

OSHEROFF, J. A., TEICH, J.M., MIDDLETON, B., STEEN, E.B., WRIGHT, A., DETMER, D.E. A Roadmap for National Action on Clinical Decision Support. Journal of the American Medical Informatics Association. p. 141–145. 2007.

PAWLAK, Z. In: Rough Sets, International Journal of Computer and information Sciences, 1982.

PAWLAK, Z. In: Rough Sets, Rough relations and Rough functions. Fundamenta Informaticae, 27, 1996.

PISA, I. T.; GALINA, A. C.; LOPES, P. R. L.; BARSOTTINI, C. G. N.; ROQUE, A. C. Lepidus R3: implementação de sistema de apoio à decisão médica em arquitetura distribuída usando serviços web. In: IX Congresso Brasileiro de Informática em Saúde - CBIS2004, 2004, Ribeirão Preto, SP. Anais do IX Congresso Brasileiro de Informática em Saúde, 2004.

RIBEIRO, H. C. S. Introdução aos sistemas especialistas. Rio de Janeiro: LCT – Livros Técnicos e Científicos Editora S.A., 1987

ROSETTA (A Rough Sets Toolkit for Analysis of Data) disponível em: <http://www.idi.ntnu.no/~aleks/rosetta>.

ROSSI, B. M.; AGUIAR JÚNIOR, S.; FERREIRA, F. O.; NAKAGAWA, W. T.; LOPES, A. Como eu trato: Câncer colorretal. Revista Brasileira de Medicina: Oncologia, São Paulo, v. 67, p.4-9, abr. 2010.

RUSSEL, S; NORVIG, P. Artificial Intelligence: A Modern Approach. EUA, Prentice-Hall, 1995.

SARTIPI, K.; ARCHER, N. P.; YARMAND, M. H. Challenges in Developing Effective Clinical Decision Support Systems. In: JAO, C. S. (Ed.). Efficient Decision Support Systems: Practice and Challenges in Biomedical Related Domain. Rijeka, Croácia: InTech, 2011. Pags 3 – 20.

SASSI, R.J., SILVA, L.A., HERNANDEZ, E. del M. An Hybrid Architecture for the Knowledge Discovery in Databases: Rough Sets Theory and artificial neural nets Self-



III Simpósio Internacional de Gestão de Projetos (III SINGEP) II Simpósio Internacional de Inovação e Sustentabilidade (II S2IS)

Organizing Maps”. International Conference on Information Systems and Technology Management. Congresso Internacional de Gestão de Tecnologia e Sistemas de Informação (4º CONTECSI). 31 a 1 de junho, 2007. São Paulo, Brasil.

ULTSCH, A. In: Self-Organizing Neural Networks perform different from statistical K-means Clustering. In Proc. GfKI, Basel, Swiss, 1995.

VESANTO, J.; ALHONIEMI, E. In: Clustering of the Self-organizing Map. IEEE Transaction on Neural Network, vol.11, May, 2000.

WEISS, S, M.; KUKIKOWSKI, C, A. Guia pratico para projetar sistemas especialistas. Rio de Janeiro, Brasil, LTC – Livros Técnicos e Científicos S.A. , 1988.

WORLD HEALTH ORGANIZATION. The Global Burden of Disease: 2004 Update. Geneva: World Health Organization, 2008.