Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

FATORES DE RISCO E TESTES NEUROPSICOLÓGICOS NO APOIO AO DIAGNÓSTICO DE SÍNDROMES DEMENCIAIS: IDENTIFICAÇÃO DE ASPECTOS RELEVANTES ATRAVÉS DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA

RISK FACTORS AND NEUROPSYCHOLOGICAL ASSESSMENTS IN THE SUPPORT OF DIAGNOSIS OF DEMENTIAL SYNDROMES: IDENTIFICATION OF RELEVANT ASPECTS THROUGH MACHINE LEARNING TECHNIQUES

Leonard Barreto Moreira[1]
Annabell Del Real Tamariz
Joyce Vieira Fetterman

RESUMO - Este estudo teve como objetivo principal identificar os fatores de risco e testes de rastreio mais relevantes no processo de diagnóstico de pacientes com suspeita clínica de demências na triagem médica, com apoio de técnicas de aprendizagem de máquina. O conjunto de dados utilizado foi composto por 605 registros médicos, com 18 diferentes atributos, de pacientes atendidos pelo Centro de Alzheimer e Parkinson no município de Campos dos Goytacazes/RJ. O método de particionamento do conjunto de dados foi o *holdout*, com amostragem estratificada. Posteriormente, foi aplicado o algoritmo de classificação supervisionado baseado na técnica de Naïve Bayes, cuja implementação encontra-se disponível na ferramenta Weka. Observaram-se os resultados numéricos dos modelos de acordo com as probabilidades estimadas pelo classificador, os quais evidenciaram os fatores de risco idade e gênero, bem como os testes de rastreio, denominados Mini Exame do Estado Mental e o Teste do Relógio, como os mais relevantes para o diagnóstico das patologias tratadas na pesquisa.

Palavras-chave: Mineração de dados. Diagnóstico Médico. Síndromes Demenciais. Aprendizagem de máquina

[1] E-mail: leonardbarreto@gmail.com

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

ABSTRACT - This study aimed to identify the most advanced risk factors and screening tests in the treatment of patients with clinical suspicion of medical screening, with the support of machine learning techniques. The data set was made by 605 medical records, with 18 different attributes, of patients attended by the Alzheimer's Center and Parkinson's in Campos dos Goytacazes/RJ. The method of partitioning the data set was holdout, with stratified sampling. Subsequently, we applied the supervised classification algorithm based on the Naïve Bayes technique, whose implementation is available in the Weka tool. We observed the numerical results of the models according to the probabilities estimated by the classifier, which showed the age and gender risk factors, as well as the screening tests, called Mini Mental State Examination and the Clock Test, as the most relevant to the diagnosis of pathologies treated in the research.

Keywords: Data Mining. Medical Diagnostic. Dementia Syndromes. Machine learning.

INTRODUÇÃO

O envelhecimento da população é um acontecimento mundial, que se dá de forma diferenciada em cada região. Conforme dados do Banco Mundial (IHME, 2013), a partir da década de 1970, a expectativa de vida mundial cresceu, aproximadamente, 20 anos. O Brasil, neste mesmo período, se posicionou entre os países da América Latina que mais realizaram progressos na expectativa de vida da população, com um aumento de 30 anos. De acordo com o censo realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em 1960, 3,3 milhões de brasileiros tinham 60 anos ou mais e representavam 4,7 por cento da população. Já no Censo de 2010, essa representação passou para 20,5 milhões de brasileiros, aproximadamente, 10,8 por cento da população. Estes números seguem a tendência mundial de crescimento da população idosa.

Segundo dados do Fundo de População das Nações Unidas (2012) , em 2012 a população mundial de pessoas com 60 anos ou mais totalizava em 800 milhões, representando 11,5 por cento da população. Prevê-se que em 2050 o número de pessoas nesta faixa etária esteja em torno de dois bilhões, correspondendo a 21,8 por cento das pessoas no mundo. A estimativa da Organização Mundial da Saúde (OMS) indica que o

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

Brasil será o sexto país em número de idosos em 2050, quando deve chegar a 58 milhões de pessoas com 60 anos ou mais.

Com o crescimento da população idosa, ocorre o aumento de problemas vinculados aos declínios no funcionamento fisiológico, associados tanto a processos evolutivos normais, quanto, para alguns indivíduos, a processos patológicos (OWSLEY, 2011) . No âmbito da saúde mental, o aumento da proporção de idosos com algum tipo de demência tem estreita relação com a ampliação de indivíduos nesta faixa etária. O número de pessoas com demência no mundo em 2016 foi estimado em 47 milhões, devendo chegar a 75,62 milhões em 2030 e 135,46 milhões em 2050 (ALZHEIMER'S DISEASE INTERNATIONAL, 2015) . Considera-se esta uma das principais causas de incapacidade durante o envelhecimento, gerando transtornos não só para o indivíduo como também para a família e a sociedade (FERRI, 2012) .

A cognição humana altera naturalmente ao longo do tempo, podendo ser observados declínios na velocidade de processamento da informação, na capacidade da memória de trabalho e na função de inibição cognitiva, isto é, no controle executivo. Alguns adultos idosos têm dificuldades cognitivas, as quais ultrapassam aquilo que poderia ser considerado normal. Em alguns casos, essas dificuldades não são tão graves que justifiquem um diagnóstico de demência, mas, em outros casos, revelam-se persistentemente progressivas (ZELINSKI; DALTON; HINDIN, 2011) .

O impacto econômico da demência é enorme, com custo em todo o mundo estimado na faixa de 818 bilhões de dólares em 2014, o equivalente a mais de 1% do produto interno bruto global. Estima-se que em 2018 os custos com as demências gire em torno de US$ 1 trilhão, podendo chegar à quantia de US$ 2 trilhões em 2030 (ALZHEIMER'S ASSOCIATION, 2017) . Nos Estados Unidos, por exemplo, os custos totais relacionados à doença de Alzheimer somente são superados pelos relacionados ao câncer e às doenças coronarianas (WORLD HEALTH ORGANIZATION, 2017) . Nesse sentido, reconhece-se que o envelhecimento não seja somente uma questão de segurança social e bem-estar, mas, sobretudo, de desenvolvimento global e de política econômica (UNITED NATIONS POPULATION FUND, 2012) .

Dentro do cenário apresentado, o presente trabalho visa a evidenciar os fatores de risco e testes de rastreio mais relevantes para o processo de diagnóstico de pacientes com suspeita clínica de demência na triagem médica, com uso de técnicas de aprendizagem de máquina.

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

DEMÊNCIA, FATORES DE RISCO E DIAGNÓSTICO

A demência não é caracterizada, essencialmente, como uma doença, mas, sim, como um termo genérico que descreve uma grande variedade de sintomas relacionados ao declínio cognitivo do indivíduo, impactando sua capacidade de realizar simples tarefas cotidianas (ADI, 2013). São numerosas as doenças do sistema nervoso central que causam demências. Dentre as diversas causas, as principais são a doença de Alzheimer (DA), e as doenças cerebrovasculares (demência vascular - DV), sendo observados casos de etiologias mistas (demência mista - DM), como DA associada à DV (CARAMELLI, 2016) . Cabe ressaltar outra condição clínica que é o transtorno cognitivo leve (TCL), situação em que um indivíduo apresenta algum grau de perda cognitiva (como o esquecimento, por exemplo) quando comparado com outras pessoas da mesma idade, mas que não se enquadram no diagnóstico de demência (MACHADO, 2016) . Dentre os diversos tipos, a doença de Alzheimer é a mais prevalente, respondendo por 60% a 70% dos casos (ALZHEIMER'S DISEASE INTERNATIONAL, 2015) .

Nesse sentido, a contribuição da epidemiologia para pesquisa em demência tem sido de grande importância, principalmente com relação aos fatores de risco, com o estabelecimento de taxas de incidência, de prevalência e demonstração desses fatores. A identificação dos fatores de risco é de vital importância, devido às potenciais implicações para a prevenção ou tratamento da doença, especialmente para os casos que são passíveis de reversão. Estratégias de saúde pública baseadas na eliminação de riscos já reduziram a incidência de doenças, mesmo quando as causas são desconhecidas. Historicamente, essa tem sido uma importante função da pesquisa epidemiológica (FREITAS; PY, 2016; RITCHIE *et al*., 2013) .

Desse modo, diversos trabalhos científicos relacionados à demência sugerem uma etiologia multifatorial, em que fatores clínicos, genéticos e ambientais, eventualmente correlacionados através de complexas interações, modulariam o risco de desenvolvimento para a doença (CARAMELLI, 2016) .

A *Alzheimer's Disease International* (ADI) descreve os fatores de risco para as demências e os categoriza em modificáveis e não modificáveis. Os fatores de risco modificáveis estão relacionados aos aspectos passíveis de serem mudados, prevenidos

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

ou tratados, diferentemente dos fatores não modificáveis, sobre os quais não se tem controle, impossibilitando o tratamento. Dentre os fatores modificáveis, podem-se destacar Fatores de risco vascular, tabagismo, obesidade, hipertensão, hipercolesterolemia, diabetes, síndrome metabólica, inatividade física, dieta e nutrição, ingestão de sódio, consumo de álcool, depressão, inatividade cognitiva ou baixo nível de escolaridade. Idade, história familiar, síndrome de *Down* e genética podem ser citados como fatores ditos não modificáveis (PRINCE *et al*., 2016).

A idade é, indiscutivelmente, um dos principais fatores de risco para a maioria das síndromes demenciais. Particularmente para a DA, além das etiologias acima, o gênero feminino é considerado um fator de risco, principalmente após os 80 anos de idade. Porém, como as mulheres possuem expectativa de vida maior que a dos homens, este fator pode ser questionável (MACHADO, 2016) . Já para a DV, os homens são mais afetados do que as mulheres (SPERANZA; MOSCI, 2016) .

Dentre os fatores não modificáveis, além da idade, outro forte fator de risco é a história familiar. Pesquisas indicam que a existência de algum familiar com a doença torna a pessoa mais propensa a desenvolvê-la, e que este risco é maior se mais de um membro da família possuir a patologia (ADI, 2012). Em indivíduos com, pelo menos, um parente de primeiro grau com a doença, este risco é 3,5 vezes maior. Geralmente, a história familiar está mais associada à DA de início precoce do que de início tardio (MACHADO, 2016) .

Diversos estudos sugerem que o controle de alguns fatores modificáveis pode diminuir a incidência das demências (JACQMIN-GADDA *et al*., 2013) . Estes fatores podem ser subdivididos em sociodemográficos, clínicos, ambientais e de estilo de vida (DENNY *et al*., 2010) . A escolaridade é apontada como um dos fatores sociodemográficos mais relevantes, de forma isolada ou correlacionada com fatores não modificáveis, como a idade e o gênero. Esta associação é alvo de diversas pesquisas, como pode ser observado nos trabalhos de Herrera *et al*. (2002) e Campbell *et al*. (2013) . Os referidos trabalhos demonstram que a baixa escolaridade associada ao aumento da idade relaciona-se com o crescimento da prevalência de demências. Herrera *et al*. (2002) relacionam também o gênero feminino aos fatores mencionados. Em contrapartida, indivíduos com alto nível educacional e submetidos a atividades de lazer mentalmente estimulantes são menos suscetíveis ao risco de demência (PRINCE *et al*., 2016) .

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

O diagnóstico das síndromes demenciais é eminentemente clínico, baseado em avaliação objetiva do desempenho cognitivo e funcional. A avaliação funcional do paciente com suspeita de demência inicia-se com anamnese, sendo seguida por uma bateria de exames cognitivos que buscam evidências de déficit na memória que acarretem interferência significativa sobre o desempenho do indivíduo em suas atividades sociais e ocupacionais (MACHADO, 2016). Feita a avaliação, são aplicados, quando necessários, os critérios para o diagnóstico de demência, que são os indicadores dos sintomas cognitivos, comportamentais e psiquiátricos que caracterizam cada uma das diferentes formas de demência.

Neste contexto, os exames neuropsicológicos constituem-se de ferramentas para o auxílio no processo de investigação das demências, avaliando as habilidades cognitivas (exemplo: fluência verbal), funcionais (exemplo: higiene pessoal) e comportamentais (exemplo: agitação). Dentre os diversos instrumentos e metodologias existentes para auxílio no processo de investigação, controle e evolução das demências, serão citadas as técnicas comumente adotadas internacionalmente. Adicionalmente, os testes recomendados, validados e empregados no Brasil, conforme trabalhos de CHAVES *et al*. (2011) e recomendações do Departamento Científico de Neurologia Cognitiva e do Envelhecimento da ABN, serão descritos no decorrer deste artigo. Para cada teste, serão apresentados também a escala utilizada e, quando couber, os respectivos pontos de corte que podem auxiliar a identificar o(s) nível(is) de comprometimento da doença.

Este conjunto sistematizado de ferramentas, utilizado para auxiliar o processo de diagnóstico de demências, é categorizado de forma sucinta em função dos domínios Cognitivo (Mini Exame do Estado Mental - MEEM, Fluência Verbal, Teste do Relógio, Lista de Figuras, Lista de palavras do CERAD (*Consortium to Estabilish a Registry for Alzheimer's Disease*) e o Exame cognitivo de Cambridge - CAMCOG), Funcional (Atividades Básicas da Vida Diária – AVD e Atividades Instrumentais da Vida Diária - AIVD), Comportamental (Escala de depressão Geriátrica - GDS) e os pertencentes aos domínios Cognitivo e Funcional (Avaliação Clínica da Demência – CDR).

O diagnóstico da causa da doença, por sua vez, depende de investigação complementar constituída, essencialmente, por exames laboratoriais e de neuroimagem estrutural. Entretanto, uma avaliação clínica cuidadosa, incluindo anamnese detalhada, exames físico e neurológico, associados a determinações bioquímicas e de neuroimagem, podem possibilitar maior acurácia no diagnóstico diferencial

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

(GALLUCCI NETO; GARCIA TAMELINI; FORLENZA, 2005; SPERANZA; MOSCI, 2016) .

É crescente o avanço de pesquisas direcionadas ao diagnóstico de demências. O processo de investigação diagnóstica é inerentemente difícil, dada a diversidade de doenças, sintomas, exames, além da complexidade da fisiologia humana. Especificamente, em relação às síndromes demenciais, a inexistência de um exame específico para refutar ou determinar o subtipo potencializa ainda mais esse problema, demandando, assim, uma extensa investigação clínica através de uma série de consultas, avaliações e exames (MOREIRA; NAMEN, 2016) . Dessa forma, tanto os especialistas do domínio quanto a sociedade em geral podem se beneficiar do desenvolvimento de ferramentas que permitam minimizar a complexidade do processo de diagnóstico de doenças.

A criação de sistemas computacionais inteligentes na área de saúde tem sido um dos desafios tecnológicos mais ambiciosos e controversos de nossa era. A pesquisa e o desenvolvimento desses sistemas datam da década de 1970, através dos primeiros sistemas de apoio à decisão implantados em hospitais e grandes instituições. Desde então, observa-se uma profunda transformação de tais sistemas, particularmente no diz respeito à sua aplicação. Neste contexto, podem-se citar os sistemas de suporte à decisão clínica (*Clinical Decision Support Systems* - CDSS), que, em essência, visam auxiliar especialistas no processo de investigação médica, possibilitando diagnósticos mais precisos (SHORTLIFFE; CIMINO, 2014).

O diagnóstico médico em CDSS é geralmente apoiado por diversos métodos e modelos matemáticos, e o uso destes geralmente depende de sua finalidade, impactando na escolha de uma metodologia ou de um conjunto de métodos. Considerada uma das técnicas com vasto potencial na área da saúde, a mineração de dados é uma metodologia emergente na área de descoberta de conhecimento e tem sido largamente utilizada em atividades de previsão diagnóstica, extração de padrões na descoberta de conhecimento em grandes conjunto de dados (LÓPEZ *et al*., 2013; PROVOST; FAWCETT, 2013) .

MATERIAIS E MÉTODOS

A presente pesquisa, classificada como descritiva quantitativa, utilizou dados provenientes de informações contidas nos prontuários de 605 pacientes atendidos pelo

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Centro de Alzheimer e Parkinson (CDAP) em Campos dos Goytacazes/RJ, coletadas a partir da aplicação desenvolvida. As informações dos pacientes consistem em dados demográficos, clínicos e laboratoriais. Os dados relativos à identificação do paciente foram desconsiderados no trabalho, permitindo total anonimato.

O desenvolvimento desta foi apoiado pelas seguintes etapas: (i) seleção dos dados, (ii) pré-processamento e transformação dos dados, (iii) construção dos modelos e (iv) avaliação da performance.

Na etapa (i) foi realizada a coleta e integração dos dados contidos nos prontuários (em papel) com informações parciais dos pacientes existentes em uma planilha eletrônica. Os dados coletados e integrados na primeira etapa foram analisados, sendo adotado como critério de seleção para o processo de mineração o quantitativo de pacientes diagnosticados com algum tipo de demência. Deste subconjunto, buscou-se os registros médicos de pacientes em estágio inicial referentes ao primeiro atendimento. O quantitativo de pacientes diagnosticados por patologia, bem como os percentuais associados, é apresentado na Tabela 1.

Tabela 1. Quantitativo e respectivos percentuais de pacientes diagnosticados com os diversos tipos de demência contidos no conjunto de dados D3.

Diagnóstico	Quantitativo	Percentual
Normal	2	0,33%
Transtorno Cognitivo Leve (TCL)	97	16,03%
Doença de Alzheimer (DA)	290	47,93%
Demência vascular (DV)	53	8,76%
Demência mista (DM)	73	12,07%
Outros	90	14,87%

Fonte: Elaborado pelos autores

Na etapa (ii), pré-processamento e transformação dos dados, foram realizadas ações para adequar os dados às escalas e pontos de corte indicados na literatura médica e por especialistas do domínio. Posteriormente, o atributo classe *diagnóstico* foi transformado para binário, onde 1 (um) indica que o paciente é acometido pela doença e

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

0 (zero), caso contrário. A Tabela 2 apresenta informações detalhadas sobre os atributos selecionados para etapa (iii).

Tabela 2. Relação de atributos, e respectivas características, selecionados da base de dados após a etapa de pré-processamento

Atributo e Referência (escala)	Estados	Descrição
tabagismo	0: negativo 1: positivo 2: Não informado	Indica se paciente é (foi) tabagista ou não
etilismo	0: negativo 1: positivo 2: Não informado	Indica se paciente é (foi) etilista ou não
estado_civil	0: Não informado 1: solteiro (a) 2: casado (a) / morando junto 3: viúvo (a) 4: Separado (a) / divorciado (a)	Estado civil do paciente
idade	0: <65 anos 1: >= 65 anos	Idade do paciente
gênero	1: masculino 2: feminino	Gênero
escolaridade	0: analfabeto 1: 1 a 4 anos de escolaridade 2: 5 a 8 anos de escolaridade 3: 9 ou mais anos de escolaridade	Educação (em anos)
meem (BRUCKI et al., 2003)	0: (0-17] 1: [18-26] 2: [27-30]	Resultado do Mini Exame do Estado Mental
avd (DUARTE; DE ANDRADE; LEBRÃO, 2007)	0: grande dependência 1: dependência moderada 2: independência	Resultado da avaliação de Atividades de Vida Diária

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

	Atributo e Referência (escala)	Estados	Descrição
	Aivd (LAWTON et al., 1982)	1: independência 2: dependência parcial 3: dependência total	Resultado da avaliação de Atividades Instrumentais de Vida Diária
0	gds (YESAVAGE, 1988)	0: normal 1: depressão 2: depressão grave	Resultado da avaliação da Escala de Depressão Geriátrica
1	Camcog (PARADELA; LOPES; LOURENÇO, 2009)	A: alto M: moderado B: baixo	Resultado do índice de comprometimento cognitivo no teste CAMCOG
2	cdr (MACEDO MONTAÑO; ANDREONI; RAMOS, 2013)	0: saudável 0,5: questionável 1: leve 2: moderada 3: grave	Resultado do teste referente a Escala clínica de demência (*clinical dementia rating*)
3	fluência_verbal (CHAVES et al., 2011)	0: <= 10 palavras 1: > 10 palavras	Quantidade de palavras evocadas em 1 minuto
4	diabetes	0: negativo 1: positivo	Indica se o paciente possui diabetes
5	has	0: negativo 1: positivo	Indica se o paciente possui hipertensão arterial
6	Cerad (BERTOLUCCI et al., 2001)	0: menos de 13 pontos 1: 13 ou mais pontos	Pontuação obtida no teste do CERAD
7	Relogio (SHULMAN; SHEDLETSKY; SILVER, 1986)	0: inabilidade absoluta para representar o relógio 1: O desenho tem algo a ver com o relógio mas com desorganização visuo-espacial grave; 2: Desorganização visuo-espacial moderada	Resultado do Teste do Relógio

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

Atributo e Referência (escala)	Estados	Descrição
	que leva a uma marcação de hora incorreta, perseveração, confusão erquerda-direita, números faltando, números repetidos, sem ponteiros, com ponteiros em excesso 3: Distribuição visuo-espacial correta com marcação errada da hora; 4: Pequenos erros espaciais com dígitos e hora corretos; 5: Relógio perfeito.	
8 diagnóstico	0: negativo 1: positivo	Indica se o paciente possui ou não determinado tipo de demência.

Fonte: Elaborado pelos autores

A construção dos modelos preditivos, etapa (iii), consistiu, inicialmente, na designação do atributo alvo (ou classe), denominado *diagnóstico*. Em seguida, o conjunto de dados foi particionado através do método *holdout*, que divide os dados em dois subconjuntos, mutuamente exclusivos, chamados de conjunto de treinamento e conjunto de teste. O tamanho das partições pode ser dividido em diversos esquemas, sendo comum a designação de 2/3 dos dados para treinamento e o restante para o conjunto de teste. Cabe salientar que a divisão do conjunto de dados seguiu a estratégia de amostragem estratificada, em que as proporções da classe do conjunto original são mantidas nos conjuntos de teste e treinamento. A construção do modelo de mineração de dados foi realizada por intermédio do classificador ingênuo de Bayes (Naïve Bayes), disponível na ferramenta Weka (*Waikato Environment for Knowledge Analysis*).

Por fim, após a construção do modelo, os resultados foram avaliados (etapa iv) segundo a probabilidade associada a cada um dos fatores de risco e exames de rastreio obtidos pelo modelo.

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

RESULTADOS E DISCUSSÃO

De abordagem simples e semântica clara, o classificador probabilístico Naïve Bayes se caracteriza pela suposição de que os atributos são condicionalmente independentes, dado o rótulo da classe (WITTEN; FRANK; HALL, 2011). Cabe salientar que, mesmo aplicado em situações em que não há independência condicional dos atributos, caso da presente pesquisa, esse método tem obtido resultados bastante satisfatórios (ABRAHAM; SIMHA; IYENGAR, 2007; VEMBANDASAMY; SASIPRIYA; DEEPA, 2015) .

As probabilidades de cada valor do atributo e da classe ocorrer são estimadas com base nas respectivas frequências no conjunto de treinamento para cada rótulo de classe. A partir da multiplicação destas probabilidades, obtém-se a classificação de um novo registro, rotulado com a classe de maior valor proveniente desta multiplicação. A Tabela 3 apresenta os quantitativos e as respectivas probabilidades referentes aos valores dos atributos mais representativos para que um registro seja classificado na classe de interesse, que indica o diagnóstico de DA, TCL, DV ou DM. Por limitações de espaço, foram ocultadas as probabilidades dos demais valores de cada atributo.

Tabela 3. Modelo gerado pela técnica naïve bayes para as patologias DA, TCL, DM e DV.

Atributo	**Valor**		**Probabilidades**			
			Classes			
			DA	**TCL**	**DM**	**DV**
tabagismo	0	Negativo	**0,74**	**0,80**	**0,96**	**0,95**
etilismo	0	Negativo	**0,82**	**0,88**	**0,98**	**0,98**
estado civil	2	Casado / morando junto	0,40	0,36	0,55	**0,74**
idade	1	65 anos ou mais de idade	**0,95**	**0,84**	**0,99**	**0,99**
gênero	2	Feminino	**0,64**	**0,72**	**0,77**	0,49
escolaridade	1	1 a 4 anos de escolaridade	**0,48**	**0,56**	**0,67**	0,57

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

Atributo	Valor		Probabilidades			
			Classes			
			DA	TCL	DM	DV
meem	0	0 a 17 pontos na avaliação	**0,59**	0,12	**0,69**	0,59
	1	18 a 26 pontos na avaliação	0,36	**0,65**	0,29	0,37
avd	2	Independência	**0,75**	**0,92**	**0,74**	**0,81**
aivd	1	Independência	0,39	**0,90**	0,20	0,47
gds	0	Normal	0,49	0,54	**0,80**	**0,80**
camcog	A	Alto	0,27	0,04	**0,95**	**0,92**
	B	Baixo	0,39	**0,81**	0,02	0,06
cdr	0	Saudável	0,03	0,24	0,00	0,03
	0.5	Questionável	0,243	0,56	0,22	0,33
	1	Leve	0,36	0,14	0,28	0,40
	2	Moderada	0,23	0,02	0,38	0,22
	3	Grave	0,12	0,02	0,10	0,00
fluencia_verbal	1	Mais de 10 palavras/min	0,57	**0,89**	0,43	0,47
diabetes	0	Negativo	0,20	0,45	0,23	0,11
	1	Positivo	**0,78**	**0,53**	**0,76**	**0,88**
	2	Não informado	0,00	0,01	0,00	0,00
has	1	Positivo	**0,90**	**0,81**	**0,98**	**0,99**
	2	Não informado	0,00	0,01	0,00	0,00
cerad	0	Menos de 13 pontos nas 3	**0,81**	**0,67**	**0,87**	**0,80**

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

Atributo	Valor		Probabilidades			
			Classes			
			DA	TCL	DM	DV
		tentativas				
relogio	0	Inabilidade absoluta para representar o relógio	**0,55**	0,16	**0,71**	**0,83**

Fonte: Elaborado pelos autores

De acordo com os dados da Tabela 3, das características ligadas aos fatores de risco, como idade, gênero, escolaridade, depressão, diabetes e hipertensão arterial (*has*), o fator idade, para idosos a partir de 65 anos, foi característica comum em pelo menos 84,9 por cento das pessoas diagnosticadas com alguma das patologias do presente estudo. Outro fator categorizado como não modificável, o gênero feminino, sugere uma alta predominância de casos em pacientes com DA ou TCL bem como a proporcionalidade de casos entre homens e mulheres para DV, confirmando, assim, estudos que indicam esse comportamento. No entanto, o alto número de casos para pacientes acometidos por DM do sexo feminino (77,0 por cento) contrasta com o equilíbrio de casos de ambos os gêneros, conforme o trabalho de Pinheiro, Carvalho e Luppi (2013) .

Em relação aos outros fatores, considerados modificáveis, a probabilidade de um paciente possuir hipertensão arterial (*has*) ou diabetes é alta para todas as patologias, confirmando estudos apontados por Machado (2016) e Ritchie e Tuokko (2011) . Por fim, do conjunto de pessoas acometidas pelas referidas doenças, a baixa escolaridade é apontada como um fator determinante: o analfabetismo é mais característico para 19 por cento de indivíduos com DA, enquanto que para pacientes que tiveram entre 1 a 4 anos de estudo, esta taxa é consideravelmente superior em todas as patologias. Cabe ressaltar que, apesar de considerado pelos referidos autores como elementos clínicos relevantes para o diagnóstico de demências, a diferença entre indivíduos dignosticados com DV com baixa e alta escolaridades não foi tão acentuada quanto nas demais patologias, sugerindo que pessoas com 9 ou mais anos de estudo têm uma probabilidade de aproximadamente 30 por centro de serem acometidos por DV.

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

No outro conjunto de atributos encontram-se os resultados das avaliações dos testes neuropsicológicos, como o MEEM, AVD, AIVD, GDS, Camcog, CDR, Fluência verbal, Lista de palavras do CERAD e o teste do relógio. Deste subconjunto, o MEEM apresenta-se como característica mais relevante para indivíduos acometidos por DA, DM e DV, uma vez que em 59, 69 e 59 por cento, respectivamente, dos casos o resultado do teste foi ruim (0 a 17 pontos) nestes grupos. Para TCL, particularmente, o resultado considerado mediano para esse teste de rastreio (*meem=1*) foi mais expressivo, denotando um menor comprometimento das funções cognitivas em indivíduos acometidos por essa enfermidade.

A maioria absoluta dos pacientes portadores de TCL, DA, DM e DV é considerada independente no teste de AVD. No entanto, especificamente para as últimas três patologias citadas, um ligeiro percentual de indivíduos apresenta grandes dificuldades (*avd=0*) para realizar atividades de autocuidado.

Em relação às capacidades de administração do ambiente de vida dentro e fora do lar, avaliadas pelo AIVD, cerca de 90 por cento dos indivíduos com TCL são considerados independentes (*aivd=1*). Dentre os que apresentam algum tipo de dependência neste exame (*aivd=2 ou 3*), as maiores taxas encontram-se em pessoas acometidas por DM ou DV.

Cerca de 80 por cento das pessoas portadoras de DM e DV não apresentam quadros depressivos (*gds=0*), seguidas, em menor escala (49 e 55 por cento), por aquelas com DA e TCL, respectivamente. No entanto, quadros depressivos (*gds=1*) são mais característicos em DA (40 por cento) e, em menor intensidade, para indivíduos com TCL (36 por cento).

Entre as pessoas diagnosticadas com DM e DV, a probabilidade destas apresentarem um comprometimento cognitivo alto no Camcog é de, aproximadamente, 95 e 92 por cento, respectivamente, em ambas as doenças. Já para os pacientes diagnosticados com TCL, é altamente provável (81,3 por cento) o não comprometimento cognitivo (*camcog=B*) nesse grupo, de acordo com o citado exame.

Os resultados relacionados com o teste CDR indicam a inexistência de uma categoria predominante para determinação de alguma das patologias citadas, exceto pela ligeira maioria observada para o grau de gravidade 'questionável' (*cdr=0.5*) do CDR para TCL. Ainda sobre a referida doença, observa-se uma taxa relativamente superior (24 por cento) para indivíduos saudáveis (*cdr=0*) em relação às demais patologias, sugerindo que as áreas de cognição e comportamento não são afetadas em 24 por cento dos casos de pacientes com suspeita de TCL.

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Dentre os indivíduos avaliados em relação à fluência verbal, aproximadamente 89 por cento dos que conseguiram realizar o teste com sucesso foram diagnosticados com TCL, ou seja, a capacidade de produção de palavras em um determinado período de tempo não é afetada em indivíduos acometidos por esta patologia. No entanto, para as demais doenças, este teste não apresentou resultados conclusivos.

Os altos percentuais associados ao baixo desempenho no teste de lista de palavras do CERAD (*cerad=0*) mostraram-se determinantes para o diagnóstico positivo em todas as patologias.

Em relação ao teste de relógio, as altas taxas associadas aos resultados considerados alterados (pontuações de 0 a 2), especialmente com predominância em situações em que o indivíduo não apresenta habilidade para representar o relógio (*relogio=0*), sugerem uma alta incidência destas doenças neste grupo de pacientes. Cabe ressaltar que tal comportamento é menos acentuado para TCL, visto que as pontuações 4 e 5 no referido teste apresentam percentuais ligeiramente maiores das demais patologias na mesma faixa de escore, o que, provavelmente, indica uma menor influência desse teste de rastreio para o diagnóstico de TCL.

CONSIDERAÇÕES FINAIS

Por intermédio desses modelos de classificação, foi possível verificar os fatores e testes neuropsicológicos mais relevantes para predizer o grau de acometimento de pacientes com suspeita clínica de demência. Das características ligadas aos fatores de risco, a idade foi apontada como um aspecto crítico para o diagnóstico em todas as patologias objetos da pesquisa, corroborando assim com estudos da respectiva área. Dentre o conjunto de atributos considerados modificáveis, a hipertensão, diabetes e o grau de escolaridade surgem como fatores determinantes para o diagnóstico de síndromes demenciais. No entanto, apesar de citado na literatura como um aspecto importante para a investigação das demências, a escolaridade não se mostra tão decisiva para indivíduos acometidos por DV. Tal característica é compartilhada com o atributo gênero, um fator não modificável, para a mesma patologia.

Em relação aos testes de rastreio, notou-se que o teste do relógio tem influência para confirmar ou refutar a presença das doenças tratadas nessa pesquisa. Já o Mini-Exame do Estado Mental (MEEM), considerado na literatura a escala de avaliação

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

cognitiva mais amplamente utilizada para diagnósticos de demências, foi mais determinante para o diagnóstico de DA.

Por fim, os modelos evidenciam a influência do Exame Cognitivo de Cambridge (CAMCOG) na investigação clínica de indivíduos acometidos por DM e DV.

REFERÊNCIAS

ABRAHAM, R.; SIMHA, J. B.; IYENGAR, S. S. **A comparative analysis of discretization methods for medical datamining with Naïve Bayesian classifier**. Proceedings - 9th International Conference on Information Technology, ICIT 2006. **Anais**...2007

ALZHEIMER'S ASSOCIATION. 2017 Alzheimer's Disease Facts and Figures. **Alzheimers Dement**, v. 13, p. 325–373, 2017.

ALZHEIMER'S DISEASE INTERNATIONAL. **World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trendsAlzheimer's Disease International**. London: Alzheimer's Disease International. Disponível em: <https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf>.

BERTOLUCCI, P. H. F. et al. Applicability of the CERAD neuropsychological battery to Brazilian elderly. **Arquivos de Neuro-Psiquiatria**, v. 59, p. 532–536, 2001.

BRUCKI, S. M. D. et al. Sugestões para o uso do mini-exame do estado mental no Brasil (Suggestions for the use of mini-mental state in Brazil). **Arquivos de Neuro-Psiquiatria**, v. 61, n. 3 B, p. 777–781, 2003.

CAMPBELL, N. L. et al. **Risk factors for the progression of mild cognitive impairment to dementiaClinics in Geriatric Medicine**, 2013.

CARAMELLI, P. Avaliação Clínica e Complementar para o Estabelecimento do Diagnóstico de Demência. In: **Tratado de geriatria e gerontologia**. 4. ed. Rio de Janeiro: Guanabara-Koogan, 2016.

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

CHAVES, M. L. F. et al. **Avaliação cognitiva, comportamental e funcional: Doença de AlzheimerDementia e Neuropsychologia**, 2011.

DENNY, J. C. et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. **Bioinformatics**, v. 26, 2010.

DUARTE, Y. A. DE O.; DE ANDRADE, C. L.; LEBRÃO, M. L. O Índex de Katz na avaliação da funcionalidade dos idosos (The Katz Index in assessing the functionality of the elderly). **Revista da Escola de Enfermagem**, v. 41, n. 2, p. 317–325, 2007.

FERRI, C. P. Envelhecimento populacional na América Latina: demência e transtornos relacionados. **Revista Brasileira de Psiquiatria**, v. 34, n. 2, p. 371–374, 2012.

FREITAS, E. V. DE; PY, L. **Tratado de Geriatria e Gerontologia (Geriatrics and Gerontology Treaty)**. 4. ed. Rio de Janeiro: Guanabara-Koogan, 2016.

GALLUCCI NETO, J.; GARCIA TAMELINI, M.; FORLENZA, O. V. Diagnóstico diferencial das demências. **Revista de Psiquiatria Clinica**, v. 32, n. 3, p. 119–130, 2005.

HERRERA, E. et al. Epidemiologic survey of dementia in a community-dwelling Brazilian population. **Alzheimer Disease and Associated Disorders**, v. 16, n. 2, p. 103–108, 2002.

JACQMIN-GADDA, H. et al. 20-Year prevalence projections for dementia and impact of preventive policy about risk factors. **European Journal of Epidemiology**, v. 28, n. 6, p. 493–502, 2013.

LAWTON, M. P. et al. A research and service oriented multilevel assessment instrument. **Journal of Gerontology**, v. 37, n. 1, p. 91–99, 1982.

LÓPEZ, V. et al. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. **Inf. Sci.**, v. 250, p. 113–141, 2013.

Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

MACEDO MONTAÑO, M. B. M.; ANDREONI, S.; RAMOS, L. R. Clinical Dementia Rating independently predicted conversion to dementia in a cohort of urban elderly in Brazil. **International Psychogeriatrics**, v. 25, n. 2, p. 245–251, 2013.

MACHADO, J. C. B. Doença de Alzheimer. In: **Tratado de Geriatria e Gerontologia**. 4. ed. Rio de Janeiro: Guanabara-Koogan, 2016.

MOREIRA, L. B.; NAMEN, A. A. Sistema preditivo para a doença de Alzheimer na triagem clínica. **Journal of Health Informatics**, v. 8, n. 3, p. 87–94, 2016.

OWSLEY, C. Aging and vision. **Vision Research**, v. 51, n. 13, p. 1610–1622, 2011.
PARADELA, E. M. P.; LOPES, C. DE S.; LOURENÇO, R. A. Adaptação para o português do Cambridge Cognitive Examination-Revised aplicado em um ambulatório público de geriatria (Portuguese adaptation of the Cambridge Cognitive Examination-Revised applied in a public outpatient clinic of geriatrics). **Cadernos de Saúde Pública**, v. 25, n. 12, p. 2562–2570, 2009.

PINHEIRO, J. S.; CARVALHO, M. F. C.; LUPPI, G. Interação medicamentosa e a farmacoterapia de pacientes geriátricos com síndromes demenciais. **Revista Brasileira de Geriatria e Gerontologia**, v. 16, p. 303–314, 2013.

PRINCE, M. et al. World Alzheimer Report 2016 Improving healthcare for people living with dementia. Coverage, Quality and costs now and in the future. **Alzheimer's Disease International (ADI)**, p. 1–140, 2016.

PROVOST, F.; FAWCETT, T. **Data Science for Business: What You Need to Know About Data Mining and Data-analytic Thinking**. 1. ed. Sebastopol, CA: O'Reilly Media, 2013.

RITCHIE, L. J.; TUOKKO, H. Clinical decision trees for predicting conversion from cognitive impairment no dementia (CIND) to dementia in a longitudinal population-based study. **Archives of Clinical Neuropsychology**, v. 26, n. 1, p. 16–25, 2011.

RITCHIE, M. D. et al. Genome- and phenome-wide analyses of cardiac conduction identifies markers of arrhythmia risk. **Circulation**, v. 127, 2013.



Temas em Saúde
Volume 18, número 4
ISSN 2447-2131
João Pessoa, 2018

Artigo

SHORTLIFFE, E. H.; CIMINO, J. J. **Biomedical Informatics Computer Applications in Health Care and Biomedicine**. 4. ed. London: Springer-Verlag, 2014.

SHULMAN, K. I.; SHEDLETSKY, R.; SILVER, I. L. The challenge of time: Clock - drawing and cognitive function in the elderly. **International Journal of Geriatric Psychiatry**, v. 1, n. 2, p. 135–140, 1986.

SPERANZA, A. C. C.; MOSCI, T. Diagnóstico Diferencial das Demências. In: **Tratado de geriatria e gerontologia**. 4. ed. Rio de Janeiro: Guanabara-Koogan, 2016.

UNITED NATIONS POPULATION FUND. **Ageing in the Twenty-First Century: A Celebration and A ChallengeUnited Nations Population Fund**. New York: United Nations Population Fund (UNFPA), 2012. Disponível em: <https://www.unfpa.org/public/home/publications/pid/11584>.

VEMBANDASAMY, K.; SASIPRIYA, R.; DEEPA, E. Heart Diseases Detection Using Naive Bayes Algorithm. **International Journal of Innovative Science, Engineering & Technology**, v. 2, n. 9, p. 441–444, 2015.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. 3. ed. San Francisco, CA, USA: Morgan Kaufmann, 2011.

WORLD HEALTH ORGANIZATION. **World health statistics 2017: monitoring health for the SDGs, Sustainable Development Goals**. Geneva: [s.n.]. Disponível em: <http://www.who.int/gho/publications/world_health_statistics/2017/EN_WHS2017_TOC.pdf?ua=1>.

YESAVAGE, J. A. Geriatric Depression Scale (GDS). **Psychopharmacol. Bull.**, v. 24, n. 1–2, p. 709–711, 1988.

ZELINSKI, E. M.; DALTON, S. E.; HINDIN, S. Cognitive Changes in Healthy Older Adults. **Journal of the American Society on Aging**, v. 35, n. 2, p. 13–20, 2011.