



## Diagnóstico de patologias da coluna vertebral utilizando ensemble com opção de rejeição

### Diagnosis of spinal column pathologies using ensemble with rejection option

### Diagnóstico de patologías de la columna vertebral utilizando ensemble con opción de rechazo

Reginaldo Pereira Fernandes Ribeiro<sup>1</sup>, Ajalmar Rego da Rocha Neto<sup>2</sup>,  
Thiago Alves Rocha<sup>2</sup>

1 Mestrando, Instituto Federal do Ceará – IFCE, Fortaleza (CE), Brasil.

2 Professor, Instituto Federal do Ceará – IFCE, Fortaleza (CE), Brasil.

Autor correspondente: Reginaldo Pereira Fernandes Ribeiro  
*E-mail:* reginaldo.fernands@gmail.com

#### Resumo

**Objetivo:** propor uma nova abordagem para tomada de decisão com opção de rejeição em comitês de classificadores. **Método:** O método desenvolvido contempla técnicas de classificação utilizando ensembles com a abordagem de Opção de Rejeição usando o Índice de Gini (IG) como métrica de confiança. Estabelecemos limiares baseados na distribuição da pureza das porcentagens obtidas de cada classe, permitindo que o modelo se abstenha de prever amostras de difícil classificação em diagnósticos médicos relacionados a doenças da coluna vertebral. **Resultados:** O modelo proposto superou os comparativos, atingindo 97.55% de acurácia e rejeitando 61.69% das amostras no cenário mais conservador. A curva de Acurácia e Rejeição destacou sua superioridade. **Conclusão:** A definição de intervalos de valores do IG oferece flexibilidade ao ajustar a rigidez do comitê, além disso revelando potencial para otimizar comitês de classificação em diversas aplicações, proporcionando maior confiabilidade no reconhecimento de padrões.

**Descritores:** Coluna Vertebral; Confiança; Opção de Rejeição.

#### Abstract

**Objective:** To propose a new approach for decision-making with rejection option in classifier committees. **Method:** The developed method encompasses classification



techniques using ensembles with the Rejection Option approach, employing the Gini Index (GI) as a confidence metric. We established thresholds based on the distribution of purity percentages obtained from each class, allowing the model to abstain from predicting samples that are difficult to classify in medical diagnostics related to cervical column diseases. **Results:** The proposed model outperformed comparisons, achieving 97.55% accuracy and rejecting 61.69% of samples in the most conservative scenario. The Accuracy and Rejection curve highlighted its superiority. **Conclusion:** Defining GI value ranges offers flexibility in adjusting the committee's rigidity, further revealing potential for optimizing classifier committees in various applications, providing greater reliability in pattern recognition.

**Keywords:** Cervical Column; Confidence; Reject Option

## Resumen

**Objetivo:** Proponer un nuevo enfoque para la toma de decisiones con opción de rechazo en comités de clasificadores. **Método:** El método desarrollado abarca técnicas de clasificación utilizando ensamblajes con el enfoque de Opción de Rechazo, empleando el Índice de Gini (IG) como métrica de confianza. Establecimos umbrales basados en la distribución de los porcentajes de pureza obtenidos de cada clase, permitiendo que el modelo se abstenga de predecir muestras difíciles de clasificar en diagnósticos médicos relacionados con enfermedades de la columna vertebral. **Resultados:** El modelo propuesto superó las comparaciones, alcanzando una precisión del 97.55% y rechazando el 61.69% de las muestras en el escenario más conservador. La curva de Precisión y Rechazo destacó su superioridad. **Conclusión:** La definición de intervalos de valores del IG ofrece flexibilidad para ajustar la rigidez del comité, además de revelar un potencial para optimizar comités de clasificación en diversas aplicaciones, proporcionando mayor fiabilidad en el reconocimiento de patrones.

**Descriptores:** Columna Vertebral; Confianza; Opción de Rechazo

## Introdução

A coluna vertebral, constituída por vértebras e discos intervertebrais, é o pilar central do corpo humano. Responsável pela sustentação, mobilidade e proteção da



medula espinhal, sua saúde é fundamental para o bem-estar e a qualidade de vida. Em humanos, é composto por 33 vértebras que junto com o crânio, as costelas e o esterno, essas vértebras constituem o sistema esquelético axial (1). No entanto, diversas patologias podem acometer essa estrutura, causando dores intensas, limitações funcionais e comprometendo a qualidade de vida.

O diagnóstico preciso e oportuno dessas doenças é crucial para a implementação de tratamentos eficazes e a prevenção de complicações. Contudo, o processo tradicional de diagnóstico, baseado em exames físicos e análises de imagem, frequentemente enfrenta limitações. A subjetividade na interpretação dos exames, a sobrecarga de trabalho dos profissionais de saúde e a complexidade de certos quadros clínicos dificultam a obtenção de diagnósticos precisos e tempestivos (1).

A aplicação de modelos de aprendizado de máquina no diagnóstico de doenças humanas, especialmente no contexto de patologias da coluna vertebral, oferece suporte significativo aos especialistas médicos na identificação precoce de doenças com base em sintomas. No entanto, a complexidade de algumas condições, como Espondilolistese e Hérnia de Disco, que apresentam sintomas semelhantes, requer avanços tecnológicos para decisões diagnósticas mais precisas (2).

Os avanços na pesquisa para o tratamento de Espondilolistese e Hérnia de Disco nas últimas décadas indicam a identificação de fatores de risco associados a essas condições. No entanto, a tecnologia de ponta é necessária para aprimorar as decisões diagnósticas com significativa precisão (3). A motivação principal deste trabalho de pesquisa reside na concepção de um sistema de inferência confiável e na oferta de diretrizes para prevenir a seleção de padrões menos imprecisos na avaliação clínica de distúrbios da coluna vertebral.

O objetivo deste trabalho propor uma nova técnica de classificação com opção de rejeição em comitês de classificadores para aprimorar a confiabilidade dos diagnósticos médicos relacionados a doenças da coluna vertebral. Em particular, exploramos a aplicação do Índice de Gini como métrica para avaliar a confiança das classificações. Dessa forma, visando identificar e rejeitar diagnósticos menos confiáveis, contribuindo assim para a robustez e eficácia do processo de tomada de decisão médica.



Este estudo busca não apenas apresentar uma nova proposta teórica, mas também validar empiricamente sua eficácia, portanto, fornecendo uma perspectiva promissora para aprimorar a confiabilidade dos sistemas de diagnóstico automatizado em medicina, com um foco específico nas doenças da coluna vertebral.

## Trabalhos Relacionados

A confiabilidade é uma métrica importante para modelos que usam abordagens baseadas em rejeição. Essas abordagens têm vantagens significativas em várias áreas, como detecção de fraudes, direção autônoma, diagnóstico médico, sistemas de recomendação e análise de sentimento (4). A confiabilidade das classificações desempenha um papel crucial no aumento da segurança e da precisão em vários cenários.

Para minimizar os possíveis erros obtidos por classificação de instâncias, os ensembles surgem como uma abordagem confiável para refinar a precisão das classificações. Esses modelos combinam a previsão de vários modelos individuais, aproveitando a diversidade para obter um resultado mais robusto. Entre os classificadores de ensemble amplamente utilizados, destacam-se os das abordagens *Voting*, *Bagging*, *Boosting* e *Stacking* (5).

Mesmo com a popularização dos métodos de comitês de classificadores, a possibilidade de erros de classificação é comum, já que no mundo real, as variáveis relacionadas aos problemas de aprendizado de máquina podem influenciar negativamente na tomada de decisão. Nesse contexto, a abordagem de rejeição, inicialmente proposta por CHOW (6), resultou em diversos estudos investigativos sobre o uso dessa técnica para estimar a confiança em problemas de classificação com rejeição. Fumera (7) e Rocha Neto et al. (2) aplicaram essa abordagem utilizando minimização de risco empírico, obtendo resultados promissores na identificação de amostras de difícil classificação.

Vários estudos têm explorado a aplicabilidade da regra de decisão de Bayes em problemas de classificação com rejeição cuja confiança é estimada com base na probabilidade a posteriori de pertencer a cada classe (8). Outra abordagem relevante é a Probabilidade Máxima a Posteriori (MAP), usada para estimar a confiança em problemas de classificação com rejeição através do trade-off entre taxa de erro e rejeição (6).



No contexto de métricas de espaço, pesquisas exploraram o uso das distâncias Euclidiana e de Mahalanobis para avaliar a confiança em classificação do tipo. Por exemplo, em k-vizinhos mais próximos, um limite de longitude entre os vizinhos pode ser usado para rejeitar amostras muito distantes (10).

Por fim, em relação às redes neurais, de acordo com Cordella et al. em 1997 (11) e Villon et al. 2020 (12), a confiança pode ser definida como a probabilidade do neurônio vencedor ou a diferença entre os dois neurônios principais. Essas abordagens possibilitaram a rejeição de amostras com baixa confiança, alcançando resultados promissores em termos de precisão e robustez do sistema.

Este estudo visa apresentar uma nova abordagem de classificação com opção de rejeição específicos para o problema de diagnóstico de doenças da coluna vertebral em um contexto multiclasse. Apresentamos uma abordagem inovadora que utiliza o Índice de Gini (IG) como limiar de rejeição em modelos de comitês de classificação.

O principal objetivo é propor uma nova abordagem para tomada de decisão com opção de rejeição em comitês de classificadores, para aprimorar a confiabilidade dos diagnósticos médicos relacionados a doenças da coluna vertebral em situações desafiadoras com padrões de difícil categorização. Para tanto, empregamos simulações com o modelo *Voting*, formado por classificadores base como *AdaBoost* (AdB), *Support Vector Machine* (SVM), *XGBoost* (XGB), *K-Nearest Neighbors* (KNN), *Multilayer Perceptron* (MLP), *Gaussian Naive Bayes* (GNB) e *Random Forest* (RF).

Para avaliar os resultados usamos métricas como acurácia e taxa de rejeição das classificações, além da análise gráfica da relação entre as taxas de acerto e de rejeição.

## Métodos

A eficácia da classificação da abordagem proposta foi investigada utilizando o conjunto de dados de patologia da coluna vertebral disponível no repositório de aprendizado de máquina da UCI<sup>1</sup>. O conjunto de dados foi desenvolvido pelo Dr. Henrique da Mota no Grupo de Pesquisa Aplicada em Ortopedia (GARO) do Centro

---

<sup>1</sup> UCI Machine Learning Repository



Médico-Cirúrgico de Reabilitação *des Massues*, Lyon, França (13). Cada registro no conjunto de dados contém seis formas e orientações da pelve e coluna lombar, com atributos diagnósticos, incluindo incidência pélvica, inclinação pélvica, inclinação sacral, ângulo de lordose lombar, raio pélvico e grau de espondilolistese.

Conforme mostrado na Tabela 1, o conjunto de dados não é balanceado, ou seja, a proporção do número de instâncias de classes diferentes não é igual.

**Tabela 1** – Detalhes dos registros do conjunto de dados para cada classe de destino.

Tipo de Registros (Pacientes)	Quantidade
Espondilolistese	150
Hérnia de Disco	60
Normais	100
Total	310

Para este estudo diferentes técnicas de aprendizado de máquina foram empregadas para abordar o problema de classificação de patologias. Em um total de sete modelos, onde todos foram submetidos a uma Grid Search para determinar parâmetros que possuem um melhor desempenho em acurácia. Todos os modelos aqui citados pertencem a biblioteca *Scikit-learn*, com exceção do *XGBoost* que foi obtido a partir da *Distributed (Deep) Machine Learning Community* (DMLC).

O classificador *Adaboost* (ADB), uma abreviação para *Adaptive Boosting*, é um meta-modelo<sup>2</sup> de aprendizado de máquina que cria um classificador forte a partir de muitos aprendizes fracos por meio do método de *boosting*. A implementação do ADB geralmente utiliza árvores de decisão como seus aprendizes fracos padrão, comumente chamados de *decision stumps* (14).

Em seguida vem o classificador *XGBoost* (XGB), uma abreviação para *Extreme Gradient Boosting* que é uma implementação eficiente e escalável do algoritmo de *gradient boosting* (15). Ele se destaca por sua capacidade de lidar com conjuntos de dados grandes e complexos, sendo amplamente utilizado em competições de ciência de dados.

O classificador *Support Vector Classifier* (SVC) é um modelo que busca encontrar o hiperplano de decisão ótimo para separar instâncias de diferentes classes

<sup>2</sup> Um modelo construído com base nas saídas de outros modelos, visando melhorar o desempenho preditivo ou a capacidade de generalização.



em um espaço de alta dimensão. É amplamente utilizado em problemas de classificação, especialmente quando os dados não são linearmente separáveis (16). O SVC utiliza uma abordagem baseada em vetores de suporte para encontrar a fronteira de decisão que maximiza a margem entre as classes.

Na abordagem de distância de padrões o classificador *k-Nearest Neighbors* (KNN) é um algoritmo de aprendizado de máquina baseado em instâncias. Ele classifica uma instância com base nas classes das *k* instâncias mais próximas no espaço de características (17).

Em relação às redes neurais usamos o classificador *Multi-layer Perceptron* (MLP) que é uma arquitetura de rede neural artificial composta por várias camadas de neurônios (18). É capaz de aprender padrões complexos nos dados, sendo especialmente eficaz em tarefas de classificação.

No contexto probabilístico usamos o classificador *Gaussian Naive Bayes* (GNB) é uma implementação do algoritmo *Naive Bayes* que assume que os atributos são distribuídos de acordo com uma distribuição gaussiana (19). Este classificador é especialmente eficaz em problemas de classificação com características contínuas e que não possuem parâmetros relevantes para realização de uma *Grid Search*.

O classificador *Random Forest* (RF) é uma técnica de ensemble que constrói várias árvores de decisão durante o treinamento e as combina para obter uma previsão mais robusta (17). O RF utiliza a técnica de *bagging* para treinar as árvores de decisão com amostragem *bootstrap*, resultando em um modelo robusto.

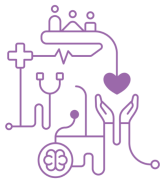
Estes classificadores foram utilizados na construção de um *ensemble Voting Classifier* com saída baseada no voto majoritário simples para tarefas de classificação. Cada classificador foi ajustado por meio de *Grid Search* para otimizar seus hiperparâmetros, resultando em configurações específicas para cada modelo.

A Tabela 2 mostra todos os parâmetros encontrados por *Grid Search* para execução dos experimentos:

**Tabela 2** - Configuração de parâmetros dos modelos obtidos por *Grid Search*.

Classificador	Parâmetros
ADB	n_estimators = 50, learning_rate = 1.0, algorithm = 'SAMME'
XGB	n_estimators = 100, learning_rate = 0.1, max_depth = 3, subsample = 0.8, colsample_bytree = 1.0
SVC	C = 10, gamma = 'scale', kernel = 'linear'
KNN	n_neighbors = 11, metric = 'manhattan', weights = 'distance'





<b>MLP</b>	activation = 'relu', alpha = 1e-05, hidden_layer_sizes = [64], learning_rate = 'constant', max_iter = 2000, solver = 'lbfgs', tol = 1e-05
<b>GNB</b>	não possui parâmetros específicos a serem ajustados.
<b>RF</b>	n_estimators = 150, random_state = 2, max_depth = 15, bootstrap = True, criterion = 'entropy', min_samples_leaf = 1, min_samples_split = 10

O uso de Comitês de Classificadores, também conhecido como *Ensemble Learning*, é uma abordagem valiosa no campo de aprendizado de máquina. Essa técnica envolve a combinação de vários modelos de previsão, denominados "membros" ou "componentes base", para aprimorar o desempenho global e a capacidade de generalização do sistema. Ao invés de depender de um único modelo para tomar decisões, o *Ensemble* utiliza a fusão das previsões de diferentes modelos, resultando em uma previsão final mais robusta e precisa.

Um método comum de combinação é o *Majority Voting*, onde os classificadores base emitem suas previsões e a classe que recebe a maioria das previsões é escolhida como a previsão final. A aplicação eficaz do *Hard Voting* destaca-se como uma técnica amplamente reconhecida no contexto de *Ensemble Learning* (25,26).

## Metodologia

Exploramos o uso de *ensembles* como classificadores, combinados com a estratégia de opção de rejeição, utilizando o Índice de Impureza de Gini (IG). Essa abordagem busca aprimorar a capacidade de classificação e ajuste de custo baseado na detecção de erros em problemas de reconhecimento de padrões.

A técnica de *Ensemble* com Opção de Rejeição visa melhorar a robustez de modelos de *machine learning* por meio da integração de uma estratégia de rejeição de previsões incertas. A abordagem é particularmente útil em contextos nos quais a confiança nas previsões é uma consideração crítica, como patologias na coluna vertebral.

Nosso método requer a especificação do número de classes, a quantidade de pesos, e as previsões estimadas para cada classe dos modelos individuais. A incerteza é avaliada utilizando métricas como o IG, e uma série de valores baseados no peso da rejeição para o cálculo do risco empírico determinando o custo da rejeição.

Para determinar os valores que representam o peso da rejeição consideramos o IG máximo em relação ao número de classes. A variação desses valores é ajustada





para refletir a incerteza desejada, resultando em uma lista de pesos representados por  $w_r$  (6) e o seu limite representado por  $w_r max$ , assim como  $c$  representa o número de classes mostrado na Equação 1:

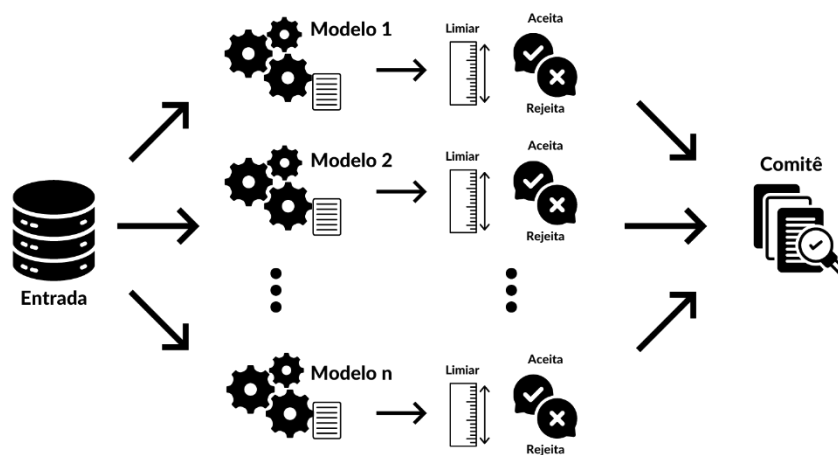
$$= \frac{c-1}{c} \quad (1)$$

O IG é calculado para avaliar a incerteza nas previsões (22). Ele representa a medida de impureza das distribuições de probabilidade associadas às classes onde  $c$  é o número de classes e o  $v$  é o vetor contendo as probabilidades para cada  $i$ -ésima classe (Equação 2). Quanto maior o IG, maior a incerteza.

$$IG = 1 - \sum_{i=1}^c v_i^2 \quad (2)$$

A função de rejeição aplica o critério de rejeição com base no IG. Se o IG excede o limiar de rejeição, a previsão é rejeitada e atribuído o valor -1. Caso contrário, a classe prevista é determinada pelo índice de probabilidade máximo da  $i$ -ésima classe (Equação 3) e representado na Figura 1.

**Figura 1** – Diagrama do Fluxo de Rejeição ou Classificação



$$\hat{y}(x) = \{-1, \text{ se } IG > w_r \quad \text{Classe } i, \text{ caso contrário} \quad (3)$$

Para decidir qual classe foi definida, a função de voto majoritário simples realiza uma votação por maioria simples para determinar a classe final em caso de rejeição. Se houver empate na contagem de votos, a opção de rejeição é acionada, retornando



o valor -1. A função de rejeição de voto aplica o critério de rejeição a todas as previsões dos modelos do *ensemble*. A votação por maioria é então realizada, resultando na classe final prevista, juntamente com cenários de rejeição.

## Resultados

Para avaliar os resultados obtidos pelo nosso modelo de Opção de Rejeição (RO\_VOTING), realizamos comparações entre os classificadores base. Aplicamos o método em todos os modelos no formato convencional, ou seja, na saída do classificador. Esses modelos incluem *AdaBoost* (RO\_ADB), *XGBoost* (RO\_XGB), *Support Vector Machine* (RO\_SVM), *k-Nearest Neighbors* (RO\_KNN), *Multi-layer Perceptron* (RO\_MLP), *Gaussian Naive Bayes* (RO\_GNB), *Random Forest* (RO\_RF), e o próprio modelo de *Voting* (RO\_VOTING). Todos foram aplicados ao conjunto de dados Coluna Cervical em seu modelo de três classes.

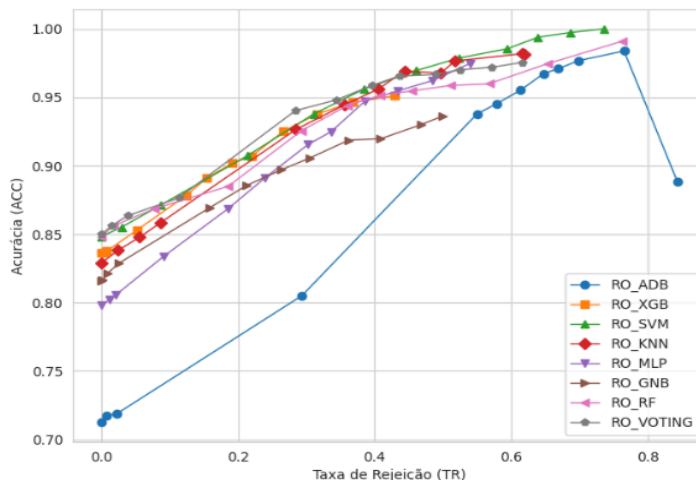
A técnica *Holdout* foi utilizada para a separação dos conjuntos de treino e teste na proporção 80/20. Os experimentos foram realizados 20 vezes, as métricas utilizadas para avaliação foram a média de acerto, que representa a precisão média nos testes após a rejeição, e a média de rejeição no subconjunto de teste. A média de rejeição é a taxa média de rejeição em relação ao número de execuções. A curva Acurácia e Rejeição (AR) é um gráfico de desempenho dos classificadores. Cada ponto na curva AR representa a média de acerto e rejeição para um determinado valor de limiar ótimo associado a um valor de  $w_r$ . Para demonstrando os resultados fizemos dois gráficos AR: um com a Acurácia (ACC) e outro com Acurácia Escalonada<sup>3</sup> (ACCE), os resultados podem ser vistos no Gráfico 1 e 2.

---

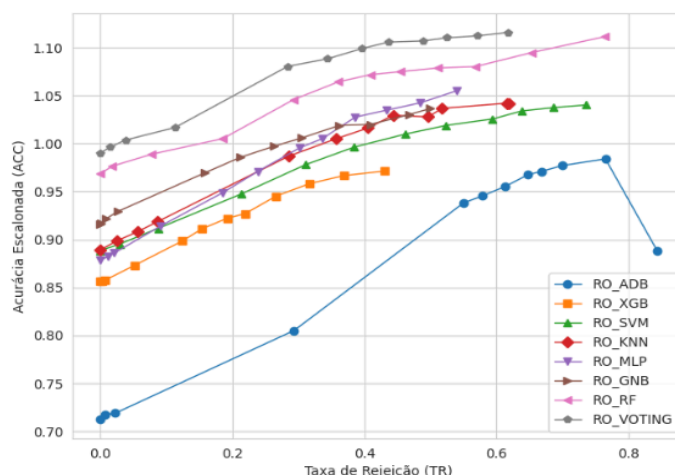
<sup>3</sup> O escalonamento de acurácia em um gráfico refere-se ao ajuste da escala do eixo y para melhor visualizar as variações na acurácia dos dados apresentados.



**Gráfico 1 - Curva AR entre os Classificadores.**



**Gráfico 2 - Curva AR Escalonada entre os Classificadores.**



Assim como a Tabela 3 com os valores de acurácia e taxa de rejeição que são representados no Gráfico 2 acima:

**Figura 2 - Valores Médios de Acurácia e Taxa de Rejeição nos Classificadores (%)**

wr	RO_ADB		RO_XGB		RO_SVC		RO_KNN		RO_MLP		RO_GNB		RO_RF		RO_VOTING	
	A	R	A	R	A	R	A	R	A	R	A	R	A	R	A	R
0.06	88.83	84.35	95.11	42.98	100.00	73.63	98.16	61.85	97.47	53.95	93.62	49.92	99.11	76.37	97.55	61.69
0.11	98.38	76.53	94.64	36.85	99.74	68.71	98.18	61.61	96.22	48.47	93.00	46.77	97.45	65.32	97.20	57.18
0.17	97.68	69.92	93.79	31.61	99.38	63.87	97.65	51.77	95.46	43.39	91.97	40.89	95.99	56.77	97.00	52.50
0.22	97.08	66.85	92.50	26.61	98.55	59.44	96.76	49.68	94.74	38.63	91.87	36.21	95.86	51.13	96.67	48.87
0.28	96.73	64.76	90.68	21.94	97.86	52.34	96.93	44.35	92.49	33.63	90.56	30.48	95.47	45.48	96.55	43.63
0.33	95.52	61.37	90.19	19.19	96.95	46.13	95.61	40.56	91.55	30.24	89.70	26.29	95.15	40.89	95.87	39.60
0.39	94.54	57.90	89.11	15.40	95.62	38.47	94.46	35.65	89.09	23.87	88.55	21.21	94.41	36.05	94.81	34.44
0.44	93.79	55.00	87.84	12.42	93.77	31.05	92.65	28.47	86.87	18.55	86.95	15.89	92.56	29.27	94.02	28.39
0.50	80.51	29.35	85.29	5.16	90.74	21.37	85.84	8.63	83.39	9.11	82.89	2.58	88.49	18.55	87.65	11.29
0.56	71.90	2.26	83.76	0.73	87.14	8.71	84.78	5.56	80.58	2.10	82.12	0.81	86.87	7.74	86.35	3.79
0.61	71.74	0.73	83.65	0.32	85.50	2.90	83.83	2.42	80.20	1.13	81.66	0.16	85.62	1.77	85.60	1.45
0.67	71.29	0.00	83.63	0.00	84.84	0.00	82.90	0.00	79.84	0.00	81.61	0.00	84.84	0.00	85.00	0.00



Ao analisar os limiares de rejeição ( $w_r$ ) em 0.06 e 0.67 para os diversos classificadores e nas áreas onde os mesmos podem ser comparados, identificamos padrões significativos relacionados à acurácia e à quantidade de instâncias rejeitadas. No limiar mais conservador de 0.06, observamos que o RO\_ADB apresenta uma acurácia de 88.83%, baixa em relação aos outros, destacando-se pela sua grande quantidade de previsões com alta confiança no limiar 0.11. Contudo, é importante notar que, nesse ponto, a taxa de rejeição é muito alta, indicando que muitas instâncias foram rejeitadas, levando o RO\_ADB a apresentar o pior resultado no *trade-off* acurácia e rejeição.

Ao aumentar o limiar para 0.67, observamos uma mudança nas estratégias dos classificadores, refletindo em menores taxas de rejeição. Nesse cenário menos conservador, o RO\_ADB adota uma postura liberal, e não rejeita nenhuma das instâncias, alcançando uma acurácia de 71.29%, indicando uma grande quantidade de erros. Em relação aos outros classificadores apresenta a maior rejeição de padrões, alcançando 84.35% no limiar mais conservador.

Em contraste, o RO\_XGBoost (RO\_XGB) mantém um desempenho notável em ambas as situações, com acurácias de 95.11% em 0.06 e 83.63% em 0.67. Essa consistência sugere a capacidade do RO\_XGB de se adaptar a diferentes limiares de rejeição, mantendo uma boa acurácia mesmo em cenários mais conservadores, onde atingiu a menor taxa de rejeição entre os classificadores.

Vale ressaltar que o RO\_Random Forest (RO\_RF), apesar de sua acurácia elevada de 99.11% em 0.06, apresenta uma taxa de rejeição de 76.37% e no limiar 0.67 apresenta uma acurácia de 84.84%, o que demonstra uma ótima adaptação, mas uma alta rejeição.

Surpreendentemente, o classificador RO\_VOTING se destaca como uma escolha versátil, apresentando uma acurácia de 97.55% e 61.69% de rejeição no limiar 0.06. Além disso, apresenta ACC de 85% e nenhuma rejeição no  $w_r$  0.67. Esses resultados indicam que o RO\_VOTING mantém um bom equilíbrio entre acurácia e taxa de rejeição, conseguindo pontuar melhor em quase todos os limiares, assim como mostra o Gráfico 2 que teve seu valor de acurácia escalonado para visualizar melhor o desempenho de cada classificador.



## Conclusão

O uso de técnicas de classificação de aprendizado de máquina desempenha um papel crucial em diversos domínios da área médica, e o diagnóstico de patologias não é exceção. A classificação de dados biomédicos é considerada uma tarefa complexa e sensível. Nesse contexto, este artigo propõe uma nova abordagem para tomada de decisão com opção de rejeição em comitês de classificadores, empregando o Índice de Incerteza de Gini (IG) nos classificadores consolidados como XGBoost, MLP, KNN e GNB. Os experimentos foram conduzidos em um conjunto de dados relacionados ao diagnóstico de patologia da coluna vertebral. Inicialmente, os algoritmos selecionados foram treinados e testados, e posteriormente, aplicamos a técnica de rejeição usando o IG como limiar e uma lista de valores ( $w_r$ ) para possibilitar a abstenção do resultado apresentado pelos classificadores.

Os resultados mostrados no Gráfico 2 e na Tabela 3 indicam que o classificador RO\_VOTING demonstra melhor desempenho em praticamente todos os limiares, com base nas métricas de acurácia e rejeição apresentados pelos classificadores. Assim, observa-se que de maneira geral, a eficiência da classificação melhora ao evitar erros decorrentes de decisões desafiadoras. Tal fenômeno é esperado em modelos que incorporam o algoritmo de rejeição, devido ao aprimoramento no desempenho de classificação tendo como consequência um maior número de rejeições dos padrões no conjunto de teste.

Para trabalhos futuros, os autores pretendem aplicar essa metodologia em outros problemas, além de incorporar a rejeição no processo de aprendizagem do classificador, possibilitando a criação de modelos capazes de aprender o valor ótimo de rejeição.

## Referências

1. Reshi AA, Ashraf I, Rustam F, Shahzad HF, Mehmood A, Choi GS. Diagnosis of vertebral column pathologies using concatenated resampling with machine learning algorithms. PeerJ Comput Sci. 22 de julho de 2021;7:e547.
2. Rocha Neto AR, Sousa R, de A. Barreto G, Cardoso JS. Diagnostic of Pathology on the Vertebral Column with Embedded Reject Option. Em: Vitrià J, Sanches JM,



- Hernández M, organizadores. Pattern Recognition and Image Analysis. Berlin, Heidelberg: Springer; 2011. p. 588–95.
3. Nanglia S, Ahmad M, Ali Khan F, Jhanjhi NZ. An enhanced Predictive heterogeneous ensemble model for breast cancer prediction. Biomed Signal Process Control. 1º de fevereiro de 2022;72:103279.
  4. Zhang XY, Xie G, Li XC, Mei T, Liu CL. A Survey on Learning to Reject. Proc IEEE. 2023.
  5. Mienye ID, Sun Y. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. IEEE Access. 2022;10:99129–49.
  6. Chow C. On optimum recognition error and reject tradeoff. IEEE Trans Inf Theory. 1970.
  7. Bartlett PL, Wegkamp MH. Classification with a Reject Option using a Hinge Loss. J Mach Learn Res. 1º de junho de 2008;9:1823–40.
  8. Fukunaga K. Introduction to statistical pattern recognition. 2. ed. San Diego [u.a.]: Acad. Press; 1990.
  9. Dubuisson B, Masson M. A statistical decision rule with incomplete knowledge about classes. Pattern Recognit. 1º de janeiro de 1993;26(1):155–65.
  10. Hellman ME. The Nearest Neighbor Classification Rule with a Reject Option. IEEE Trans Syst Sci Cybern. julho de 1970;6(3):179–85.
  11. Cordella LP, Foggia P, Sansone C, Tortorella F, Vento M. Classification reliability and its use in multi-classifier systems. Em: Del Bimbo A, organizador. Image Analysis and Processing. Berlin, Heidelberg: Springer; 1997. p. 46–53.
  12. Villon S, Mouillot D, Chaumont M, Subsol G, Claverie T, Villéger S. A new method to control error rates in automated species identification with deep learning algorithms. Sci Rep. 2020.



13. Guilherme Barreto AN. Vertebral Column [Internet]. UCI Machine Learning Repository; 2005. Disponível em: <https://archive.ics.uci.edu/dataset/212>
14. Ferreira AJ, Figueiredo MAT. Boosting Algorithms: A Review of Methods, Theory, and Applications. 2012.
15. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Em: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery; 2016. p. 785–94. Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939785>
16. Homenda W, Luckner M, Pedrycz W. Classification with rejection based on various SVM techniques. Em: 2014 International Joint Conference on Neural Networks (IJCNN). Beijing, China: IEEE; 2014. p. 3480–7. Disponível em: <https://ieeexplore.ieee.org/document/6889655>
17. Tan PN, Steinbach M, Karpatne A, Kumar V. Introduction to Data Mining. Pearson; 2019.
18. Gamas Sousa R, Rocha Neto AR, Cardoso JS, Barreto GA. Robust classification with reject option using the self-organizing map. Neural Comput Appl. 1º de outubro de 2015;26(7):1603–19.
19. Anand V, KiranBala B, Srividhya S, C. K, Younus M, Rahman MH. Gaussian Naïve Bayes Algorithm: A Reliable Technique Involved in the Assortment of the Segregation in Cancer. Mob Inf Syst. 17 de junho de 2022;2022:1–7.
20. Ballabio D, Todeschini R, Consonni V. Chapter 5 - Recent Advances in High-Level Fusion Methods to Classify Multiple Analytical Chemical Data. Em: Cocchi M, organizador. Data Fusion Methodology and Applications. Elsevier; 2019. p. 129–55. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780444639844000053>
21. Dogan A, Birant D. A Weighted Majority Voting Ensemble Approach for Classification. Em: 2019 4th International Conference on Computer Science and Engineering (UBMK). 2019. p. 1–6.





**CBIS'24**

**XX Congresso Brasileiro de Informática em Saúde**  
08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

22. Yuan Y, Wu L, Zhang X. Gini-Impurity Index Analysis. IEEE Trans Inf Forensics Secur. 2021.