UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

**Inteligência Artificial Aplicada ao Auxílio no
Diagnóstico da Dor Pélvica Crônica**

VINICIUS OLIVERIO

ORIENTADOR: PROF. DR. OMERIO BENEDICTO POLI NETO.

Ribeirão Preto - SP
Jan/2018

UNIVERSIDADE DE SÃO PAULO

FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

PROGRAMA DE PÓS-GRADUAÇÃO EM GINECOLOGIA E OBSTETRÍCIA

**Inteligência Artificial Aplicada ao Auxílio no
Diagnóstico da Dor Pélvica Crônica**

VINICIUS OLIVERIO

Tese apresentada ao Programa de Pós-Graduação em Ginecologia e Obstetrícia para a obtenção do título de Doutor em Ciências. Área de concentração: Biologia da Reprodução.

Orientador: Prof. Dr. Omero Benedicto Poli Neto.

Ribeirão Preto - SP
Janeiro/2018

Dados Internacionais de Catalogação na Publicação

Oliverio, Vinicius

Inteligência artificial aplicada ao auxílio no diagnóstico da dor pélvica crônica / Vinicius Oliverio. - - Ribeirão Preto, 2018.
58 p.

Orientador: Omero Benedicto Poli Neto

Tese (Doutorado – Programa de pós-graduação em Ginecologia e Obstetrícia) – Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto, 2018.

1. Dor pélvica crônica. 2. Inteligência artificial. 3. Diagnóstico automático. 4. Aprendizagem de máquina. 5. Ginecologia e obstetrícia. I. Título

AGRADECIMENTOS

Ao meu orientador Prof. Dr. Omero Benedicto Poli Neto pela participação ativa e direta nesta etapa importante da minha vida profissional, sempre me guiando na realização deste trabalho científico, meu eterno agradecimento.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) pelo financiamento deste projeto de pesquisa.

A minha família, a minha namorada Karolini pelo apoio e a todas as pessoas que participaram, contribuindo para realização deste trabalho, direta ou indiretamente, meu agradecimento.

EPÍGRAFE

“Nós somos o que pensamos.

Tudo o que somos surge de nossos pensamentos.

Com o pensamento, construímos e destruimos o mundo.

O pensamento nos segue como uma carroça segue a parelha de bois.

Nós somos o que pensamos.

Nossa imaginação pode nos causar mais danos que o pior inimigo.

Mas, uma vez que conseguimos controlar o pensamento, temos a ajuda mais poderosa, que nem mesmo nossos pais podem nos dar.”

Sidarta Gautama

RESUMO

A dor pélvica crônica (DPC) é uma condição clínica comum, inclusive no Brasil. Estudos do nosso núcleo de pesquisa evidenciam uma prevalência em torno de 11%. Ela é, dentre outros agravos, causa recorrente de procura a serviços de saúde, falta ao trabalho e baixa produtividade. Com a inserção definitiva da mulher no mercado de trabalho a doença tem impactado negativamente no desempenho econômico desse grupo, afetando diretamente a atividade econômica domiciliar. Todavia, o diagnóstico da causa desta condição clínica não é trivial e erros podem ocorrer devido à falta de conhecimento sobre o assunto, portanto, a criação de um sistema que possa auxiliar no diagnóstico e na identificação das causas desta condição clínica é de grande importância para que os erros de diagnósticos sejam minimizados e para que esta condição seja tratada da melhor maneira possível. Contudo, o presente trabalho tem como objetivo o desenvolvimento de um sistema de inteligência artificial que auxilie no diagnóstico da DPC em seu estágio inicial, minimizando erros e tempo de tratamento por meio da mineração de dados em uma base médica previamente preenchida com casos já diagnosticados como auxílio a aprendizagem do sistema.

Palavras-chave: Dor Pélvica Crônica, Inteligência Artificial, Diagnóstico automático, Aprendizagem de Máquina, Ginecologia e obstetrícia.

ABSTRACT

Chronic pelvic pain (CPP) is a common clinical condition, including in Brazil. Studies held by our research group shows evidences of a prevalence around 11%. It is, among other illness, a recurrent cause of lack of productivity, work miss and utilization of healthcare services. As the women is each time more inserted in the labor force this condition is impacting negatively the economic performance of this group, affecting directly the domestic economic activity. However, the diagnostic of the cause of this condition is not trivial, and mistakes can happen due to the lack of knowledge about the matter, therefore, the creation of a system that may help in the diagnosing and identification of the causes of this clinical condition is of great importance to the decrease of mistaken diagnoses leading to a better treatment of the condition. However, the current work has as objective the development of an artificial intelligence system that help in the CPP diagnose in its initial stage, decreasing mistakes and treatment time by using data mining in a medical database previously filled with already diagnosed cases as auxiliary to the learning of the system.

Keywords: Chronic Pelvic Pain, Artificial Intelligence, Automatic Diagnose, Machine Learning, Gynecology and obstetrics.

LISTA DE FIGURAS

Figura 2.1 - Árvore de Decisão.....	10
Figura 2.2 - Exemplo SVM.	11
Figura 2.3 - Exemplo complexo SVM.	12
Figura 2.4 - Transformação SVM.	13
Figura 2.5 - Algoritmo Adaboost (Adaptado de [47]).	14
Figura 2.6 - Exemplo k -NN.	15
Figura 4.1 - Experimentação PT4.....	28
Figura 4.2 - Seleção de Atributos.	29
Figura 4.3 - Comparação das Performances.....	33
Figura 4.4 - Comparação das Sensibilidades.....	34
Figura 4.5 - Comparação das Especificidades.	36
Figura 4.6 - Modelo do Diagnóstico de Endometriose.....	46

LISTA DE TABELAS

Tabela 2.1 - Jogar Tênis.	9
Tabela 2.2 - Exemplo de um conjunto de dados multi-rótulo.	16
Tabela 2.3 - Conjunto transformado utilizando PT1.	17
Tabela 2.4 - Conjunto transformado utilizando PT2.	17
Tabela 2.5 - Conjunto transformado utilizando PT3.	17
Tabela 2.6 - Conjunto para endometriose transformado utilizando PT4.	18
Tabela 2.7 - Conjunto para DIP transformado utilizando PT4.	18
Tabela 2.8 - Conjunto para constipação transformado utilizando PT4.	19
Tabela 2.9 - Conjunto para bexiga dolorosa transformado utilizando PT4.	19
Tabela 2.10 - Conjunto transformado utilizando PT5.	19
Tabela 3.1 - Comparativo de Algoritmos [56].	24
Tabela 4.1 - Precisão do Diagnóstico.....	32
Tabela 4.2 - Sensibilidades.	34
Tabela 4.3 - Especificidades.	35
Tabela 4.4 - Análise das Variáveis.	38

LISTA DE ABREVIATURAS E SIGLAS

DPC – *Dor Pélvica Crônica*

IA – *Inteligência Artificial*

HC – *Hospital das Clínicas*

FMRP – *Faculdade de Medicina de Ribeirão Preto*

DIP – *Doença Inflamatória Pélvica*

PT1 – *Método de transformação de problemas 1*

PT2 – *Método de transformação de problemas 2*

PT3 – *Método de transformação de problemas 3*

PT4 – *Método de transformação de problemas 4*

PT5 – *Método de transformação de problemas 5*

SVM – *Support Vector Machines*

Adaboost – *Adaptative Boosting*

kNN – *k-Nearest Neighbors*

InfoGain – *Information Gain Based Attribute Selection*

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO.....	1
1.1 Contexto.....	1
1.1.1 Dor Pélvica Crônica	1
1.1.2 Inteligência Artificial	3
1.1.3 Mineração de Dados.....	4
1.2 Motivação e Objetivos	5
1.3 Organização do Trabalho.....	5
CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA.....	6
2.1 Considerações Iniciais.....	6
2.2 Aprendizagem de Máquina	6
2.2.1 Aprendizagem Supervisionada	7
2.2.1.1 Naive Bayes	8
2.2.1.2 C4.5.....	10
2.2.1.3 Support Vector Machines	11
2.2.1.4 Adaboost.....	14
2.2.1.5 k-Nearest Neighbors	15
2.3 Problemas multi-rótulo.....	16
2.4 Seleção de Atributos	20
2.4.1 Information Gain Based Attribute Selection.....	20
2.5 K-fold Cross Validation	21
2.6 Considerações Finais	22
CAPÍTULO 3 - TRABALHOS CORRELATOS.....	23
3.1 Diagnóstico Automático	23
3.2 Considerações Finais	25
CAPÍTULO 4 - DIAGNOSTICANDO A CAUSA DA DPC	26
4.1 Considerações Iniciais.....	26
4.2 Metodologia e Desenvolvimento	26
4.2.1 Validação dos resultados	27

4.3 Metodologia e Desenvolvimento	27
4.3.1 Métodos de transformação	27
4.3.2 Seleção de Atributos	28
4.3.3 Performances dos Algoritmos	32
4.4 Estudo de caso: Diagnóstico da Endometriose	36
4.5 Considerações Finais	47
CAPÍTULO 5 - CONCLUSÕES E TRABALHOS FUTUROS.....	48
CAPÍTULO 6 - PRODUÇÕES CIENTÍFICAS	50
REFERÊNCIAS.....	66
APÊNDICES	71

Capítulo 1

INTRODUÇÃO

Neste capítulo é introduzida para o leitor a ideia geral do trabalho, mostrando sua relevância e a motivação para a sua execução, assim como o objetivo que se busca ao final do trabalho. O capítulo está estruturado da seguinte maneira: Contexto, onde é discutida a dor pélvica crônica, a inteligência artificial e mineração de dados, onde são apresentados alguns conceitos introdutórios. Além disso, é apresentada também a motivação para o desenvolvimento de um trabalho de investigação que aborda o diagnóstico automático da dor pélvica crônica, bem como, uma descrição mais precisa dos objetivos desta pesquisa.

1.1 Contexto

1.1.1 Dor Pélvica Crônica

A dor pélvica crônica (DPC) é uma condição clínica frequente e definida como dor localizada na região inferior do abdome (abaixo da cicatriz umbilical) e/ou na pelve, não exclusivamente menstrual ou associada à doença neoplásica, persistente por pelo menos seis meses e intensa o suficiente para causar incapacidade funcional, necessitando de tratamento clínico ou cirúrgico [1]. Possui difícil controle, porque frequentemente não se identifica a causa primária. A etiologia não é clara e, usualmente, resulta de uma complexa interação entre os sistemas gastrointestinal, urinário, ginecológico, musculoesquelético, neurológico, psicológico e endócrino, influenciado ainda por fatores socioculturais [2]. Alguns estudos têm tentado identificar fatores de risco para a DPC, mas os resultados não são conclusivos. Há evidências que fatores como abuso de drogas ou álcool, abortos, fluxo menstrual aumentado, doença inflamatória pélvica, cesáreas, patologias pélvicas e comorbidades psicológicas (transtornos

de humor, somatização, hipocondria, dentre outros) estejam associadas com um aumento do risco para a doença [3]. Em estudo recente realizado pelo nosso grupo identificamos como fatores independentemente associados à DPC na população de Ribeirão Preto: dispareunia, cirurgia abdominal prévia, depressão, dismenorréia, ansiedade, atividade sexual atual, dor lombar, constipação, sintomas urinários irritativos e baixo nível educacional [4]. Ainda mais recentemente, na população de São Luís - MA identificamos os seguintes: cirurgia abdominal prévia, dispareunia, dismenorréia, alcoolismo, migrânea, dor lombar.

A prevalência estimada de DPC é de 3,8% em mulheres varia 4% a 24% em mulheres na idade reprodutiva [5-8], prevalência superior à da enxaqueca, asma e lombalgia. Em Ribeirão Preto a prevalência estimada é de 11,5% [4]. Há um impacto direto na sua vida conjugal, social e profissional, transformando a DPC em um sério problema de saúde pública [9-11]. Cerca de 60% das mulheres com a doença nunca receberam o diagnóstico específico e 20% nunca realizaram qualquer investigação para elucidar a causa da dor [12]. Em unidades de cuidados primários, 39% das mulheres queixam-se de dor pélvica. É responsável por 40 a 50% das laparoscopias ginecológicas, 10% das consultas ginecológicas e, aproximadamente, 12% das histerectomias [13]. Não sabemos sua real prevalência em países em desenvolvimento, como o Brasil, mas estima-se que seja superior àquela encontrada em países desenvolvidos. Adicionalmente, implica custo direto superior a dois bilhões de dólares por ano nos Estados Unidos e, certamente, traz um impacto considerável no cenário socioeconômico nacional, tanto decorrentes dos elevados gastos infringidos ao sistema único de saúde [14] quanto àqueles secundários às perdas decorrentes das repercussões ocupacionais negativas que estão brevemente discutidas abaixo.

Paralelo a isso, temos observado que a base do conhecimento médico está em contínua expansão, e o seu desenvolvimento está diretamente relacionado à execução de pesquisas científicas de qualidade. Uma das ferramentas que tem permitido essa expansão considerável na execução de pesquisas são os recursos implementados pela área de tecnologia da informação. A informática viabiliza, com a criação de sistemas de registro e gerenciamento de formulários eletrônicos, a captação e o armazenamento confiável de dados clínicos selecionados, para que estudos científicos sejam realizados [18]. Adicionalmente, o uso de formulário eletrônico estruturado contribui para redução de erros em coletas de dados, diminuição da quantidade de papel e os seus custos relativos, e principalmente, potencializa as buscas e comparações das informações, melhorando a investigação clínica e atividades de ensino [19]. Além disso, com a tecnologia atual é possível acessá-los remotamente e de forma segura através da internet, anulando as barreiras geográficas [20].

1.1.2 Inteligência Artificial

O objetivo da Inteligência Artificial (IA) é o desenvolvimento de paradigmas ou algoritmos que requeiram máquinas para realizar tarefas cognitivas, para as quais os humanos atualmente são melhores. Um sistema de IA deve ser capaz de fazer três coisas [21]:

- Armazenar conhecimento;
- Aplicar o conhecimento armazenado para resolver problemas;
- Adquirir novo conhecimento através da experiência.

Além de ser capaz de fazer tais coisas, um sistema de IA deve possuir três componentes fundamentais: representação, raciocínio e aprendizagem [21]. A representação é a maneira que é utilizada para representar o conhecimento genérico sobre um domínio do problema, o raciocínio, é a habilidade de resolver problemas, e por fim, a aprendizagem é um processo de mudança de comportamento obtido através da experiência construída por fatores emocionais, neurológicos, relacionais e ambientais [22].

Aprendizado de Máquina então pode ser visto como o processo de mudança de estrutura, programa ou dado, baseado em entradas ou em resposta a informações externas, de maneira que o desempenho futuro de um algoritmo em alguma tarefa melhore [23]. Segundo Tom Mitchell [24], para se dizer que uma máquina aprende ela deve, a partir de uma experiência E , com respeito a uma classe de tarefas T e medida de performance P , melhorar sua performance na tarefa T , medida por P , com essa experiência E .

Tradicionalmente, métodos de aprendizado de máquina indutivos são classificados em três tipos: aprendizado supervisionado, aprendizado não supervisionado e aprendizado semissupervisionado. A forma e a presença (ou ausência) de supervisão são aspectos primordiais para caracterizar cada uma das três diferentes abordagens. De forma resumida, na aprendizagem supervisionada, um conjunto de dados rotulados é fornecido ao algoritmo para que ele aprenda a classificar novos dados com base nesses exemplos fornecidos (*classification*), então, para este tipo de aprendizagem o que se deseja é construir classificadores. Na aprendizagem não supervisionada não é fornecido este conjunto de dados rotulados, portanto o que se deseja é agrupar os dados de acordo com sua similaridade (*clustering*), o algoritmo de aprendizagem não supervisionada procura encontrar padrões nos dados a serem agrupados. Por fim na aprendizagem semissupervisionada também se tem um conjunto de exemplos fornecidos, mas este conjunto é limitado, portanto em uma das

abordagens deste tipo de paradigma é a realimentação do conjunto de treinamento com dados classificados pelo próprio algoritmo, as classificações realizadas pelo algoritmo são acrescentadas ao conjunto de treinamento e o algoritmo é treinado novamente, mas agora com mais dados no conjunto de treinamento.

Outro aspecto relevante ao presente trabalho no que se refere à inteligência artificial é a mineração de dados, essa mineração é realizada utilizando métodos de aprendizagem de máquina supracitados, porém para que os dados de uma base de conhecimento possam ser utilizados para mineração é necessária a sua preparação. Conjuntos de dados não são a representação ideal do mundo, mesmo que estes dados tenham sido adquiridos de maneira confiável eles não estão livres de incerteza, distorção e ruído. Este ruído é inerente ao mundo real, não somente a erros ou procedimentos incorretos [25], e.g. no caso de bases de conhecimento de consultas médicas o paciente pode omitir algum fato ou responde-lo de maneira incorreta.

1.1.3 Mineração de Dados

Mais do que nunca o mundo é sobrecarregado com dados. A quantidade de dados gravados cresce rapidamente e de acordo com [26] o universo digital será 44 vezes maior em 2020 que era em 2009, totalizando 35 zettabytes de dados. Como o volume de dados cresce, a proporção dele que é compreendida decresce. Escondida em todos estes dados está a informação, potencialmente útil de onde alguns campos de pesquisa podem tirar proveito [27].

A mineração de dados se trata de resolver problemas analisando todos os tipos de dados presentes nas bases de dados. Pode ser definida como o processo de descoberta de padrões em dados, e por definição, este processo deve ser automático ou semiautomático [27]. Este processo de mineração de dados utiliza técnicas de aprendizagem de máquina de modo a atingir seu objetivo de identificar padrões em dados.

Também é necessário dizer que para se construir um bom minerador (algoritmo de mineração de dados) é necessário preparar os dados para que possam ser utilizados em conjunto com estas técnicas. A preparação e pesquisa dos dados levam a um entendimento dos mesmos o que permite se descobrir qual o modelo correto a ser construído [28].

1.2 Motivação e Objetivos

Como visto anteriormente, o correto diagnóstico da causa da DPC é um aspecto importante na saúde pública e o desenvolvimento de um sistema capaz de auxiliar neste processo é também de grande relevância. Assim o principal objetivo deste trabalho é investigar, implementar e avaliar métodos que permitam auxiliar o profissional de saúde no diagnóstico da causa da DPC. Para tanto, se utilizou a base de dados do HC de casos previamente diagnosticados pelos médicos responsáveis. Têm-se então como objetivos específicos:

- Desenvolver um sistema dotado de inteligência capaz de auxiliar no diagnóstico das causas da DPC.
- Investigar e implementar métodos para preparar dados existentes nesta base com a finalidade de treinar outros algoritmos com estas informações;
- Investigar e implementar métodos para o auxílio do diagnóstico de novos pacientes probabilisticamente.
- Investigar e implementar métodos para identificar padrões (identificar a correlação entre as variáveis) nos dados contidos nesta base de dados.

1.3 Organização do Trabalho

A organização da sequência deste texto está assim estruturada, no Capítulo dois, a fundamentação teórica básica é apresentada, abordando os temas Aprendizagem de Máquina, Problemas Multi-rótulo, Métodos de Seleção de Atributos e Validação dos Resultados. O Capítulo três traz os trabalhos correlatos mais relevantes encontrados no campo do diagnóstico automático de doenças utilizando inteligência artificial. No capítulo quatro a proposta e a metodologia de desenvolvimento do trabalho são mostradas, também é vista a implementação onde os algoritmos implementados são explicados e demonstrados e a experimentação realizada é demonstrada. No capítulo cinco a conclusão é discutida.

Capítulo 2

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo a fundamentação teórica essencial para a compreensão do trabalho é apresentada. Neste sentido, são apresentados conceitos básicos relativos à Aprendizagem de máquina, problemas Multi-rótulo, seleção de atributos e validação dos resultados.

2.1 Considerações Iniciais

Neste capítulo é dada a fundamentação teórica básica para a compreensão deste trabalho de doutorado. Primeiramente é abordado o tema de Aprendizagem de Máquina onde são discutidos alguns paradigmas e algoritmos, posteriormente o conceito de aprendizagem multi-rótulo é introduzido, a seleção de atributos é então discutida e por fim uma discussão a respeito da validação dos resultados é realizada. Depois de abordada a fundamentação necessária para a compreensão do projeto de investigação apresentado neste texto, o desenvolvimento do projeto é descrito e explicado em mais detalhes.

2.2 Aprendizagem de Máquina

Por não haver uma definição única aceita por toda a comunidade, não se pretende neste texto definir o Aprendizado de máquina de maneira rígida. Em geral, Aprendizado de Máquina pode ser visto como o processo de mudança de estrutura, programa ou dado, baseado em entradas ou em resposta a informações externas, de maneira que o desempenho futuro de um algoritmo em alguma tarefa melhore [23]. Segundo [24], para se dizer que uma

máquina aprende ela deve, a partir de uma experiência E , com respeito a uma classe de tarefas T e medida de performance P , melhorar sua performance na tarefa T , medida por P , com essa experiência E .

Tradicionalmente, métodos de aprendizado de máquina indutivos são classificados em 3 abordagens: aprendizado supervisionado, aprendizado não supervisionado e aprendizado semissupervisionado. A forma e a presença (ou ausência) de supervisão são aspectos primordiais para caracterizar cada uma das três diferentes abordagens. Em todos os tipos de aprendizado tem-se como base um conjunto de dados $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, onde $X = \{x_1, x_2, \dots, x_n\}$ é o conjunto de vetores m -dimensionais $x_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,m}\}$ que representam instâncias de entrada, e $Y = \{y_1, y_2, \dots, y_n\}$ é o conjunto de saídas, onde cada saída (ou rótulo) y_i está vinculado a um vetor x_i . O processo de aprendizado normalmente é definido pela aproximação de uma função que tome x_i como entrada e forneça y_i como saída ($f(x_i) \rightarrow y_i$). Para uma dada instância i , se o rótulo y_i é conhecido, então diz-se que esta é uma instância rotulada. Mas, caso o rótulo y_i não seja conhecido, então diz-se que esta é uma instância não rotulada. Sendo bastante genérico, pode-se dizer que D é formado por dois subconjuntos D_r e D_n ($D = D_r \cup D_n$). O subconjunto D_r é formado por dados rotulados, já o subconjunto D_n é formado por dados não rotulados. A próxima subseção dá uma visão geral sobre Aprendizagem Supervisionada, que é a abordagem utilizada no presente trabalho.

2.2.1 Aprendizagem Supervisionada

Nesta abordagem, normalmente se deseja construir um classificador. Assim, o subconjunto D_r é utilizado para supervisionar o processo de aprendizado. Neste sentido, numa primeira etapa o modelo de aprendizado (classificador) é treinado com base em D_r e aprende a prever o valor de y para qualquer vetor x . Assim, com base no modelo treinado, os dados presentes em D_n podem ser automaticamente rotulados.

Supondo, por exemplo, que se queira aprender a classificar um carro como sendo um “carro familiar” ou “não familiar”. Supondo também que se tenha um conjunto D_r de dados já rotulados definindo exemplos de carros familiares e não familiares. Então, para se definir um classificador, o que se pode fazer é encontrar um conjunto de características compartilhadas pelos exemplos positivos (carros familiares), e que não sejam compartilhadas pelos exemplos negativos (carros não familiares). Assim, através do conjunto de características encontradas nos exemplos positivos e negativos é possível fazer uma predição: Dado um carro que não tenha sido visto antes (ou seja, uma instância de D_n) e, checando a descrição que se aprendeu,

é possível dizer se um novo automóvel que se queira classificar é um “carro familiar” ou não [32].

Em outras palavras, na aprendizagem supervisionada o algoritmo aprende com a ajuda de um supervisor, que possui conhecimento sobre o problema, através de uma etapa de treinamento. Este conhecimento é normalmente representado por um conjunto de treinamento D_t [22]. Com base neste conjunto de treinamento o algoritmo pode determinar qual a saída esperada quando novas entradas forem testadas. O conjunto D_n é tradicionalmente utilizado para se testar a precisão do classificador, por este motivo, D_n é comumente chamado de conjunto de testes.

Dentre os vários algoritmos de aprendizagem supervisionada podemos citar as Redes Neurais [22], Árvores de Decisão [33], Máquinas de Vetores de Suporte [34], Naive Bayes [35], entre outros. Com a finalidade de ilustrar este tipo de aprendizagem serão demonstrados os algoritmos utilizados na experimentação do presente trabalho.

2.2.1.1 Naive Bayes

Também chamado de Classificador Naive Bayes, esse é um método de aprendizado supervisionado, onde um conjunto de exemplos de treinamento deve ser fornecido como entrada ao algoritmo, e uma nova instância (representada por um vetor a_i) é apresentada para que seja classificada. O treinamento deste classificador consiste na definição da função alvo $f(x) \rightarrow v$ capaz de mapear corretamente o valor alvo, ou classificação, para uma dada instância [24].

O Naive Bayes se baseia na suposição simplificada que os atributos são condicionalmente independentes dado o valor alvo (v). Em outras palavras, a suposição é que dado o valor alvo da instância i , a probabilidade da conjunção dos elementos de a_i ($\{a_{i,1}, a_{i,2}, \dots, a_{i,m}\}$) é dada pelo produto das probabilidades dos atributos individuais: $P(a_1, a_2 \dots a_n | v_j) = \prod_i P(a_i, v_j)$. Portanto a abordagem utilizada pelo Naive Bayes é demonstrada na equação 2.1 [24].

Classificador Naive Bayes:

$$v_{nb} = \arg \max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j) \quad (2.1)$$

Onde v_{nb} denota o valor alvo de saída dado pelo Naive Bayes. Como dito anteriormente, o método de aprendizagem Naive Bayes envolve um passo de aprendizagem

no qual os vários termos $P(v_j)$ e $P(a_i|v_j)$ são estimados, baseados nas suas frequências no conjunto de exemplos de treinamento e no teorema de Bayes, isto é feito da seguinte forma:

$$P(a_i | v_j) = \frac{P(v_j | a_i)P(a_i)}{P(v_j)} \quad (2.2)$$

Para ilustrar o processo de aprendizagem temos os dados da Tabela 2.1 que serão classificados utilizando o algoritmo Naive Bayes. Este exemplo trata de um conjunto de dados onde se deve dar a probabilidade do dia ser um bom dia para se jogar tênis dada a condição climática do mesmo (adaptado de [24]).

Tabela 2.1 - Jogar Tênis.

<i>Visual</i>	<i>Temperatura</i>	<i>Umidade</i>	<i>Vento</i>	<i>Jogar?</i>
Ensolarado	Quente	Alta	Fraco	Sim
Ensolarado	Quente	Alta	Forte	Não
Nublado	Quente	Alta	Fraco	Sim
Chuvoso	Frio	Alta	Fraco	Não

Supondo que os dados da Tabela 2.1 foram fornecidos ao Naive Bayes como conjunto de exemplos de treinamento e seja também fornecida a nova instância a ser classificada (Ensolarado,Frio,Alta,Forte). Neste exemplo, seguindo a descrição do algoritmo dada anteriormente, as instâncias são descritas por 4 atributos, ou seja, $x = \{a_1, a_2, a_3, a_4\}$, e o conjunto V de classes é composto por dois elementos $V = \{\text{Sim}, \text{Não}\}$. Para a nova instância dada o Naive Bayes vai calcular utilizando a equação (2.1) a probabilidade de Jogar = Sim e Jogar = Não para então decidir qual será o valor de saída, conforme os cálculos a seguir.

- $P(\text{Sim}|\text{Ensolarado},\text{Frio},\text{Alta},\text{Forte}) = P(\text{Sim}) * P(\text{Ensolarado}|\text{Sim}) * P(\text{Frio}|\text{Sim}) * P(\text{Alta}|\text{Sim}) * P(\text{Forte}|\text{Sim}) = 2/4 * 1/2 * 0/2 * 2/2 * 0/2 = 0.5 * 0.5 * 0.01 * 1 * 0.01 = 0.000025.$
- $P(\text{Não}|\text{Ensolarado},\text{Frio},\text{Alta},\text{Forte}) = P(\text{Não}) * P(\text{Ensolarado}|\text{Não}) * P(\text{Frio}|\text{Não}) * P(\text{Alta}|\text{Não}) * P(\text{Forte}|\text{Não}) = 2/4 * 1/2 * 1/2 * 2/2 * 1/2 = 0.5 * 0.5 * 0.5 * 1 * 0.5 = 0.0625.$

Tendo estas probabilidades calculadas o Naive Bayes tem como saída o mais provável que neste caso é “Não”. Portanto para esta nova instância a classificação realizada pelo Naive Bayes é que a instância de entrada pertence à classe “Não”. Caso também se queira saber qual

a probabilidade real das duas classes pode-se normalizar as probabilidades da seguinte maneira:

- $P(\text{Sim}|\text{Ensolarado}, \text{Frio}, \text{Alta}, \text{Forte}) = 0.000025 / (0.000025 + 0.0625) = 0.03\%$
- $P(\text{Não}|\text{Ensolarado}, \text{Frio}, \text{Alta}, \text{Forte}) = 0.0625 / (0.000025 + 0.0625) = 99.97\%$

O Naive Bayes nem sempre traz uma estimativa de probabilidade muito precisa de cada classe, mas tende a ter um nível de acerto muito bom quanto à ordenação das classes mais prováveis. O índice de acerto das classificações em alguns domínios se compara ao de Redes Neurais e Árvores de Decisão [24]. A razão pela qual a probabilidade estimada pelo Naive Bayes não ser sempre precisa é justamente a simplificação feita por ele que desconsidera as relações entre os atributos, assumindo que eles são condicionalmente independentes dada a classe.

2.2.1.2 C4.5

O aprendizado por árvores de decisão [33] é um dos métodos mais utilizados para inferência indutiva. É um método para aproximar funções de valores discretos tolerante a ruídos e capaz de aprender expressões disjuntivas [24]. Um dos algoritmos da família das árvores de decisão mais utilizados é o C4.5 [41], que será utilizado no presente trabalho.

No aprendizado por árvores de decisão a função aprendida é representada por meio de uma árvore de decisão, que também pode ser representada como um conjunto de regras *if-then* para facilitar sua compreensão. Uma árvore de decisão gerada para o exemplo da tabela 2.1 pode ser visto na figura 2.1.

Figura 2.1 - Árvore de Decisão.

Para gerar a árvore o algoritmo utiliza uma abordagem “*top-down*”, iniciando a construção partindo do topo, respondendo a seguinte questão “qual atributo deve estar na raiz da árvore?”. Para responder esta questão, cada atributo é testado utilizando um teste estatístico para descobrir quão bem este atributo consegue sozinho classificar os exemplos, então o atributo que se sai melhor neste teste é colocado na raiz da árvore. Uma aresta descendente da raiz é criada para cada valor possível deste atributo com os respectivos nós ligados a este atributo, sendo este nó novamente selecionado através de um teste estatístico levando em consideração o ponto da árvore no qual ele se encontra. Este processo é repetido até que as possibilidades estejam cobertas [24]. O produto final do algoritmo, como dito anteriormente, é um conjunto de regras e a árvore de decisão criada.

2.2.1.3 Support Vector Machines

O Support Vector Machines (SVM) [34] é um classificador linear que realiza a classificação construindo hiperplanos em um espaço multidimensional que separam as instâncias de acordo com sua classe. É um algoritmo baseado em fronteiras de decisão, onde objetos que pertencem a uma mesma classe pertencem a um mesmo lado do plano. A figura 2.2 mostra um exemplo de divisão realizada em um conjunto de dados de modo a classificá-los em dois grupos.

Figura 2.2 - Exemplo SVM.

Como o SVM é um classificador linear, ele somente consegue separar as instâncias utilizando uma reta [44], e como pode ser visto no exemplo da figura 2.2 os objetos podem facilmente ser divididos em dois grupos utilizando uma reta, entretanto nem sempre isto é possível, as vezes é necessária uma curva para realizar a divisão, como pode ser visto no exemplo da figura 2.3.

Figura 2.3 - Exemplo complexo SVM.

Como o exemplo da figura 2.3 mostra, os dados não podem ser divididos utilizando uma simples reta, então o SVM utiliza uma função matemática para distanciar os dois grupos de modo a poder realizar a divisão, esse processo é chamado de mapeamento ou transformação [44]. O resultado deste processo realizado no exemplo mostrado na figura 2.3 pode ser visto na figura 2.4.

Figura 2.4 - Transformação SVM.

Esse processo de transformação, ilustrado na figura 2.4, facilita a classificação pois ao invés de criar uma função que resulte numa curva complexa para classificar os diferentes objetos o SVM pode encontrar uma linha que realize esta classificação [44].

2.2.1.4 Adaboost

O algoritmo Adaboost (abreviação de “*Adaptive Boosting*”) [45] é um algoritmo de impulsioneamento (“*Boosting*”), que é uma técnica baseada na ideia de criar uma regra de predição precisa combinando várias regras relativamente fracas e imprecisas. O Adaboost foi o primeiro algoritmo prático de impulsioneamento, e é amplamente utilizado e aplicado em diversos campos [46]. A figura 2.5 mostra o funcionamento do algoritmo Adaboost.

Dado: $(x_1, y_1), \dots, (x_m, y_m)$ onde $x_i \in \mathcal{X}$, $y_i \in \{-1, +1\}$.

Inicialize: $D_1(i) = 1/m$ para $i = 1, \dots, m$.

Para $t = 1, \dots, T$:

- Treine um classificador fraco utilizando a distribuição D_t .
- Pegue a hipótese fraca $h_t : \mathcal{X} \rightarrow \{-1, +1\}$.
- Meta: selecionar h_t com um baixo erro:

$$\varepsilon_t = \Pr_{i \sim D_t} [h_t(x_i) \neq y_i].$$

- Escolha $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right)$.

- Atualize, para $i = 1, \dots, m$:

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

onde Z_t é um fator de normalização (escolhido de modo que D_{t+1} seja uma distribuição)

Retorne a hipótese final:

$$H(x) = \text{sgn} \left(\sum_{t=1}^T \alpha_t h_t(x) \right).$$

Figura 2.5 - Algoritmo Adaboost (Adaptado de [47]).

A figura 2.5 mostra o pseudocódigo para o Adaboost, nele se vê que o algoritmo recebe m exemplos de treinamento rotulados $(x_1, y_1), \dots, (x_m, y_m)$ onde os x_i pertencem a um domínio $\mathcal{X}$, e os rótulos $y_i \in \{-1, +1\}$. Em cada rodada $t = 1, \dots, T$, uma distribuição D_t é computada para cada um dos m exemplos de treinamento, e um algoritmo de aprendizagem qualquer é aplicado para encontrar uma hipótese $h_t: \mathcal{X} \rightarrow \{-1, +1\}$, onde o objetivo do algoritmo de aprendizagem é encontrar uma hipótese com um baixo erro ε_t referente a D_t . A hipótese final ou combinada H computa a função sinal de uma combinação balanceada das hipóteses [46].

$$F(x) = \text{sgn} \left(\sum_{t=1}^T \alpha_t h_t(x) \right). \quad (2.3)$$

Isto é equivalente a dizer que H é computada como um balanceamento das hipóteses h_t onde para cada hipótese é aplicado um peso α_t [46].

2.2.1.5 k-Nearest Neighbors

O algoritmo k -Nearest Neighbors (k -NN) [43] é um algoritmo baseado em instâncias, que ao contrário dos algoritmos tradicionais de aprendizagem que constroem uma descrição explícita da função alvo baseado nos exemplos de treinamento, simplesmente armazena os exemplos de treinamento e quando uma nova instância deve ser classificada, o algoritmo analisa sua relação (similaridade) com os exemplos de treinamento para determinar a classe à qual pertence. Por atrasar o processo até que uma nova instância tenha que ser classificada os algoritmos de aprendizagem baseados em instâncias são geralmente chamados de “preguiçosos” [24].

O algoritmo baseado em instâncias mais básico e utilizado é o k -NN. Este algoritmo assume que todas as instâncias correspondem a pontos em um espaço n -dimensional, então utiliza alguma medida de distância para determinar quais pontos (a quantidade depende do k escolhido) são mais similares, e portanto, estão mais próximos ao item que se deseja classificar [24]. Se o $k=1$ então será levado em consideração o vizinho mais próximo, se $k=3$ então serão levados em consideração os 3 vizinhos mais próximos, e assim por diante. Como pode ser observado na figura 2.6.

Figura 2.6 - Exemplo k -NN.

A figura 2.6 mostra um exemplo onde cada símbolo “+” é um exemplo positivo (pertence à classe que se deseja classificar) e cada símbolo “-” é um exemplo negativo (não pertence à classe que se deseja classificar). O símbolo “?” representa uma nova instância que se deseja classificar, no exemplo temos um $k=3$, portanto esta nova instância será classificada como pertencente à classe, pois dois dos seus três vizinhos mais próximos pertencem a classe.

2.3 Problemas multi-rótulo

Os algoritmos tradicionais de aprendizagem são preparados para trabalhar com um único rótulo (classe), ou seja, são bons em definir se um dado pertence ou não a uma classe, entretanto o problema do diagnóstico automático da DPC é um problema multi-rótulo pois um paciente pode ter uma ou mais doenças que combinados causam esta condição.

No início a classificação multi-rótulo foi motivada principalmente pelas tarefas de categorização de textos e diagnósticos médicos, como trata o presente trabalho, entretanto hoje é utilizada em diversos problemas de classificação.

Em [30] e [31], os autores fizeram um survey de abordagens de classificação multi-rótulo e mostraram os métodos de transformação de problemas mais utilizados para adaptar problemas multi-rótulo a algoritmos tradicionais, os cinco métodos que podem ser utilizados no problema proposto no presente trabalho serão abordados nesta seção. A tabela 2.2 mostra um conjunto de dados adaptado de [30] que será utilizado para demonstração destes métodos de transformação.

Tabela 2.2 - Exemplo de um conjunto de dados multi-rótulo.

<i>Ex.</i>	<i>Endometriose</i>	<i>DIP</i>	<i>Constipação</i>	<i>Bexiga Dolorosa</i>
1	X			X
2			X	X
3	X			
4		X	X	

A tabela 2.2 mostra 4 linhas (pacientes) e 4 colunas (doenças). O paciente da linha 1 tem endometriose e síndrome da bexiga dolorosa, o paciente da linha 2 tem constipação e

síndrome da bexiga dolorosa, o paciente da linha 3 tem somente endometriose e o paciente da linha 3 tem DIP (doença inflamatória pélvica) e constipação. Todos os exemplos posteriores utilizam a tabela 2.2 como base.

Os primeiros dois métodos de transformação de problemas foram descritos em [30], [31] e [36]. No primeiro método de transformação (PT1) a abordagem é aleatoriamente descartar os múltiplos rótulos de cada instância multi-rótulo deixando-as somente com um rótulo cada, enquanto o segundo método de transformação (PT2) é feito pelo simples descarte de cada instância multi-rótulo do conjunto de dados. As tabelas 2.3 e 2.4 mostram o conjunto de dados da tabela 2.2 utilizando os métodos PT1 e PT2 respectivamente.

Tabela 2.3 - Conjunto transformado utilizando PT1.

<i>Ex.</i>	<i>Endometriose</i>	<i>DIP</i>	<i>Constipação</i>	<i>Bexiga Dolorosa</i>
1	X			
2				X
3	X			
4			X	

Tabela 2.4 - Conjunto transformado utilizando PT2.

<i>Ex.</i>	<i>Endometriose</i>	<i>DIP</i>	<i>Constipação</i>	<i>Bexiga Dolorosa</i>
3	X			

Estes dois métodos de transformação não foram utilizados pois descartam muita informação que pode ser útil para o processo de diagnóstico, e ao descartar esta informação o algoritmo pode ignorar o diagnóstico que pode ser a real causa da DPC em determinado paciente, tornando o diagnóstico impreciso.

O terceiro método de transformação (PT3), descrito em [30], [31], [36] e [37], considera cada diferente conjunto de rótulos das instancias multi-rótulo existente no conjunto de dados e cria um novo rótulo combinando os rótulos originais. A tabela 2.5 mostra o resultado desta transformação.

Tabela 2.5 - Conjunto transformado utilizando PT3.

<i>Ex.</i>	<i>Endometriose</i>	<i>DIP</i>	$\wedge$	<i>Constipação</i>	$\wedge$	<i>Bexiga Dolorosa</i>	$\wedge$

	<i>Constipação</i>	<i>Bexiga Dolorosa</i>	<i>Endometriose</i>
1			X
2		X	
3	X		
4		X	

Este método tem como resultado um conjunto de dados com um grande conjunto de rótulos, que pode não abranger todas as combinações possíveis de rótulos, o que poderia atrapalhar a precisão do classificador e inibir o diagnóstico de algumas combinações de doenças.

O quarto método de transformação (PT4) foi demonstrado em [30], [31], [36], [38], [39] e [40], e consiste em dividir o conjunto original em n conjuntos (onde n é o número de classes) contendo todos os exemplos do conjunto original rotulados como l se os exemplos originais continham l e $\neg l$ caso contrário. Então um classificador de rótulo único é construído para cada novo conjunto de dados. As tabelas 2.6, 2.7, 2.8 e 2.9 mostram os novos conjuntos criados por PT4.

Tabela 2.6 - Conjunto para endometriose transformado utilizando PT4.

<i>Ex.</i>	<i>Endometriose</i>	$\neg$ <i>Endometriose</i>
1	X	
2		X
3	X	
4		X

Tabela 2.7 - Conjunto para DIP transformado utilizando PT4.

<i>Ex.</i>	<i>DIP</i>	$\neg$ <i>DIP</i>
1		X
2		X
3		X
4	X	

Tabela 2.8 - Conjunto para constipação transformado utilizando PT4.

<i>Ex.</i>	<i>Constipação</i>	<i>¬Constipação</i>
1		X
2	X	
3		X
4	X	

Tabela 2.9 - Conjunto para bexiga dolorosa transformado utilizando PT4.

<i>Ex.</i>	<i>Bexiga Dolorosa</i>	<i>¬Bexiga Dolorosa</i>
1	X	
2	X	
3		X
4		X

Esta abordagem é uma das duas abordagens utilizadas para a experimentação pois não ignora nenhum dado e permite que a classificação seja realizada utilizando algoritmos de rótulo único, tais como o Naive Bayes.

A última abordagem que será discutida (PT5) foi demonstrada em [30] e consiste em decompor cada exemplo e n novos exemplos (onde n é o número de classes deste exemplo), então um algoritmo de rótulo único é utilizado para a classificação neste conjunto de dados transformado. A tabela 2.10 mostra o novo conjunto de dados criado por PT5.

Tabela 2.10 - Conjunto transformado utilizando PT5.

<i>Ex.</i>	<i>Endometriose</i>	<i>DIP</i>	<i>Constipação</i>	<i>Bexiga Dolorosa</i>
1	X			
1				X
2			X	
2				X
3	X			
4		X		
4			X	

Esta abordagem também foi utilizada na experimentação pois não ignora nenhum dado e também permite que os dados sejam classificados utilizando um método de classificação de rótulo único.

2.4 Seleção de Atributos

O processo de seleção de atributos é considerado um passo importante do processo de classificação pois melhora a performance do classificador e reduz a complexidade computacional removendo informação redundante ou irrelevante [48]. Neste contexto o presente trabalho utiliza tais técnicas para melhorar o percentual de acerto do diagnóstico e também para verificar quais variáveis são realmente importantes para o processo de classificação de cada uma das possíveis causas da DPC. O método de seleção de atributos utilizado no presente trabalho é o InfoGain (*Information Gain Based Attribute Selection*) que será apresentado na subseção 2.4.1.

2.4.1 Information Gain Based Attribute Selection

O InfoGain [24] é um algoritmo que cria um ranking dos atributos ou subconjuntos de atributos calculando o ganho de informação deste atributo em relação a uma classe. O ganho de informação é uma medida de comparação entre duas distribuições de probabilidade, o algoritmo avalia um atributo X medindo o quanto ela agrega em informação em relação a uma classe Y como pode ser visto na fórmula 2.4 [47].

$$I(X) = H(P(Y)) - H(P(Y/X)) \quad (2.4)$$

$I(X)$ é o ganho de informação do atributo X , $H(P(Y))$ é a entropia condicional da distribuição de probabilidade de Y e $H(P(Y/X))$ é a entropia condicional da distribuição de probabilidade da probabilidade condicional de Y dado X .

De modo resumido a equação 2.4 mede a diferença entre a distribuição marginal de Y assumindo que este é independente do atributo X e a distribuição condicional de Y assumindo

que ele é dependente de X . Se Y for independente de X este terá um valor pequeno de ganho de informação e vice-versa [47].

Este foi o algoritmo escolhido pois se provou muito eficiente em melhorar a performance de classificadores e em comparação com outros métodos de seleção se sobressai em relação à performance final do algoritmo. Em [47] os autores comparam 6 métodos de seleção de atributos (CfsSubsetEval, Componentes Principais, ConsistencySubsetEval, InfoGainAttributeEval, OneRAttributeEval e ReliefAttributeEval), onde primeiramente os atributos foram selecionados utilizando estes algoritmos e posteriormente se utilizou os atributos selecionados em 4 algoritmos de classificação diferentes (J48, Naive Bayes, IBK e Decision Table) comparando suas performances. De todos os testes o algoritmo InfoGain somente não foi o melhor quando se utilizou o classificador IBK, mas mesmo neste caso ele foi o segundo melhor.

2.5 K-fold Cross Validation

O k-fold cross validation é um método de validação de performance de algoritmos, onde para cada execução do algoritmo, k casos aleatórios são removidos da base de treinamento para posterior validação [57]. Nesta etapa de validação se utiliza o conhecimento adquirido pelo algoritmo para classificar estes casos que foram removidos da base e comparar a classe à qual o caso pertence com a classe estimada pelo algoritmo.

Por exemplo, caso o conjunto de dados tenha 200 casos e se utilize um k igual a 10, o conjunto de dados é dividido então em 10 blocos de 20 casos, então na primeira rodada o primeiro bloco de 20 é retirado para validação e o algoritmo é treinado com os outros 9 blocos de 20 casos (180 casos no total), e assim sucessivamente até que todos os blocos tenham sido utilizados para ambas as finalidades. Ao final a média das taxas de acerto é calculada e utilizada como performance do algoritmo.

A escolha do k é uma escolha arbitrária, um k maior implica em uma melhor validação, e um k menor implica em uma melhor aprendizagem (pois o conjunto de treinamento será maior) [58], desta maneira cada caso deve ser avaliado e o k escolhido de acordo com o conjunto de dados existente. Este é o método utilizado no presente trabalho para validação dos resultados encontrados.

2.6 Considerações Finais

Neste capítulo foram mostrados os conceitos fundamentais para o entendimento do trabalho passando pelos tipos de aprendizagem, onde alguns algoritmos utilizados pelo presente trabalho foram discutidos, o conceito de problema multi-rótulo, o tema seleção de atributos e por fim a validação pelo método k-fold cross validation.

Capítulo 3

TRABALHOS CORRELATOS

Neste capítulo são abordados alguns trabalhos correlatos importantes para o entendimento do que tem sido pesquisado nesta área de detecção de contradições e importantes para o presente trabalho.

Na área de diagnóstico automático foram encontrados vários artigos que utilizam IA aplicada a área médica, sendo que são descritos aqui alguns dos que estão mais próximos do domínio desta tese de doutorado, ou seja, diagnóstico automático de doenças.

3.1 Diagnóstico Automático

Na área de diagnóstico automático foram encontrados diversos trabalhos recentes, dentre eles serão discutidos nesta seção os trabalhos [49], [54], [56]. Em [49] os autores utilizam o algoritmo C4.5 de árvores de decisão para o diagnóstico da tuberculose. Neste estudo foram estudados 238 pacientes adultos com efusão pleural, que foram diagnosticados com tuberculose (64 pacientes) ou com efusão maligna (174 pacientes), destes pacientes foram colhidas 12 variáveis, sexo, idade, temperatura, tamanho da efusão, lateralidade da efusão, contagem total de leucócitos, contagem de células brancas diferenciais, glicose, proteína, desidrogenase de lactato (LDH), ADA e pH.

O algoritmo foi implementado utilizando o software Weka 3.4 [50]. Dos 12 atributos disponíveis o algoritmo selecionou 4 atributos como preditivos para tuberculose, ADA ($\geq 40\text{U/L}$), idade (≤ 35 anos), temperatura ($\geq 37.8^\circ\text{C}$) e contagem de células vermelhas pleurais

($\leq 5 \times 10^9/L$). Com a utilização deste sistema os autores conseguiram um índice de acerto de 97.45%, com 92.2% de sensibilidade e 98.3% de especificidade.

Os autores afirmam que baseados em dados clínicos disponíveis foi possível desenvolver um modelo de decisão muito simples e preciso para o diagnóstico de tuberculose e efusão maligna utilizando o algoritmo C4.5. Como limitações os autores citam que não foram incluídos outros tipos de efusão pleural não tuberculosas, tais como, pneumonia.

No trabalho [54], os autores utilizam o algoritmo Naive Bayes para diagnosticar a diabetes tipo 2, como justificativa os autores utilizam o dado de que 3.8 milhões de pessoas morrem anualmente dos problemas provenientes da diabetes e um sistema de diagnóstico da diabetes poderia auxiliar no diagnóstico desta condição, especialmente em áreas remotas onde não há médicos e o resultado do diagnóstico pode ser imediato.

Os autores utilizaram um conjunto de dados com 415 pacientes e 10 atributos cada, idade, histórico familiar, sexo, fumante, faz utilização de álcool, fadiga, sede, frequência de urinação, altura e peso.

O algoritmo foi desenvolvido no software Matlab [55] e em seus testes os autores conseguiram um índice de acerto de 95%, desta forma concluem que o classificador Naive Bayes pode ser muito útil no diagnóstico da diabetes, e seus resultados sugerem que este sistema pode ser importante no auxílio aos médicos no diagnóstico.

Em [56] os autores fazem um comparativo de 15 algoritmos também no diagnóstico de diabetes. Foram utilizados os algoritmos Ensemble based Diabetes Diagnoser, Artificial Neural Network, Naive Bayes, Support Vector Machine, J48 Graft, K-Nearest Neighbor, Decorate, END, Random Forest, Bagging, Multi class classifier, Decision stump, Multi boost, User Classifier e Random Tree.

Os autores utilizaram uma base de dados com 10 variáveis (idade, histórico familiar, sexo, fumante, faz utilização de álcool, fadiga, sede, frequência de urinação, altura e peso) de 400 pacientes. Os algoritmos foram implementados utilizando o software Matlab [55] e o resultado da precisão dos algoritmos pode ser visto na tabela 3.1.

Tabela 3.1 - Comparativo de Algoritmos [56].

<i>Algorithms used</i>	<i>Correctly classified</i>	<i>Incorrectly classified</i>
<i>Ensemble based Diabetes Diagnoser</i>	98,60%	1,40%
<i>Artificial Neural Network</i>	96,00%	4,00%
<i>Naïve Bayes</i>	95,00%	5,00%
<i>Support Vector Machine</i>	94,00%	6,00%
<i>J48 Graft</i>	91,49%	8,51%

<i>K- nearest neighbor</i>	91,23%	8,77%
<i>Decorate</i>	91,23%	8,77%
<i>END</i>	91,23%	8,77%
<i>Random forest</i>	90,97%	9,03%
<i>Bagging</i>	89,69%	10,31%
<i>Multi class classifier</i>	89,69%	10,31%
<i>Decision stump</i>	88,65%	11,35%
<i>Multi boost</i>	88,65%	11,35%
<i>User Classifier</i>	88,65%	11,35%
<i>Random Tree</i>	88,40%	11,60%

Como pode ser observado na tabela 3.1 os autores conseguiram uma precisão de 98% com um algoritmo composto (Ensemble based Diabetes Diagnoser), entretanto mesmo com algoritmos simples como o Naive Bayes a precisão foi alta. Desta forma os autores afirmam que é possível desenvolver uma ferramenta de diagnóstico para diabetes com uma boa precisão.

Já em [59] os autores fazem um resumo de 29 trabalhos que utilizam inteligência artificial para o diagnóstico de doenças como artrite reumatoide, câncer, doenças do pulmão, doenças cardíacas, diabetes, hepatite, Alzheimer, doenças do fígado, dengue e parkinson.

Destes trabalhos em 6 dos 8 melhores trabalhos apresentados são utilizados os mesmos algoritmos utilizados no presente trabalho, sendo que nos diversos trabalhos demonstrados o algoritmo Naive Bayes foi o que apresentou a melhor performance em vários casos, em alguns chegou à taxa de acerto de 96% como no caso do diagnóstico de hepatite e também no caso de doenças relacionadas ao fígado.

3.2 Considerações Finais

Neste capítulo alguns trabalhos correlatos foram discutidos, onde foi possível observar o que já foi realizado na área de diagnostico automático de doenças e as tendências desta área, corroborando a escolha dos algoritmos escolhidos no presente trabalho.

Capítulo 4

DIAGNOSTICANDO A CAUSA DA DPC

Neste capítulo é apresentada a metodologia utilizada, a implementação dos métodos utilizados e a experimentação realizada com a finalidade da validação do trabalho.

4.1 Considerações Iniciais

Neste capítulo é apresentada a proposta de trabalho de doutorado desenvolvida. Também é apresentada a metodologia onde são detalhadas as tarefas propostas e como elas foram realizadas para se alcançar os objetivos. Por fim a implementação dos métodos propostos é demonstrada bem como a experimentação destes métodos.

4.2 Metodologia e Desenvolvimento

Como anteriormente descrito, a proposta deste trabalho é investigar, implementar e avaliar métodos que permitam diagnosticar a causa da DPC em pacientes do HC-FMRP. A experimentação consistiu de 3 partes, primeiramente a escolha do método de transformação a ser utilizado, posteriormente a seleção dos atributos e pôr fim a execução dos algoritmos nas bases de dados geradas.

Todos os passos foram executados no software Weka 3.8 [50], que é um software que contém diversos algoritmos de inteligência artificial e tratamento de dados disponíveis tanto em sua interface gráfica quanto como API para integração com outros sistemas. Este software

foi desenvolvido e é mantido pelo grupo de aprendizagem de máquina da Universidade de Waikato na Nova Zelândia.

4.2.1 Validação dos resultados

Como dito anteriormente, a validação da performance dos algoritmos se deu pelo método da validação cruzada (k-fold cross validation) [57], onde para cada execução do algoritmo, k casos aleatórios são removidos da base de treinamento para posterior validação. No presente trabalho o k utilizado foi 10, portanto em cada execução 10 casos eram removidos da base de treinamento para que pudessem ser utilizados para a validação.

4.3 Metodologia e Desenvolvimento

A experimentação foi realizada em etapas, primeiramente foram comparados os resultados dos métodos de transformação PT4 e PT5 com a finalidade de descobrir qual se mostra mais adequado para o problema, posteriormente realizou-se a seleção de atributos para que somente as variáveis com capacidade preditiva para cada uma das doenças fosse utilizada e finalmente utilizando o método de transformação mais adequado foram comparadas as precisões de 5 algoritmos diferentes com a finalidade de descobrir qual se adaptava melhor ao problema resultando em um maior índice de acerto.

4.3.1 Métodos de transformação

Como dito na seção 2.3 os métodos escolhidos para a experimentação foram o PT4 e PT5 pois ambos não implicavam em perda de informação da base de dados. Para a execução do método PT5 como demonstrado na seção 2.3 cada caso que continha mais de um diagnóstico gerou n casos, onde n é o número de diagnósticos daquele caso, desta forma houve um aumento no número de casos na base de dados.

Após a alteração na base o algoritmo Naive Bayes foi executado para que a sua performance pudesse ser utilizada para comparar a eficácia dos dois métodos de transformação. Para o método PT5 o percentual de acerto obtido foi de 45%.

Para a execução do método PT4 da base de dados original foram geradas 7 bases de dados, um para cada possível diagnóstico (constipação, DIP, endometriose, neuralgias, síndrome da bexiga dolorosa, síndrome da congestão pélvica e síndrome do intestino irritável) e para cada uma das bases o algoritmo Naive Bayes foi executado e o resultado pode ser visto na figura 4.1.

Figura 4.1 - Experimentação PT4.

Como pode ser visto em todos os casos a performance foi melhor que a execução do algoritmo no método PT5, sendo a precisão média de 80%, desta forma chegou-se a conclusão que o método mais adequado para o presente problema é o método PT4, portanto este é o método final utilizado para a construção do classificador.

4.3.2 Seleção de Atributos

O segundo passo realizado foi a seleção de atributos, neste passo todas as bases de dados geradas no método PT4 foram submetidas ao algoritmo InfoGain para que este possa selecionar os atributos mais adequados para cada caso. Inicialmente haviam 108 variáveis que

estão descritas no apêndice A e após a execução da seleção de atributos este número foi bastante reduzido em alguns casos, o resultado desta seleção pode ser visto na figura 4.2.

Figura 4.2 - Seleção de Atributos.

Como pode ser observado na figura 4.2, uma quantidade diferente de atributos foi selecionada para cada caso, entretanto em todos os casos a quantidade de atributos foi menos de 30% da quantidade total de atributos disponíveis que era de 109. Os atributos selecionados para cada diagnóstico estão descritos abaixo:

- Constipação
  - ANT_fantasias_sexuais
  - INT_dor_pelvica_subitos_cessam_espontaneamente
  - INT_SI_constipacao_intestinal
  - INT_SI_diarreia
  - INT_volume_fluxo
  - INT_Diagnostico
- DIP
  - ANT_doencas
  - INT_durante_ato_sexual_dor_vaginal_prof_ou_dor_pelvica

-
- INT_EF_dor_palpacao_anexos
  - INT_EF_sinais_colpите_cervicite
  - INT_intervalo_intermenstrual_dias
  - ND_pelvica_apos_ato_sexual_horas_dias
  - Tem_DPC
  - INT_Diagnostico
  - Endometriose
    - ANT_atividade_sexual
    - ANT_atv_fisica_regular_freq
    - ANT_desejo
    - ANT_espіotomias
    - ANT_excitacao
    - ANT_forceps
    - ANT_idade_menopausa
    - ANT_incisao_cirurgica
    - ANT_n_gestacoes
    - ANT_n_partos
    - ANT_status_menstrual
    - Auto_medicacao
    - SCORE_escolaridade
    - Estado_Civil
    - Idade_CN
    - INT_ASF_vitima_abuso_emocional_humilhacao_insulto
    - INT_contracepcao
    - INT_dismenorreia
    - INT_dor_pelvica_molhora_repouso
    - INT_dor_pelvica_piora_apos_atv_fisica_prolongada
    - INT_dor_pelvica_subitos_cessam_espontaneamente
    - INT_duracao_fluxo_menstrual
    - INT_EF_alteracao_septo_reto_vaginal
    - INT_EF_carnett
    - INT_EF_dor_referida
    - INT_EF_vulvodinea
    - INT_SI_constipacao_intestinal

-
- INT_SI_distensao_abdominal
    - ND_colicas_durante_menstruacao
    - Tem_DPC
    - Tipo_dor
    - Tratamento_medico_tentou_anteriormente
    - INT_Diagnostico
  - Neuralgias
    - ANT_n_partos
    - INT_dor_pelvica_piora_apos_atv_fisica_prolongada
    - INT_EF_alodinea
    - INT_EF_pto_gatilho
    - INT_EF_sensibilidade_area
    - INT_SI_inicio_mudanca_frequencia_evacuacao
    - Tipo_dor
    - INT_Diagnostico
  - Bexiga Dolorosa
    - SCORE_escolaridade
    - INT_duracao_fluxo_menstrual
    - INT_SI_diarreia
    - INT_SU_Geral
    - INT_TB_escore_numeric
    - Total_PHQ
    - INT_Diagnostico
  - Congestão Pélvica
    - ANT_atividade_sexual
    - ANT_forceps
    - Fatores_piora_DPC
    - Idade_CN
    - INT_apos_ato_sexual_dor_latejamento
    - INT_contracepcao
    - INT_dor_pelvica_altera_lado
    - INT_EF_alteracao_septo_reto_vaginal
    - INT_EF_palpacao_massas
    - INT_EF_sensibilidade_area

- INT_EF_vaginismo
- INT_SI_diarreia
- INT_SI_inicio_mudanca_forma_fezes
- INT_SI_inicio_mudanca_frequencia_evacuacao
- Tratamento_medico_tentou_anteriormente
- INT_Diagnostico
- Intestino Irritável
  - INT_dor_pelvica_subitos_cessam_espontaneamente
  - INT_SI_dor_melhora_defecacao
  - INT_SI_inicio_mudanca_forma_fezes
  - INT_SI_inicio_mudanca_frequencia_evacuacao
  - Tem_DPC
  - Total_PHQ
  - Total_SPIN
  - INT_Diagnostico

4.3.3 Performances dos Algoritmos

Após a escolha do método PT4 como mais apropriado ao problema e da execução da seleção de atributos, as bases de dados foram modificadas para conter somente os atributos selecionados e após esta etapa foi realizada a execução dos algoritmos nas diversas bases de dados onde foi possível chegar ao resultado ilustrado na tabela 4.1.

Tabela 4.1 - Precisão do Diagnóstico.

	<i>Naive Bayes</i>	<i>C4.5</i>	<i>SVM</i>	<i>AdaBoost</i>	<i>kNN</i>
<i>Constipação</i>	92.17%	92.46%	92.75%	92.75%	91.3%
<i>DIP</i>	90.72%	91.59%	92.17%	91.01%	89.27%
<i>Endometriose</i>	76.30%	66.95%	67.82%	70.14%	66.08%
<i>Neuralgias</i>	93.04%	92.17%	91.59%	91.59%	91.59%
<i>Bexiga Dolorosa</i>	84.05%	87.24%	89.27%	86.37%	84.05%
<i>Congestão Pélvica</i>	83.47%	91.59%	91.01%	91.03%	85.79%

<i>Intestino Irritável</i>	99.71%	99.42%	99.71%	100%	99.71%
<i>Precisão Média</i>	87.78%	88.77%	89.19%	89.02%	86.83%

Como pode ser visto a precisão dos algoritmos foi similar. O algoritmo que teve a melhor performance na média foi o SVM, tendo a melhor performance em 3 dos 7 possíveis diagnósticos, entretanto a média de precisão de todos os algoritmos foi similar não excedendo uma diferença de 3 pontos percentuais do pior para o melhor. Este resultado pode ser visto graficamente na figura 4.3.

Figura 4.3 - Comparação das Performances.

Outras informações importantes extraídas dos testes foram a sensibilidade e especificidade de cada um dos classificadores. As sensibilidades resultantes podem ser vistas na tabela 4.2 e graficamente na figura 4.4.

Tabela 4.2 - Sensibilidades.

	<i>Naive Bayes</i>	<i>C4.5</i>	<i>SVM</i>	<i>AdaBoost</i>	<i>kNN</i>
<i>Constipação</i>	0.441	0.529	0.529	0.529	0.441
<i>DIP</i>	0.355	0.194	0.258	0.194	0.29
<i>Endometriose</i>	0.782	0.554	0.605	0.633	0.633
<i>Neuralgias</i>	0.6	0.12	0.24	0.28	0.32
<i>Bexiga Dolorosa</i>	0.796	0.429	0.429	0.449	0.49
<i>Congestão Pélvica</i>	0.241	0	0	0	0.138
<i>Intestino Irritável</i>	1	1	1	1	0.97
<i>Precisão Média</i>	0.602	0.403	0.437	0.440	0.468

Figura 4.4 - Comparação das Sensibilidades.

Como pode ser visto as sensibilidades não foram similares entre os classificadores. O algoritmo com as melhores sensibilidades foi o Naive Bayes, tendo a maior média de sensibilidades, sendo o melhor em 6 dos 7 possíveis diagnósticos. As especificidades podem ser vistas na tabela 4.3 e graficamente na figura 4.5.

Tabela 4.3 - Especificidades.

	<i>Naive Bayes</i>	<i>C4.5</i>	<i>SVM</i>	<i>AdaBoost</i>	<i>kNN</i>
<i>Constipação</i>	0.974	0.968	0.971	0.971	0.965
<i>DIP</i>	0.962	0.987	0.987	0.981	0.952
<i>Endometriose</i>	0.662	0.763	0.732	0.753	0.682
<i>Neuralgias</i>	0.956	0.984	0.969	0.966	0.963
<i>Bexiga Dolorosa</i>	0.848	0.946	0.97	0.932	0.899
<i>Congestão Pélvica</i>	0.899	1	0.994	0.997	0.924
<i>Intestino Irritável</i>	0.997	0.994	0.997	1	1
<i>Precisão Média</i>	0.898	0.948	0.945	0.942	0.912

Figura 4.5 - Comparação das Especificidades.

Como pôde ser visto as especificidades foram similares, entretanto o algoritmo com as melhores especificidades foi o C4.5, tendo a maior especificidade média, sendo o melhor em 3 dos 7 casos.

4.4 Estudo de caso: Diagnóstico da Endometriose

Após a realização dos testes acima, foi realizado um novo experimento focado no diagnóstico da endometriose. Com base nas 33 variáveis escolhidas pelo algoritmo de seleção de atributos removendo manualmente com o auxílio de um especialista e realizando testes foi possível chegar no número de 12 variáveis sem que a performance do algoritmo fosse prejudicada, estas 12 variáveis selecionadas foram:

- ANT_epsiotomias
  - Histórico de episiotomias
- ANT_forceps
  - Histórico de forceps
- INT_EF_carnett
  - Sinais de Carnett
- INT_EF_sinais_colpите_cervicite
  - Sinais de colpите e cervicite
- INT_fluxo_menstrual
  - Volume do fluxo menstrual
- ND_profunda_durante_ato_sexual_cat
  - Intensidade da dispareunia
- Deflag_menstrual
  - Piora relacionada a menstruação
- Relaciona
  - Fatores que relaciona o inicio dos sintomas
- Deflag_ovulacao
  - Piora relacionada a ovulação
- Neuralgias
  - Presença de dor neuropática
- Síndrome_Bexiga_Dolorosa
  - Presença de escores altos de sintomas urinários
- ANT_infertilidade
  - Infertilidade diagnosticada
- INT_Diagnostico
  - Diagnóstico de endometriose

Com estas variáveis o algoritmo foi executado novamente para garantir que a performance não foi prejudicada e o resultado foi uma taxa de acerto de 74.4%, o que não é tão impactante visto que com as 33 variáveis a taxa de acerto foi de 76.3%, portanto uma redução de somente 1.9%. A tabela 4.4 mostra a análise das variáveis contidas na base de dados, nesta tabela é possível verificar o perfil das pacientes contidas na mesma.

Tabela 4.4 - Análise das Variáveis.

Variável	Endometriosis (n=144)	No Endometriosis (n=154)
ANT_atividade_sexual		
Sim	137/144 [95%]	151/154 [98%]
ANT_atv_fisica_regular_freq		
1 a 2 vezes por semana	5/144 [3.4%]	5/154 [3.2%]
3 a 4 vezes por semana	12/144 [8.3%]	13/154 [8.4%]
5 a 7 vezes por semana	14/144 [9.7%]	9/154 [5.8%]
nao realiza	102/144 [71%]	119/154 [77.2%]
raramente	11/144 [7.6%]	8/154 [5.4%]
ANT_cirurgia_abdominal		
Sim	100/144 [70%]	114/154 [74%]
ANT_coitarca (média)	17.00	16.58
ANT_complicacoes_puerperais		
Sim	9/144 [6.3%]	17/154 [11%]
ANT_desejo		
Sim	38/144 [26.4%]	50/154 [32.4%]
ANT_epsiotomias		
Sim	28/144 [19.4%]	81/154 [52.6%]
ANT_excitacao		
Sim	116/144 [80.6%]	50/154 [32.4%]
ANT_fantasias_sexuais		
Sim	61/144 [42.4%]	110/154 [71.4%]
ANT_forceps		
Sim	2/144 [1.4%]	136/154 [88.3%]
ANT_freq_sexual		
1 a 2 vezes por semana	47/144 [32.6%]	45/154 [29.2%]
3 a 4 vezes por semana	30/144 [20.8%]	30/154 [19.5%]
5 a 7 vezes por semana	6/144 [4.2%]	8/154 [5.2%]
menos de 1 vez por semana	30/144 [20.8%]	40/154 [26%]
nenhuma	31/144 [21.6%]	31/154 [20.1%]
ANT_cafeina		
Sim	0/144 [0%]	0/154 [0%]
ANT_tabagismo		
Sim	0/144 [0%]	0/154 [0%]
ANT_etilismo		
Sim	0/144 [0%]	0/154 [0%]
ANT_drogas_ilicitas		
Sim	0/144 [0%]	0/154 [0%]
ANT_idade_menarca (média)	12.24	12.52
ANT_idade_menopausa (média)	0.0	0.0
ANT_status_menstrual		
menacme	144/144 [100%]	154/154 [100%]
ANT_impulso_sexual		

	Sim	99/144 [68.8%]	88/154 [57.2%]
ANT_incisao_cirurgica			
	nenhuma	46/144 [32%]	63/154 [41%]
	transversal	45/144 [31.2%]	40/154 [26%]
	longitudinal	17/144 [11.8%]	14/154 [9%]
	laparoscopia	28/144 [19.5%]	18/154 [11.6%]
	oblíqua	0/144 [0%]	7/154 [4.5%]
	outras	8/144 [5.5%]	12/154 [7.9%]
ANT_lubrifi_adequada		112/144	
	Sim	[78.8%]	104/154 [67.6%]
ANT_masturbacao			
	Sim	37/144 [25.7%]	34/154 [22%]
ANT_n_abortos (média)		0.3819	0.3442
ANT_n_cesareas (média)		0.7222	1.019
ANT_n_gestacoes (média)		1.0	2.396
ANT_n_parceiros (média)		2.0	2.981
ANT_n_partos (média)		1.0	2.026
ANT_orgasmo		106/144	
	Sim	[73.7%]	104/154 [67.6%]
ANT_sonhos_eroticos			
	Sim	60/144 [41.7%]	53/154 [34.4%]
Auto_medicao		106/144	
	Sim	[73.7%]	104/154 [67.6%]
Com_quem_vive_cat			
	companheiro e filhos	56/144 [38.8%]	87/154 [56.5%]
	apenas companheiro	37/144 [25.7%]	26/154 [16.8%]
	com os pais	18/144 [12.5%]	7/154 [4.5%]
	com os filhos	16/144 [11.1%]	17/154 [11%]
	sozinha	4/144 [2.7%]	3/154 [1.9%]
	outros	13/144 [9.2%]	14/154 [9.3%]
SCORE_escolaridade (média)		3.472	2.662
Estado_Civil			
	casado	67/144 [46.5%]	86/154 [55.9%]
	solteiro	32/144 [22.2%]	21/154 [13.6%]
	relacionamento estavel	30/144 [20.8%]	31/154 [20.1%]
	divorciado	8/144 [5.5%]	7/154 [4.6%]
	separado	5/144 [3.4%]	4/154 [2.6%]
	viuvo	2/144 [1.6%]	2/154 [1.2%]
	outro	0/144 [0%]	3/154 [2%]
Fatores_melhora_DPC			
	analgésicos	61/144 [42.2%]	52/154 [33.8%]
	nenhum	34/144 [23.6%]	42/154 [27.2%]
	deitar	0/144 [0%]	8/154 [5.2%]
	outros	49/144 [34.2%]	52/154 [33.8%]

Fatores_piora_DPC			
	nenhum	43/144 [29.8%]	31/154 [20.1%]
	relação sexual	11/144 [7.6%]	16/154 [10.3%]
	exercício	0/144 [0%]	9/154 [5.8%]
	bexigacheia	0/144 [0%]	3/154 [1.9%]
	estresse	3/144 [2%]	0/154 [0%]
	exercício e relação sexual	3/144 [2%]	0/154 [0%]
	relação sexual e bexiga cheia	3/144 [2%]	0/154 [0%]
	outros	81/144 [56.6%]	95/154 [61.9%]
FPS_Valor (média)		5.653	6
Idade_CN (média)		36.28	38.27
INT_apos_ato_sexual_dor_latejamento			
	Sim	65/144 [45.2%]	95/154 [61.7%]
INT_ASF_vitima_abuso_emocional_humilhacao_insulto			
	Sim	19/144 [13.2%]	28/154 [18.3%]
	No	119/144 [82.7%]	121/154 [78.5%]
	Não respondeu	6/144 [4.1%]	5/154 [3.2%]
INT_ASF_vitima_abuso_fisico_sexual			
	Sim	12/144 [8.3%]	20/154 [13%]
	No	125/144 [86.9%]	129/154 [83.8%]
	Não respondeu	7/144 [4.8%]	5/154 [3.2%]
INT_contracepcao			
	nenhum	50/144 [34.9%]	34/154 [22%]
	pílula	46/144 [32%]	36/154 [23.3%]
	laqueadura	14/144 [9.7%]	27/154 [17.5%]
	condom	9/144 [6.2%]	0/154 [0%]
	injetável	9/144 [6.2%]	22/154 [14.2%]
	vasectomia	5/144 [3.4%]	9/154 [5.8%]
	outro	11/144 [7.6%]	26/154 [17.2%]
ND_colicas_durante_menstruacao (média)		5.771	3.448
INT_dismenorreia_tripla_checagem			
	leve	11/144 [7.6%]	17/154 [11%]
	moderada	35/144 [24.3%]	32/154 [20.7%]
	nenhuma	20/144 [13.8%]	60/154 [39%]
	severa	78/144 [54.3%]	45/154 [29.3%]
INT_dismenorreia_dupla_checagem			
	leve	24/144 [16.6%]	27/154 [17.5%]
	moderada	53/144 [36.8%]	41/154 [26.6%]
	nenhuma	20/144 [13.8%]	60/154 [39%]
	severa	47/144 [32.8%]	26/154 [16.9%]
INT_dispa_local			
	fundo vaginal	75/144 [52.3%]	72/154 [46.7%]
	não se aplica	49/144 [34%]	46/154 [29.8%]
	parede vaginal e fundo vaginal	10/144 [7%]	7/154 [4.5%]
	introito vaginal	4/144 [2.7%]	4/154 [2.6%]

introito vaginal e fundo vaginal	3/144 [2%]	4/154 [2.6%]
parede vaginal	0/144 [0%]	7/154 [4.5%]
outro	3/144 [2%]	14/154 [9.3%]
INT_dispareunia		
nenhuma das anteriores	70/144 [48.6%]	75/154 [48.7%]
não informado	20/144 [13.8%]	16/154 [10.3%]
dor genitopelvica importante	17/144 [11.8%]	27/154 [17.5%]
medo importante de penetração	8/144 [5.5%]	7/154 [4.5%]
outros	29/144 [20.3%]	29/154 [19%]
INT_dor_pelvica_altera_lado		
Sim	18/144 [12.5%]	15/154 [9.7%]
INT_dor_pelvica_molhora_repouso		
Sim	59/144 [41%]	67/154 [43.5%]
INT_dor_pelvica_piora_apos_atv_fisica_prolongada		
Sim	52/144 [36%]	75/154 [48.7%]
INT_dor_pelvica_subitos_cessam_espontaneamente		
Sim	48/144 [33.3%]	56/154 [36.3%]
INT_duracao_fluxo_menstrual (média)	34.24	19.73
INT_durante_ato_sexual_dor_vaginal_prof_ou_dor_pelvica		
Sim	84/144 [58.3%]	105/154 [68.2%]
INT_EF_alodinea		
Sim	5/144 [3.4%]	10/154 [6.5%]
INT_EF_alteracao_septo_reto_vaginal		
Sim	11/144 [7.6%]	5/154 [3.2%]
INT_EF_carnett		
intensidade da dor aumenta	11/144 [7.6%]	34/154 [22%]
	133/144	
dor diminui/não muda	[92.4%]	120/154 [78%]
INT_EF_dor_assoalho_pelvico		
Sim	30/144 [20.8%]	38/154 [24.6%]
INT_EF_dor_palpacao_anexos		
Sim	54/144 [37.5%]	62/154 [40.2%]
INT_EF_dor_referida		
mesmo território de inervação	18/144 [12.5%]	25/154 [16.3%]
outro território de inervação	4/144 [2.8%]	4/154 [2.6%]
	122/144	
nenhuma	[84.7%]	125/154 [81.1%]
INT_EF_dor_superficie_perineo		
Sim	13/144 [9%]	25/154 [16.2%]
INT_EF_hernia		
Sim	2/144 [1.4%]	1/154 [0.6%]
INT_EF_IMC (média)	26.08	26.58
INT_EF_palpacao_massas		
Sim	13/144 [9%]	6/154 [3.9%]
INT_EF_pto_gatilho		
Sim	20/144 [13.8%]	39/154 [25.3%]
INT_EF_sensibilidade_area		

	anestesia	0/144 [0%]	1/154 [0.6%]
	anestesia e hipoestesia	0/144 [0%]	1/154 [0.6%]
	hiperestesia	5/144 [3.4%]	7/154 [4.5%]
	hipoestesia	1/144 [0.7%]	7/154 [4.5%]
	inalterada	125/144 [86.8%]	121/154 [78.8%]
	inalterada e hipoestesia	0/144 [0%]	1/154 [0.6%]
	não informado	13/144 [9.1%]	16/154 [10.4%]
INT_EF_sinais_colpite_cervicite			
	Sim	2/144 [1.4%]	10/154 [6.5%]
INT_EF_vaginismo			
	Sim	6/144 [4.1%]	8/154 [5.2%]
INT_EF_vulvodinea			
	Sim	5/144 [3.4%]	9/154 [5.8%]
INT_fluxo_menstrual			
	amenorréia	19/144 [13.2%]	43/154 [27.9%]
	irregular	51/144 [35.4%]	31/154 [20.1%]
	regular	74/144 [51.4%]	80/154 [52%]
INT_intervalo_intermenstrual_dias (média)		28.0	28.0
SCORE_importancia_dpc (média)		0.9444	0.9675
INT_SI_constipacao_intestinal			
	sem queixas referentes à constipação	95/144 [66.1%]	91/154 [59%]
	necessidade de fazer força em 25% das defecações	22/144 [15.3%]	28/154 [18.2%]
	necessidade de fazer força em 25% das defecações e sensação de evacuação incompleta em 25% das defecações	13/144 [9.1%]	11/154 [7.1%]
	sensação de evacuação incompleta em 25% das defecações	6/144 [4.1%]	0/154 [0%]
	necessidade de fazer força em 25% das defecações e evacuação raramente ocorre sem uso de laxativos	3/144 [2%]	13/154 [8.4%]
	outros	5/144 [3.4%]	11/154 [7.3%]
INT_SI_diarreia			
	perda de fezes moles ou aquosas sem dor em ao menos 75% das evacuações	4/144 [2.8%] 140/144	8/154 [5.2%]
	sem queixas referentes à diarreia	[97.2%]	146/154 [94.8%]
INT_SI_distensao_abdominal			
	sem queixas referentes à distensão ou gases	91/144 [63.2%]	95/154 [61.6%]
	distensão abdominal visível ou sensação de distensão ou gases em excesso	40/144 [27.7%]	40/154 [25.9%]
	outros	13/144 [9.1%]	19/154 [12.5%]
INT_SI_dor_melhora_defecacao			
	Sim	26/144 [19.5%]	25/154 [16.3%]
INT_SI_inicio_mudanca_forma_fezes			
	Sim	12/144 [8.5%]	11/154 [7.3%]
INT_SI_inicio_mudanca_frequencia_evacuacao			
	Sim	13/144 [9%]	12/154 [7.8%]
INT_sono			

	Sim	52/144 [36.2%]	81/154 [52.6%]
INT_SU_Geral		101/144	
	sem queixas	[70.1%]	84/154 [54.8%]
	infecção urinária de repetição	5/144 [3.4%]	3/154 [1.9%]
	noctúria > 2	4/144 [2.8%]	4/154 [2.6%]
	perde urina ao tossir, espirrar ou rir	3/144 [2%]	6/154 [3.8%]
	sensação de bexiga cheia após urinar	2/144 [1.6%]	3/154 [1.9%]
	outros	29/144 [20.1%]	54/154 [35%]
INT_TB_escore (média)		6.222	8.649
INT_volume_fluxo			
	aumentado	48/144 [33.4%]	37/154 [24%]
	não se aplica	20/144 [13.8%]	42/154 [27.2%]
	normal	58/144 [40.3%]	64/154 [41.5%]
	reduzido	17/144 [11.8%]	11/154 [7.3%]
	reduzido e aumentado	1/144 [0.7%]	0/154 [0%]
Intensidade_dor_vas (média)		59.21	63.66
ND_ao_sentar (média)		1.847	2.065
ND_articular_muscular (média)		1.667	1.636
ND_bexiga_cheia (média)		2.646	3.377
ND_difer_colicas_antes_menstruacao (média)		3.951	2.714
ND_enxaqueca (média)		2.521	2.916
ND_inguinal_mudanca_decubito (média)		1.507	1.773
ND_lombar (média)		2.84	3.273
ND_ovulacao (média)		2.667	2.136
ND_pelvica_apos_ato_sexual_horas_dias (média)		2.965	4.091
ND_pos_menstrual (média)		3.938	2.942
ND_profunda_durante_ato_sexual (média)		3.736	4.909
ND_urinar (média)		0.8611	1.169
ND_vaginal_queimacao_apos_ato_sexual (média)		1.444	1.779
Profissao_cat			
	aposentada	2/144 [1.6%]	3/154 [1.9%]
	dependente	5/144 [3.4%]	2/154 [1.3%]
	desempregada	8/144 [5.5%]	5/154 [3.2%]
	do lar	31/144 [21.5%]	41/154 [26.6%]
	não informado	0/144 [0%]	1/154 [0.6%]
	remunerada	98/144 [68%]	102/154 [66.4%]
Religiao			
	não informado	54/144 [37.7%]	64/154 [41.8%]
	catolicismo	47/144 [32.6%]	36/154 [23.3%]
	protestantismo	28/144 [19.4%]	36/154 [23.3%]
	sem religião	6/144 [4.1%]	6/154 [3.8%]
	espiritismo	4/144 [2.8%]	4/154 [2.6%]
	outros	5/144 [3.4%]	8/154 [5.2%]
Tem_DPC		134/144	
	Sim	[93.1%]	145/154 [94.2%]

Tempo_DPC_anos (média)		5.09	4.903
Tempo_DPC_meses (média)		1.028	2.006
Total_PHQ (média)		4.625	5.338
Total_SPIN (média)		2.361	2.994
Total_SRQ_A (média)		7	8.071
Total_SRQ_B (média)		0.4514	0.5779
Tratamento_medico_tentou_anteriormente			
nenhum		32/144 [22.2%]	36/154 [23.3%]
analgésicos		28/144 [19.4%]	44/154 [28.5%]
analgésicos e anticoncepcional oral		8/144 [5.5%]	4/154 [2.6%]
analgésicos e antidepressivos		0/144 [0%]	5/154 [3.2%]
analgésicos, anticoncepcional oral e ginecologista		4/144 [2.8%]	0/154 [0%]
anticoncepcional oral		3/144 [2%]	0/154 [0%]
analgésicos e cirurgia		2/144 [1.6%]	0/154 [0%]
outros		67/144 [46.5%]	65/154 [42.4%]
Deflag_coito			
Sim		39/144 [27%]	61/154 [39.6%]
Deflag_menstrual			
Sim		79/144 [55%]	44/154 [28.5%]
Deflag_espont			
Sim		71/144 [49.4%]	98/154 [63.6%]
Deflag_exercicio			
Sim		19/144 [13.1%]	26/154 [16.8%]
Relaciona			
cirurgia		19/144 [13.1%]	39/154 [25.5%]
nada		102/144 [71%]	80/154 [51.9%]
outros		20/144 [13.8%]	20/154 [12.9%]
parto		1/144 [0.7%]	10/154 [6.5%]
perda		1/144 [0.7%]	4/154 [2.6%]
violência		1/144 [0.7%]	1/154 [0.6%]
Deflag_ovulacao			
Sim		5/144 [3.4%]	0/154 [0%]
Dor_visceral			
Sim		88/144 [61.1%]	50/154 [32.4%]
Dor_neuropatica			
Sim		5/144 [3.4%]	18/154 [11.6%]
Dor_somatica			
Sim		17/144 [11.8%]	25/154 [16.2%]
Dor_indefinida			
Sim		38/144 [26.4%]	72/154 [46.7%]
DIP			
Sim		10/144 [6.9%]	21/154 [13.6%]
Constipation			
Sim		8/144 [5.5%]	19/154 [12.3%]
Neuralgias			
Sim		1/144 [0.7%]	14/154 [9%]

Síndrome.Bexiga.Dolorosa	Sim	8/144 [5.5%]	28/154 [18.1%]
Síndrome.Congestão.Pélvica	Sim	8/144 [5.5%]	20/154 [12.9%]
Síndrome.Intestino.Irritável	Sim	14/144 [9.8%]	15/154 [9.7%]
ant_infertilidade	Sim	24/144 [16%]	8/154 [5.2%]
ant_depressao	Sim	23/144 [16%]	42/154 [27.2%]
ant_lombalgia	Sim	29/144 [20.2%]	39/154 [25.3%]
ant_enxaqueca	Sim	36/144 [25%]	45/154 [29.2%]
ant_TPM	Sim	31/144 [21.6%]	27/154 [17.5%]
ant_dor_cronica	Sim	60/144 [41.7%]	79/154 [51.2%]

Com estas 12 variáveis o algoritmo C4.5 foi executado com a finalidade de extrair um modelo visual (árvore de decisão) que pudesse auxiliar os médicos a realizar o diagnóstico da endometriose com somente estas 12 questões básicas. Este modelo pode ser observado na figura 4.6.

Figura 4.6 - Modelo do Diagnóstico de Endometriose.

4.5 Considerações Finais

Com base nos resultados obtidos nos experimentos apresentados, pode-se observar que os métodos propostos puderam ser utilizados para classificar os pacientes com uma boa performance. Também foi possível reduzir a quantidade de questões necessárias para o diagnóstico de cada uma das doenças o que melhora a qualidade do atendimento e evita exames desnecessários.

Capítulo 5

CONCLUSÕES E TRABALHOS FUTUROS

A dor pélvica crônica (DPC) é uma condição clínica incapacitante, que necessita de tratamento clínico ou cirúrgico. Sendo de difícil controle sua etiologia não é clara e, usualmente, resulta de uma complexa interação entre vários sistemas, sendo ainda influenciada por fatores socioculturais. Desta forma sua causa é de difícil identificação o que faz com que se demore a encontrar sendo necessários diversos exames e o preenchimento de formulários extensos contendo o histórico dos pacientes.

Neste contexto a presente tese apresentou um estudo investigativo sobre o diagnóstico automático da DPC utilizando inteligência artificial. O desenvolvimento de tal sistema de inteligência artificial capaz de diagnosticar a causa da DPC é de grande importância, podendo causar um impacto positivo na vida das pacientes que podem ter seu diagnóstico identificado mais cedo.

Durante o trabalho de pesquisa, foram investigados vários temas os quais deram suporte para a proposta, desenvolvimento e implementação dos algoritmos de diagnóstico com base em algoritmos de aprendizado de máquina clássicos e amplamente conhecidos. Além destes métodos, foram também implementados, testados e avaliados mecanismos de seleção de atributos, validação de resultados e transformação de problemas multi-rótulos.

Como pode ser visto, o sistema de IA desenvolvido teve uma boa performance diagnóstica considerando que este diagnóstico pode ser realizado com algumas questões simples, evitando custos desnecessários, como exames e até pequenas cirurgias. Este diagnóstico também permite que a doença possa ser identificada mais cedo e o início do tratamento possa ser realizado em unidades básicas de saúde. Assim sendo, a utilização de sistemas de inteligência artificial podem ser úteis em todos os níveis do sistema de saúde, evitando diagnósticos tardios, e também podendo auxiliar no treinamento de estudantes e médicos, com a criação de sistemas de tutoria inteligentes como feito em [51]. Os benefícios

deste tipo de treinamento é que o paciente é menos exposto até que o profissional tenha habilidade suficiente para realizar um diagnóstico confiável.

Desta forma, foi possível gerar um modelo que pode auxiliar médicos no processo de diagnose de doenças complexas como por exemplo a endometriose, ajudando a evitar erros de diagnóstico e também a necessidade de exames complexos, os resultados obtidos foram satisfatórios e os métodos propostos são promissores. A utilização de algoritmos de inteligência artificial no diagnóstico médico é dita uma das áreas mais promissoras para a análise de dados médicos. Estes algoritmos têm sido utilizados em oncologia, patologias de fígado, doenças da tireoide, reumatologia, cardiologia, ginecologia, etc. [52].

Entretanto a utilização de sistemas de inteligência artificial pode trazer alguns riscos, por exemplo, quando algo é amplamente utilizado no cotidiano sua existência é esquecida e seu funcionamento é tomado como certo. A utilização de algoritmos semissupervisionados (que se retroalimentam) podem levar a comportamentos inesperados e até, no caso de sistemas de classificação, a erros de diagnóstico. Um sistema supervisionado baseado em dados não confiáveis terá uma classificação não confiável ou um conjunto de dados enviesado levará a uma classificação enviesada [53].

Como trabalho futuro seria possível indicar a realização de um acompanhamento do sistema atual em um ambiente real com a validação dos diagnósticos e a inclusão da retroalimentação e re-treinamento do sistema a cada diagnóstico corretamente realizado e confirmado. Desta forma seria possível avaliar a possibilidade da utilização do sistema em um ambiente real e validar se os resultados encontrados no presente trabalho condizem com a realidade encontrada em ambulatorios.

Capítulo 6

PRODUÇÕES CIENTÍFICAS

Algumas produções científicas foram geradas a partir deste trabalho, estas produções foram publicadas e apresentadas em congressos e journals da área de ciência da computação e informática aplicada a área da saúde. As três publicações geradas são:

- Oliverio, Vinicius; POLI-NETO, OMERO BENDICTO . Case Study: Classification Algorithms Comparison for the Multi-Label Problem of Chronic Pelvic Pain Diagnosing. In: 2017 IEEE 33rd International Conference on Data Engineering (ICDE), 2017, San Diego. 2017 IEEE 33rd International Conference on Data Engineering (ICDE), 2017. p. 1507-1509. (Qualis A1 em Ciência da Computação)
- Oliverio, Vinicius; Bendicto Poli-Neto, Omero . Artificial Intelligence Applied in the Multi-label Problem of Chronic Pelvic Pain Diagnosing. Lecture Notes in Computer Science. 1ed.: Springer International Publishing, 2016, p. 80-85.
- SILVA, R. P. E. ; OLIVERIO, V. ; SILVA, J. C. R. E. ; REIS, F. J. C. ; NOGUEIRA, A. A. ; POLI NETO, O. B. . The Importance of Computing Systems in Chronic Pelvic Pain Research. INTERNATIONAL JOURNAL OF COMPUTER AND INFORMATION TECHNOLOGY, v. 2, p. 255-260, 2013.

Este capítulo contém estas publicações em formato original com a finalidade de reforçar a relevância científica do presente trabalho.

The Importance of Computing Systems in Chronic Pelvic Pain Research

Ricardo Perussi-e-Silva, Vinicius Oliverio, Julio Cesar Rosa-e-Silva, Francisco Jose Candido-dos-Reis, Antonio Alberto Nogueira, Omero Benedicto Poli-Neto.

Department of Obstetrics and Gynecology.
 Universidade de São Paulo, USP.
 Ribeirão Preto, Brasil.

Abstract — Chronic pelvic pain is a common clinical condition with negative consequences for many aspects of women's life. The clinical presentation is heterogeneous and the involvement of several body systems impairs the identification of the exact etiology of the problem. At the same time, a clinical treatment of good quality depends on the professional and the learning process is slow. Besides, biomedical aspects predominate in the educational process during the graduation in health sciences, which are inadequate due to the biopsychosocial approach that this condition requires. It is believed that the acquisition, storage and adequate processing of information related to the different aspects of the clinical history of these women may help professionals to better understand this disease. Objectives: 1- to evaluate the appropriate filling out of the form regarding the clinical variables and identification of the professional responsible for patient care; 2- to evaluate the satisfaction of the user regarding free and manual filling out or the new electronic and structured method for a new case along time. Methods: the project was approved by the Research Ethics Committee of HCFMRP and the implantation of the system was approved by the Committee for the Analysis of Medical Records and Deaths of HCFMRP. The process of development and implantation of the system was based on the principles of software engineering, with the adoption of the classical life cycle starting with analysis of the problem, survey of requisites, project, codification, tests, and maintenance. The Filemaker Pro software was used for the development of the form, together with the Filemaker Server 11 for database management. One-hundred new cases were evaluated in a first phase in an alternate manner: 50 of them elaborated with the manual system and 50 with the electronic system. The number of data that were not provided, the number of erasures or inadequately filled out data, the time of execution of a new case and the final identification of the professional responsible were determined. A structured questionnaire was also applied for an objective evaluation of user satisfaction, with analysis and implantation of suggestions for an adequate layout. In a second phase, 50 new cases were elaborated using the new electronic system and were also evaluated by the users. Results: As important as usability were the significant results obtained in terms of the reduction of consumables since the reports to be inserted in the medical records of the patients were reduced by 50%. The mean time needed to fill out the traditional (manual) form was less than the mean time needed to fill out the electronic form, although this difference can be considered to be of little

significance from a clinical viewpoint. Conclusion: The electronic form is filled out within a time similar to that needed for a traditional form, it generates a completely legible report of fewer pages, without erasures or missing information, and contains the precise identification of the professional in 100% of cases. The users also stated that they were satisfied with the system.

Keywords-component; Pelvic Pain; Computing; Medical; Research.

I. INTRODUCTION

A. General considerations

Chronic pelvic pain is a common clinical condition worldwide and is defined as a pain in the lower abdomen and/or in the pelvis, persisting for at least six months and intense enough to cause functional incapacity, demanding clinical or surgical treatment [1].

This condition is difficult to control because frequently the primary cause is not identified. The etiology is not clear and, usually results from a complex interaction between the gastrointestinal, urinary, gynecological, musculoskeletal, neurological, psychological and endocrine systems, also influenced by sociocultural facts [2].

The disease implies a direct expense exceeding two billion dollars per year in the United States and has a considerable impact on the socioeconomic national scenario, directly due to elevated expenses and secondarily due to the incapacity caused by this condition [3].

At the same time, the importance of computing has been increasing over the years in a wide variety of fields, including medicine, in which the use of computational systems is extensive. Ranging from X-ray scanners to computerized robotic surgeries, computing has been used in many tasks in medicine. One of these tasks is the possibility to enable the acquisition and reliable storage of selected clinical data that can be very useful for scientific research, including the field of chronic pelvic pain, by the creation of systems that record and

manage structured forms [4]. Another important task would be to obtain artificial intelligence systems that would help with the diagnosis of the real cause of this condition.

On this basis, the present paper presents a discussion about computing systems and their importance for chronic pelvic pain research, showing what they can do to improve the reliability of acquired data, enabling the interaction between different databases and systems and also leading to more reliable investigations based on clinical data.

By discussing these questions the goal of this paper is to promote reflection among clinical professionals and researchers about the benefits of computing systems, such as the structured forms and artificial intelligence systems described here, for supporting the development of good quality research.

B. Structured forms

The use of structured forms in medicine as stated in [5] “reduces the ambiguity of natural language format reports and enhances the precision, clarity and value of clinical documents”, and having a precise and clear data is essential to research in a manner that the results depends on the data that is being studied. Other studies as [6], [7] and [8] also reinforces the idea that by using structured reports the data acquired by it is more precise and clear than regular natural language format reports.

Structured forms are computer forms that can be configured to request any type of information, and all data that are inputted into them are controlled and can be easily manipulated allowing, among other features, searching, storing and comparing with other similar elements. These features facilitate clinical research and teaching.

Other benefits of the use of structured forms are cost reduction since the storage of electronic data is cheaper and involves a reduction of the amount of paper needed [5], and improved communication, with the possibility of obtaining better replies to questions that really matter [7]. There is also the possibility of accessing these data anywhere, enabling easy sharing of information [8], which is an important feature when thinking about geographically distant research contributors.

There are many systems that allow the assembly of structured forms, such as Oracle RDC and Medidata Rave. One of these systems, called Filemaker, is the worldwide leader in easy-to-use database software that can be easily operated by non-technical people [9]. This software was used by the Center for Epidemiological Research in Women's Health (NUPEM) of the University of São Paulo to create a structured form for medical consultation of patients with chronic pelvic pain. This software has enabled a better control of the information for these patients and has provided more precise data [10], as can be seen in the “Methods” section. It has also provided the basis for the creation of another system, which uses artificial intelligence to help with the diagnosis of the cause of chronic pelvic pain.

Traditionally, the paper form used by the medical service contains 4 pages of questionnaire. Erasures, inadequate information and even incomplete filling out of the form were constant occurrences. The development of a form that would guide a correct filling out within an acceptable period of time for the conditions of the hospital service, as well as the reduction of the amount of paper in the records, and mainly an easier later investigation of the database were the major reasons for the development of the new system. The objectives of the present study were: 1- to evaluate user satisfaction with the utilization of this new proposal as opposed to free and manual recording of the data of the first medical visit, and 2- to describe the eventual benefits of the use of the electronic structured form compared to the traditional model employed thus far.

II. MATERIALS AND METHODS

The study was conducted at the Chronic Pelvic Pain and Gynecologic Video-endoscopy Center of the Department of Gynecology and Obstetrics, Faculty of Medicine of Ribeirão Preto, University of São Paulo. The project was approved by the local Ethics Committee and by the Committee for the Analysis of Medical Records and Deaths (protocol nº 6694/2010).

A. Development and implantation of the structured form

For the development and implantation of the multicenter system we adopted the cascade life cycle of software engineering, justified by the high degree of detailing of the entry, processing and exit requisites [11]. In order to obtain better acceptance of the technology by the users, we developed a software with an intuitive interface that would facilitate the search for information. In parallel, we insured safety of the information by defining access rules through virtual private networks (VPN) and users and passwords with privileged access to the system. The Filemaker Server 11 solution of Filemaker Co. was used as the database management system (DBMS).

A five point Likert scale, where 1 = unsatisfied and 5 = highly satisfied, was used to assess the quality of the software and to compare the traditional method (paper form) to the electronic method (structured form). We analyzed user satisfaction regarding screen layout, flow of form filling out, menu organization, clarity of the information requested, and software operability. We also evaluated user preference for one of the systems based on the guidelines contained in the form and the legibility of the report generated. We described and compared the mean time for the filling out and general feasibility of the information. All professionals were informed about the installation of the new system but no meeting was offered before its implementation. Access support was provided daily by the persons responsible for the development of the system at the place where the clinical activities were performed.

In a first phase, 100 new cases were evaluated between August 2011 and March 2012 in an alternate manner: 50 of them elaborated with the manual system and 50 with the

electronic system. All cases were the responsibility of gynecology and obstetrics medical residents, particularly of those in the 3rd and 4th year of specialization. The professional was supposed to fill out the information of a new case obligatorily using the electronic system to answer the questionnaire. Preferentially, the same professional processed a new case using the traditional system (manual filling out of a form printed by the print shop of the Hospital) and the subsequent one using the electronic system (finally generating a report pre-formatted by the persons responsible for the outpatient clinic, printed and attached to the medical records of the patient). In this phase we determined the number of data not provided, the number of erasures or of improperly filled out data, the time of processing of a new case, and the presence of the final identification of the professional responsible. We also applied a structured questionnaire in order to evaluate in an objective manner the satisfaction of the users with the filling out methods, with analysis and implantation of suggestions regarding adequacy of the layout. This questionnaire and its pertinent results are presented in Table 1.

In a second phase, after the implantation of suggestions for layout adequacy and functionality, 50 additional new cases were processed from April to July 2012 using only the electronic system. At that time, we also reevaluated user satisfaction using the same questionnaires.

B. Statistical analysis

The Agostino and Pearson test was used to determine the normality of numerical variables. Variables with a normal distribution were then analyzed by the Mann-Whitney test and variables with non-normal distribution were analyzed by the Kolmogorov-Smirnov test. The curve of user satisfaction was represented by sigmoid dose-response graphs. The level of significance was set at $\alpha=5\%$. All analyses were performed with the aid of the GraphPad Prism 5® software.

III. RESULTS AND DISCUSSION

No absence of information was detected in the electronic form since the variables were configured as obligatory fields, so that a visit could not be completed without a completely filled out form. In contrast, with the manual method there was a median number of 9 absent data (range: 3-18) out of a total of 107 data considered essential (8.4%) ($P<.0001$). No inadequate, or erased data or data incoherent with the field were detected in the electronic form since all fields of the form had previously defined values and, when necessary, pertinent information was provided for guidance. With the manual method, there was a median of 15 occurrences (range: 9-23) of erased data or data provided in inappropriately filled out fields out of a total of 107 data considered essential (14.0%) ($P<.0001$). In the present study, we did not determine the reliability of the information obtained.

The mean time needed to fill out the traditional (manual) form was less than the mean time needed to fill out the electronic form, i.e., 44.6 ± 4.7 minutes versus 48.7 ± 5.4 minutes

($P<.0001$), although this difference is of little significance from a clinical viewpoint. The reports consisted of two printed pages, whereas the standard form (free text) previously used consisted of 4 pages, with a consequent 50% reduction of the number of pages inserted in the medical records of the patients. We observed that in 24 new cases processed in the first phase using the free and manual text (48%, 24/50) it was necessary to let the professional know at the end of the process that he had forgotten to identify himself or had done so in an illegible manner, a fact that did not occur with the electronic system since identification was the first condition required to authenticate the use of the system ($P<.0001$).

The degree of user satisfaction with the system was directly dependent on the time of exposure to the system and presented an important inflexion of the curve soon after the implantation of the suggestions made by the users themselves (Fig.1). According to the technology acceptance model (TAM), what determines the utilization of the system by an individual is an easy use and mainly the positive results provided by this new technological method [12].

Figure 1: Curve for the degree of satisfaction with screen layout and flow of system filling out (question 1), organization of menus and of list of information (question 2), system with clear and understandable information (question 3), learning with the new system (question 4), operationality of the system (question 5), satisfaction with the performance of the application (question 6), interest in the system (question 7), possibility of reporting all the desired information in the system (question 8) and preference for the manual method or the informatized method (question 9). The legibility of system was investigated in questions 10 and 11, but the electronic form was absolutely superior. Because of this, they are not presented at Figure.

The subjective item considered to be most relevant by the professional was the fact that the structured electronic form not only permits the digitization of the information by the users, but also provides guidance for the correct filling out of each variable and its indexing according to previously stipulated requisites. Additional benefits are the economy of paper and the reduced number of sheets attached to the medical records of the patients, which used to be four and are now a maximum of two. Clarity and reliability of the final reports printed in the medical records are also observed since there is control of the

complete recording of all the variables considered essential for a precise diagnosis.

The information acquired with this structured form is also being used to create another system that uses artificial intelligence to help diagnose the real cause of chronic pelvic pain in these patients. The information acquired with this form is stored in the FileMaker database and is accessed by the artificial intelligence system that handles these data, in an attempt to help diagnose the cause of this condition.

A. *Future perspectives: Artificial Intelligence*

The goal of artificial intelligence (AI) is the development of paradigms and algorithms that require machines to accomplish cognitive tasks for which humans are currently better. AI systems have to be capable of doing three things [13]:

- Store knowledge;
- Apply the stored knowledge to solve problems;
- Acquire new knowledge through experience.

Besides the capacity of doing such things, AI systems must have three fundamental components: representation, reasoning and learning [14]. Representation is used to represent the knowledge about a problem's domain, reasoning is the ability to solve problems, and finally, learning is the process of change of behavior obtained through experience built by emotional, neurological, relational and environmental factors [15].

According to Mitchell [16], to say that a machine can learn it should, from an experience E, regarding a class of tasks T and performance measure P, enhance its performance in T, measured by P, with the acquired experience E.

AI systems are present in many fields of medicine, such as image processing, therapy planning, medical database mining, disease diagnosis, etc, and, as stated in Section III, there is another system that helps research on chronic pelvic pain in the NUPEM. This system uses AI to help diagnose the real cause of chronic pelvic pain. The purpose of the system is not to replace the doctor or any other professional, but to help doctors to diagnose the cause of this complex condition, especially resident doctors, because, as mentioned earlier, this systems is being used in the University of São Paulo. The system can also be used to teach; e.g. with this type of system medical students can be asked to diagnose a situation and then compare their diagnosis to that made by the system and check whether they made a wrong diagnosis and why.

Chronic pelvic pain causes a lot of direct expenses and incapacitates women for work. Therefore it is extremely important to find the cause of this condition as soon as possible, and this type of AI system can help with this task.

IV. CHRONIC PELVIC PAIN RESEARCH

There are two basic research methods: quantitative research and qualitative research, which are defined by [17] as follows.

Qualitative research is based on a process of inquiry whose goal is the understanding of a social or human problem from

different perspectives. The qualitative research is conducted in a natural setting and involves a process of building a complex scenario of the phenomenon of interest.

In contrast, quantitative research is the investigation of a known problem, based on testing a theory, measured with numbers, and analyzed using statistical techniques. The goal of this method of research is to determine whether the predictive generalizations of a theory hold true.

The chronic pelvic pain research carried out by NUPEM is a quantitative type of research and therefore is based on experiments conducted on a collection of data. On this basis, it is clear that the results of this type of research depend on the quality of the acquired data, and database systems together with structured forms provides the tools for the acquisition of more precise, good quality data.

Computing systems also provide a set of tools to handle these data. Imagine searching for a single characteristic among a large number of patients who filled out a regular paper form. It is easy to see that this task would be more difficult than searching a computer database that can be easily queried for specific data. In addition a computer system that handles these characteristics for a purpose can be set up to help the researchers.

As said before, this system also simplifies the sharing of information. With the advent of the World Wide Web the structured forms and the databases can be accessed from anywhere safely, breaking geographic barriers [18] and permitting the exchange of information among researchers over the world, democratizing research.

V. CONCLUSIONS

The electronic form is filled out within a time similar to that needed for a traditional form, it generates a completely legible report of fewer pages, without erasures or missing information, and contains the precise identification of the professional in 100% of cases. The users also stated that they were satisfied with the system, despite some initial signs of resistance. The best evaluations were obtained after the more significant considerations of the users were inserted in the system.

As could be seen, the use of computing enhances research by enabling the acquisition of better quality, more reliable data. It also improves the sharing of information and the handling of acquired data, providing tools for all of these purposes.

We showed here a successful case of the use of structured forms together with database systems in the University of São Paulo, with great improvement of filling time, data quality and interaction between systems.

REFERENCES

- [1] ACOG, "ACOG Practice Bulletin No. 51. Chronic pelvic pain", *Obstet Gynecol.* Mar 2004.
- [2] F. M Howard, "Chronic pelvic pain.", *Obstet Gynecol.* Mar 2003.

- [3] F. M. Blyth, L. M. March, A. J. Brnabic, M. J. Cousins, "Chronic pain and frequent use of health care.", *Pain*, v. 111, n. 1, p. 51-58, September 2004.
- [4] E. A. Tolley and A. S. Headley, "Meta-analyses: what they can and cannot tell us about clinical research", *Cur Opin Clin Nutr Met Care*, v. 8, p. 177-181, 2005.
- [5] R. Hussein, U. Engelmann, A. Schroeter and H. Meinzer, "DICOM structured reporting. Part 1. Overview and Characteristics.", *Radiographics*, v. 24, n. 3, pp. 891-896, May 2004.
- [6] R. Hussein, U. Engelmann, A. Schroeter and H. Meinzer, "DICOM structured reporting. Part 2. Problems and Challenges in Implementation.", *Radiographics*, v. 24, n. 3., pp. 891-896, May 2004.
- [7] S. R. Y. Geist and J. R. Geist, "Improvement in medical consultation responses with a structured request form.", *Journal of Dental Education* v. 72, n. 5553-561, May 2008.
- [8] E. C. King, "The Importance of Good Clinical Data Management and Statistical Programming Practices to Reproducible Research.", CCTST, University of Cincinnati. Available at: <http://cctst.uc.edu/sites/default/files/files/Slides%20King%202012-01-20.pdf>, accessed on 12/24/2012.
- [9] FileMaker Inc., "About FileMaker, Inc.". Available at: <http://www.filemaker.com/company/>, accessed on 12/24/2012.
- [10] R. P. Silva, O. Poli-Neto, "Formulário eletrônico estruturado de informações aplicadas à pesquisa sobre dor pélvica crônica.", XIII Congresso Brasileiro de Informática em Saúde, November 2012, in press.
- [11] R. S. Pressman, "Engenharia de Software", Markron Books, São Paulo, 1995.
- [12] F. D. Davis, R. P. Bagozzi, P. R. Warshaw, "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models". *Management Science*, v.35, n.8, p.982-1003, 1989.
- [13] A. P. Sage, "Concise Encyclopedia of Information Processing in Systems and Organizations", New York, Pergamon, 1990.
- [14] S. Haykin, "Neural Networks: A Comprehensive Foundation", Prentice Hall, 2nd ed, 842p, 1998.
- [15] J. N. Nilsson, "Introduction to Machine Learning: An Early Draft of a Proposed TextBook", Stanford University, Stanford, United States.
- [16] T. M. Mitchell, "Machine Learning", McGraw-Hill, 421p, 1997.
- [17] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd ed., SAGE Publications, 2008, 296 p.
- [18] L. B. Afrin, V. Kuppaswamy, B. Slater, R. K. Stuart, "Electronic clinical trial protocol distribution via the World-Wide Web: A prototype for reducing costs and errors, improving accrual, and saving trees.", *J Am Med Inform Assoc*, v. 4, n. 1, p. 25-35, Jan-Feb 1997.

Table 1. User satisfaction degree (medical professional) according to Lickert scale.

Question		Score					P		
Screen layout, flow, and ease of operation									
1. Were you satisfied with the layout and sequence of presentation?	Dissatisfied	1	2	3	4	5	1 st phase 3.1±0.9	2 nd phase 4.8±0.4	<.0001
2. Were you satisfied with the organization of menus and lists of information?	Dissatisfied	1	2	3	4	5	3.8±0.9	5.0±0.2	<.0001
3. Did you think that the information contained in the system is clear and understandable ?	Very difficult	1	2	3	4	5	4.3±0.7	4.8±0.4	<.0001
4. How do you rate your learning of the new system ?	Very difficult	1	2	3	4	5	3.4±1.0	4.3±0.3	<.0001
5. Would you be able to operate the system by yourself without help ?	Very difficult	1	2	3	4	5	2.7±0.8	4.0±0.6	<.0001
System performance									
6. Were you satisfied with the performance of the system ?	Dissatisfied	1	2	3	4	5	3.3±0.7	4.9±0.4	<.0001
General assessment									
7. Are you interested in this information system ?	Not interested	1	2	3	4	5	2.7±0.8	4.8±0.4	<.0001
8. Do you think that the system permits you to report everything you want or need ?	Dissatisfied	1	2	3	4	5	4.3±0.7	4.7±0.5	.0192
9. Which method do you prefer to process a new case ?	Free text	1	2	3	4	5	This method	4.3±0.6	<.0001
10. Were you satisfied with the legibility of the free text ?	Dissatisfied	1	2	3	4	5	2.0±0.9	Very satisfied	---
11. Were you satisfied with the legibility of the electronic report ?	Dissatisfied	1	2	3	4	5	4.5±0.6	4.9±0.3	.0004

Note: The results are presented by mean±standard deviation. Mann Whitney test was performed since some variable not presented normal distribuit

Artificial Intelligence Applied in the Multi-Label Problem of Chronic Pelvic Pain Diagnosing

Vinicius Oliverio and Omero Bendicto Poli-Neto

Universidade de São Paulo, Ribeirão Preto - SP, 14049-900, Brazil
{vinicius.oliverio@usp.br,polineto@fmrp.usp.br}

Abstract. Chronic pelvic pain is a common clinical condition with negative consequences in many aspects of women's life. The clinical presentation is heterogeneous and the involvement of several body systems impairs the identification of the exact etiology of the problem. At the same time, a clinical treatment of good quality depends on the professional and the learning process is slow. The goal of the paper is to show techniques used to create an artificial intelligence system capable of indicating the probable causes of this condition in order to help the doctors in the diagnosing process. This system uses a supervised learning algorithm along with multi-label problem modeling techniques and attribute selection algorithms to achieve the desired goal.

Keywords: Artificial Intelligence, Machine Learning, Chronic Pelvic Pain, Diagnosis, Multi-label.

1 Introduction

Chronic pelvic pain (CPP) is a common clinical condition worldwide and is defined as a pain in the lower abdomen and/or in the pelvis, persisting for at least six months and intense enough to cause functional incapacity, demanding clinical or surgical treatment [1].

This condition is difficult to control because frequently the primary cause is not identified. The etiology is not clear and, usually results from a complex interaction between the gastrointestinal, urinary, gynecological, musculoskeletal, neurological, psychological and endocrine systems, also influenced by sociocultural facts [2].

The disease implies a direct expense exceeding two billion dollars per year in the United States and has a considerable impact on the socioeconomic national scenario, directly due to elevated expenses and secondarily due to the incapacity caused by this condition [3].

At the same time, the importance of computing has been increasing over the years in a wide variety of fields, including medicine, in which the use of computational systems is extensive. Ranging from X-ray scanners to computerized robotic surgeries, computing has been used in many tasks. One of these tasks is the possibility to enable the acquisition and reliable storage of selected clinical data that can be very useful for scientific research, including the field of chronic

pelvic pain, by the creation of systems that record and manage structured forms [4]. Another important task would be to obtain artificial intelligence systems that would help with the diagnosis of the real cause of this condition using the databases created by those computational structured forms.

By discussing these questions, the goal of this paper is to promote reflection among clinical professionals and researchers about the benefits of computing systems, such as artificial intelligence systems, for supporting the development of good quality research and helping professionals in understanding better the diseases.

1.1 Naive Bayes

The Naive Bayes algorithm was chosen to be the classifier algorithm because bayesian learning algorithms that calculate explicit probabilities for hypotheses, are among the most practical approaches to certain types of learning problems [6]. In [9] a detailed study comparing the Naive Bayes classifier to other learning algorithms is provided, and it shows that the Naive Bayes classifier is competitive with decision trees, neural networks and other learning algorithms in many cases and in some cases it outperforms these other methods.

The Naive Bayes classifier bases itself in the simplified supposition that the attributes are conditionally independent given the target value (the class y). In other words, the supposition that the target value of one instance i , the probability of the conjunction of the elements of x_i ($\{x_{i,1}, x_{i,2}, \dots, x_{i,m}\}$) is given by the multiplication of the probabilities of the individual attributes: $P(x_1, x_2, \dots, x_i | y_j) = \prod_i P(x_i, y_j)$.

The Naive Bayes not always brings a very accurate probability of each class, but tends to have a good success rate in finding the most probable class. The reason which the estimated probability given by the naive bayes algorithm is not always precise is because of the simplification made by it when it disregards the relations between the attributes, assuming they are conditionally independent given the class.

2 Modeling

The chronic pelvic pain diagnose problem is a multi-label problem since each patient can have one or more diseases that are causing the condition. In the present paper each instance (patient) can have up to 7 diagnoses (labels), these possible diseases are constipation, pelvic inflammatory disease (PID), endometriosis, neuralgias (it covers ilioinguinal neuralgia and pudendal neuralgia), painful bladder syndrome, pelvic congestion syndrome and irritable bowel syndrome.

As it is a multi-label problem, it requires a multi-label strategy (modeling) since the classic Naive Bayes algorithm is a single-label classification algorithm. The next subsection shows how this problem can be solved.

2.1 Multi-label classification

In multi-label classification, the examples are associated with one or more labels. In the beginning, multi-label classification was mainly motivated by the tasks of text categorization and medical diagnosis, for instance, in medical diagnosis, a patient may be suffering for example from endometriosis and constipation at the same time [10].

In [10] and [11], the authors made a survey of multi-label classification approaches and they show some problem transformation methods for adapting a multi-label problem to a single-label algorithm, the two used methods will be discussed.

The first chosen problem transformation method (called PT4) was used in several papers, such as, [10], [11], [12] and [13], and it consists in splitting the original data set into n data sets (where n is the number of classes) containing all examples of the original, labeled as 1 if the labels of the original example contained 1 and as -1 otherwise. Then it is necessary to build one single label classifier to each data set. This approach is one of the two used in this paper in the experimentation process, it does not ignore any data and allows the classification to be done using a single label classifier such as Naive Bayes.

The second chosen approach (called PT5) was shown in [10] and consists of decomposing each example that has more than one label in n examples (where n is the number of labels that this example has). Then it learns one single-label classifier from the transformed data set. This approach was also one of the two used in this paper in the experimentation process, it also does not ignore any data and allows the classification to be done using Naive Bayes as well.

3 Experimentation

The training data set used to train the algorithm is a database of real diagnosed cases from the Clinics Hospital of Ribeirão Preto of the University of São Paulo (HCFMRP/USP). It consists of 346 instances (patients) with 106 attributes from the form filled by the doctors in each medical consultation of those patients, these forms contains the full history of the patient from the first consultation until the final diagnose.

The experimentation process consisted in four experiments, the first and the second one using the PT5 method, and the third and fourth experiment using the PT4 method. Since there are 106 attributes (symptoms, examinations, etc), it is not known if all attributes are important for the diagnosis of all the diseases, so in the experiments two and four after the problem transformation method was applied, the database passed through the Correlation-based Feature Subset Selection algorithm [14] in order to discover which attributes were relevant to the classification process, this way, it is possible to check which classification will be better, the one with all attributes or the one with selected attributes.

In the second experiment the attribute selection algorithm was executed once in the transformed data set, and in the fourth experiment for each of the resulting

data sets the attribute selection algorithm was executed in order to find out which variables are relevant to each one of the diseases (classes).

In the end of all experiments the resulting data sets were provided to the Naive Bayes classifier to train the algorithm, and its accuracy was validated using k-fold cross-validation [15] method, using a 10-fold size. The attribute selection algorithm and the Naive Bayes classifier were made using the Weka 3.9 software [16].

4 Results

In the experiments 2 and 4, the first step after the problem transformation methods were executed, was to run an attribute selection algorithm in order to reduce the amount of attributes used in the classification process. In experiment 2 the attribute selection algorithm selected 33 attributes, and in experiment 4 it selected 6 attributes for the constipation classifier, 7 for the PID classifier, 33 for the endometriosis classifier, 7 for the neuralgias classifier, 6 for the painful bladder classifier, 15 for the pelvic congestion classifier and 7 for the irritable bowel classifier.

As shown before, the first two experiments used PT5 and a single classifier. Initially the accuracy of the algorithm was very low, having a success rate of 23.57% (experiment 1), then after the attribute selection step the success rate increased to 45.5% (experiment 2). The result of the consecutive two tests (experiments 3 and 4), that used the PT4 method, can be seen in figure 1.

Fig. 1. Results from experiments 3 and 4.

As shown in figure 3 the accuracy of all classifiers used to process the data sets created by PT4 are better than the one classifier used for PT5. Also, in almost all

cases the classifier that used only the selected attributes had better results than the classifiers that used all attributes, the only case that it didn't is the classifier for Endometriosis. One more interesting case is the case of the Irritable Bowel Syndrome, that achieved a success rate of 99.71%, this is explainable because this disease has a known set of characteristics that determine if the patient has it or not, and the selected attributes were exactly the ones used to determine it.

Another important result from the tests were the sensitivity and specificity of the classifier in experimentations 3 and 4, and these values can be seen on tables 1 and 2.

Table 1. Sensitivity and specificity of the classifiers on experiment 3.

	Constipation	PID	Endometriosis	Neuralgias	Painful Bladder	Pelvic Congestion	Irritable Bowel
Sensitivity	0.353	0.548	0.687	0.64	0.816	0.345	0.697
Specificity	0.839	0.777	0.818	0.847	0.784	0.87	0.923

Table 2. Sensitivity and specificity of the classifiers on experiment 4.

	Constipation	PID	Endometriosis	Neuralgias	Painful Bladder	Pelvic Congestion	Irritable Bowel
Sensitivity	0.441	0.355	0.782	0.6	0.796	0.241	1
Specificity	0.974	0.962	0.662	0.956	0.848	0.889	0.997

Though the specificity in almost all cases were better than the sensitivity it is also good for this diagnosing problem. Despite the fact that the multidisciplinary approach is fundamental, the specific diagnose of the cause of the CPP is necessary for a good therapeutic planning. In this point there is another relative difficulty that is the multiplicity of symptoms. Many patients presents diverse symptoms that might suggest the commitment of several systems. The ability of the professional to list the most significant may optimize the process of complementary investigation. Consequently, besides propitiating a rational path to the diagnose, having a good specificity might improve the effectiveness-cost, avoiding the request of multiple exams, that in most of the cases are unnecessary.

5 Final Considerations

This paper shows a Naive Bayes classifier adapted to the multi-label problem of diagnosing the Chronic Pelvic Pain etiology, i.e., the diseases that are causing it. It was possible to achieve a good classification accuracy by using a problem transformation method and a attribute selection algorithm, in some cases the algorithm reached a success rate greater than 99%. However the algorithm in

almost all cases had a better specificity but as already said it might reduce the cost for the health systems as it shows a path to the diagnose avoiding expenses with unnecessary exams, and also reduce the time that it takes for the doctor do diagnose the real cause of the condition.

As a future work, a benchmark of different classifiers could be done in order to discover which one is the one that fits the best in the Chronic Pelvic Pain diagnose problem, also a benchmark of different attribute selection algorithms probably could improve the classification accuracy.

References

1. ACOG: ACOG Practice Bulletin No. 51. Chronic pelvic pain. *Obstetrics & Gynecology*. vol. 103, 589–605 (2004)
2. Howard, F.M.: Chronic pelvic pain. *Obstetrics & Gynecology*. vol. 101, 594–611 (2003)
3. Blyth, F.M., March, L.M., Brnabic, A.J., Cousins, M.J.: Chronic pain and frequent use of health care. *Pain*, vol. 111, 51–58 (2004)
4. Tolley, E.A., Headley, A.S.: Meta-analyses: what they can and cannot tell us about clinical research. *Cur Opin Clin Nutr Met Care*. vol. 8, 177–181 (2005)
5. Nils John Nilsson. *Introduction to Machine Learning: An Early Draft of a Proposed TextBook*. Stanford University (1997)
6. Mitchell, Tom M.: *Machine Learning*. McGraw-Hill Science/Engineering/Math (1997)
7. Haykin, S.: *Neural Networks: A Comprehensive Foundation*. Prentice Hall (1998)
8. John, G.H., Langley, P.: Estimating Continuous Distributions in Bayesian Classifiers. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers, 338–345 (1995).
9. Michie, D., Spiegelhalter, D. J., Taylor, C. C.: *Machine learning, neural and statistical classification*. Ellis Horwood (1994)
10. Tsoumakas, G., Katakis, I.: Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 1–13 (2007)
11. Carvalho, A.C.P.L.F., Freitas, A.A.: A Tutorial on Multi-label Classification Techniques. *Foundations of Computational Intelligence*, vol. 5, 177–195 (2009)
12. Boutell, M.R., Luo, J., Shen, X., Brown, C.M.: Learning multi-label scene classification. *Pattern Recognition*, vol. 37, 1757–1771 (2004)
13. Li, T., Ogihara, M.: Detecting emotion in music. In: *Proceedings of the International Symposium on Music Information Retrieval* (2003)
14. Hall, M.A.: *Correlation-based Feature Selection for Machine Learning* (Doctoral thesis). University of Waikato, Hamilton, New Zealand (1999)
15. Lachenbruch, P.A., Mickey, M.R.: Estimation of error rates in discriminant analysis. *Technometrics*, vol. 10, 1–12 (1968)
16. Hall, M.A., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, vol. 11, issue 1 (2009)

Case Study: Classification Algorithms Comparison for the Multi-Label Problem of Chronic Pelvic Pain Diagnosing

Vinicius Oliverio
Universidade de São Paulo
Ribeirão Preto - SP, 14049-900, Brazil
Email: vinicius.oliverio@usp.br

Omero Bendicto Poli-Neto
Universidade de São Paulo
Ribeirão Preto - SP, 14049-900, Brazil
Email: polineto@fmrp.usp.br

Abstract—Chronic pelvic pain is a common clinical condition with negative consequences in many aspects of women's life. The clinical presentation is heterogeneous and the involvement of several body systems impairs the identification of the exact etiology of the problem. At the same time, a clinical treatment of good quality depends on the professional and the learning process is slow. The goal of this paper is to compare different classification algorithms for diagnosing the probable cause of the chronic pelvic pain in those women. A multi-label problem modeling technique and an attribute selection algorithm were used to prepare the problem to the comparison of the algorithms Naive Bayes, C4.5, Support Vector Machines, AdaBoost and KNN.

Keywords—Artificial Intelligence, Machine Learning, Chronic Pelvic Pain, Diagnosis, Multi-label.

I. INTRODUCTION

Chronic pelvic pain (CPP) is a common clinical condition worldwide and is defined as a pain in the lower abdomen and/or in the pelvis, persisting for at least six months and intense enough to cause functional incapacity, demanding clinical or surgical treatment [1].

This condition is difficult to control because frequently the primary cause is not identified. The etiology is not clear and, usually results from a complex interaction between the gastrointestinal, urinary, gynecological, musculoskeletal, neurological, psychological and endocrine systems, also influenced by sociocultural facts [2].

The disease implies a direct expense exceeding two billion dollars per year in the United States and has a considerable impact on the socioeconomic national scenario, directly due to elevated expenses and secondarily due to the incapacity caused by this condition [3].

At the same time, the importance of computing has been increasing over the years in a wide variety of fields, including medicine, in which the use of computational systems is extensive. Ranging from X-ray scanners to computerized robotic surgeries, computing has been used in many tasks. One of these tasks is the possibility to enable the acquisition and reliable storage of selected clinical data that can be very useful for scientific research, including the field of chronic pelvic pain, by the creation of systems that record and manage structured forms [4]. Another important task would be to obtain artificial intelligence systems that would help with the diagnosis of the

real cause of this condition using the databases created by those computational structured forms.

A discussion of how a system can be designed and built to help solving this diagnosing problem was shown in [5]. This paper intends to compare different classification algorithms to determine which one is the best suited for this problem, by comparing the performance of them in this classification problem.

II. CLASSIFICATION ALGORITHMS

The chosen algorithms to this comparison are Naive Bayes, C4.5, Support Vector Machines (SVM), AdaBoost and KNN. They were chosen based on the survey paper [6]. This survey paper presents the top 10 algorithms in data mining, identified by the IEEE International Conference on Data Mining (ICDM) in december 2006, and as the paper states those 10 algorithms are the most influential data mining algorithms in the research community. The 10 algorithms presented in [6] are C4.5, K-means, SVM, Apriori, EM, PageRank, AdaBoost, KNN, Naive Bayes and CART, and the 5 that were adapted to the present problem are Naive Bayes, C4.5, Support Vector Machines (SVM), AdaBoost and KNN.

III. MODELING

The chronic pelvic pain diagnose problem is a multi-label problem since each patient can have one or more diseases that are causing the condition (for each x there is one or more y). In the present paper each instance (patient) can have up to 7 diagnoses (labels), these possible diseases are constipation, pelvic inflammatory disease (PID), endometriosis, neuralgias (it covers ilioinguinal neuralgia and pudendal neuralgia), painful bladder syndrome, pelvic congestion syndrome and irritable bowel syndrome. Since the classic classification algorithms are single-label classification algorithms, the data set had to be adapted, so for each possible diagnose one exclusive data set was created where each instance belongs to a binary class ("has the disease" or "doesn't have the disease") [5].

IV. EXPERIMENTATION

The training data set used to train the algorithms is a database of real diagnosed cases from the Clinics Hospital of Ribeirão Preto of the University of São Paulo

(HCFMRP/USP). It consists of 346 instances (patients) with 106 attributes from the form filled by the doctors in each medical consultation of those patients, these forms contains the full history of the patient from the first consultation until the final diagnose.

As said before this training data set was split into 7 data sets, one for each possible disease, containing a binary class showing if that person has or not that determined disease. Each of these 7 data sets were passed through the Correlation-based Feature Subset Selection algorithm [9] in order to discover which attributes were relevant to the classification process of each disease.

The algorithms performance were validated using k-fold cross-validation [10] method, with a 10-fold size. The classification algorithms and experiments were made using the Weka 3.9 software [11].

V. RESULTS

First of all, each algorithm was tested in each data set, the results where gathered and shown on table I.

TABLE I. CLASSIFICATION SUCCESS RATE.

	Naive Bayes	C4.5	SVM	AdaBoost	KNN
Constipation	92,17%	92,46%	92,75%	92,75%	91,30%
PID	90,72%	91,59%	92,17%	91,01%	89,27%
Endometriosis	71,30%	66,95%	67,82%	70,14%	66,08%
Neuralgias	93,04%	92,17%	91,59%	91,59%	91,59%
Painful Bladder	84,05%	87,24%	89,27%	86,37%	84,05%
Pelvic Congestion	83,47%	91,59%	91,01%	91,30%	85,79%
Irritable Bowel	99,71%	99,42%	99,71%	100,00%	99,71%
Avg. Acc.	87,78%	88,77%	89,19%	89,02%	86,83%

As can be seen the accuracy of the algorithms were similar. The algorithm that performed better was the SVM, having the greater average accuracy rate and being the better in 3 out of 7 tests. This result can be seen graphically on figure 1.

Fig. 1. Graphical classification comparison.

Another important information taken from the tests were the sensibility and specificity of each of the classifiers. The resulting sensibilities can be seen on table II and graphically on figure 2.

TABLE II. SENSIBILITIES.

	Naive Bayes	C4.5	SVM	AdaBoost	KNN
Constipation	0,441	0,529	0,529	0,529	0,441
PID	0,355	0,194	0,258	0,194	0,29
Endometriosis	0,782	0,554	0,605	0,633	0,633
Neuralgias	0,6	0,12	0,24	0,28	0,32
Painful Bladder	0,796	0,429	0,429	0,449	0,49
Pelvic Congestion	0,241	0	0	0	0,138
Irritable Bowel	1	1	1	1	0,97
Avg. Sen.	0,602	0,403	0,437	0,440	0,468

Fig. 2. Graphical sensibilities comparison.

As can be seen the sensibilities were not similar among the classifiers. The algorithm with the best sensibilities was the Naive Bayes algorithm, having the greater average sensibility, being the better in 6 out of 7 cases and having no sensibility equal to 0. The resulting specificities can be seen on table III and graphically on figure 3.

TABLE III. SPECIFICITIES.

	Naive Bayes	C4.5	SVM	AdaBoost	KNN
Constipation	0,974	0,968	0,971	0,971	0,965
PID	0,962	0,987	0,987	0,981	0,952
Endometriosis	0,662	0,763	0,732	0,753	0,682
Neuralgias	0,956	0,984	0,969	0,966	0,963
Painful Bladder	0,848	0,946	0,97	0,932	0,899
Pelvic Congestion	0,889	1	0,994	0,997	0,924
Irritable Bowel	0,997	0,994	0,997	1	1
Avg. Spec.	0,898	0,948	0,945	0,942	0,912

As can be seen, the specificities were very alike, however the algorithm with the best specificities was the C4.5 algorithm, having the greater average specificity and being the better in 3 out of 7 cases.

Fig. 3. Graphical specificities comparison.

Though the specificity in almost all cases were better than the sensitivity it is also good for this diagnosing problem. Despite the fact that the multidisciplinary approach is fundamental, the specific diagnose of the cause of the CPP is necessary for a good therapeutic planning. In this point there is another relative difficulty that is the multiplicity of symptoms. Many patients presents diverse symptoms that might suggest the commitment of several systems. The ability of the professional to list the most significant may optimize the process of complementary investigation. Consequently, besides propitiating a rational path to the diagnose, having a good specificity might improve the effectiveness-cost, avoiding the request of multiple exams, that in most of the cases are unnecessary.

VI. FINAL CONSIDERATIONS

This paper shows a comparison of classification algorithms on the multi-label problem of diagnosing the Chronic Pelvic Pain etiology, i.e., the diseases that are causing it. It was possible to achieve a good classification accuracy with all algorithms, in some cases the algorithms reached a success rate of 100%. However the algorithm with best accuracy was the Support Vector Machine algorithm, having the greater average accuracy rate and being the better in 3 out of 7 tests.

It was also possible to verify that the different algorithms had different sensibilities and specificities, the best sensibilities where achieved using Naive Bayes, having the greater average sensibility, being the better in 6 out of 7 cases and having no sensibility equal to 0, and the best specificitiess was the C4.5 algorithm, having the greater average specificity and being the better in 3 out of 7 cases.

All algorithms in almost all cases had a better specificity, however as already said it might reduce the cost for the health systems as it shows a path to the diagnose avoiding expenses with unnecessary exams, and also reduce the time that it takes for the doctor do diagnose the real cause of the condition.

REFERENCES

- [1] ACOG. (2004) ACOG Practice Bulletin No. 51. Chronic pelvic pain. *Obstetrics & Gynecology*. vol. 103, pp 589-605
- [2] F.M. Howard. (2003) Chronic pelvic pain. *Obstetrics & Gynecology*. vol. 101, pp 594-611
- [3] F.M. Blyth, L.M. March, A.J. Brnabic, M.J. Cousins. (2004) Chronic pain and frequent use of health care. *Pain*, vol. 111, pp 51-58
- [4] E.A. Tolley, A.S. Headley. (2005) Meta-analyses: what they can and cannot tell us about clinical research. *Cur Opin Clin Nutr Met Care*. vol. 8, pp 177-181
- [5] V. Oliverio and O. B. Poli-Neto. (2016) Artificial Intelligence Applied in the Multi-Label Problem of Chronic Pelvic Pain Diagnosing. *Lecture Notes in Computer Science*, vol. 10069, pp 80-85, Springer
- [6] W. Xindong, et. al. (2008) Top 10 Algorithms in data mining. *Knowledge and Information Systems*. vol. 14, issue 1, pp 137
- [7] Tom M. Mitchell. (1997) *Machine Learning*. McGraw-Hill Science/Engineering/Math.
- [8] V. Vapnik. (1995) *The Nature of Statistical Learning Theory*. Springer, New York.
- [9] M.A. Hall. (1999) *Correlation-based Feature Selection for Machine Learning (Doctoral thesis)*. University of Waikato, Hamilton, New Zealand
- [10] P.A. Lachenbruch, M.R. Mickey. (1968) Estimation of error rates in discriminant analysis. *Technometrics*, vol. 10, pp 1-12
- [11] M.A. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten. (2009) *The WEKA Data Mining Software: An Update*. *SIGKDD Explorations*, vol. 11, issue 1

REFERÊNCIAS

- [1]. ACOG. ACOG Practice Bulletin No. 51. Chronic pelvic pain. *Obstet Gynecol.* 2004;103:589-605.
- [2]. Howard FM. Chronic pelvic pain. *Obstet Gynecol.* 2003;101:594-611.
- [3]. Latthe, P., Mignini, L., Gray, R., Hills, R., Khan, K.. Factors predisposing women to chronic pelvic pain: systematic review. *Bmj.* 2006;332:749-55.
- [4]. Silva G.P.O.G.; Nascimento A.L.; Michelazzo D.; Alves-Junior F.F.; Rocha M.G.; Rosa-e-Silva J.C.; et al. High prevalence of chronic pelvic pain in women in Ribeirao Preto, Brazil and direct association with abdominal surgery. *Clinics (Sao Paulo).* 2011;66:1307-12.
- [5]. Latthe, P.; Latthe, M.; Say, L.; Gulmezoglu, M.; Khan, K.S. WHO systematic review of prevalence of chronic pelvic pain: a neglected reproductive health morbidity. *BMC public health.* 2006;6:177.
- [6]. Zondervan, K.T.; Yudkin, P.L.; Vessey, M.P.; Dawes, M.G.; Barlow, D.H.; Kennedy, S.H. The prevalence of chronic pelvic pain in women in the United Kingdom: a systematic review. *Br J Obstet Gynaecol.* 1998;105:93-9.
- [7]. Zondervan, K.T.; Yudkin, P.L.; Vessey, M.P.; Dawes, M.G.; Barlow, D.H.; Kennedy, S.H. Prevalence and incidence of chronic pelvic pain in primary care: evidence from a national general practice database. *Br J Obstet Gynaecol.* 1999;106:1149-55.
- [8]. Grace, V.M.; Zondervan, K.T. Chronic pelvic pain in New Zealand: prevalence, pain severity, diagnoses and use of the health services. *Aust N Z J Public Health.* 2004;28:369-75.
- [9]. Romao, A.P.; Gorayeb, R; Romao, G.S.; Poli-Neto, O.B.; Dos Reis, F.J.; Rosa-e-Silva, J.C.; et al. Chronic pelvic pain: multifactorial influences. *J Eval Clin Pract.* 2011;17:1137-9.
- [10]. Romao, A.P.; Gorayeb, R; Romao, G.S.; Poli-Neto, O.B.; Dos Reis, F.J.; Rosa-e-Silva, J.C.; et al. High levels of anxiety and depression have a negative effect on quality of life of women with chronic pelvic pain. *International journal of clinical practice.* 2009;63:707-11.
- [11]. Souza, P.P.; Romao, G.S.; Rosa-e-Silva, J.C.; Candido Dos Reis, F.; Nogueira, A.A.; Poli-Neto, O.B. Qualitative research as the basis for a biopsychosocial approach to women with chronic pelvic pain. *Journal of psychosomatic obstetrics and gynaecology.* 2011;32:165-72.

-
- [12]. Zondervan, K; Barlow, D.H. Epidemiology of chronic pelvic pain. *Baillieres Best Pract Res Clin Obstet Gynaecol.* 2000;14:403-14.
- [13]. Broder, M.S.; Kanouse, D.E.; Mittman, B.S.; Bernstein, S.J. The appropriateness of recommendations for hysterectomy. *Obstet Gynecol.* 2000;95:199-205.
- [14]. Blyth, F.M.; March, L.M.; Brnabic, A.J.; Cousins, M.J. Chronic pain and frequent use of health care. *Pain.* 2004;111:51-8.
- [15]. Walsh, A.; Kelly, J.; Johnson, S.; Rajkumar, S.; Bennetts, K. Performance problems of patients with chronic low-back pain and the measurement of patient-centered outcome. *Spine (Phila Pa 1976).* 2004;29:87-93.
- [16]. Persson, E.; Rivano-Fischer, M.; Eklund, M. Evaluation of changes in occupational performance among patients in a pain management program. *J Rehabil Med.* 2004;36:85-91.
- [17]. Nickel, R.; Silvado, C.E; Germiniani, F.M.; Paola, L.; Silveira, N.L.; Souza, J.R.; et al. Quality of life issues and occupational performance of persons with epilepsy. *Arq Neuropsiquiatr.* 2012;70:140-4.
- [18]. Tolley, E.A.; Headley, A.S. Meta-analyses: what they can and cannot tell us about clinical research. *Cur Opin Clin Nutr Met Care*, v. 8, p. 177-181, 2005.
- [19]. Hussein, R.; et al. DICOM structured reporting: Part 1. Overview and characteristics. *Radiographics*, v. 24, n. 3, p. 891-6, May-Jun 2004.
- [20]. Afrin, L.B.; et al. Electronic clinical trial protocol distribution via the World-Wide Web: a prototype for reducing costs and errors, improving accrual, and saving trees. *J Am Med Inform Assoc*, v. 4, n. 1, p. 25-35, Jan-Feb 1997.
- [21]. Sage, A.P. *Concise Encyclopedia of Information Processing in Systems and Organizations*. New York: Pergamon, 1990.
- [22]. Haykin, S. *Redes Neurais: Princípios e Prática*. 2º Edicao. Porto Alegre: Bookman, 2001. 900p.
- [23]. Nilsson, J.N. *Introduction to Machine Learning: An Early Draft of a Proposed TextBook*. 1997. Stanford University, Stanford, Estados Unidos.
- [24]. Mitchell, T.M. *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997. 421p.
- [25]. Dorian, P. *Data Preparation for Data Mining*. San Francisco: Morgan Kaufmann, 1999. 539p.
- [26]. Koff, W.; Gustafson P. “Data rEvolution”, CSC Leading Edge Forum, 2011. Disponível em: http://assets1.csc.com/innovation/downloads/LEF_2011Data_rEvolution.pdf
- [27]. Witten, I.H.; Frank, E. “Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations”. Morgan Kaufmann, 351 p, 1999.

-
- [28]. Dorian, P. "Data Preparation for Data Mining". Morgan Kaufmann, 513 p, 1999.
  - [29]. Warren, J.W.; Morozov, V.; Howard, F.M. Could chronic pelvic pain be a functional somatic syndrome? *Am J Obstet Gynecol.* 2011;205:199.e1-5.
  - [30]. Tsoumakas, G.; Katakis, I. "Multi-Label Classification: An Overview". *International Journal of Data Warehousing and Mining*, David Tanar (Ed.), Idea Group Publishing, 3(3), pp. 1-13, 2007.
  - [31]. Carvalho, A.C.P.L.F.; Freitas, A.A.: A Tutorial on Multi-label Classification Techniques. *Foundations of Computational Intelligence*, vol. 5, pp. 177-195, 2009.
  - [32]. Alpaydin, E. *Introduction to Machine Learning*, 2nd ed. MIT press. 2010. 579p.
  - [33]. Quinlan, J.R. Induction of Decision Trees. In *Machine Learning Journal*, volume 1, issue 1, pages 81-106. Kluwer Academic Publishers, 1986.
  - [34]. Cortes, C; Vapnik, V. Support-Vector Networks. In *Machine Learning Journal*, volume 20, issue 3, pages 273-297. Kluwer Academic Publishers, 1995.
  - [35]. John, G.H.; Langley, P. Estimating Continuous Distributions in Bayesian Classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. Pages 338-345. Morgan Kaufmann Publishers, San Mateo, 1995.
  - [36]. Boutell, M.R.; Luo, J.; Shen, X.; Brown, C.M. Learning multi-label scene classification. *Pattern Recognition*, vol. 37, pp. 1757-1771, 2004.
  - [37]. Diplaris, S.; Tsoumakas, G.; Mitkas, P.; Vlahavas, I. Protein Classification with Multiple Algorithms. In: *Proceedings of the 10th Panhellenic Conference on Informatics*, 2005.
  - [38]. Goncalves, T.; Quaresma, P. A Preliminary Approach to the Multilabel Classification Problem of Portuguese Juridical Documents. In: *Proceedings of the 11th Portuguese Conference on Artificial Intelligence*, 2003.
  - [39]. Lauser, B.; Hotho, A. Automatic multi-label subject indexing in a multilingual environment. In: *Proceedings of the 7th European Conference in Research and Advanced Technology for Digital Libraries*, 2003.
  - [40]. Li, T.; Ogihara, M. Detecting emotion in music. In: *Proceedings of the International Symposium on Music Information Retrieval*, 2003.
  - [41]. Quinlan, J.R. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.
  - [42]. Aha, D.W.; Kibler, D.; Albert, M.K. Instance-based Learning Algorithms. *Machine Learning*, vol. 6, issue 1, pp. 37-66, 1991.

-
- [43]. Fix, E.; Hodges, J. Discriminatory analysis, nonparametric discrimination: consistency properties. Tech. Rep. 4, USAF School of Aviation Medicine, Randolph Field, Texas. 1951.
  - [44]. Hill, T.; Lewicki P. STATISTICS: Methods and Applications. StatSoft, Tulsa, 2007.
  - [45]. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, vol. 55, pp. 119–139, 1997.
  - [46] Schapire, R.E. Explaining Adaboost. *Empirical Inference*, Springer Berlin Heidelberg, pp. 37-52, 2003.
  - [47] Yildirim, P. Filter Based Feature Selection Methods for Prediction of Risks in Hepatitis Disease. *International Journal of Machine Learning and Computing*, vol. 5, pp. 258-263, 2015.
  - [48] L. Ma; et. al.. Evaluation of Feature Selection Methods for Object-Based Land Cover Mapping of Unmanned Aerial Vehicle Imagery Using Random Forest and Support Vector Machine Classifiers. *International Journal of Geoinformation*, vol. 6, issue 2, 2017.
  - [49] Porcel, J.M.; et. al. A decision tree for differentiating tuberculous from malignant pleural effusions. *Respiratory Medicine*, vol. 102, pp. 1159-1164, 2008.
  - [50] Hall, M.A., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, vol. 11, issue 1, 2009.
  - [51] Bennet, C.C., Hauser, K.: Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial Intelligence in Medicine*, vol. 57, pp. 9–19, 2013.
  - [52] Kononenko, I.: Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, vol. 23, pp. 89–109, 2001.
  - [53] Yudkowsky, E.: Artificial Intelligence as a Positive and Negative Factor in Global Risk. *Global Catastrophic Risks*, pp. 308–345, 2008.
  - [54] Sarwar, A., Sharma, V.: Intelligent Naive Bayes Approach to Diagnose Diabetes Type-2. *Special Issue of International Journal of Computer Applications on Issues and Challenges in Networking, Intelligence and Computing Technologies*, pp 14-16, 2012.
  - [55] MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.
  - [56] Sethi, H., Goraya, A., Sharma, V.: Artificial Intelligence based Ensemble Model for Diagnosis of Diabetes. *International Journal of Advanced Research in Computer Science*, vol. 8, issue 4, 2017.

- [57] Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*. vol. 14, pp 1137–1145, 1995.
- [58] Anguita, D.; et. al.: The 'K' in K-fold Cross Validation. *The 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp 441-446, 2012.
- [59] Dinu, A.J., Ganesan, R., Joseph, F., Balaji, V.: A study on Deep Machine Learning Algorithms for diagnosis of diseases. *International Journal of Applied Engineering Research*, vol. 12, number 17, pp 6338-6346, 2017.

APÊNDICES

APÊNDICE A.

Variáveis contidas na base de dados:

- Aspectos Gerais:
  - Com_quem_vive_cat
  - Escolaridade
  - Estado_civil
  - Profissao_cat
  - Religiao
  - Idade_CN
- Quantificação da Dor:
  - Numero_episodios_DPC
  - Tempo_DPC_anos
  - Tempo_DPC_meses
  - Tipo_dor
  - FPS_valor
- Significações:
  - INT_ME_importancia_DPC
  - Total_PHQ
  - Total_SPIN
  - Total_SQR_A
  - Total_SQR_B
- Associações:
  - Aparecimento
  - Fato_relaciona_DPC
  - Fatores_melhora_DPC
  - Fatores_piora_DPC
- Especificações:
  - ND_ao_sentar
  - ND_articular_muscular
  - ND_bexiga_cheia
  - ND_colicas_durante_menstruacao
  - ND_difer_colicas_antes_menstruacao
  - ND_enxaqueca
  - NDinguinal_mudanca_decubito
  - ND_lombar
  - ND_ovulacao
  - ND_pelvica_apos_ato_sexual_horas_dias
  - ND_pos_menstrual
  - ND_profunda_durante_ato_sexual
  - ND_urinar
  - ND_vaginal_queimacao_apos_ato_sexual
  - INT_sono

-
- Antecedentes:
    - ANT_cirurgia_abdominal
    - ANT_incisao_cirurgica
    - ANT_status_menstrual
    - ANT_idade_menarca
    - ANT_idade_menopausa
    - ANT_atividade_sexual
    - ANT_coitarca
    - ANT_n_abortos
    - ANT_n_cesareas
    - ANT_n_gestacoes
    - ANT_n_parceiros
    - ANT_n_partos
    - ANT_complicacoes_puerpuraais
    - ANT_forceps
    - ANT_episiotomias
    - INT_ASF_vitima_abuso_emocional_humilhacao_insulto
    - INT_ASF_vitima_abuso_fisico_sexual
  - Queixas urinarias
    - INT_SU_Geral
    - INT_TB_escore
  - Hábitos:
    - Atividade_fisica
    - Tabagismo
    - Etilismo
    - Cafeina
    - Drogas_ilicitas
  - Padrão Sexual:
    - ANT_freq_sexual
    - ANT_excitacao
    - ANT_desejo
    - ANT_impulso_sexual
    - ANT_lubrif_adequada
    - ANT_orgasmo
    - ANT_sonhos_eroticos
    - ANT_masturbacao
    - ANT_fantasias_sexuais
    - INT_dispa_local
    - INT_dispareunia
    - INT_dispareu
    - INT_contracepcao
    - INT_outro_parceiro_sexual
  - Exame Físico:
    - INT_EF_alodinea
    - INT_EF_alteracao_septo_reto_vaginal
    - INT_EF_carnett
    - INT_EF_dor_assoalho_pelvico
    - INT_EF_dor_palpitacao_anexos
    - INT_EF_dor_referida
    - INT_EF_dor_superficie_perineo

-
- INT_EF_hernia
    - INT_EF_IMC
    - INT_EF_palpacao_massas
    - INT_EF_pto_gatilho
    - INT_EF_sensibilidade_area
    - INT_EF_sinais_colpите_cervicite
    - INT_EF_vaginismo
    - INT_EF_vulvodinea
  - Padrão Menstrual:
    - INT_fluxo_menstrual
    - INT_escape_menstrual
    - INT_duracao_fluxo_menstrual
    - INT_volume_fluxo
    - INT_intervalo_intermenstrual_dias
    - INT_dismenorreia
  - Tratamentos Anteriores
    - Tratamento_medico_tentou_anteriormente
    - Auto_medicao
  - Queixas Digestórias
    - INT_SI_constipacao_intestinal
    - INT_SI_diarreia
    - INT_SI_distensao_abdominal
    - INT_SI_dor_melhora_defecacao
    - INT_SI_inicio_mudanca_forma_fezes
    - INT_SI_inicio_mudanca_frequencia_evacuacao
  - Queixas Vasculares
    - INT_dor_pelvica_altera_lado
    - INT_dor_pelvica_molhora_repouso
    - INT_dor_pelvica_piora_apos_atv_fisica_prolongada
    - INT_dor_pelvica_subitos_cesam_espontaneamente
    - INT_apos_ato_sexual_dor_latejamento
    - INT_durante_ato_sexual_dor_vaginal_prof_ou_dor_pelvica