

Modelo de Aprendizado de Máquina para prever alocação de pacientes com lombalgia crônica para unidades de práticas integradas em um sistema de cuidados de saúde baseados em valor.

Machine Learning model to predict back pain patient allocation for integrated health practice units in a value based health care system

Autores

Laboratórios e instituição

*e-mail: preencher

Resumo

A lombalgia crônica é responsável por grande parte dos custos globais de saúde necessitando cuidados especializados por longo tempo. A imensa variabilidade de pacientes em seus itinerários terapêuticos entre hospitais e provedores torna os cuidados de saúde baseados em valor potencialmente menos viáveis. Para avançar mais efetivamente na gestão e nos esforços de prevenção secundária, terciária e quaternária, a adoção de ferramentas, métodos e técnicas de ciência de dados, como algoritmos de aprendizado de máquina supervisionados e não supervisionados se faz necessária, dado que, estes modelos têm se mostrado confiáveis na previsão de resultados específicos em várias doenças neurológicas. Entretanto, o desenvolvimento de estratégias de gerenciamento das unidades de prática integrada ainda não foram testados com auxílio destas técnicas. Neste estudo, fez-se uso de base de dados oriunda de 6 diferentes questionários para avaliação dos pacientes e aplicação de técnicas de correlação entre variáveis ordinais e nominais para determinação de quais questões apresentam maior efeito sobre, principalmente, a lombalgia. Ademais, fez-se uso de 5 diferentes algoritmos de aprendizado de máquina para realizar a predição da ausência ou presença de dor na lombar nos pacientes. Como resultado, selecionou-se 7 variáveis para alimentar os modelos de predição, dos quais o que apresentou melhores acurácias foi o *XBoost Classifier*, com acurácia, precisão e recall de 0.8 e F1-Score de 0.78. O trabalho conduzido permite a redução no número de questões atualmente aplicadas e auxilia no processo de triagem, a partir da resposta fornecida de cada paciente.

Palavras-chave: lombalgia crônica, Saúde Baseada em Valor, aprendizado de máquina

Abstract

Chronic low back pain is responsible for a large part of the global health costs requiring specialized care for a long time. The immense variability of patients in their therapeutic itineraries between hospitals and health care providers has been making the value-based healthcare potentially less viable. In order to effectively advance in the field of secondary, tertiary and quaternary prevention and management and efforts, the adoption of data science tools, as supervised and unsupervised algorithms is necessary, taking into account that these models have been shown reliable in predicting results in various neurological diseases. Nevertheless, the development of management strategies for integrated practice units has not yet been testes with the aid of theses techniques. In this study, it was made use of a database consisted of questions from six different questionnaires to assess the patients. The initial analysis was performed through correlation techniques applied for nominal and ordinal variables to determine which questions have the greatest influence on low back pain condition. Furthermore, five different machine learning algorithms were applied to predict the absence or presence of low back pain. The outcome was consisted of seven variables, those which were used as input to the machine learning models. The models were assessed according to their metrics, and the one which resulted in best performance was the XBoost Classifier, with accuracy, precision and recall with a value of 0.8 and F1-Score of 0.78. The developed work allows the reduction in the number of questions currently applied and assists in the screening process, based on the answers provided by each patient.

Key words: Chronic low back pain, machine learning, Value-Based Healthcare

1 Introdução

Uma das questões primordiais propostas por Michael E. Porter no livro “*Redefining Health Care*” foi o porquê do modelo de gestão competitivo estar falhando no sistema de cuidados em saúde. O autor argumenta, por exemplo, que ao longo da história na economia, a competição no mercado privado é uma das maiores forças para a melhoria na qualidade e nos custos dos produtos e serviços. Entretanto, isto não foi o observado no caso da área de saúde, no qual a competição estava apenas agindo sobre os custos, estes os quais exclusivamente cresciam e não necessariamente a qualidade dos serviços prestados melhoravam ¹.

A necessidade de um sistema de saúde baseado em valor, surgiu da constatação do excessivo e insustentável custo do atual sistema praticado ². Esta análise inicialmente foi realizada principalmente no sistema de saúde americano, o qual apresentou um paradoxo fundamental relacionado ao aumento no conhecimento em biomedicina, o qual foi protagonista em inovações em terapias e em procedimentos cirúrgicos além de gerenciamento de condições as quais anteriormente eram fatais, mas, entretanto, este sistema começou a apresentar problemas em questões básicas relacionados à qualidade, aos resultados atingidos para os pacientes e aos custos ³.

Neste mesmo cenário, têm-se que em 2006 a *National Academy of Medicine* estabeleceu as bases para uma medicina baseada em evidências para o propósito de prover uma base de confiança para os líderes nacionais em cuidados de saúde, de modo a permitir a geração de um sistema que permitisse gerar valor real para os pacientes e para a sociedade. A finalidade de avançar até um “*Learning Health System*” rapidamente emergiu e foi definido como um sistema em que a ciência, informática, estímulos e cultura são alinhados para melhoria e inovação contínua ⁴.

No cenário mundial, a problemática do aumento de custos é também citada em um relatório elaborado pela empresa Deloitte em 2019, o qual apontou que o gasto mundial em saúde foi estimado para crescer a uma taxa anual de 5,4 % entre 2018-2022, comparada com um aumento de 2,9% entre 2013-2017. ⁵. Esta estimativa foi baseada principalmente no fortalecimento do dólar em relação ao euro e a outras moedas; à expansão da cobertura de assistência médica nos países em desenvolvimento; ao envelhecimento da população; ao surgimento de novos tratamentos e tecnologias em saúde e do aumento dos custos com mão de obra no setor de saúde. ⁶

No campo específico da saúde cirurgia neurológica, têm-se que as opções diagnósticas e terapêuticas disponíveis apresentam um cenário de alta complexidade no início do século XXI. Da mesma forma, também é variado o espectro de resultados possíveis porque existem muitas dimensões de interpretação. Ainda influenciados pelo contexto social, pacientes e médicos estão sobrecarregados com informações da era digital e deste modo, a tomada de decisão hoje é cardinal e crítica. O recurso contemporâneo para essa demanda desafiante é a aplicação das tecnologias de sistemas de gestão da informação, como a inteligência artificial.

Avanços recentes em inteligência artificial (IA) estão criando novas oportunidades para personalizar intervenções de saúde baseadas em tecnologia para pacientes com dor crônica. Ferramentas presentes no campo de IA - ambientes de aprendizagem inteligente, geração de narrativas interativas, modelagem de usuário e treinamento adaptativo - podem ser utilizadas para modelar a aprendizagem e o envolvimento de pacientes com dor crônica e fornecer suporte personalizado em tecnologias de saúde adaptativas. Muitas dessas tecnologias surgiram de aplicações centradas em atividades humanas para educação, treinamento e entretenimento. No entanto, sua aplicação na melhoria da saúde, até o momento, tem sido comparativamente limitada.

Um exemplo de estudo que faz uso de técnicas e ferramentas estatísticas foi o desenvolvido por Depintor et al. (2016), no qual estimou-se a prevalência de algias vertebrais crônicas em indivíduos com 15 ou mais anos de idade e procurou identificar fatores associados. Este trabalho fez uso da Regressão de Cox (ou modelo de Riscos Proporcionais). Para as análises bivariadas, as associações estatísticas foram determinadas através do teste de Log-Rank. Para as variáveis ordinais o teste de chi-quadrado foi utilizado para encontrar tendências e as análises foram performadas utilizando o software STATA 13.0.

A pesquisa desenvolvida entrevistou 826 participantes e o resultado indicou que a prevalência de algias vertebrais crônicas foi estimada em 22 % com intervalo de confiança, ao nível de significância de 5 %, de 19,3 % – 25,0 %. Os fatores associados com as algias vertebrais crônicas foram: sexo feminino, 30 ou mais anos de idade, quatro anos ou menos de escolaridade, sintomas compatíveis com ansiedade e esforço físico intenso durante a ocupação principal.⁷

Algias vertebrais crônicas, fazem parte da categoria de dores classificadas como dores crônicas, as quais afetam aproximadamente 20 % da população mundial. Dor crônica primária é definida por uma dor que persiste por mais de 3 meses e que possui significativo impacto no

bem estar emocional, sendo uma forte causa de angústia, de desmoralização e deficiência funcional nos pacientes, o que a torna uma das principais fontes de sofrimento ⁸.

Atualmente têm-se dois sistemas de classificação de dores bem estabelecidas, denominadas de *STarT Back* e *McKenzie*. O método *McKenzie* faz uso do histórico de sintomas do paciente e das dores apresentadas a partir da realização de determinados movimentos e os classifica em 3 diferentes grupos de acordo com a sua síndrome. Por sua vez, o *STarT Back* classifica os pacientes em alto, médio e baixo risco de desenvolvimento de sintomas persistentes que o desabilitam, tomando como base fatores físicos e psicossociais. Estes dois métodos são exemplos de abordagens que não consideram apenas a base anatômica para a realização do diagnóstico ⁹.

Uma revisão da literatura da aplicação de algoritmos de ML para dores nas costas e na lombar foi realizado por Tagliaferri et al. (2020), no qual são selecionados 48 artigos para a avaliação das abordagens e comparação destes com os métodos já estabelecidos e previamente citados de *STarT Back* e *McKenzie*. Dos 48 artigos selecionados, 45 utilizaram amostras com tamanho menor que 1000, 19 de fizeram uso de menos de 5 parâmetros no modelo final, 13 aplicaram múltiplos modelos e atingiram alta acurácia e 25 avaliaram a dor lombar através de classificação binária (paciente possui ou não possui a dor) ¹⁰.

A partir do apresentado dos estudos e do panorama apresentado, têm-se que simplificar a complexidade não é uma solução e, portanto, a ciência aplicada aos dados, compreendida como conhecimento humano agregado às tecnologias digitais é a melhor alternativa disponível para tomada de decisão na área de cirurgia neurológica. Discutir opções baseadas em dados é uma decisão mais importante que uma incisão.

Ilustramos as oportunidades fornecidas pelas tecnologias adaptativas impulsionadas pela IA para cuidados preventivos de saúde para pacientes com dor crônica, descrevendo uma visão de como as futuras intervenções preventivas de saúde para esse grande grupo de pacientes podem ser realizadas dentro e fora da clínica especializada.

2. Objetivo

Neste estudo, a partir da utilização da análise exploratória de dados, pré-processamento e aplicação de modelos de aprendizagem de máquina sobre a base de dados disponível, buscou-se a seleção dos principais fatores que estão correlacionados com a presença de dor lombar crônica, de dor nas costas e dor nas pernas.

3 Procedimento metodológico

Coleta dos dados

Os dados disponíveis são resultados da avaliação clínica em uma unidade de cuidados integrados multidisciplinares especializados em dor crônica de fevereiro a dezembro de 2019. Na amostra disponível foram incluídos 240 pacientes submetidos a avaliação clínica. Os dados oriundos dos questionários são de natureza sócio-demográfica e clínica, os quais estão enumerados a seguir:

- 1 - Questionário básico de sintomas da coluna vertebral (Protocolo de Gotemburgo),
- 2 - Inventário Breve de Dor (BPI),
- 3 - Índice Oswestry 2.0 de Incapacidade,
- 4 - Questionário de Incapacidade de Roland Morris (RMDQ),
- 5 - Questionário para avaliação da qualidade de vida 12-Item Short Form Health Survey (SF-12)
- 6 - Questionário para Diagnóstico de Dor Neuropática 4 (DN-4).

Pré processamento dos dados

As questões presentes nos questionários foram divididas em duas categorias, quais sejam, questões com respostas de caráter nominal (todas convertidas para o formato binário) e questões com respostas de caráter ordinal. Esta abordagem permite que a observação da correlação entre as variáveis binárias seja realizada através de tabela cruzada (também denominada de tabela de contingência) e que a correlação entre variáveis binárias e variáveis de caráter ordinal seja realizada através do coeficiente de correlação de ponto bisserial.

A base de dados inicial, é resultante da junção dos 6 questionários previamente citados. e composta por 118 variáveis (vetores coluna). Adiciona-se que sobre as variáveis que

apresentam respostas de caráter binário, utilizou-se a técnica de *One Hot Encoder* para gerar novas colunas.

Após a aplicação da técnica de *One Hot Encoder* para variáveis binárias, obteve-se 146 vetores coluna, e dentre estes, 101 são de caráter binário e 45 de caráter ordinal. Es conjunto de dados pode então ser submetido à análises de correlação, construção de gráficos e utilizado nos algoritmos de aprendizado de máquina.

Para a aplicação do modelo de aprendizado de máquina para classificação do paciente quanto à presença ou não de dor lombar, removeu-se as os pacientes que não preencheram esta opção, correspondente a dor de número 30 no questionário Inventário Breve de Dor. Como resultado, têm-se 138 pacientes na base de dados final.

Sabe-se que com esta quantidade de vetores coluna (146), e apenas 138 pacientes, têm-se um caso de dados esparsos, nos quais está presente o fenômeno de *Curse of Dimensionality*, introduzido por Bellman, (1957). A esparsidade dos dados se torna um problema para obter resultados com significância estatística em um modelo de aprendizado de máquina, porque o número de observações deveriam crescer exponencialmente com a dimensão (número de variáveis ou fatores) ¹².

Para reduzir o número de variáveis preditoras, utilizou-se os resultados derivados das análises de correlação entre variáveis binárias e entre variáveis binárias e variáveis de caráter ordinal, de modo a selecionar aquelas que apresentam os maiores valores absolutos.

Adicionalmente, utilizou-se a análise de PCA com 2 componentes para visualização das observações pertencentes ao grupo de pessoas que possuem dor na região lombar e as que assinalaram a opção de não apresentarem esta dor.

Correlações e validação da estatística

A seguir nesta seção são descritos os métodos para correlação entre variáveis binárias e variáveis binárias e variáveis de caráter ordinal, as quais foram utilizadas para seleção de variáveis.

Correlação entre variáveis binárias

Para avaliar a correlação entre as variáveis que se apresentam com caráter binário, fez-se uso da ferramenta de tabela cruzada (também denominada de tabela de contingência), a qual sua estrutura está presente na Tabela 1:

Tabela 1 – Exemplo de tabela cruzada

		Variável B	
		0 (não)	1 (sim)
Variável A	0 (não)	a	b
	1 (sim)	c	d

Na Tabela 1 têm-se um exemplo de tabela cruzada, na qual a, b, c e d são números inteiros. Esta estrutura de tabela permite avaliar, se a mudança de proporção entre não/sim da variável A possui correlação com a variável B.

Para avaliar a relevância estatística dos resultados, utilizou-se o teste de hipótese de chi-quadrado:

H0: Independência – A não depende de B

H1: Dependência – A depende de B

O cálculo do chi-quadrado se dá através da seguinte expressão:

$$\chi^2 = \frac{(ad - bc) * N}{n_1 * n_2 * n_3 * n_4}$$

Em que:

a, b, c e d, são as contagens de observações presentes na Tabela 1

$n_1 : a + b$

$n_2 : c + d$

$n_3 : a + c$

$n_4 : b + d$

$N : n_1 + n_2 + n_3 + n_4$

Pode-se interpretar este teste de hipótese como a diferença na distribuição de frequência da variável A devido a inserção da variável B. Deste modo, quando o valor do chi-quadrado calculado é maior ou igual ao chi-quadrado tabelado, ao adotar nível de significância de 5 %, pode-se rejeitar a hipótese nula da independência e considerar a hipótese alternativa de dependência da variável A em relação à variável B ¹³

Correlação entre questões binárias e questões com resposta de caráter ordinal

A correlação entre questões que se apresentam dicotômicas e questões que apresentam em escala ordinal, foi realizada através do coeficiente de ponto bisserial.

Considera-se a divisão de um grupo em (0 – não) e (1 – sim) e, portanto, pode-se calcular o coeficiente de ponto bisserial de acordo com a seguinte expressão:

$$r_{pb} = \frac{(M_1 - M_0)}{s_n} * \sqrt{\frac{n_1 n_0}{n^2}}$$

Em que:

s_n : Desvio padrão considerando todos os dados da população

M_0 e M_1 : Média da variável que apresenta caráter ordinal para os pacientes que responderam a questão de forma negativa (0 – não) e de forma positiva (1 - sim), respectivamente

n_1 e n_0 : Número de pessoas pertencentes a cada um dos grupos (0 – não) e (1 – sim)

A avaliação da significância estatística deste coeficiente é realizada pelo teste de hipótese do coeficiente de correlação de Pearson, dado que a correlação de ponto bisserial é um caso específico daquela para uma das variáveis sendo dicotômica. Neste teste de hipótese têm-se:

H0: $r_{pb} = 0$

H1: $r_{pb} \neq 0$

O teste utiliza a distribuição t de *student*, de modo que o valor de t calculado se dá por:

$$t = \frac{r_{pb} * \sqrt{n - 2}}{\sqrt{1 - r_{pb}^2}}$$

Em que:

r_{pb} : Coeficiente de correlação de ponto bisserial

n : Número de observações

De acordo com o critério estatístico adotado, se o valor de t calculado for superior ao valor de t tabelado ao adotar nível de significância de 5 %, rejeita-se a hipótese nula que aponta o coeficiente como nulo ¹⁴

Deste modo, neste documento foram apresentados apenas as correlações que passaram por este teste.

Amostragem

A técnica de amostragem utilizada foi a de K dobras estratificadas (*StratifiedKFold*), com 10 subdivisões ($K = 10$), de modo a manter a porcentagem da base original de amostras para cada classe.

Modelo de aprendizado de máquina supervisionado

Foram avaliadas as performances dos seguintes modelos de aprendizado de máquina supervisionado para classificação:

- Regressão Logística com regularização do tipo l1 e l2 (*elasticnet*), com fator de regularização l1 de no valor de 0.8. O algoritmo de otimização utilizado para obtenção dos coeficientes foi o lbfgs (*Limited-memory Broyden–Fletcher–Goldfarb–Shanno*).

- Rede neural com 1 camada oculta composta por 5 neurônios com a função de ativação *ReLU* (unidade linear retificada) e com a função de ativação *sigmoid* na camada de saída. A rede foi treinada através de tamanhos discretos de amostras no valor de 5 observações por 200 épocas.
- *Random Forest*: Fez-se uso de 50 árvores de decisão, com profundidade média de 8 níveis. O critério escolhido para realização do split foi o da impureza de gini.
- *SVM (Support Vector Machine)*: Fez-se uso do kernel de Função de Base Radial, a qual permite, através das ferramentas de produto escalar e da expansão da série de Taylor, obter uma relação entre as observações em uma dimensão infinita. As probabilidades geradas por este modelo foram calibradas através do método do dimensionamento de Platt, o qual aplica a função regressão logística sobre as probabilidades originais geradas, conforme detalhado em ¹⁵.
- *XBGClassifier*: Implementação do método de *Gradient Boosting* com maior velocidade e design. Faz uso dos denominados *weak learners*, árvores de decisão com apenas um nó e duas folhas, para através de método de *ensemble* (junção de vários modelos), obter ao final um modelo otimizado e robusto para, no caso deste estudo, ser utilizado em classificação ¹⁶.

Métricas de avaliação do modelo

Para avaliar a performance dos modelos quanto à classificação correta dos pacientes quanto à presença de lombalgia crônica, utilizou-se no as métricas Acurácia, Precisão (*Precision*) , Revocação (*Recall*) e *F1-Score*, calculadas a partir do treinamento dos modelos realizados com validação cruzada.

Todas estas métricas podem ser calculadas a partir da matriz de confusão, a qual a sua estrutura é apresentada na Tabela 2.

Tabela 2 - Exemplo de matriz de confusão

		Real	
		0 (não)	1 (sim)
Predição	0 (não)	TP	FN
	1 (sim)	FP	TP

A partir do exposto na Tabela 2, define-se, portanto, as seguintes métricas:

$$Acurácia = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

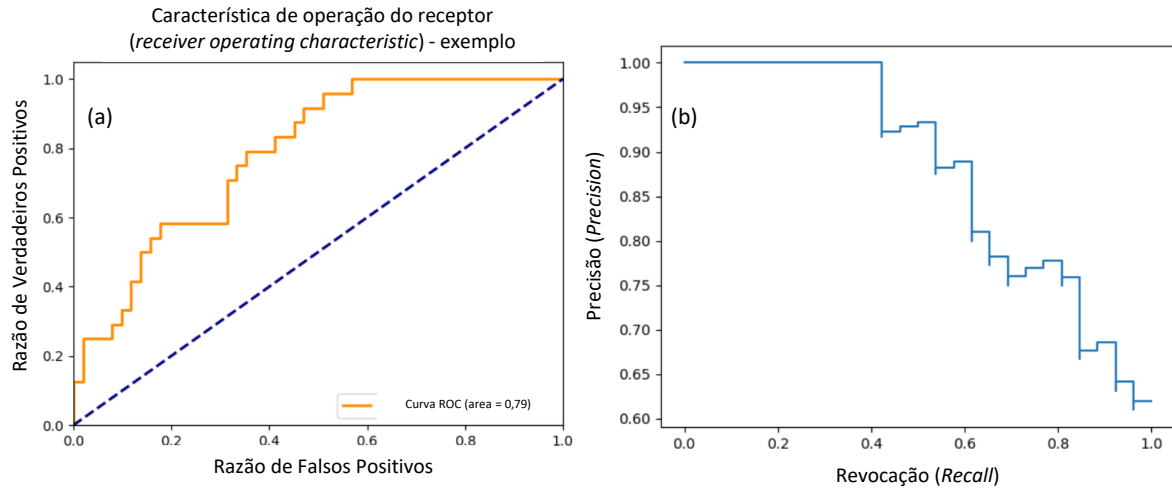
$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

A métrica *Precision* calcula a razão entre o número de positivos corretamente preditos em relação a todos positivos preditos. Entretanto, a métrica *Recall* é particularmente mais importante considerado-se o foco de análise deste trabalho, dado que o cálculo é realizado a partir da razão entre o valor total de Verdadeiros Positivos em relação a soma de Verdadeiros Positivos e Falsos Negativos ¹⁷.

A razão da utilização da mesma se encontra no fato de que um modelo de aprendizado de máquina para predição da ocorrência ou não de lombalgia crônica em pacientes deve apresentar baixa quantidade de Falsos Negativos e consequentemente um alto valor de *Recall*. Ademais, têm-se que a métrica F1-Score permite obter uma média harmônica entre *Precision* e *Recall* ¹⁸.

A avaliação do modelo também foi realizada a partir das curvas Figura 1 (a) área sobre a curva ROC (*Area Under Curve* – AUC) e *Precision-Recal* Figura 1 (b).

Figura 1 – Exemplo de curvas ROC e *Precision-Recall*



Fonte: Adaptado de ¹⁹

A Curva ROC permite observar como ocorre a distribuição entre a Razão de Verdadeiros Positivos (TPR) e a Razão de Falsos Positivos (FPR), para diferentes *thresholds*, o qual é definido entre 0 e 1.

O cálculo de ambos é definido por:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

O mesmo princípio está presente na curva de *Precision x Recall*.

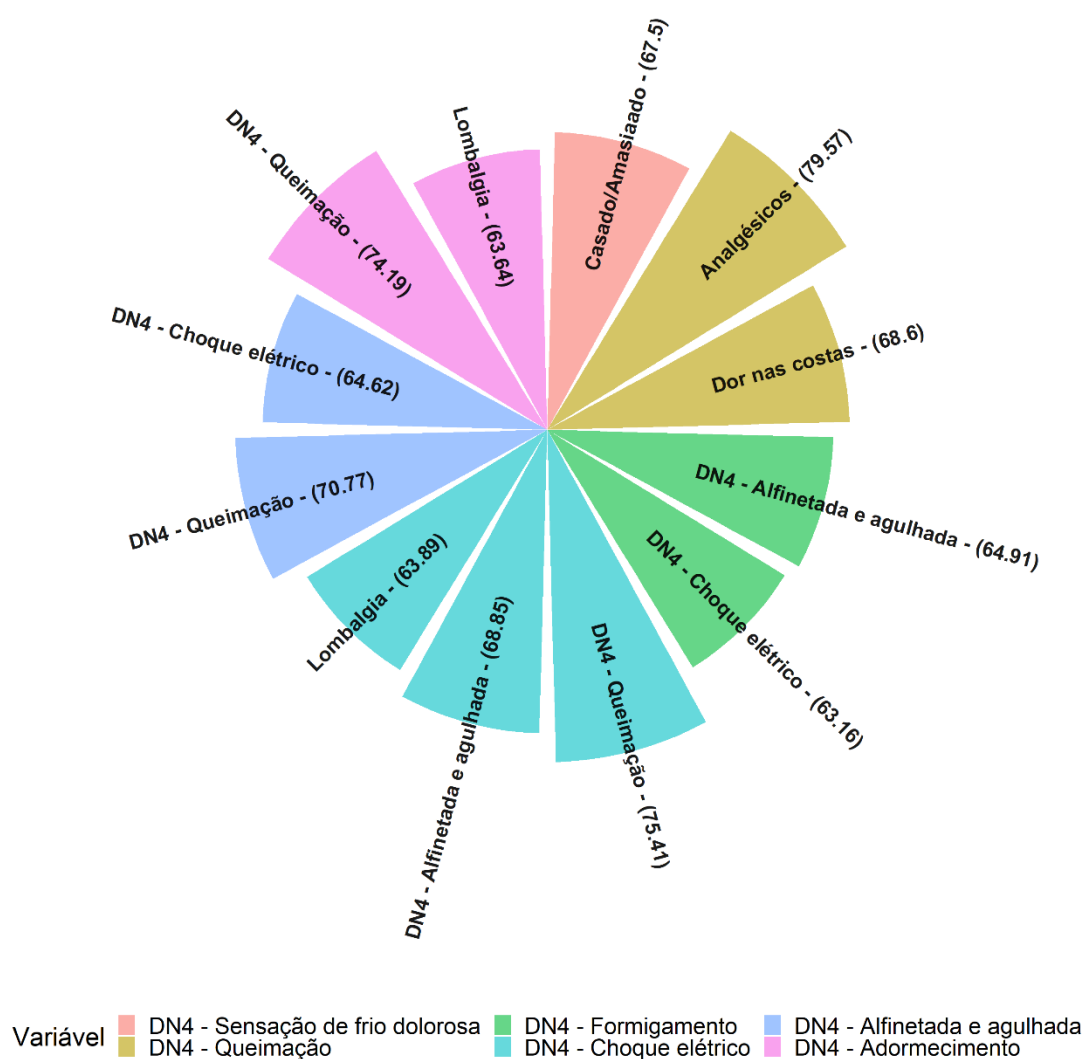
4 Resultados e discussão

Correlações

Principais correlações das variáveis do Questionário para Diagnóstico de Dor Neuropática 4 (DN4)

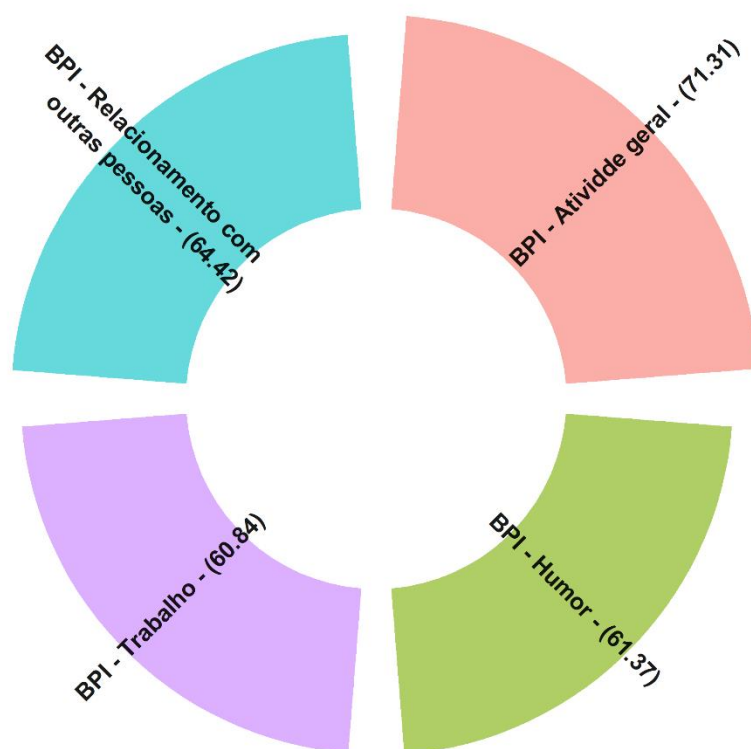
Na Figura 2 são apresentadas as principais correlações com as variáveis presentes no questionário para diagnóstico de dor neuropática 4

Figura 2 – Principais correlações com as variáveis presentes no Questionário para Diagnóstico de Dor Neuropática 4 (DN4)



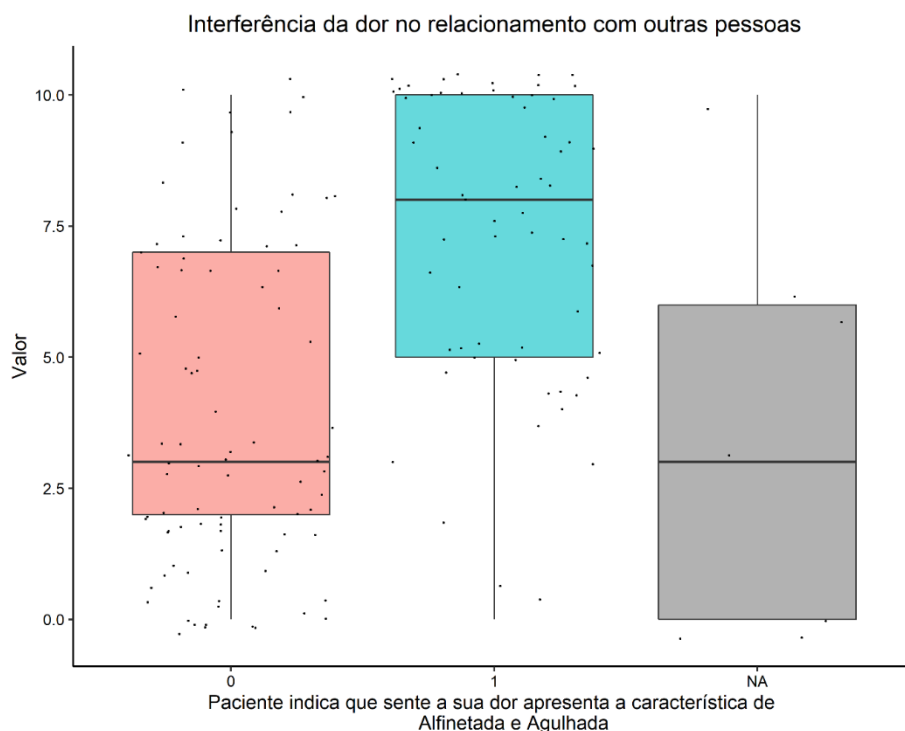
Observa-se da Figura 2 que 79,19 % dos pacientes que assinalaram que apresentam o sintoma de adormecimento na região em que sentem dor, também marcaram que a dor apresenta a característica de queimação. Destaca-se também que 68,6 % das pessoas que disseram apresentar o sintoma de queimação na região em que sentem dor, marcaram a opção de sentirem dor nas costas.

Figura 3 – Principais correlações entre a questão do questionário DN4, sobre apresentar ou não a característica de alfinetada e agulhada na região da dor



A Figura 3 apresenta quais foram as questões de caráter ordinal que apresentam maior correlação com a questão de caráter binário, se o paciente apresenta ou não o sintoma de alfinetada e agulhada na região de sua dor. Através da Figura 4 têm-se um exemplo do porquê a correlação no valor de 64,42 para a questão BPI – Relacionamento com outras pessoas.

Figura 4 – Gráfico de *boxplot* para visualização da distribuição dos dados da questão 2-b do questionário DN4 com a interferência da dor no relacionamento com outras pessoas



Na Figura 4 constata-se que os pacientes que assinalaram apresentarem o sintoma de alfinetada e agulhada na região da dor, também responderam valores maiores para a interferência desta dor no relacionamento com outras pessoas, como é observado na maior densidade de pontos, e no *boxplot* de cor azul, correspondente à opção 1 da questão binária.

Ressalta-se que apenas esta questão (2 – b – Alfinetada e agulhada) do Questionário para diagnóstico de Dor Neuropática apresentou um número superior a 30 pessoas que responderam sim ou não à questão. As demais questões do questionário DN4 não apresentaram significância estatística no coeficiente de correlação ponto bisserial ou apresentaram dados desbalanceados entre sim ou não, ou seja, uma proporção muito maior de pessoas respondeu a opção sim ou não em relação à outra alternativa.

Principais correlações entre as variáveis lombalgia, dor nas pernas e dor nas costas com as questões presentes nos questionários Índice Oswestry 2.0 de Incapacidade e Qualidade de Vida SF-12

Variáveis de caráter ordinal

Na Tabela 3 pode-se visualizar quais foram as principais variáveis de caráter ordinal correlacionadas com a Lombalgia oriundas do questionário Oswestry 2.0 de Incapacidade e Qualidade de vida SF-12.

Tabela 3 – Questões de caráter ordinal dos questionários Oswestry 2.0 de Incapacidade e Qualidade de vida SF-12 mais fortemente correlacionados com a Lombalgia

Ordinal	Coefficiente de correlação	Valor-p
SF12_M1_1_v1	0,43	6,63E-05
Oswestry1_v1	0,38	5,46E-04
Oswestry2_v1	0,37	6,07E-04
SF12_M1_8_v1	0,35	1,25E-03
Oswestry9_v1	0,32	3,33E-03
Oswestry4_v1	0,31	4,13E-03
Somatoria_Oswestry	0,31	4,78E-03
Oswestry10_v1	0,31	5,08E-03
SF12_M1_10_v1	-0,30	5,92E-03
SF12_M1_12_v1	-0,36	1,07E-03
SF12_M1_11_v1	-0,37	7,40E-04
SF12_M1_9_v1	-0,38	5,77E-04
SF12_M1_2_v1	-0,40	1,97E-04
SF12_M1_5_v1	-0,46	1,53E-05
SF12_M1_6_v1	-0,47	1,07E-05
SF12_M1_3_v1	-0,48	9,29E-06
SF12_M1_7_v1	-0,49	4,68E-06
SF12_M1_4_v1	-0,51	1,59E-06

Nota-se que a primeira questão presente no questionário Qualidade de vida SF-12, codificada como SF12_M1_1_ apresentou o maior coeficiente de correlação de ponto bisserial, no valor de 0,43. Esta questão é redigida de forma a possibilitar ao paciente informar em geral, qual a qualidade de sua saúde, em uma escala que vai de: “fraca”, “razoável”, “boa”, “muito boa” e “excelente”, enumeradas como 5,4,3,2 e 1, respectivamente. Deste modo, constata-se que pacientes que informaram possuírem a condição de lombalgia também possuem maior probabilidade de responderem esta questão com a opção de qualidade de vida mais próxima de “razoável” ou “fraca”.

A variável que, em módulo apresentou o maior valor foi a correlacionada à questão codificada como SF12-M1_4_v1 no valor de -0,51, a qual pergunta ao paciente, se, nas últimas 4 semanas, quanto tempo teve no seu trabalho outras atividades diárias regulares o problema de

realizar menos do que queria como consequência do seu estado físico. A resposta é composta por uma escala que vai de “sempre”, “a maior parte do tempo”, “algum tempo”, “pouco tempo” e “nunca”, enumeradas como 1,2,3,4 e 5 respectivamente. Este resultado indica que pacientes com Lombalgia possuem maior tendência a indicarem que “sempre” ou “a maior parte do tempo” realizou menos do que queria no trabalho ou outras atividades diárias regulares, como consequência do seu estado físico.

Ademais, deve-se citar a correlação da condição de lombalgia com a questão SF12_M1_3_v1 (valor de -0,48), a qual indica que a saúde atual do paciente o limita na atividade de subir vários lances de escada e também com a questão SF12_M1_6_v1 (valor de -0,47), respectiva a quanto o paciente teve de problemas emocionais que o levaram a realizar menos do que queria em atividades diárias regulares.

Complementa-se ainda que a questão codificada como Oswestry1_v1 mensura a intensidade da dor do paciente no momento, em uma escala que vai de 1 a 6. Esta variável apresentou coeficiente de correlação no valor de 0,38 o que aponta que pacientes com lombalgia possuem maior tendência de possuírem dores mais intensas.

A Tabela 4 traz as duas questões mais correlacionadas com a questão de dor nas pernas.

Tabela 4 - Questões de caráter ordinal dos questionários RMDQ, Oswestry e SF-12 mais fortemente correlacionados com a dor nas pernas

Ordinal	Coefficiente de correlação	Valor-p
Oswestry1_v1	0,30	8,79E-04
SF12_M1_3_v1	-0,30	8,77E-04

A pergunta codificada como Oswestry1_v1 questiona qual a intensidade da dor que sente no momento. Desta forma, pacientes que possuem dor nas pernas também possuem maior probabilidade de assinalarem positivo para esta questão. Além desta, a questão presente no questionário de Qualidade de vida SF-12, codificada como SF12-M1-3-v1, relacionada a quanto a saúde do paciente o limita em subir vários lances de escada, está correlacionada com a presença de dor nas pernas.

A Tabela 5 apresenta quais questões estão mais correlacionadas com a presença de dor nas costas no geral em pacientes.

Tabela 5 -- Questões de caráter ordinal do questionário Qualidade de vida SF-12 mais fortemente correlacionados com a dor nas costas

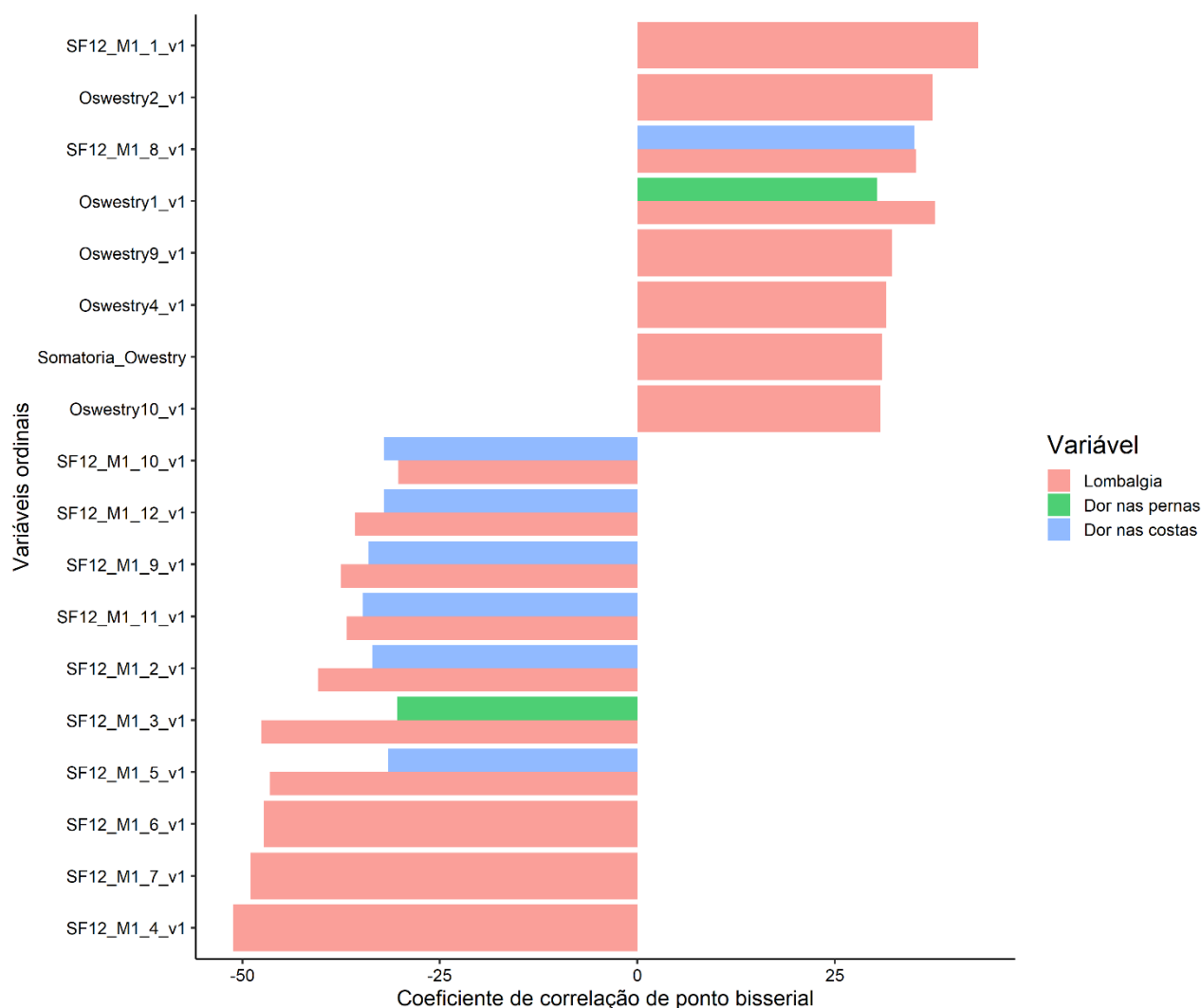
Ordinal	Coefficiente de correlação	Valor-p
SF12_M1_2_v1	-0,34	1,55E-04
SF12_M1_5_v1	-0,32	3,84E-04
SF12_M1_8_v1	0,35	7,57E-05
SF12_M1_9_v1	-0,34	1,25E-04
SF12_M1_10_v1	-0,32	3,03E-04
SF12_M1_11_v1	-0,35	8,81E-05
SF12_M1_12_v1	-0,32	3,05E-04

Constata-se que a pergunta do questionário Qualidade de vida SF-12 codificada como SF12_M1_8_V1, a qual questiona quanto a como a dor interferiu com o trabalho normal do paciente nas últimas 4 semanas, em uma escala de: “absolutamente nada”, “um pouco”, “moderadamente”, “bastante” e “imenso”, enumeradas como 1,2,3,4 e 5 respectivamente, está correlacionada com a presença ou não de dor nas costas nos pacientes. Desta forma aqueles que responderam que possuem dor nas costas possuem maior tendência a indicarem que a dor interferiu “bastante” ou “imensamente” em seu trabalho normal nas últimas 4 semanas.

A pergunta codificada como SF12_M1_11_v1 traz a questão quanto a quanto tempo, durante as últimas 4 semanas o paciente se sentiu triste ou deprimido com as seguintes respostas: “sempre”, “a maior parte do tempo”, “algum tempo”, “pouco tempo” e “nunca”, enumeradas como 1,2,3,4 e 5. Deste modo, dado o coeficiente de correlação negativo -0,35, conclui-se que pacientes que possuem dor nas costas possuem maior probabilidade de responder esta questão com as opções “sempre” e “maior parte do tempo”.

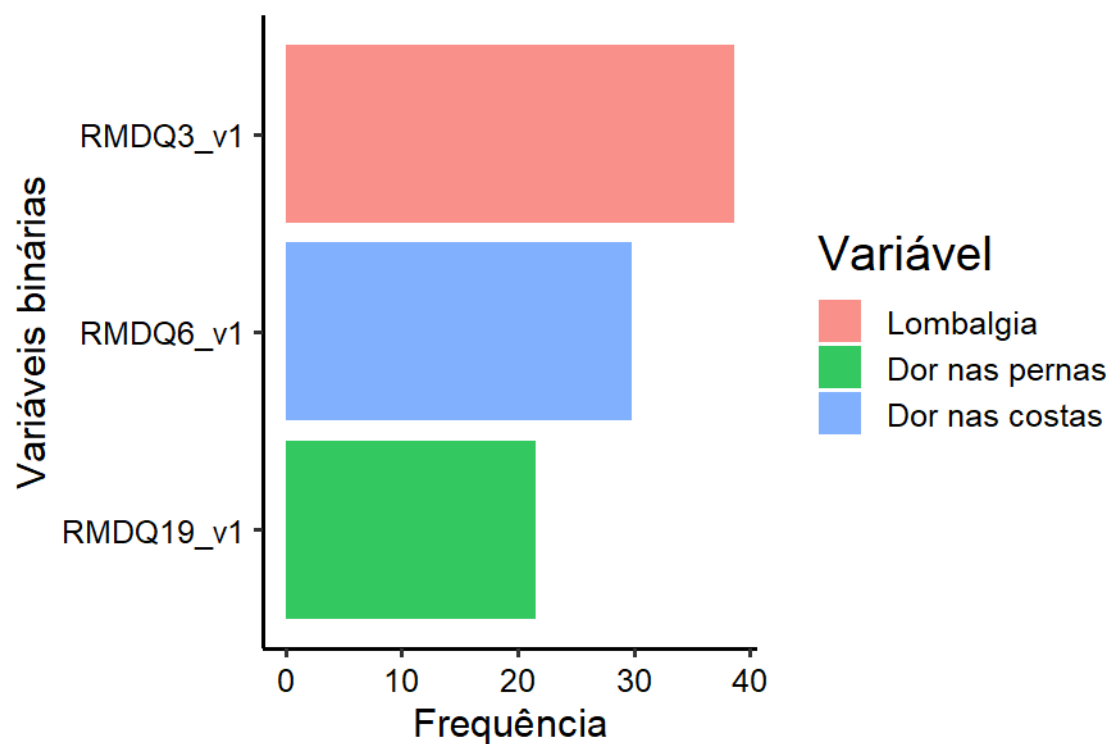
A Figura 5 traz um resumo visual das correlações correlações presentes na Tabela 3, Tabela 4 e Tabela 5 e comentados previamente na forma de texto.

Figura 5 - Questões de caráter ordinal dos questionários, Oswestry e SF-12 mais fortemente correlacionados com a Lombalgia, Dor nas pernas e Dor nas costas



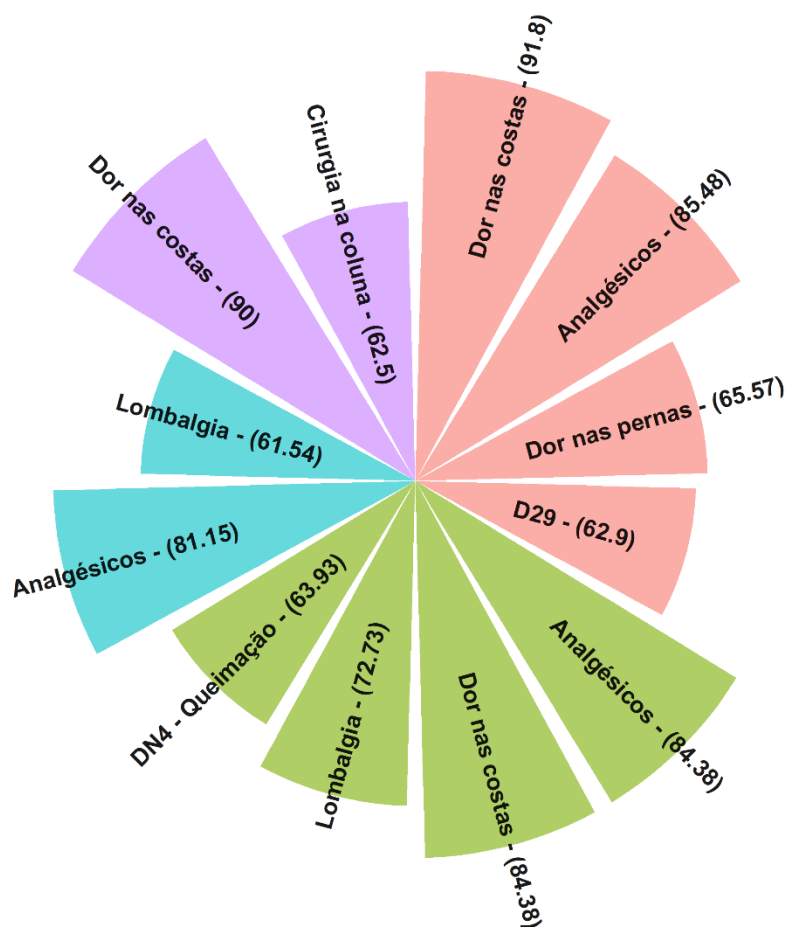
Variáveis de caráter binário

O Problema aqui é que não há variável que apresenta relevância na frequência, porque sempre são frequências abaixo de 50 %, se o valor é menor que 50 %, então não há como apontar que se o paciente indicou possuir lombalgia também maior probabilidade de por exemplo ficar em casa a maior parte do tempo por causa das costas (Questão 1 do RMDQ).



Principais correlações entre as variáveis binárias dor nas costas, lombalgia, dor nas pernas e estar ou não desempregado.

Figura 6 - Principais correlações das variáveis binárias com lombalgia, dor nas costas, dor nas pernas e estar ou não desempregado



Variável ■ Lombalgia ■ Dor nas pernas ■ Dor nas costas ■ Desempregado

Na Figura 6, é possível visualizar quais as questões são mais correlacionadas com as variáveis lombalgia, dor nas pernas, dor nas costas e estar ou não desempregado. Nota-se que 65,57 % das pessoas que marcaram a opção de apresentar dor na região lombar (lombalgia) também marcaram que possuem dor nas pernas.

Importante mencionar também que 63,93 % dos pacientes que apresentam dor nas pernas também marcaram a opção no Questionário DN4 que a sua dor apresenta a característica de queimação (DN – queimação). Nota-se também a alta porcentagem de pacientes que fazem uso de analgésico, sendo 84,38 % para aqueles que possuem dor nas pernas e 90 % para quem possui dor nas costas.

Ressalta-se também que não houve nenhuma correlação maior que 60 e com nível de significância de 5% para a variável que considera se o paciente já realizou alguma cirurgia na coluna

A partir das correlações apresentadas, definiu-se os seguintes critérios para seleção de variáveis preditoras que são utilizadas no modelo de aprendizado de máquina:

- De acordo com a tabela cruzada, aquelas variáveis que apresentam porcentagem maior que o valor de 60 % de correlação.
- Da correlação entre a variável binária Lombalgia (se possui ou não), selecionar apenas as que apresentam correlação superior a 0,5 ou inferior a -0,5.

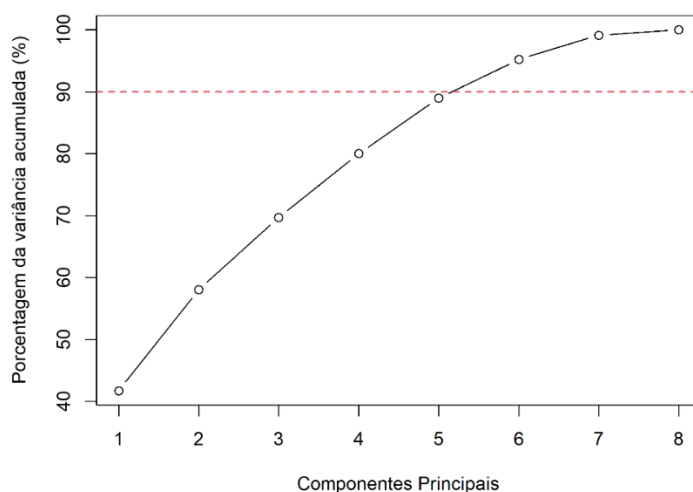
Deste modo, as variáveis selecionadas foram:

- Gotemburgo – Se faz ou não uso de Analgésicos
- Gotemburgo – Se possui dor nas costas
- Gotemburgo – Se possui dor nas pernas
- Gotemburgo – Qual sexo (Masculino ou feminino)
- BPI – Se possui dor na região 29 da questão 2 do questionário de Inventário Breve de Dor
- Gotemburgo – Há quanto tempo está com dor nas pernas
- BPI – Intensidade da dor no momento
- Short Form Health Survey – Realizou menos do que queria nas atividades diárias devido ao problema de dor

Análise de Componentes Principais

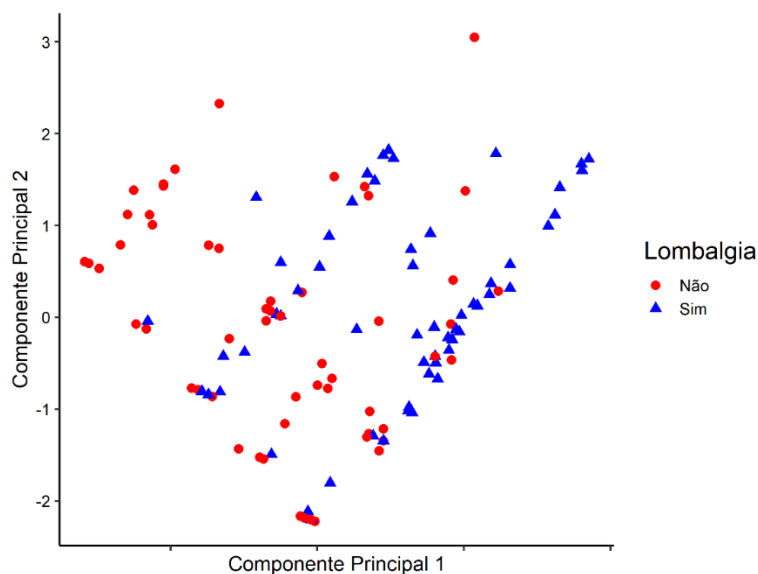
Da análise de componentes principais, pode-se visualizar na Figura 7 o gráfico de Scree Plot, o qual permite inferir qual a porcentagem da variância explicada de acordo com o número de componentes principais utilizados. Desta forma, observa-se que 2 componentes conseguem explicar até aproximadamente 59 % da variabilidade dos dados e com 5 componentes, explica-se até 94 %.

Figura 7 - *Scree plot* oriundo da análise de componentes principais



Dado que 2 componentes principais são suficientes para explicar até aproximadamente 59 % da variabilidade dos dados, pode-se visualizar na Figura 8 como os dados ficam distribuídos em duas dimensões.

Figura 8 -Visualização em dois eixos da distribuição dos casos de Lombalgia com a utilização dos 2 primeiros componentes principais



Nota-se da Figura 8 que com apenas 2 dimensões há grande presença de sobreposição entre os casos positivos e negativos.

Modelo de aprendizado de máquina

Da aplicação dos 5 modelos de aprendizado de máquina previamente apresentados na seção de Procedimento metodológico, têm-se na Figura 9 o resultado da curva ROC obtida e na Figura 10 da curva Precision-Recall, considerando então a variação do *threshold* para predição entre 0 e 1..

Figura 9 – Curva ROC para os 5 diferentes modelos de aprendizado de máquina

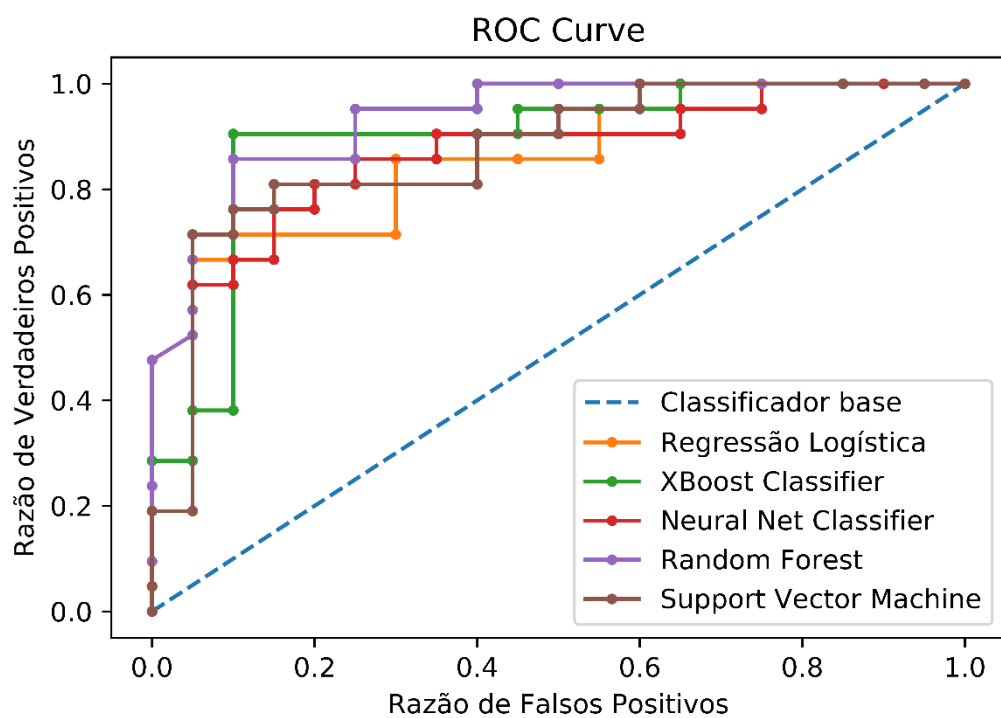
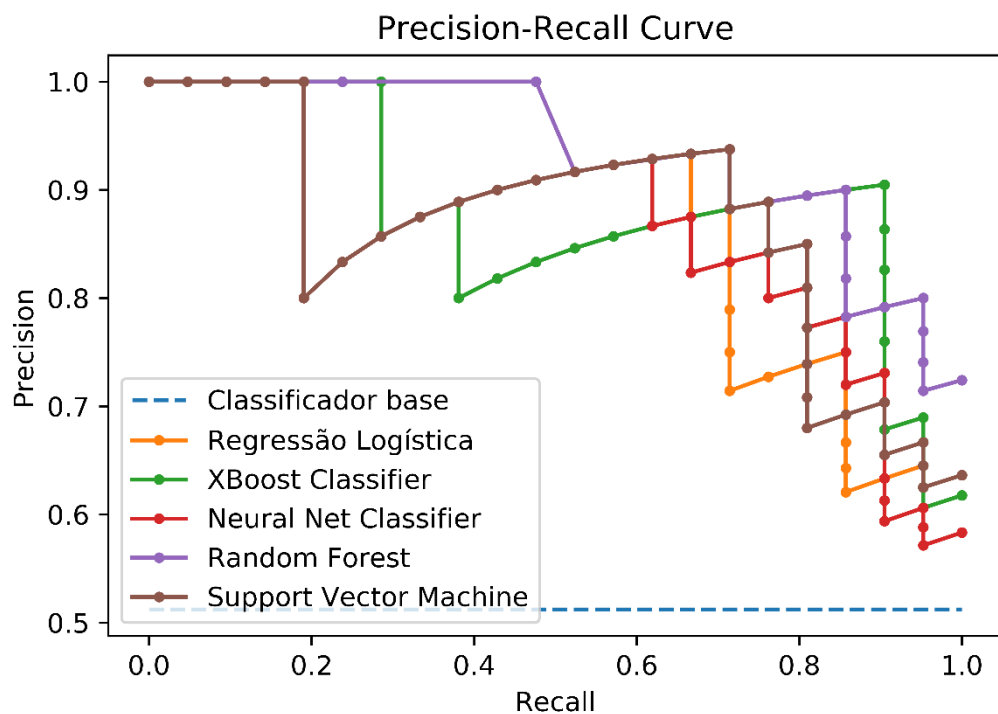


Figura 10 - Curva Precision-Recall para os 5 diferentes modelos de aprendizado de máquina



Para o valor de *threshold* usualmente utilizado, padrão, de 0,5, têm-se na Tabela 6 as métricas finais obtidas com a aplicação de 5 diferentes modelos de aprendizado de máquina quanto à performance na classificação dos pacientes quanto à presença de Lombalgia.

Tabela 6 - Resultados das métricas de performance dos modelos de aprendizado de máquina testados

	Acurácia	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Regressão Logística	0.780256	0.815357	0.72619	0.732547
Rede Neural	0.707317	0.827153	0.809699	0.818333
Random Forest	0.778755	0.767641	0.740476	0.735991
Support Vector Machine	0.785348	0.875595	0.690476	0.721166
XBoost Classifier	0.808635	0.808065	0.804701	0.789293

Nota-se portanto que com a utilização do *threshold* padrão de 0.5, têm-se que o algoritmo que o modelo de rede neural apresentou o maior valor de *Recall* e *F1-Score* e o modelo construído a partir do algoritmo de *XBoost* apresentou o maior valor de acurácia, assim como o segundo maior valor de *Recall* e *F1-Score*. Deve-se mencionar que, a performance destes modelos podem ser melhoradas através da coleta de bases de dados maiores e com uma menor proporção de valores faltantes.

Ressalta-se que a tarefa realizada e exposta, da descoberta de quais são as principais variáveis que possuem os maiores valores de correlação com a dor na lombar, permite a redução do número de questões dos questionários aplicados. Desta forma, tem-se que um dos principais benefícios da adoção desta medida é a maior probabilidade de os pacientes responderem a questionários com um menor número de questões em relação aos atuais, contribuindo assim positivamente para uma geração de base de dados com melhor qualidade e integridade dos dados e auxiliando no processo de triagem dos pacientes.

5 Conclusões

6 Agradecimentos

7 Referências

1. Porter ME, Teisberg EO. *Redefining Health Care - Creating Value-Bases Competition on Results.*; 2006.
2. Curfman GD, Morrissey S, Drazen JM. High-Value Health Care — A Sustainable Proposition. *N Engl J Med.* 2013;369(12):1163-1164. doi:10.1056/nejme1310884
3. Smith M, Saunders R, Stuckhardt L, McGinnis JM. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America.* Vol 51.; 2014. doi:10.5860/choice.51-3277
4. National Academy of Medicine. Artificial Intelligence in Health Care. Published online 2018.
5. Stephanie Allen P, Hammett DR, Vettori E de, et al. 2019 Global health care outlook Shaping the future. *Des Issues.* Published online 2019:41. <https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-hc-outlook-2019.pdf><http://www.ncbi.nlm.nih.gov/books/NBK2665/><http://dx.doi.org/10.1016/j.bios.2016.09.038><https://pdfs.semanticscholar.org/956b/6ee61>
6. Medici AC, Monitor UH, Market L. Por André C . Medici. 2019;(December). doi:10.13140/RG.2.2.27521.40807
7. Depintor JDP, Bracher ESB, Cabral DMC, Eluf-Neto J. Prevalence of chronic spinal pain and identification of associated factors in a sample of the population of São Paulo, Brazil: cross-sectional study. *Sao Paulo Med J.* 2016;134(5):375-384. doi:10.1590/1516-3180.2016.0091310516
8. Kennedy R, Abd-Elsayed A. The International Association for the Study of Pain (IASP) Classification of Chronic Pain Syndromes. *Pain.* Published online 2019:1101-1103. doi:10.1007/978-3-319-99124-5_234
9. Werneke MW, Edmond S, Young M, Grigsby D, McClenahan B, McGill T. Association between changes in function among patients with lumbar impairments classified according to the STarT Back Screening Tool and managed by McKenzie credentialed physiotherapists. *Physiother Theory Pract.* 2020;36(5):589-597. doi:10.1080/09593985.2018.1490839
10. Tagliaferri SD, Angelova M, Zhao X, et al. Artificial intelligence to improve back pain outcomes and lessons learnt from clinical classification approaches: three systematic

- reviews. *npj Digit Med*. 2020;3(1). doi:10.1038/s41746-020-0303-x
11. Bellman RE. *Dynamic Programming*. Princeton University Press; 1957.
 12. Venkat N. The Curse of Dimensionality. Published online 2010:169-181. doi:10.1201/ebk0824740993-10
 13. NCSS. Contingency Tables Square Test. In: *NCSS - Statistical Software*. ; 2019:1-39. https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Contingency_Tables-Crosstabs-Chi-Square_Test.pdf
 14. Bedrick EJ. Biserial Correlation. *Encycl Biostat*. Published online 2005. doi:10.1002/0470011815.b2a10007
 15. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. *ICML 2005 - Proc 22nd Int Conf Mach Learn*. 2005;(1999):625-632. doi:10.1145/1102351.1102430
 16. Browlee J. A Gentle Introduction to XGBoost for Applied Machine Learning. Published 2016. Accessed September 28, 2020. <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>
 17. Powers DMW. Evaluation: From Precision , Recall and F-Factor to ROC , Informedness , Markedness & Correlation. 2007;(December). Commonly used evaluation measures including Recall, Precision, F-Factor and Rand Accuracy are biased and should not be used without clear understanding of the biases, and corresponding identification of chance or base case levels of the statistic. Using t
 18. Sasaki Y. The truth of the F-measure The truth of the F-measure. 2015;(January 2007):1-6.
 19. Scikit-Learn Developers. Metrics and scoring: quantifying the quality of predictions. Published 2020. Accessed October 28, 2020. https://scikit-learn.org/stable/modules/model_evaluation.html