AVALIAÇÃO E INTERPRETAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA NA CLASSIFICAÇÃO BINÁRIA E MULTICLASSE DE PATOLOGIAS DA COLUNA VERTEBRAL

Resumo: Este estudo avaliou oito algoritmos de aprendizado de máquina na classificação binária e multiclasse de patologias da coluna vertebral utilizando o conjunto de dados Vertebral Column. O modelo MLP apresentou o melhor desempenho, com AUC de 0,947. As análises interpretáveis indicaram que o atributo *degree_spondylolisthesis* foi o mais relevante. Os resultados demonstram o potencial do aprendizado de máquina no apoio ao diagnóstico assistido de alterações vertebrais.

Palavras-chave: Inteligência Artificial Explicável; Orange Data Mining; Patologias de Coluna Vertebral;

INTRODUÇÃO

A coluna vertebral é a principal estrutura de sustentação do corpo humano, formada por vértebras, discos intervertebrais, nervos, músculos e medula espinhal (PESSIN, 2022). No total, são 26 ossos vertebrais no adulto e, junto ao crânio, costelas e esterno, compõem o esqueleto axial (RESHI et al., 2021). Além de permitir movimentos, sua função é proteger a medula espinhal, sendo fundamental para a qualidade de vida. No entanto, essa estrutura pode sofrer alterações que causam dor, limitações físicas e prejuízos ao bem-estar (SILVA et al., 2022; RIBEIRO; NETO; ROCHA, 2024).

Os Transtornos Discais da Coluna (TDC) afetam os discos intervertebrais e podem ser provocados por fatores como traumas, má postura, doenças metabólicas, infecções ou condições congênitas (PESSIN, 2022; SILVA et al., 2023). Essas alterações incluem degeneração, hérnia e compressões, gerando sintomas como dor, rigidez muscular, limitação de movimentos e, em casos mais graves, déficits neurológicos. As regiões mais atingidas são a torácico-lombar e a lombossacral, sendo essas doenças causas frequentes de atendimento médico e afastamento no Brasil (SILVA et al., 2023, LI et al, 2023).

No Brasil, doenças osteomusculares figuram entre as principais causas de incapacidade laboral e concessão de benefícios previdenciários. Entre essas condições, destacam-se as patologias da coluna vertebral, como dorsalgias e transtornos dos discos intervertebrais, frequentemente associadas ao afastamento temporário ou permanente das atividades laborais (SOUZA; SANTANA, 2012; BRASIL, 2026). Entre as principais Patologias da Coluna Vertebral (PVC) analisadas neste trabalho estão a hérnia de disco e a espondilolistese.

A hérnia de disco caracteriza-se pela ruptura do anel fibroso, ocasionando o deslocamento do núcleo pulposo e a consequente compressão de estruturas adjacentes. Essa condição ocorre com maior frequência entre os 40 a 50 anos de vida, acometendo aproximadamente 2% a 3% da população, com maior incidência no sexo masculino. Entre as principais manifestações clínicas encontram-se dor intensa, limitação funcional e alterações no alinhamento corporal. Seus principais fatores de risco incluem hábitos ergonômicos inadequados, desalinhamentos biomecânicos, desequilíbrios musculares e predisposição genética (CASEMIRO; VIEIRA, 2021; PESSIN, 2022).

Já a espondilolistese consiste no deslizamento de uma vértebra em relação à adjacente, ocasionando desalinhamento da coluna, predominantemente na região lombar. Pode provocar dor, parestesia e déficits sensoriais decorrentes da compressão nervosa. Com maior prevalência em indivíduos jovens, a doença é multifatorial, por exemplo, sendo causada por sedentarismo, obesidade, tabagismo, alterações anatômicas e predisposição genética (SILVA et al., 2022; LAURINDO et al., 2024; PESSIN, 2022).

O diagnóstico dessas condições envolve exames clínicos e de imagem, como raio-X, ressonância magnética e tomografia (RODRIGUES et al., 2024). Apesar disso, esses métodos podem apresentar limitações, como subjetividade na análise e alta demanda sobre os profissionais, dificultando diagnósticos rápidos e precisos (RESHI et al., 2021; RIBEIRO; NETO; ROCHA, 2024). Diante desse cenário, a Inteligência Artificial (IA) fornece diversos algoritmos e modelos com potencial capacidade de reconhecimento de padrões de PVC para servir como ferramenta de apoio na área da saúde.

A IA constitui uma subárea da Ciência da Computação dedicada ao desenvolvimento de sistemas capazes de simular comportamentos inteligentes na resolução de problemas complexos (COPPIN, 2015). Dentro desse campo, o Aprendizado de Máquina (AM) refere-se à capacidade dos sistemas de aprimorar seu desempenho a partir da experiência, mediante a análise de dados, sendo aplicável a problemas de

classificação e regressão (FACELI et al., 2025). Em tarefas preditivas, esses sistemas podem classificar ou estimar resultados com base em dados previamente rotulados, contribuindo para diagnósticos mais eficientes e para o acompanhamento e tratamento dos pacientes (LECUN; HUANG; BOTTOU, 2004; LOGITHASAN et al., 2022; SHAILAJA; SEETHARAMULU; JABBAR, 2018).

Dentre os algoritmos clássicos de Aprendizado de Máquina (AM) aplicados a tarefas de classificação, este estudo avalia AdaBoost, *Decision Tree* (DT), *Gradient Boosting* (GB), *k-Nearest Neighbors* (*k*-NN), *Naive Bayes* (NB), *Multi-Layer Perceptron* (MLP), *Random Forest* (RF) e *Support Vector Machine* (SVM).

Algoritmos do tipo *boosting*, como AdaBoost e GB, combinam múltiplos estimadores considerados "fracos" em uma sequência iterativa, buscando aprimorar progressivamente o desempenho preditivo do modelo. O *k*-NN realiza a classificação com base na proximidade entre instâncias semelhantes, enquanto o NB utiliza princípios probabilísticos fundamentados na hipótese de independência entre atributos. O MLP consiste em uma rede neural artificial composta por múltiplas camadas capazes de aprender padrões complexos nos dados. O DT organiza o processo de decisão em uma estrutura hierárquica ramificada, por meio de sucessivas divisões dos dados, enquanto o RF combina diversas árvores de decisão para aumentar a robustez e reduzir o risco de sobreajuste. Por fim, o SVM busca identificar o hiperplano ótimo capaz de maximizar a separação entre diferentes classes no espaço de características (PALECHOR; MANOTAS, 2019; IZBICKI, DOS SANTOS, 2020).

Embora os algoritmos de AM apresentam resultados promissores em tarefas de predição e apoio ao diagnóstico médico, muitos desses modelos possuem natureza opaca, dificultando a compreensão de seus mecanismos internos de decisão. Essa limitação pode comprometer a confiança e a interpretabilidade dos sistemas inteligentes, especialmente em aplicações clínicas que demandam maior transparência e justificativas claras para as decisões automatizadas (ADADI; BERRADA, 2018; MISHRA, 2022).

Dentre as técnicas de explicação da XAI, o *Shapley Additive Explanations* (SHAP) é uma técnica amplamente utilizada para interpretar modelos de AM e AP (VIMBI; SHAFFI; MAHMUD, 2024). SHAP é uma das ferramentas mais populares para explicar previsões, sendo aplicada em cerca de 70% dos estudos sobre diagnóstico assistido. Baseado na Teoria dos Jogos, SHAP atribui um peso, chamado valor de Shapley, a cada característica de um modelo, determinando sua influência na predição (MOLNAR, 2020). Além de ser agnóstico ao modelo, ele permite interpretar de forma clara quais fatores mais impactam os resultados, auxiliando na compreensão de modelos de "caixa-fechada". Essas qualidades tornam SHAP essencial para aprimorar a transparência e a confiabilidade dos sistemas de IA (SHAP, 2018; VIMBI; SHAFFI; MAHMUD, 2024)

Diante disso, este estudo tem como objetivo avaliar o desempenho de oito algoritmos clássicos de AM na classificação de patologias da coluna vertebral. Adicionalmente, buscou-se identificar as variáveis mais relevantes para as predições realizadas pelo modelo com melhor desempenho, utilizando o método SHAP para análise interpretativa. Outra contribuição deste trabalho consiste na utilização do *software* Orange Data Mining, que possibilita análise exploratória de dados, visualizações personalizadas e treinamento de modelos de AM por meio de programação visual. Essa abordagem facilita a investigação e aplicação de técnicas de IA por pesquisadores de diferentes áreas, especialmente em contextos relacionados à saúde (NAIK; SAMANT, 2016; POPCHEV; OROZOVA, 2023).

MATERIAL E MÉTODOS

Conjunto de dados

Este estudo utilizou o conjunto de dados "*Vertebral column*" , disponível no repositório UCI Machine Learning Repository1. O conjunto de dados contém 310 instâncias e 7 variáveis (atributos), abrangendo informações biomecânicas como incidência, inclinação e raio da pelve, inclinação sacral, ângulo de lordose lombar e grau de espondilolistese (BARRETO; ROCHA NETO, 2005).

Um médico durante sua residência médica em cirurgia da coluna vertebral no Centro Médico-Cirúrgico de Reabilitação de Massas, Lyon na França, por meio do Grupo de Pesquisa Avançada em Ortopedia (*Group of Applyed Research in Orthopaedics*, GARO), compilou o conjunto de dados "Vertebral column" com informações de pacientes saudáveis e pacientes não saudáveis com espondilolistese ou hérnia de disco (BARRETO; ROCHA NETO, 2005; ROCHA NETO et al., 2008).
O conjunto de dados "*Vertebral column*" disponibiliza duas versões relacionadas às tarefas de classificação diferentes, porém relacionadas. A primeira tarefa consistiu em classificar os pacientes em uma das três categorias: Normal (100 pacientes), Hérnia de Disco (60 pacientes) ou Espondilolistese (150 pacientes).

[1] Disponível em.<https://archive.ics.uci.edu/dataset/212/vertebral+column>

Para a segunda tarefa, as categorias Hérnia de Disco e Espondilolistese foram agrupadas em uma única categoria denominada "*abnormal*" (anormal em português). Assim, a segunda tarefa consiste em classificar os pacientes em uma das duas categorias: *normal* (100 pacientes) ou *abnormal* (210 pacientes) (RIBEIRO; NETO; ROCHA, 2024).

O conjunto de dados foi organizado em seis atributos biomecânicos. Ângulo de incidência pélvica, importante na avaliação do equilíbrio sagital da coluna vertebral, Versão pélvica, um parâmetro radiológico fundamental para avaliar o equilíbrio sagital da coluna vertebral e a orientação da pelve, Ângulo de lordose lombar, utilizado para avaliar a curvatura da região lombossacral da coluna vertebral, Inclinação sacral, determina a posição da coluna lombar, pois o plano sacral forma a base da coluna vertebral, e o grau de espondilolistese, que corresponde à quantificação percentual do deslocamento anterior ou posterior de uma vértebra em relação ao segmento adjacente (BARRETO; ROCHA NETO, 2005; RIBEIRO; NETO; ROCHA, 2024).

A incidência pélvica (PI, *pelvic incidence*) corresponde a um parâmetro anatômico específico e relativamente constante para cada indivíduo, sendo mensurada em radiografias lombossacras na incidência lateral. Esse parâmetro é definido pelo ângulo formado entre a linha que conecta o ponto médio da placa terminal superior de S1 ao centro de rotação femoral e a linha perpendicular à placa terminal superior de S1. Estudos demonstram que o PI apresenta discreto aumento ao longo da vida adulta, além de valores mais elevados em indivíduos com espondilolistese, havendo relação linear entre o aumento do PI e a gravidade do escorregamento vertebral (TEBET, 2014).

Na Figura 1(A), a incidência pélvica é representada pelo ângulo formado entre a linha traçada do centro da cabeça femoral até o ponto médio da placa terminal sacral e a linha perpendicular ao centro dessa mesma placa. A placa terminal sacral é delimitada pelo segmento compreendido entre o corno posterior do sacro e a porção anterior superior de S1, localizada no promontório sacral. Já na Figura 2(B), nos casos em que as cabeças femorais não se encontram perfeitamente sobrepostas na radiografia, o centro de cada cabeça femoral é identificado individualmente e conectado por uma linha auxiliar. O ponto médio dessa linha é então utilizado como referência para o traçado da linha em direção ao centro da placa terminal sacral, permitindo a correta determinação da incidência pélvica (TEBET, 2014).

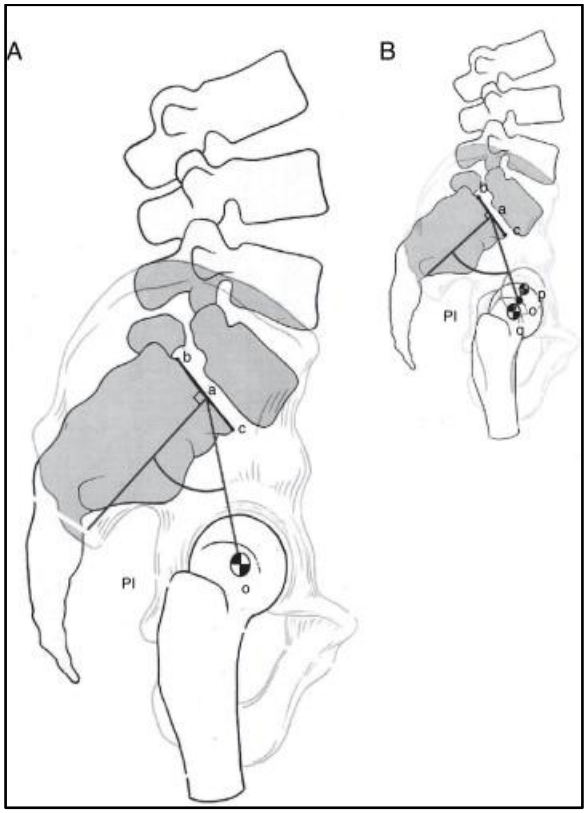


Figura 1. Representação esquemática da mensuração da PI em radiografia lombossacra lateral. (A) O PI é definido pelo ângulo formado entre a linha traçada do centro da cabeça femoral ao ponto médio da placa terminal sacral e a linha perpendicular à placa terminal superior de S1. (B) Nos casos em que as cabeças femorais não estão sobrepostas, utiliza-se o ponto médio entre os centros femorais como referência para o traçado da linha em direção à placa terminal sacral. Fonte: (TEBET, 2014).

Pré- Processamento de dados

No que se refere ao pré-processamento dos dados, não foi necessária a aplicação de técnicas adicionais nos atributos *pelvic_incidence*, *pelvic_tilt, lumbar_lordosis_angle, sacral_slope, pelvic_radius e degree_spondylolisthesis*, uma vez que o conjunto de dados é composto exclusivamente por atributos numéricos tabulares. Também foram realizados testes utilizando normalização dos dados nos intervalos [0, 1] e [-1, 1]. Entretanto, os resultados obtidos demonstraram redução no desempenho dos algoritmos avaliados quando comparados à utilização dos valores originais dos atributos, motivo pelo qual optou-se por manter os dados em sua escala original.

Além disso, o atributo alvo *class* foi convertido para o formato ordinal, permitindo sua utilização adequada nos modelos de classificação. Inicialmente, foi realizado o treinamento utilizando a versão binária do conjunto de dados, composta pelas classes Normal e

Abnormal. Em seguida, foi conduzido o treinamento da versão multiclasse, contendo as classes *Normal*, *Disk Hernia* (Hérnia de Disco) e *Spondylolisthesis* (Espondilolistese). As Figuras 2 e 3 apresentam a distribuição e o balanceamento do conjunto de dados "Vertebral Column" em suas versões binária (PVC-2C) e multiclasse (PVC-3C), respectivamente.

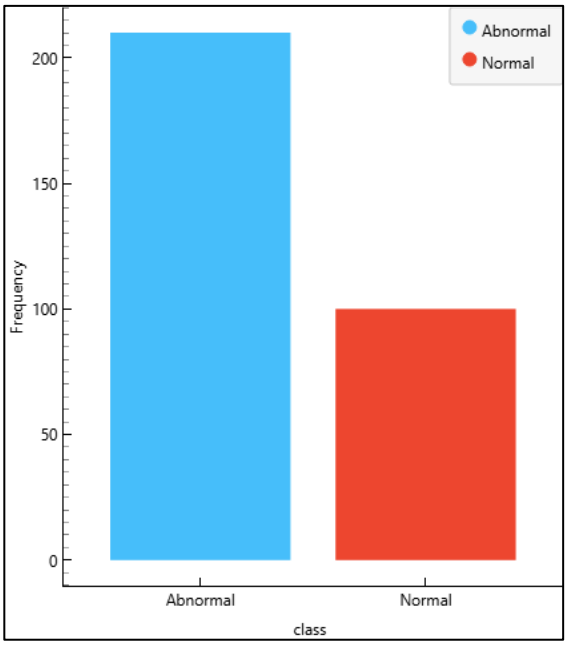


Figura 2. Distribuição original de amostras do conjunto de dados versão binária (PVC-2C).

Com o objetivo de minimizar o desbalanceamento entre as classes presentes nos conjuntos de dados, foi aplicada a *técnica Synthetic Minority Over-sampling Technique* (SMOTE) proposta por Chawla et al. (2002). Essa abordagem permitiu gerar artificialmente novas amostras das classes minoritárias, promovendo uma distribuição mais equilibrada entre as categorias e contribuindo para a melhoria do desempenho dos modelos preditivos.

No conjunto PVC-2C, originalmente composto pelas classes *Normal* (100 amostras) e *Abnormal* (210 amostras), o balanceamento resultou em um total de 410 amostras, sendo 205 pertencentes a cada classe. Já no conjunto PVC-3C, formado pelas classes Normal, Hérnia de Disco e Espondilolistese, o processo de balanceamento gerou um total de 438 amostras, distribuídas em 145 amostras da classe Normal, 147 da classe Hérnia de Disco e 146 da classe Espondilolistese.

As Figuras 4 e 5 apresentam a distribuição balanceada das amostras por classe nos conjuntos PVC-2C e PVC-3C, respectivamente.

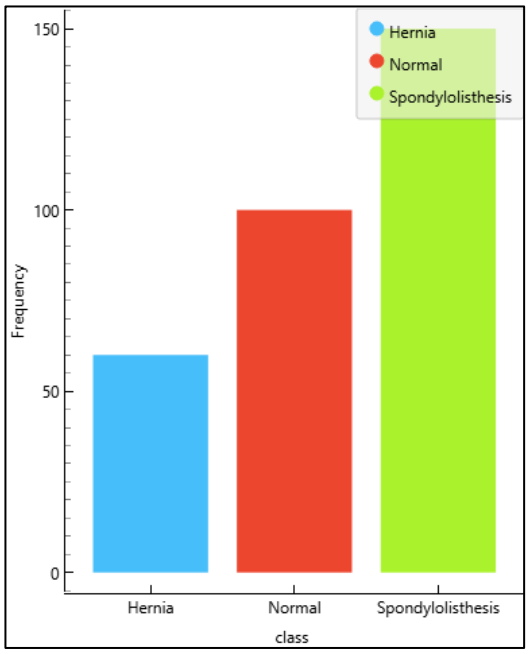


Figura 3. Distribuição original de amostras do conjunto de dados versão multiclasse (PVC-3C).

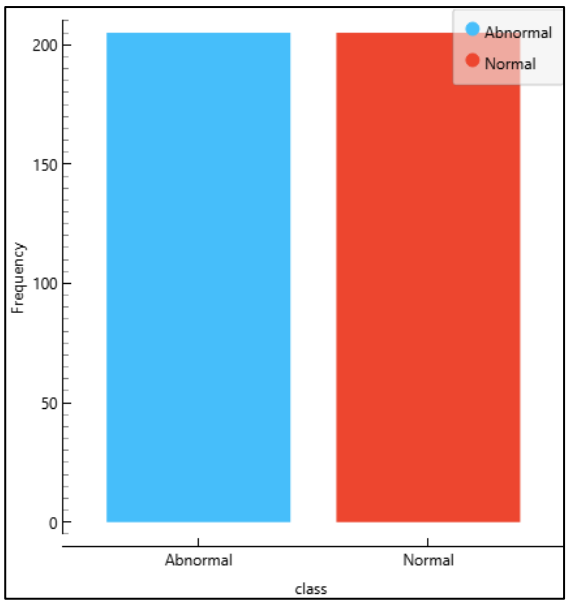


Figura 4. Distribuição balanceada de amostras do conjunto de dados versão binária (PVC-2C).

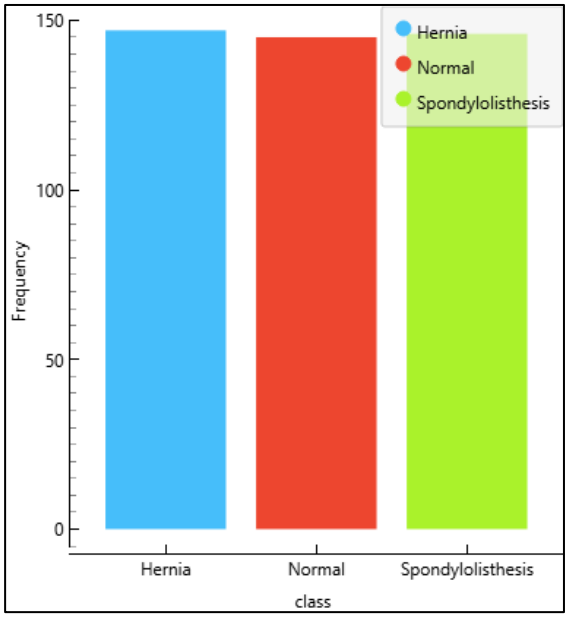


Figura 5. Distribuição balanceada de amostras do conjunto de dados versão multiclasse (PVC-3C).

Experimentos computacionais

Os experimentos computacionais foram realizados com o objetivo de avaliar o desempenho de modelos de AM na classificação dos níveis de obesidade. Para assegurar maior confiabilidade e robustez nos resultados, empregou-se a técnica de validação cruzada *k-fold* com (k = 10), associada ao método *Holdout*, no qual o conjunto de dados foi particionado em 80% para treinamento e 20% para teste. Essa estratégia contribui para aumentar a estabilidade das métricas de desempenho e minimizar a influência de particionamentos específicos dos dados sobre os resultados obtidos (PANCH; SZOLOVITS; ATUN, 2018).

Foram avaliados oito algoritmos de Aprendizado de Máquina: AdaBoost, DT, GB, k-NN, NB, MLP, RF e SVM. O MLP foi configurado com cinco neurônios na camada oculta, função de ativação ReLU, solver Adam, parâmetro de regularização alpha igual a 0,02 e limite máximo de 1000 iterações, visando equilibrar capacidade de generalização e eficiência computacional. O AdaBoost foi parametrizado com 50 estimadores baseados em árvores de decisão, taxa de aprendizado de 1,0 e método SAMME.R para classificação. O SVM utilizou kernel RBF, parâmetro C igual a 1 e limite máximo de 100 iterações. O algoritmo *k*-NN foi executado com cinco vizinhos, utilizando distância euclidiana e pesos uniformes. A RF foi configurada com 10 árvores de decisão, enquanto o GB utilizou 100 árvores e taxa de aprendizado de 0,1.

O desempenho dos modelos foi analisado por meio das métricas *Area Under the Curve* (AUC), Acurácia, Precisão, Revocação (do inglês *Recall*) e Medida F1 (do inglês *F1-Score*). A métrica AUC avalia a capacidade discriminativa do modelo entre as classes, sendo que valores mais próximos de 1 indicam melhor desempenho. A acurácia representa a proporção de classificações corretas em relação ao total de amostras avaliadas. A precisão mede a proporção de instâncias classificadas como positivas que realmente pertencem à classe positiva, sendo particularmente relevante em cenários nos quais falsos positivos podem causar impactos significativos. O *recall*, também denominado sensibilidade, avalia a capacidade do modelo em identificar corretamente os casos positivos, aspecto importante em aplicações nas quais a não detecção de indivíduos com obesidade pode acarretar consequências clínicas relevantes. Por fim, o *F1-Score* corresponde à média harmônica entre precisão e *recall*, sendo uma métrica apropriada para contextos com possível desbalanceamento entre classes, pois fornece uma avaliação equilibrada do desempenho dos modelos (IZBICKI, DOS SANTOS, 2020; FACELI et al., 2025).

RESULTADOS E DISCUSSÃO

Os modelos foram avaliados quanto à sua capacidade de classificar as patologias de coluna vertebral com precisão. A análise de desempenho foi realizada utilizando as métricas AUC, Acurácia (AC), Precisão (*Prec*), Revocação (*Recall*) e Medida F1 (F1), permitindo uma avaliação abrangente da eficácia dos algoritmos. Os desempenhos dos modelos são apresentados no intervalo 0 a 1 indicando o percentual de 0% a 100%, respectivamente.

A Tabela 1 exibe os resultados dos algoritmos avaliados na versão do conjunto de dados de duas Classes (classificação binária entre *normal* e *abnormal*), enquanto a Tabela 2 exibe os resultados quando os algoritmos são aplicados na versão de três classes (classificação multiclasse de normal, hérnia de disco e espondilolistese).

Os resultados apresentados na Tabela 1 demonstram que os modelos avaliados obtiveram desempenhos promissores na tarefa de classificação binária do conjunto de dados vertebral, com destaque para os algoritmos MLP, SVM e AdaBoost. O modelo MLP apresentou o melhor desempenho geral, alcançando a maior AUC (0,947), indicando elevada capacidade de distinção entre pacientes normais e anormais. Além disso, obteve valores consistentes de acurácia (0,829), *F1-score* (0,829), precisão (0,830) e recall (0,829), evidenciando um comportamento equilibrado entre identificação correta das classes e redução de falsos positivos e falsos negativos.

Tabela 1. Desempenho dos oito modelos na classificação binária.

Modelo	AUC	CA	F1	Prec	Recall
RF	0,897	0,817	0,817	0,817	0,817
DT	0,801	0,817	0,817	0,817	0,817
GB	0,897	0,805	0,805	0,805	0,805
k-NN	0,927	0,817	0,816	0,822	0,817
SVM	0,896	0,841	**0,841**	**0,847**	**0,841**
NB	0,870	0,780	0,779	0,787	0,780
AdaBoost	0,841	**0,841**	**0,841**	0,842	**0,841**
MLP	**0,947**	0,829	0,829	0,830	0,829

Tabela 2. Desempenho dos oito modelos na classificação multiclasse.

Modelo	AUC	CA	F1	Prec	Recall
RF	0,903	0,829	0,829	0,830	0,829
DT	0,801	0,817	0,817	0,817	0,817
GB	0,897	0,805	0,805	0,805	0,805
k-NN	0,927	0,817	0,816	0,822	0,817
SVM	0,885	**0,841**	**0,841**	**0,847**	**0,841**
NB	0,870	0,780	0,779	0,787	0,780
AdaBoost	0,841	**0,841**	**0,841**	0,842	**0,841**
MLP	**0,947**	0,829	0,829	0,830	0,829

Ainda na Tabela 1, o SVM e o AdaBoost também apresentaram resultados expressivos, ambos atingindo acurácia e *F1-score* de 0,841. O SVM obteve a maior precisão entre todos os modelos (0,847), sugerindo maior confiabilidade nas classificações positivas realizadas pelo algoritmo. Esses resultados indicam que métodos baseados em margens de separação e técnicas de *ensemble* podem ser particularmente eficazes para problemas envolvendo atributos biomecânicos da coluna vertebral.

Pela Tabela 1, observa-se que o algoritmo *k*-NN apresentou uma AUC elevada (0,927), superior à de vários modelos com maior acurácia, demonstrando boa capacidade discriminativa mesmo com métricas globais ligeiramente inferiores ao SVM e ao MLP. Já os modelos RF e DT obtiveram desempenho semelhante em acurácia e F1-score (0,817), embora o RF tenha apresentado uma AUC significativamente superior, indicando melhor estabilidade na separação entre as classes. Por outro lado, o NB apresentou os menores resultados entre os modelos avaliados, possivelmente devido à hipótese de independência entre atributos, condição raramente observada em dados biomecânicos reais.

Os resultados apresentados na Tabela 2 demonstram que os modelos de aprendizado de máquina obtiveram desempenho satisfatório na tarefa de classificação multiclasse das patologias da coluna vertebral. De maneira geral, os algoritmos apresentaram métricas equilibradas entre acurácia, precisão, *recall* e *F1-score*, indicando boa capacidade de distinguir pacientes pertencentes às classes Normal, Hérnia de Disco e Espondilolistese.

O modelo MLP na Tabela 2 destacou-se por alcançar a maior AUC (0,947), evidenciando elevada capacidade discriminativa entre as três classes avaliadas. Embora sua acurácia (0,829) tenha sido ligeiramente inferior à obtida por SVM e AdaBoost, o alto valor de AUC sugere maior robustez na separação das categorias ao longo de diferentes limiares de decisão. Esse resultado reforça o potencial das redes neurais artificiais para modelar relações complexas e não lineares entre atributos biomecânicos da coluna vertebral.

Os algoritmos SVM e AdaBoost da Tabela 2 apresentaram os melhores valores de acurácia e *F1-score* (0,841), indicando desempenho consistente na classificação multiclasse. O SVM também obteve a maior precisão (0,847), demonstrando maior confiabilidade nas predições realizadas pelo modelo. Esses resultados sugerem que métodos baseados em margens de separação e técnicas ensemble podem oferecer maior estabilidade e generalização em cenários clínicos envolvendo múltiplas patologias vertebrais.

Na Tabela 2, o RF apresentou desempenho competitivo, com AUC de 0,903 e acurácia de 0,829, confirmando sua capacidade de lidar adequadamente com dados tabulares e relações entre variáveis biomecânicas. O *k*-NN também apresentou AUC elevada (0,927), indicando boa habilidade discriminativa, embora com métricas globais ligeiramente inferiores. Em contrapartida, o NB obteve o menor desempenho entre os modelos avaliados, possivelmente devido à suposição de independência entre atributos, hipótese que raramente se sustenta em dados médicos reais, nos quais frequentemente existem correlações entre variáveis anatômicas e biomecânicas.

De forma geral, os resultados das Tabelas 1 e 2 evidenciam que modelos de aprendizado de máquina podem ser aplicados de maneira promissora no apoio ao diagnóstico de alterações da coluna vertebral, especialmente por meio de técnicas capazes de modelar relações não lineares entre atributos biomecânicos. Os elevados desempenhos obtidos por modelos como MLP, SVM e AdaBoost demonstram o potencial dessas abordagens para integrar sistemas inteligentes de apoio à decisão clínica, auxiliando profissionais da saúde no rastreamento precoce, triagem automatizada e identificação de diferentes patologias vertebrais.

Além disso, a capacidade dos modelos em realizar tanto classificação binária quanto multiclasse amplia suas possibilidades de aplicação em cenários clínicos reais, permitindo não apenas distinguir indivíduos normais e anormais, mas também identificar patologias específicas, como Hérnia de Disco e Espondilolistese. Esses achados reforçam a relevância do aprendizado de máquina no desenvolvimento de ferramentas de diagnóstico assistido mais rápidas, precisas e consistentes, reduzindo a dependência exclusiva da interpretação humana e contribuindo para processos de avaliação clínica mais eficientes e menos subjetivos.

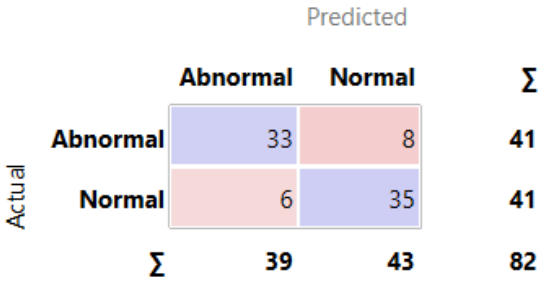


Figura 6. Matriz de confusão do modelo MLP no conjunto de dados PVC-2C

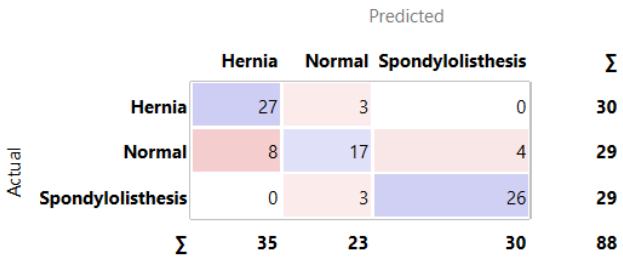


Figura 7. Matriz de confusão do modelo MLP no conjunto de dados PVC-3C

O modelo MLP foi selecionado para análise das matrizes de confusão apresentadas nas Figuras 6 e 7 por ter obtido o maior valor de AUC entre todos os algoritmos avaliados, tanto na classificação binária quanto na multiclasse. Esse resultado indica maior capacidade discriminativa do modelo em diferenciar corretamente as classes, tornando-o o algoritmo mais adequado para uma análise detalhada do comportamento das predições.

Na Figura 6, referente ao conjunto PVC-2C, observa-se que o modelo classificou corretamente 33 amostras da classe Abnormal e 35 da classe Normal, totalizando 68 classificações corretas em 82 amostras avaliadas. Os erros ocorreram principalmente na confusão entre indivíduos normais e anormais, com 8 amostras anormais classificadas como normais e 6 amostras normais classificadas como anormais. Esses resultados demonstram que o modelo apresenta boa capacidade de identificação das alterações vertebrais, embora ainda existam casos limítrofes que dificultam a separação completa entre as duas classes.

Já na Figura 7, correspondente ao conjunto PVC-3C, o MLP apresentou desempenho satisfatório na distinção entre as classes Hérnia, Normal e Espondilolistese. O modelo classificou corretamente 27 amostras de Hérnia, 17 amostras Normais e 26 amostras de Espondilolistese. Observa-se que os maiores erros ocorreram entre as classes Normal e Hérnia, além de algumas confusões entre Normal e Espondilolistese. Esse comportamento pode estar relacionado à similaridade de determinados atributos biomecânicos entre pacientes saudáveis e indivíduos em estágios menos severos das patologias vertebrais.

A Figura 8 apresenta a interpretação do impacto dos atributos sobre as decisões do modelo MLP por meio da análise SHAP para dados PVC-2C. Observa-se que o atributo degree_*spondylolisthesis* foi o mais influente no processo de classificação, apresentando os maiores valores de impacto positivo e negativo na saída do modelo. Valores elevados desse atributo contribuíram significativamente para a classificação de pacientes como anormais, enquanto valores reduzidos estiveram mais associados à classe normal. Esse comportamento sugere que o grau de espondilolistese possui forte relação com alterações biomecânicas relevantes para o diagnóstico das patologias vertebrais analisadas.

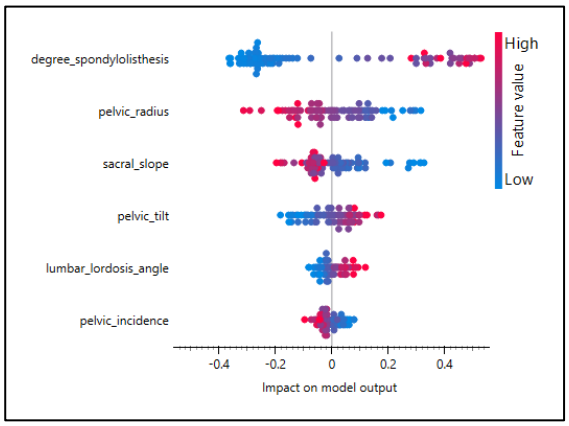


Figura 8. Distribuição do impacto dos atributos no modelo MLP para o conjunto PVC-2C utilizando análise SHAP.

Os atributos *pelvic_radius* e *sacral_slope* também demonstraram influência considerável nas decisões do modelo. Nota-se que diferentes faixas de valores desses atributos impactam a predição de maneira distinta, indicando relações complexas e não lineares entre as variáveis biomecânicas e as classes do problema. Já os atributos *pelvic_tilt*, *lumbar_lordosis_angle* e *pelvic_incidence* apresentaram menor dispersão e impacto reduzido na

saída do modelo, sugerindo menor relevância individual para a classificação binária realizada.

A Figura 9 reforça essa interpretação ao apresentar a importância dos atributos baseada na redução da AUC. Novamente, *degree_spondylolisthesis* destacou-se como o atributo mais relevante, sendo responsável pela maior redução no desempenho do modelo quando removido. Isso evidencia que o MLP depende fortemente dessa variável para diferenciar indivíduos normais e anormais. Em seguida, os atributos *pelvic_radius* e *sacral_slope* também apresentaram contribuição relevante, embora significativamente inferior ao principal atributo.

De maneira geral, os gráficos evidenciam que o modelo MLP aprendeu padrões biomecânicos coerentes com características clínicas associadas às alterações vertebrais no contexto de classificação binária. Além disso, a interpretação dos atributos contribui para aumentar a transparência e a confiabilidade do modelo, aspecto importante em aplicações médicas, pois permite compreender quais variáveis exercem maior influência nas decisões automatizadas do sistema.

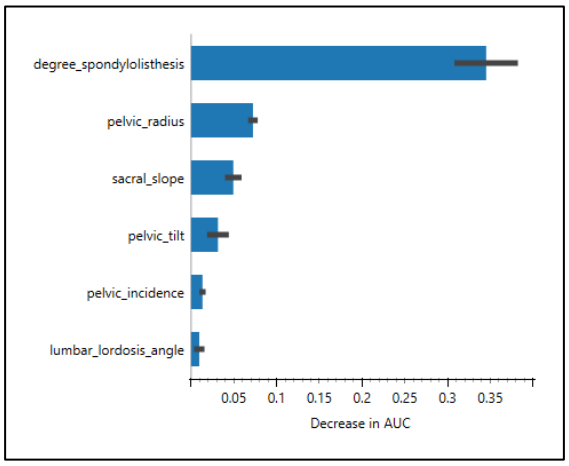


Figura 9. Importância dos atributos do modelo MLP baseada na redução da métrica AUC no conjunto PVC-2C.

A Figura 10 apresenta a análise SHAP do modelo MLP aplicada ao conjunto PVC-3C, evidenciando o impacto individual dos atributos biomecânicos sobre as predições multiclasse. Observa-se que o atributo *degree_spondylolisthesis* permaneceu como a variável mais influente no processo de classificação, apresentando ampla dispersão dos valores de impacto no modelo. Valores elevados desse atributo contribuíram principalmente para aumentar a probabilidade de determinadas classes patológicas, enquanto valores reduzidos tenderam a aproximar as predições da classe normal. Esse comportamento demonstra forte associação entre o grau de espondilolistese e as alterações biomecânicas da coluna vertebral.

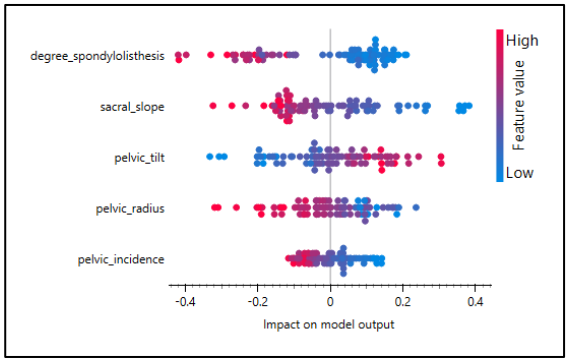


Figura 10. Distribuição do impacto dos atributos no modelo MLP para o conjunto PVC-3C utilizando análise SHAP.

Os atributos *sacral_slope, pelvic_tilt* e *pelvic_radius* também apresentaram influência significativa nas decisões do modelo. Em especial, o sacral_slope demonstrou ampla variação no impacto das predições, sugerindo que diferentes valores dessa medida biomecânica podem estar relacionados às distintas patologias presentes no conjunto PVC-3C. O atributo *pelvic_tilt* apresentou comportamento mais distribuído entre impactos positivos e negativos, indicando participação relevante na separação entre as classes. Já *pelvic_radius* mostrou influência moderada, contribuindo de forma complementar para a discriminação dos padrões vertebrais.

Por outro lado, o atributo *pelvic_incidence* apresentou menor dispersão e menor impacto sobre as saídas do modelo, sugerindo contribuição reduzida na classificação multiclasse quando comparado aos demais atributos biomecânicos analisados.

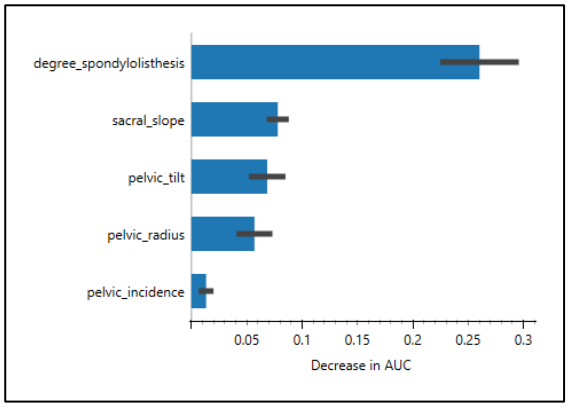


Figura 11. Importância dos atributos do modelo MLP baseada na redução da métrica AUC no conjunto PVC-2C.

A Figura 11 complementa essa análise ao apresentar a importância dos atributos baseada na redução da AUC. Novamente, *degree_spondylolisthesis* destacou-se como o atributo mais relevante para o desempenho do modelo, sendo responsável pela maior redução da AUC quando removido. Em seguida,

sacral_slope, *pelvic_tilt* e *pelvic_radius* também demonstraram importância significativa, indicando que o modelo depende da combinação dessas variáveis para realizar a diferenciação entre pacientes normais, com hérnia de disco e com espondilolistese.

CONCLUSÃO

Este trabalho teve como objetivo avaliar o desempenho de diferentes algoritmos de AM na classificação de patologias da coluna vertebral utilizando os subconjuntos de dados PVC-2C e PVC-3C. Os resultados demonstraram que modelos como MLP, SVM e AdaBoost apresentaram desempenhos promissores tanto na classificação binária quanto multiclasse, destacando-se principalmente pela elevada capacidade discriminativa, precisão e equilíbrio entre as métricas avaliadas. O modelo MLP obteve o maior valor de AUC nos dois cenários analisados, evidenciando maior robustez na diferenciação entre indivíduos normais e pacientes com alterações vertebrais. Além disso, as análises interpretáveis por meio de SHAP e importância de atributos mostraram que variáveis biomecânicas como *degree_spondylolisthesis*, *sacral_slope* e *pelvic_radius* exercem papel relevante nas decisões do modelo. Esses achados reforçam o potencial do aprendizado de máquina como ferramenta de apoio à decisão clínica, contribuindo para diagnósticos mais rápidos, precisos e menos dependentes exclusivamente da interpretação humana.

Como trabalhos futuros, sugere-se a ampliação da investigação com outros conjuntos de dados clínicos e biomecânicos contendo maior diversidade de pacientes e patologias vertebrais, permitindo avaliar a capacidade de generalização dos modelos em cenários reais. Também podem ser exploradas técnicas mais avançadas de aprendizado profundo, métodos híbridos de *ensemble* e abordagens de aprendizado explicável, visando aumentar tanto o desempenho quanto a interpretabilidade dos sistemas inteligentes. Além disso, futuras pesquisas podem integrar exames de imagem, dados clínicos e informações demográficas aos atributos biomecânicos, possibilitando o desenvolvimento de modelos mais completos e aplicáveis em ambientes hospitalares e sistemas de apoio ao diagnóstico médico.

AGRADECIMENTOS

Agradecemos ao Programa Institucional de Bolsas de Iniciação Científica Júnior (PIBIC Jr/IFCE) pelo apoio financeiro concedido mediante a bolsa de pesquisa, permitindo o desenvolvimento deste estudo.

REFERÊNCIAS

ADADI, A.; BERRADA, M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, [S.l.], v. 6, p. 52138-52160, 2018. DOI: 10.1109/ACCESS.2018.2870052.

BARRETO, G.; NETO, A. Vertebral Column. Irvine: UCI Machine Learning Repository, 2005. DOI: https://doi.org/10.24432/C5K89B.

BRASIL. Ministério da Previdência Social. Previdência Social protege segurados ao conceder 4,1 milhões de benefícios por incapacidade temporária em 2025. Brasília, DF, 27 jan. 2026. Disponível em: https://www.gov.br/previdencia/pt-br/noticias/2026/janeiro/previdencia-social-protege-segurados-ao-conceder-4-1-milhoes-de-beneficios-por-incapacidade-temporaria-em-2025. Acesso em: 26 maio 2026.

BUDNA, B. et al. JSI Sound – a machine-learning tool in Orange for simple biosound classification. In: DHIFLI, W.; DIALLO, A. B.; NGUIFO, E. M.; ZAKI, M. J. (org.). *Proceedings of the Workshop on Advances in Bioinformatics and Artificial Intelligence: Bridging the Gap*, Melbourne, 2017. Aachen: CEUR-WS.org, 2017. p. 26–29. Disponível em: https://ceur-ws.org/Vol-1942/paper4.pdf. Acesso em: 26 maio 2026.

CASEMIRO, K. G.; VIEIRA, K. V. S. Eficácia das abordagens fisioterapêuticas no tratamento conservador de hérnia de disco: revisão de literatura. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, [S.l.], v. 7, n. 10, p. 2243–2265, 2021. Disponível em: https://periodicorease.pro.br/rease/article/view/2795. Acesso em: 26 maio 2026.

CHAWLA, N. V. et al. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, [S.l.], v. 16, p. 321–357, 2002.

COPPIN, B. *Inteligência artificial*. Tradução e revisão técnica de Jorge Duarte Pires Valério. Rio de Janeiro: LTC, 2015.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, [S.l.], v. 20, n. 3, p. 273–297, 1995.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. *Inteligência artificial: uma abordagem de aprendizado de máquina*. 3. ed. Rio de Janeiro: Gen LTC, 2025.

IZBICKI, R.; DOS SANTOS, T. M. *Aprendizado de máquina: uma abordagem estatística*. São Carlos: Rafael Izbicki, 2020.

LAURINDO, A. C. A. et al. Efeitos da terapia manual no tratamento de pacientes com espondilolistese. *Brazilian Journal of Implantology and Health Sciences*, [S.l.], v. 6, n. 11, p. 2263–2272, 2024. Disponível em: https://bjihs.emnuvens.com.br/bjihs/article/view/4303. Acesso em: 26 maio 2026.

LECUN, Y.; HUANG, F. J.; BOTTOU, L. Learning methods for generic object recognition with invariance to pose and lighting. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)*. [S.l.]: IEEE, 2004. v. 2, p. II-104–II-111.

LI, J. et al. Traumatic lumbar disc herniation: a systemic case review and meta-analysis. *Brain and Spine*, [S.l.], v. 3, p. 102350, 2023. DOI: 10.1016/j.bas.2023.102350. Disponível em: https://www.sciencedirect.com/science/article/pii/S2772529423006380. Acesso em: 26 maio 2026.

LOGITHASAN, V. et al. Using machine learning to automatically measure axial vertebral rotation on radiographs in adolescents with idiopathic scoliosis. *Medical Engineering & Physics*, [S.l.], v. 107, p. 103848, 2022. DOI: 10.1016/j.medengphy.2022.103848. Disponível em: https://www.sciencedirect.com/science/article/pii/S1350453322000972. Acesso em: 26 maio 2026.

MISHRA, P. *Practical explainable AI using Python: artificial intelligence model explanations using Python-based libraries, extensions, and frameworks*. Cham: Springer, 2022.

MOLNAR, C. *Interpretable Machine Learning*. [S.l.]: Leanpub, 2020. Disponível em: https://books.google.com.br/books?id=jBm3DwAAQBAJ. Acesso em: 26 maio 2026.

NAIK, A.; SAMANT, L. Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science*, [S.l.], v. 85, p. 662–668, 2016.

PALECHOR, F. M.; MANOTAS, A. H. Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. *Data in Brief*, [S.l.], v. 25, p. 104344, 2019.

PANCH, T.; SZOLOVITS, P.; ATUN, R. Artificial intelligence, machine learning and health systems. *Journal of Global Health*, [S.l.], v. 8, n. 2, p. 020303, 2018.

PESSIN, J. *Dor na coluna vertebral: fatores determinantes entre dor na coluna vertebral e a sua relação com achados radiológicos em idosos*. São Paulo: Editora Dialética, 2022.

POPCHEV, I.; OROZOVA, D. Algorithms for machine learning with Orange system. *International Journal of Online and Biomedical Engineering*, [S.l.], v. 19, n. 4, p. 109–123, 2023.

RESHI, A. A. et al. Diagnosis of vertebral column pathologies using concatenated resampling with machine learning algorithms. *PeerJ Computer Science*, [S.l.], v. 7, p. e547, 2021. DOI: 10.7717/peerj-cs.547. Disponível em: https://peerj.com/articles/cs-547/. Acesso em: 26 maio 2026.

RIBEIRO, R. P. F.; NETO, A. R. R.; ROCHA, T. A. Diagnóstico de patologias da coluna vertebral utilizando ensemble com opção de rejeição. *Journal of Health Informatics*, [S.l.], v. 16, n. especial, 2024. DOI: 10.59681/2175-4411.v16.iEspecial.2024.1216. Disponível em: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/1216. Acesso em: 26 maio 2026.

ROCHA NETO, A. R. et al. Diagnostic of pathology on the vertebral column with embedded reject option. In: VITRIÀ, J.; SANCHES, J. M.; HERNÁNDEZ, M. (org.). *Pattern Recognition and Image Analysis*. Berlin: Springer, 2011. p. 544–551. DOI: https://doi.org/10.1007/978-3-642-21257-4_73.

RODRIGUES, M. A. et al. Análise computacional da biomecânica de movimentos da coluna vertebral. *Journal of Health Informatics*, [S.l.], v. 16, n. especial, 2024. Disponível em: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/1307. Acesso em: 26 maio 2026.

SHAILAJA, K.; SEETHARAMULU, B.; JABBAR, M. A. Machine learning in healthcare: a review. In: *International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2., 2018. Proceedings […]. Coimbatore: IEEE, 2018. p. 910–914. DOI: 10.1109/ICECA.2018.8474918.

SHAP. Welcome to the SHAP documentation. [S.l.], 2018. Disponível em: https://shap.readthedocs.io/en/latest/index.html. Acesso em: 27 maio 2026.

SILVA, C. A. et al. Conceitos e tratamentos da espondilolistese – revisão de literatura. In: *Ciências da Saúde: desafios, perspectivas e possibilidades*. [S.l.]: Editora Científica Digital, 2022. v. 5, p. 86–95. DOI: 10.37885/220408694.

SILVA, I. C. et al. Analysis of a decade of hospitalizations for spinal disc disorders in Brazil. *Research, Society and Development*, [S.l.], v. 12, n. 11, p. e94121143803, 2023. DOI: 10.33448/rsd-v12i11.43803. Disponível em: https://rsdjournal.org/index.php/rsd/article/view/43803. Acesso em: 26 maio 2026.

SOUZA, N. S. S.; SANTANA, V. S. Posição socioeconômica e duração do benefício por incapacidade devido a doenças musculoesqueléticas relacionadas ao trabalho. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 28, n. 2, 2012. DOI: 10.1590/S0102-311X2012000200011.

TEBET, M. A. Conceitos atuais sobre equilíbrio sagital e classificação da espondilólise e espondilolistese. *Revista Brasileira de Ortopedia*, [S.l.], v. 49, n. 1, p. 3-12, 2014. DOI: 10.1016/j.rbo.2013.04.011. Disponível em: https://www.sciencedirect.com/science/article/pii/S0102361613001835. Acesso em: 26 maio 2026.

VIMBI, V.; SHAFFI, N.; MAHMUD, M. Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain Informatics*, [S.l.], v. 11, n. 1, p. 10, 2024. DOI: 10.1186/s40708-024-00222-1. Disponível em: https://doi.org/10.1186/s40708-024-00222-1. Acesso em: 26 maio 2026.