Classificadores Lineares e Não-Lineares Baseados em Distância de Mahalanobis Aplicados em Cardiotocografia

Patrícia T. Leitão
Pós-Grad. em Eng. de Teleinformática
Universidade Federal do Ceará
Campus do Pici, Fortaleza
patricia.tavares@alu.ufc.br

Igor R. Sousa
Pós-Grad. em Eng. de Teleinformática
Universidade Federal do Ceará
Campus do Pici, Fortaleza
igor.sousa@alu.ufc.br

Guilherme A. Barreto
Pós-Grad. em Eng. de Teleinformática
Universidade Federal do Ceará
Campus do Pici, Fortaleza
gbarreto@ufc.br

***Resumo*—Neste trabalho, é realizada uma investigação abrangente a respeito do desempenho de cinco variantes de classificadores baseados em distância de Mahalanobis com o objetivo de classificar o estado fetal (normal, suspeito e patológico) a partir de dados de cardiotocograma: o classificador quadrático gaussiano (QG) sem regularização, QG com regularização de Tikhonov, QG com regularização de Friedman, QG com matriz de covariância agregada e QG com matriz de covariância diagonalizada. Essas variantes são avaliadas em combinação com técnicas de transformação de dados, como a transformação de Box-Cox e a análise de componentes principais, sendo comparadas ao classificador linear de mínimos quadrados e a resultados reportados na literatura. O classificador QG com regularização de Tikhonov ($\lambda = 0{,}008$) e transformação de Box-Cox ($\gamma = 0{,}3$) destacou-se por alcançar uma taxa mediana de acerto de 87,41%, valor máximo de 91,3%, especificidade de 99,7%, precisão de 95,4% e matriz de covariância com posto completo.**

***Index Terms*—cardiotocograma, distância de Mahalanobis, regularização de Tikhonov, transformação de Box-Cox.**

I. Introdução

O cardiotocograma é o registro simultâneo dos batimentos cardíacos fetais e das contrações uterinas, obtido por meio de um transdutor de ultrassom posicionado no abdômen materno. Esse exame é amplamente utilizado durante a gestação para fornecer aos obstetras informações essenciais sobre a saúde fetal. Por exemplo, ao analisar um cardiotocograma, o obstetra pode identificar sinais de sofrimento fetal, como a falta de oxigenação. Com isso, é possível realizar intervenções precoces, prevenindo a morte fetal ou o desenvolvimento de lesões neurológicas [1]. No entanto, interpretações incorretas do cardiotocograma podem levar à realização de intervenções cirúrgicas desnecessárias [2].

Nesse cenário, técnicas computacionais têm sido exploradas como ferramentas auxiliares na análise de cardiotocogramas, visando melhorar a detecção de padrões associados a diferentes condições fetais, como em [3], que aplicaram 32 classificadores a dados de cardiotocograma, destacando modelos baseados em árvores e *support vector machine* (SVM).

Este trabalho conta com o apoio financeiro do CNPq (No. 313251/2023-1) e do Centro de Referência em Inteligência Artificial - CEREIA, convênios 23067.055633/2021-19 (UFC) / 2020/09706-7 (FAPESP).

Entre as técnicas não lineares de classificação, as baseadas na distância de Mahalanobis [4] são particularmente interessantes, pois utilizam a matriz de covariância para incorporar a correlação entre atributos e apresentam menor complexidade computacional em comparação a métodos mais sofisticados, como redes neurais profundas. Isso as torna atrativas em contextos em que se desejam bons índices de classificação, mas com recursos computacionais limitados, como em dispositivos *wearable* ou dispositivos portáteis, como o Doppler Fetal Portátil S6 da *Best In Care* [5], capaz de identificar possíveis sofrimentos fetais, como hipóxia e circulatura de cordão.

Ao lidar com banco de dados, algumas técnicas de transformação podem ser aplicadas na etapa de pré-processamento dos dados. Entre elas, destacam-se a transformação de Box-Cox [6], que tem como objetivo tornar as distribuições marginais dos atributos mais próxima da normal [7], e a análise de componentes principais (PCA) [8], que gera novos atributos não correlacionados e é frequentemente usada para reduzir a dimensionalidade [9].

Apesar das vantagens dos classificadores não lineares baseados na distância de Mahalanobis, como a capacidade de incorporar correlações entre atributos e a baixa complexidade computacional, seu potencial no contexto da classificação de cardiotocogramas ainda é relativamente pouco explorado na literatura. Este trabalho busca contribuir nesse sentido ao investigar de forma abrangente o uso desses classificadores, considerando diferentes variantes e avaliando seu desempenho em conjunto com técnicas de transformação dos dados, como as transformações de Box-Cox e PCA. Entre as variantes analisadas, incluem-se classificadores lineares e não lineares. Os resultados obtidos são comparados ao classificador linear de mínimos quadrados e a resultados existentes na literatura.

Desta forma, este trabalho está inserido no contexto de sistemas biomédicos embarcados, que demandam soluções computacionalmente mais leves que técnicas avançadas como redes neurais e máquinas de vetores suporte. Neste primeiro momento, avalia-se o desempenho de tais modelos clássicos na classificação de cardiotocogramas, em contraste com abordagens mais sofisticadas.

O artigo é dividido da seguinte forma. Na Seção II, apresentam-se os classificadores baseados na distância de Mahalanobis, enquanto que na Seção III, apresenta-se o classificador linear de mínimos quadrados. Na Seção IV, descrevem-se os métodos de transformação de Box-Cox e PCA. A metodologia do trabalho é apresentada na Seção V. Por fim, resultados e conclusões são apresentados nas Seções VI e VII.

II. Classificadores baseados na distância de Mahalanobis

Classificadores que usam a distância de Mahalanobis em vez da euclidiana consideram a inversa da matriz de covariância dos dados da i-ésima classe, $\mathbf{\Sigma}_i^{-1} \in \mathbb{R}^{p\times p}$, como matriz de ponderação, em que p é o número de atributos. Assim, a distância do vetor de atributos $\mathbf{x}_n \in \mathbb{R}^{p\times 1}$ ao i-ésimo centróide $\mathbf{m}_i \in \mathbb{R}^{p\times 1}$ é dada por

$$d(\mathbf{x}_n|\mathbf{m}_i,\mathbf{\Sigma}_i) = (\mathbf{x}_n - \mathbf{m}_i)^T\, \mathbf{\Sigma}_i^{-1}\, (\mathbf{x}_n - \mathbf{m}_i)\ , \qquad (1)$$

em que a distância $d(\mathbf{x}_n|\mathbf{m}_i,\mathbf{\Sigma}_i)$ é utilizada como função discriminante $g_i(\mathbf{x}_n)$ da classe i. As classes podem ser ranqueadas (*i.e.* ordenadas) em função dos valores de suas respectivas funções discriminantes de forma que

$$\mathbf{x}_n \in \text{classe } i \quad \text{se} \quad g_i(\mathbf{x}_n) < g_j(\mathbf{x}_n),\ \forall i \neq j\,. \qquad (2)$$

A partir de (1), diferentes variações surgem para diferentes suposições sobre $\mathbf{\Sigma}_i$ a fim contornar a possível singularidade de $\mathbf{\Sigma}_i^{-1}$. A seguir, são abordados os seguintes classificadores baseados na distância de Mahalanobis: classificador quadrático gaussiano (QG) sem regularização, QG com regularização de Tikhonov (Tikho), QG com matriz de covariância agregada (Pool), QG com regularização de Friedman (Fried) e QG com matriz de covariância diagonalizada (Diag).

A. Classificador Quadrático Gaussiano sem regularização

O classificador QG sem regularização supõe que cada classe i possui uma matriz $\mathbf{\Sigma}_i$ independente, o que leva a um discriminante quadrático do tipo $g_i(\mathbf{x}_n) = \mathbf{x}_n^T\mathbf{A}_i\mathbf{x}_n + \mathbf{b}_i^T\mathbf{x}_n + c_i$, pois o desenvolvimento de (1) resulta em

$$g_i(\mathbf{x}_n) = \mathbf{x}_n^T\mathbf{\Sigma}_i^{-1}\mathbf{x}_n - 2\mathbf{m}_i^T\mathbf{\Sigma}_i^{-1}\mathbf{x}_n + \mathbf{m}_i^T\mathbf{\Sigma}_i^{-1}\mathbf{m}_i\,, \qquad (3)$$

em que $\mathbf{A}_i = \mathbf{\Sigma}_i^{-1}$, $\mathbf{b}_i = -2\mathbf{\Sigma}_i^{-1}\mathbf{m}_i$ e $c_i = \mathbf{m}_i^T\mathbf{\Sigma}_i^{-1}\mathbf{m}_i$.

B. Classificador QG com regularização de Tikhonov

O desempenho do classificador quadrático depende fortemente da invertibilidade da matriz de covariância de cada classe, que está intimamente associada ao seu posto ser completo. Uma maneira de tornar a matriz de covariância inversível, com posto completo, é aplicar a regularização de Tikhonov [10] sobre a matriz de covariância $\mathbf{\Sigma}_i$, definida como

$$\mathbf{\Sigma}_i(\lambda) = \mathbf{\Sigma}_i + \lambda\mathbf{I}_p \quad , \quad 0 \leq \lambda \ll 1\,, \qquad (4)$$

em que p é o número de atributos e λ é o coeficiente de regularização. O efeito prático da regularização de Tikhonov é adicionar um pequeno valor λ aos elementos da diagonal principal de $\mathbf{\Sigma}_i$, tornando suas colunas (ou, equivalentemente, suas linhas) vetores não colineares. Consequentemente, suas linhas tornam-se linearmente independentes e seu posto completo. Seu discriminante também é quadrático.

C. Classificador QG com matriz de covariância agregada

Outra tentativa de tornar a matriz de covariância inversível é utilizar uma única matriz agregada para as classes, como

$$\mathbf{\Sigma}_{pool} = \sum_i^K \frac{N_i}{N}\mathbf{\Sigma}_i\,, \qquad (5)$$

em que K é o número de classes, N_i é a quantidade de amostras da classe i e N é o número total de amostras. Esta regularização pode ser útil para problemas com poucas amostras ou que possuem classes desbalanceadas, evitando problemas numéricos causados por matrizes inversa mal condicionadas [11]. Assim, o desenvolvimento de (1) resulta em

$$g_i(\mathbf{x}_n) = \mathbf{x}_n^T\mathbf{\Sigma}_{pool}^{-1}\mathbf{x}_n - 2\mathbf{m}_i^T\mathbf{\Sigma}_{pool}^{-1}\mathbf{x}_n + \mathbf{m}_i^T\mathbf{\Sigma}_{pool}^{-1}\mathbf{m}_i\,. \qquad (6)$$

Sendo a matriz $\mathbf{\Sigma}_{pool}$ comum a todas as classes, o termo $\mathbf{x}_n^T\mathbf{\Sigma}_{pool}^{-1}\mathbf{x}_n$ é comum a toda função discriminante $g_i(\mathbf{x}_n)$, $1 \leq i \leq K$. Assim, este termo não influencia na classificação em (2), e (6) pode ser reescrita como

$$g_i(\mathbf{x}_n) = -2\mathbf{m}_i^T\mathbf{\Sigma}_{pool}^{-1}\mathbf{x}_n + \mathbf{m}_i^T\mathbf{\Sigma}_{pool}^{-1}\mathbf{m}_i\,, \qquad (7)$$

resultando no discriminante linear $g_i(\mathbf{x}_n) = \mathbf{b}_i^T\mathbf{x}_n + c_i$, com $\mathbf{b}_i = -2\mathbf{\Sigma}_{pool}^{-1}\mathbf{m}_i$ e $c_i = \mathbf{m}_i^T\mathbf{\Sigma}_{pool}^{-1}\mathbf{m}_i$.

D. Classificador QG com regularização de Friedman

Este método, proposto por Friedman [12], combina linearmente a matriz de covariância agregada com as matrizes de covariância das classes. Assim, esta combinação pode estimar melhor a estrutura das classes e, consequentemente, aprimorar a classificação [13]. A matriz resultante é dada por

$$\mathbf{\Sigma}_i(\beta) = \frac{(1-\beta)N_i\mathbf{\Sigma}_i + \beta\mathbf{\Sigma}_{pool}}{(1-\beta)N_i + \beta N} \quad , \quad 0 \leq \beta \leq 1\,. \qquad (8)$$

Nota-se que para os valores extremos de β,

$$\mathbf{\Sigma}_i(\beta) = \begin{cases} \mathbf{\Sigma}_i, & \text{se } \beta = 0 \\ \mathbf{\Sigma}_{pool}, & \text{se } \beta = 1 \end{cases}\,. \qquad (9)$$

Portanto, o discriminante do classificador com regularização de Friedman é quadrático para $0 \leq \beta < 1$, enquanto que para $\beta = 1$ resulta na variante do classificador QG com matriz de covariância agregada descrito na seção anterior.

E. Classificador QG com matriz de covariância diagonalizada

Este classificador assume que os atributos são descorrelacionados entre si e possuem variâncias diferentes. Assim, tem-se que a matriz de covariância $\mathbf{\Sigma}_i^* = \text{diag}(\mathbf{\Sigma}_i)$ é dada por

$$\mathbf{\Sigma}_i^* = \begin{bmatrix} \sigma_{i,1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{i,2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{i,p}^2 \end{bmatrix}\,, \qquad (10)$$

em que $\sigma_{i,j}^2$, $j = 1,\dots,p$, é a variância do j-ésimo atributo da classe i. A principal vantagem deste caso particular está no fato de sempre existir a inversa de $\mathbf{\Sigma}_i^*$ visto que esta matriz sempre tem posto completo, ou seja, os vetores que formam suas colunas (ou linhas) são mutualmente ortogonais entre si.

Vale ressaltar que esta é a mesma suposição feita pelo classificador *naive* Bayes gaussiano [14], que assume que os atributos são gaussianos e descorrelacionados. Deste modo, esta variante leva à seguinte função discriminante quadrática:

$$g_i(\mathbf{x}_n) = \mathbf{x}_n^T \boldsymbol{\Sigma}_i^* \mathbf{x}_n - 2\mathbf{m}_i^T \boldsymbol{\Sigma}_i^* \mathbf{x}_n + \mathbf{m}_i^T \boldsymbol{\Sigma}_i^* \mathbf{m}_i \,. \qquad (11)$$

É imporante salientar que os classificadores descritos nas Subseções II-A, II-B, II-D e II-E são não lineares, enquanto a variante descrita na Subseção II-C é linear.

III. Classificador linear de mínimos quadrados

Este classificador formula o problema de classificação como uma transformação linear $\mathbf{y}_n = \mathbf{W}\mathbf{x}_n$, em que o rótulo $\mathbf{y}_n \in \mathbb{R}^{K\times 1}$ é um vetor binário, de dimensão K e de norma unitária que identifica unicamente a classe de $\mathbf{x}_n$ entre as K existentes. Já $\mathbf{W} \in \mathbb{R}^{K\times p}$ é a matriz de pesos responsável pela transformação linear do vetor de atributos. Os vetores-rótulo $\mathbf{y}_n$ seguem a convenção *one-hot-encoding*, de forma que as classes são mutualmente exclusivas. Agregando todos os N vetores-rótulo $\mathbf{y}_n$ em uma matriz $\mathbf{Y} \in \mathbb{R}^{K\times N}$ e todos os N vetores de atributos $\mathbf{x}_n$ de treinamento em uma matriz $\mathbf{X} \in \mathbb{R}^{p\times N}$, obtém-se a versão matricial de $\mathbf{y}_n = \mathbf{W}\mathbf{x}_n$:

$$\mathbf{Y} \;=\; \mathbf{W}\mathbf{X}\,. \qquad (12)$$

O critério dos mínimos quadrados ordinários (MQO) consiste em minimizar o quadrado do erro de estimação $J_{\text{MQO}}(\hat{\mathbf{W}})$ cometido pela estimativa $\hat{\mathbf{W}}$ da matriz $\mathbf{W}$,

$$J_{\text{MQO}}(\hat{\mathbf{W}}) = \frac{1}{2}\text{Tr}\left\{ \left(\mathbf{Y} - \hat{\mathbf{W}}\mathbf{X}\right)^T \left(\mathbf{Y} - \hat{\mathbf{W}}\mathbf{X}\right) \right\}, \qquad (13)$$

em que Tr denota o *traço* de uma matriz. Pode-se mostrar que a estimativa de $\mathbf{W}$ que minimiza $J_{\text{MQO}}(\hat{\mathbf{W}})$ de forma ótima é

$$\hat{\mathbf{W}} = \mathbf{Y}\mathbf{X}^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1}. \qquad (14)$$

Finalmente, a classe predita para a amostra $\mathbf{x}_n$ é obtida pelo índice de maior componente do vetor $\hat{\mathbf{y}}_n = \hat{\mathbf{W}}\mathbf{x}_n$.

IV. Métodos de transformação nos dados

Nesta seção, são descritas duas transformações de dados utilizadas na classificação de padrões: a transformação de Box-Cox (BC) e a análise de componentes principais (PCA).

A. Box-Cox

Normalidade é uma importante suposição para muitas técnicas e métodos estatísticos. Porém, muitos atributos não seguem uma distribuição normal. Nestes casos, é possível realizar uma operação não linear, conhecida como transformação de Box-Cox [6] sobre um atributo qualquer, de forma que

$$x^* = \begin{cases} \dfrac{x^\gamma - 1}{\gamma}, & \text{se } 0 < \gamma < 1 \\ \ln x, & \text{se } \gamma = 0 \end{cases}, \qquad (15)$$

em que γ é um parâmetro a determinar. Com isso, as variáveis não gaussianas podem assumir uma distribuição com uma forma mais simétrica, mais próxima da normal. Testes de hipóteses podem ser executados para confirmar a *gaussianidade* da variável aleatória resultante x^*, tais como o teste de Kolmogorov-Smirnov e o de Shapiro-Wilk [15].

B. Análise de componentes principais

PCA é uma transformação linear que atua nos dados $\mathbf{x}$ para gerar um conjunto de dados cujos atributos são descorrelacionados, ou seja, em que não há correlação linear entre as componentes do novo vetor de atributos $\mathbf{z}$ [9]. Assim, a nova matriz de covariância é diagonal. A transformação dos dados é realizada de acordo com

$$\mathbf{z}_n = \mathbf{V}_\varphi \mathbf{x}_n \,, \qquad (16)$$

em que $\mathbf{V}_\varphi$ é uma matriz de ordem $\varphi \times p$ formada pelos autovetores $\mathbf{v}$ da matriz de covariância das amostras,

$$\mathbf{V}_\varphi = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \ \ldots \ \mid \mathbf{v}_\varphi]^T, \; 1 \le \varphi \le p \,. \qquad (17)$$

Os autovetores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_\varphi$ estão ordenados de acordo com a ordem decrescente dos seus autovalores da matriz de covariância do conjunto de dados original. Para $\varphi = p$, o novo vetor de atributos $\mathbf{z}_n \in \mathbb{R}^{p\times 1}$ possui mesma dimensão que o vetor de atributos $\mathbf{x}_n \in \mathbb{R}^{p\times 1}$ original. Porém, para $\varphi < p$, ocorre uma redução de dimensionalidade. Desta forma, a matriz de covariância dos dados transformados possui dimensões $\varphi \times \varphi$.

Por fim, com a aplicação do PCA e a escolha adequada de φ, pode-se descorrelacionar os dados e reduzir a dimensionalidade ao mesmo tempo que se preserva a informação relevante contida nos dados originais [9].

V. Metodologia

O conjunto de dados utilizado neste trabalho, disponível em [16], possui 2.126 amostras de cardiotocogramas com 21 atributos recomendados pela Federação Internacional de Ginecologia e Obstetrícia para melhor avaliar a saúde fetal. A rotulação dos cardiotocogramas foi realizada por três obstetras especialistas de acordo com o estado fetal (normal, suspeito ou patológico). O conjunto de dados é desbalanceado, com 1.655 (77,74%) das amostras pertencentes à classe "normal", 295 (13,88%) das amostras pertencentes à classe "suspeita" e 176 (8,28%) das amostras pertencentes à classe "patológica".

A abordagem utilizada neste trabalho consiste em aplicar as cinco variações descritas na Seção II ao conjunto de dados, além das transformações de Box-Cox e PCA, como ilustrado na Fig. 1. Os resultados são comparados com o classificador linear de mínimos quadrados.

Para treinar os classificadores, são utilizados 80% do conjunto de dados. O restante é utilizado para validar o classificador em um conjunto de teste. As matrizes de covariância $\boldsymbol{\Sigma}_i$ e os vetores média $\mathbf{m}_i$ são calculados apenas com os dados de treinamento. O mesmo ocorre com a estimativa da matriz de regressores $\hat{\mathbf{W}}$ dos mínimos quadrados e a matriz de autovetores $\mathbf{V}_\varphi$ do PCA.

São feitas 100 realizações independentes de cada um dos classificadores, em que são analisados os valores máximo, mínimo, médio, mediano e desvio-padrão da taxa de acerto (TA) do conjunto de teste, além do número de condicionamento recíproco (rcond) das matrizes de covariância.

Os classificadores que possuem hiperparâmetro(s) são submetidos a uma busca em grade para encontrar o(s) melhor(es)

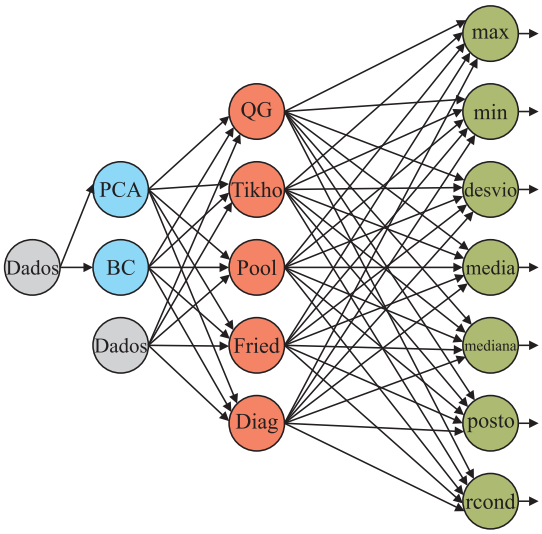


Figura 1: Abordagem utilizada no trabalho.

valor(es). A escolha das melhores configurações é realizada a partir do valor da mediana da TA. As melhores configurações de cada tipo de classificador são comparadas entre si. Destes, as realizações com maior e menor TA têm suas matrizes de confusão analisadas. Os experimentos foram realizados em um Intel Core i7-11800H (4,60 GHz), 16 GB de RAM, GPU NVIDIA RTX 3060, utilizando MATLAB R2018b.

VI. Resultados

Os resultados da análise comparativa de desempenho dos classificadores e das técnicas de transformação nos dados são apresentados na Tabela I e na Fig. 2. Percebe-se que entre os classificadores quadráticos (CQ), o de melhor acurácia utiliza a transformação de Box-Cox com $\gamma = 0{,}5$, com mediana de $86{,}59\%$ de taxa de acerto e desvio-padrão de $1{,}65\%$. Porém, as matrizes de covariância das classes não possuem posto completo e seus números de condicionamento são nulos.

Este problema é contornado pelo classificador quadrático com regularização de Tikhonov (Tikho), cujas matrizes de covariância das classes possuem posto completo e melhor número de condicionamento para todas as configurações (seja com Box-Cox, PCA ou nenhum dos dois). Dentre as diferentes configurações, a de melhor performance ocorre com o coeficiente de regularização $\lambda = 0{,}008$ e transformação de Box-Cox com $\gamma = 0{,}3$, alcançando uma mediana de $87{,}41\%$ e desvio-padrão de $1{,}65\%$. Esta configuração possui a melhor performance entre todos os classificadores analisados.

Entre as configurações do classificador com matriz de covariância agregada (Pool), a aplicação da transformada de Box-Cox com $\gamma = 0{,}8$ obteve melhor performance, com mediana de $84{,}94\%$ e desvio padrão de $1{,}57\%$. Porém, mesmo agregando as três matrizes de covariância em uma única, a matriz agregada não possui posto completo, exceto em algumas realizações ocasionais decorrentes da separação aleatória dos dados entre treinamento e teste.

Já para o classificador com regularização de Friedman (Fried), mesmo com a aplicação da transformação de Box-Cox ou PCA, a melhor configuração ocorre quando nenhuma transformação é utilizada e o coeficiente da regularização é $\beta = 0{,}9$, obtendo mediana de $85{,}18\%$ e desvio-padrão de $1{,}66\%$. Porém, a matriz de covariância regularizada não possui posto completo e seu número de condicionamento é zero.

O classificador QG com covariância diagonalizada apresentou melhora de desempenho com a transformação de Box-Cox com $\gamma = 0{,}9$, passando de $68{,}35\%$ para $73{,}65\%$ de mediana. Porém, uma vez que este classificador supõe que os atributos são descorrelacionados, a aplicação de PCA nos dados juntamente com a transformação de Box-Cox com $\gamma = 0{,}5$ resultou em um novo melhoramento: de $73{,}65\%$ para $76{,}23\%$ de mediana e $1{,}73\%$ de desvio-padrão. O coeficiente φ da análise de componentes principais é considerado aqui como um hiperparâmetro, e revelou que para o caso do classificador QG com covariância diagonalizada, a melhor classificação ocorre com $\varphi = 4$, o que representa $91{,}74\%$ da variância total dos dados originais. Assim sua matriz de covariância possui posto $= 4$ (completo) e melhores números de condicionamento recíproco (rcond) entre todas as configurações realizadas.

Um comportamento peculiar ocorre com a aplicação de PCA ao classificador QG com matriz de covariância diagonalizada. Nota-se na Fig. 2e, em um primeiro momento, que a TA aumenta conforme o aumento do número de autovetores utilizados (φ), atingindo um valor máximo. Porém, a partir deste ponto, o acréscimo de autovetores utilizados causa uma queda na TA. Isso pode ser melhor observado na Fig. 3, que mostra uma fatia do gráfico da Fig. 2e para $\gamma = 0{,}5$.

Este fato pode estar relacionado com o fenômeno de Hughes [17] (ou maldição da dimensionalidade), que está ligado à degradação da estimação da matriz de covariância decorrente da adição de atributos sem significância estatística para uma mesma quantidade de amostras. Como mencionado anteriormente na seção IV-B, $\mathbf{V}_\varphi$ é uma matriz composta pelos autovetores $\mathbf{v}$ da matriz de covariância em ordem decrescente de representabilidade dos dados. Os elementos da inversa da matriz de covariância após a transformação por PCA são apresentados na Tabela II, assim como a porcentagem da informação contida em cada autovetor.

Nota-se que a quantidade de informação contida nos autovetores cai drasticamente. A relação entre o menor e maior autovalor é extremamente elevada, resultando em um baixíssimo número de condicionamento recíproco[1]. A Fig. 3 também contém o gráfico do número de condicionamento recíproco da nova matriz de covariância de acordo com φ. Pode-se observar que a adição dos autovetores com informação estatisticamente irrelevantes pioram o condicionamento. Também é possível observar que $\varphi = 4$ possui a melhor classificação. É interessante notar a semelhança da curva de taxa de acerto presente na Fig. 3 (em marrom) com a curva decorrente do fenômeno de Hughes presente em Hughes (1968) [17].

[1]O número de condicionamento recíproco é próximo de 1 para matrizes bem-condicionadas e próximo de zero para mal-condicionadas.

Tabela I: Resultados dos classificadores.

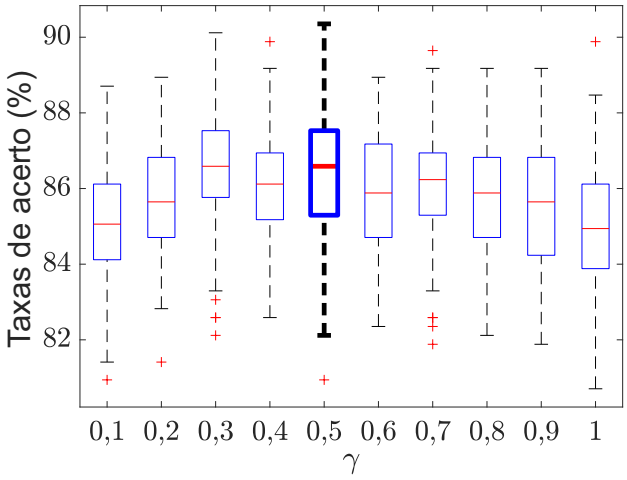
Classificador	hiperparâmetros	Média DP (%)	Mediana (%)	Mínimo Máximo (%)	POSTO classe 1	POSTO classe 2	POSTO classe 3	NÚMERO DE CONDICIONAMENTO classe 1	NÚMERO DE CONDICIONAMENTO classe 2	NÚMERO DE CONDICIONAMENTO classe 3
QG	———	85,25 1,60	85,18	80,71 89,65	19/21	19/21	20/21	0	0	$1{,}7 \cdot 10^{-18}$
QG + BC	$\gamma = 0{,}5$	86,34 1,65	86,59	80,94 90,35	19/21	19/21	20/21	0	0	$5{,}8 \cdot 10^{-18}$
QG + PCA	$\varphi = 9$ (99,78%)	82,33 1,55	82,12	78,59 86,82	9/9	9/9	9/9	$6{,}2 \cdot 10^{-3}$	$1{,}3 \cdot 10^{-3}$	$3{,}7 \cdot 10^{-4}$
Tikho	$\lambda = 0{,}004$	82,30 1,86	82,47	76,71 88,00	21/21	21/21	21/21	$2{,}7 \cdot 10^{-4}$	$2{,}2 \cdot 10^{-4}$	$7{,}1 \cdot 10^{-5}$
Tikho + BC	$\lambda = 0{,}008$ $\gamma = 0{,}3$	87,28 1,65	87,41	83,29 91,29	21/21	21/21	21/21	$2{,}8 \cdot 10^{-6}$	$2{,}3 \cdot 10^{-6}$	$7{,}8 \cdot 10^{-7}$
Tikho + PCA	$\lambda = 0{,}007$ $\varphi = 9$ (99,78%)	82,50 1,80	82,71	78,59 86,59	9/9	9/9	9/9	$4{,}0 \cdot 10^{-3}$	$1{,}3 \cdot 10^{-3}$	$3{,}5 \cdot 10^{-4}$
Pool	———	84,64 1,69	84,47	80,47 89,41	20/21			$3{,}5 \cdot 10^{-18}$		
Pool + BC	$\gamma = 0{,}8$	84,80 1,57	84,94	80,47 88,23	20/21			$2{,}6 \cdot 10^{-17}$		
Pool + PCA	$\varphi = 7$ (98,85%)	83,57 1,64	83,76	79,01 88,47	7/7			$8{,}8 \cdot 10^{-3}$		
Fried	$\beta = 0{,}9$	85,16 1,66	85,18	80,94 89,18	20/21	20/21	20/21	$3{,}3 \cdot 10^{-17}$	$3{,}3 \cdot 10^{-17}$	$3{,}3 \cdot 10^{-17}$
Fried + BC	$\beta = 1$ $\gamma = 0{,}9$	84,87 1,64	84,95	80,94 88,71	20/21			$6{,}8 \cdot 10^{-18}$		
Fried + PCA	$\beta = 0{,}8$ $\varphi = 7$ (98,85%)	83,63 1,51	83,53	80,23 87,01	7/7	7/7	7/7	$9{,}5 \cdot 10^{-3}$	$9{,}4 \cdot 10^{-3}$	$1{,}0 \cdot 10^{-2}$
Diag	———	60,42 21,28	68,35	6,12 74,35	21/21	20/21	21/21	$4{,}5 \cdot 10^{-13}$	0	$1{,}5 \cdot 10^{-11}$
Diag + BC	$\gamma = 0{,}9$	73,42 2,12	73,65	69,18 79,01	21/21	20/21	21/21	$4{,}5 \cdot 10^{-13}$	0	$1{,}2 \cdot 10^{-11}$
Diag + PCA	$\varphi = 4$ (91,74%)	64,50 2,67	64,23	58,59 72,23	4/4	4/4	4/4	0,1075	0,0996	0,0754
Diag + BC + PCA	$\gamma = 0{,}5$ $\varphi = 4$ (91,74%)	76,30 1,73	76,23	72,00 81,18	4/4	4/4	4/4	0,1324	0,0729	0,0234
MQO	———	85,60 1,57	85,65	81,18 88,70	20/21			$2{,}3 \cdot 10^{-30}$		

Por fim, os resultados destes classificadores são comparados aos do classificador linear de mínimos quadrados, que obteve mediana de $85{,}65\%$ e desvio-padrão de $1{,}57\%$. O posto da matriz de regressores não é completo, e o número de condicionamento recíproco é zero.

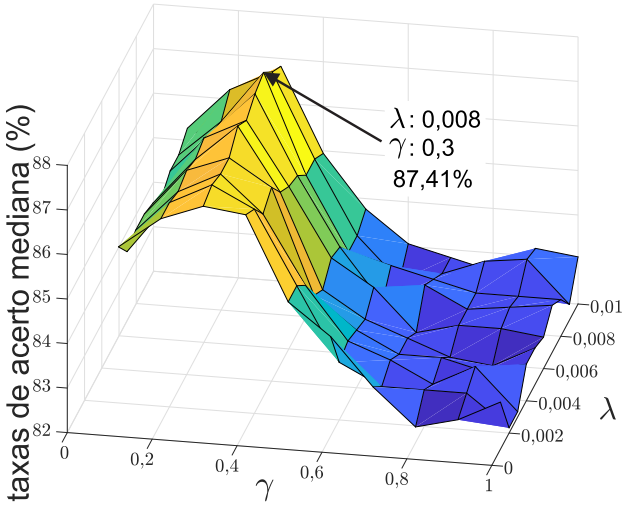
Na Fig. 4, são comparados os *boxplots* das melhores configurações de cada classificador. Pode-se perceber por esta Figura e pela Tabela I que o classificador quadrático com regularização de Tikhonov com $\lambda = 0{,}008$ e transformação de Box-Cox com $\gamma = 0{,}3$ possui o melhor desempenho, seguido pelo classificador quadrático (CQ) com transformação de Box-Cox com $\gamma = 0{,}5$. O classificador de mínimos quadrados ordinários (MQO) possui desempenho ligeiramente acima do desempenho do classificador com matriz de covariância agregada (Pool) com transformação de Box-Cox e do classificador com regularização de Friedman. O classificador QG com matriz de covariância diagonalizada, mesmo com a transformação de Box-Cox e aplicação da PCA, obtém o pior desempenho entre os classificadores analisados.

As matrizes de confusão dos melhores e do pior caso presentes na Fig. 4 são apresentadas na Fig. 5. Com $91{,}3\%$ de taxa de acerto, o melhor classificador quadrático com regularização de Tikhonov e transformação de Box-Cox com $\gamma = 0{,}3$ e $\lambda = 0{,}008$ erra apenas 37 das 425 amostras de teste. Já o pior classificador QG com covariância diagonalizada com transformação de Box-Cox e PCA, com $\gamma = 0{,}5$ e $\varphi = 4$ possui 72% de TA, errando 119 das 425 amostras de teste. Uma possível explicação para essa baixa TA em relação aos demais é o fato deste classificador desprezar as covariâncias entre classes, visto que todas as configurações do classificador QG com covariância diagonalizada (com Box-Cox e PCA ou não) possuem taxa de acerto bastante abaixo dos demais classificadores. É importante mencionar que este classificador pode apresentar desempenho superior em outros conjuntos de dados, dependendo das características estatísticas envolvidas. Por fim, o melhor classificador linear dos mínimos quadrados obtém 88,7% de TA, errando 48 das 425 amostras de teste.

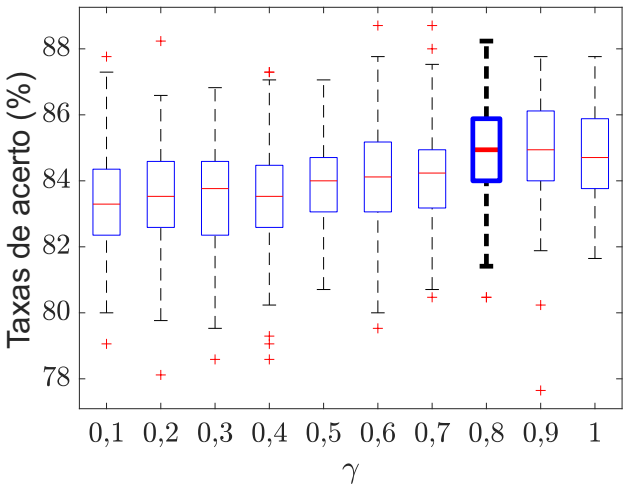
Considerando a classe "suspeito" como uma classe neutra, é possível analisar os verdadeiro-positivos (VP), verdadeiro-negativos (VN), falso-positivos (FP) e falso-negativos (FN) em relação à classificação das classes "normal" e "patológico". Desta forma, pode-se obter figuras de mérito como sensibili-



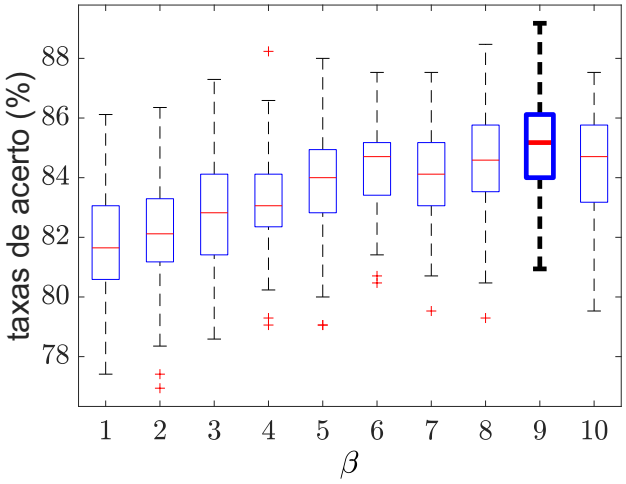
(a) QG + BC.



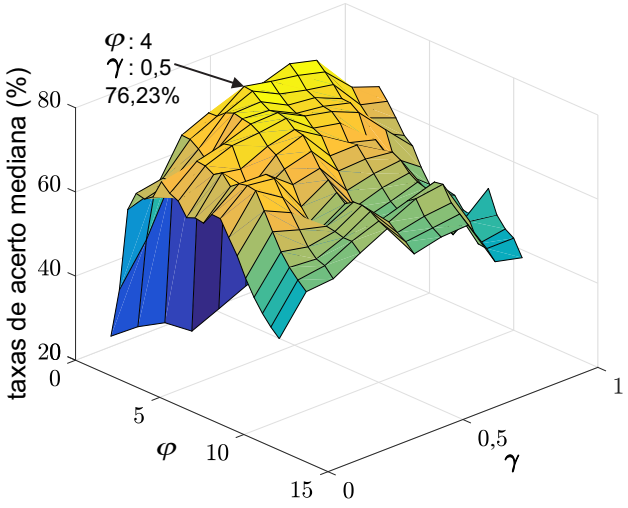
(b) Tikho + BC.



(c) Pool + BC.



(d) Fried.



(e) Diag + BC + PCA.

Figura 2: Detalhes das melhores configurações.

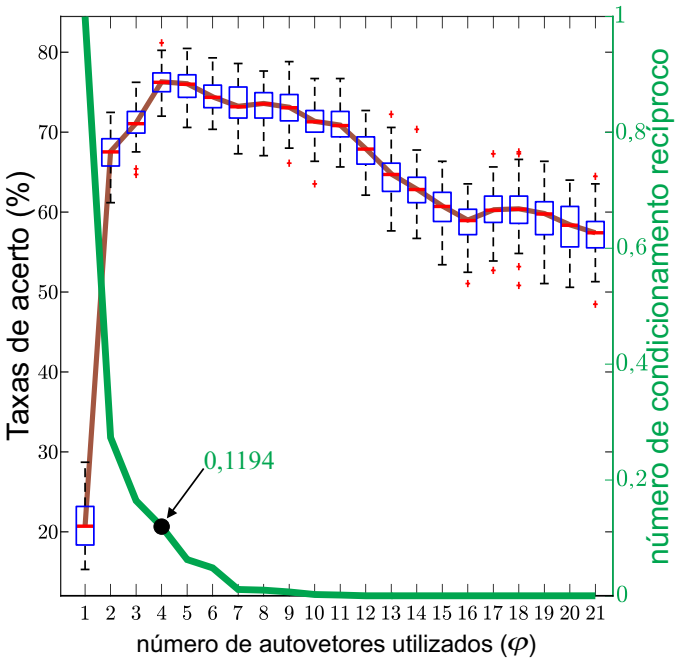


Figura 3: Diag + BC + PCA com $\gamma = 0{,}5$.

Tabela II: Valores de $1/\sigma_{i,i}^2$ após PCA e informação de cada autovetor.

i	$1/\sigma_{i,i}^2$	**info(%)**	i	$1/\sigma_{i,i}^2$	**info(%)**
1	$3{,}4 \cdot 10^{-4}$	58,9	**12**	2,3	$8{,}6 \cdot 10^{-3}$
2	$1{,}2 \cdot 10^{-3}$	16,1	**13**	3,57	$5{,}5 \cdot 10^{-3}$
3	$2{,}0 \cdot 10^{-3}$	9,7	**14**	7,6	$2{,}6 \cdot 10^{-3}$
4	$2{,}8 \cdot 10^{-3}$	7,0	**15**	$4{,}9 \cdot 10^{2}$	$4{,}0 \cdot 10^{-5}$
5	$5{,}4 \cdot 10^{-3}$	3,7	**16**	$1{,}4 \cdot 10^{5}$	$1{,}4 \cdot 10^{-7}$
6	$7{,}0 \cdot 10^{-3}$	2,8	**17**	$1{,}8 \cdot 10^{5}$	$1{,}1 \cdot 10^{-7}$
7	$3{,}1 \cdot 10^{-2}$	0,64	**18**	$3{,}4 \cdot 10^{5}$	$5{,}7 \cdot 10^{-8}$
8	$3{,}7 \cdot 10^{-2}$	0,53	**19**	$8{,}1 \cdot 10^{6}$	$2{,}4 \cdot 10^{-9}$
9	$5{,}1 \cdot 10^{-2}$	0,39	**20**	$2{,}9 \cdot 10^{8}$	$6{,}9 \cdot 10^{-11}$
10	$1{,}5 \cdot 10^{-1}$	0,13	**21**	$1{,}4 \cdot 10^{21}$	$2{,}0 \cdot 10^{-15}$
11	$2{,}7 \cdot 10^{-1}$	0,074			

dade, especificidade, precisão e F1-*score*, definidos como

$$\text{sensibilidade} = \frac{\text{VP}}{\text{VP} + \text{FN}} \tag{18}$$

$$\text{especificidade} = \frac{\text{VN}}{\text{VN} + \text{FP}} \tag{19}$$

$$\text{precisão} = \frac{\text{VP}}{\text{VP} + \text{FP}} \tag{20}$$

$$\text{F1-}score = \frac{2\text{VP}}{2\text{VP} + \text{FN} + \text{FP}}\,. \tag{21}$$

Estas figuras de mérito para os classificadores apresentados na Figura 5, considerando a classe "suspeito" como uma classe neutra, são apresentadas na Tabela III. Observa-se que, embora o classificador com regularização de Tikhonov e transformação de Box-Cox tenha sensibilidade menor que o classificador linear de mínimos quadrados, o mesmo possui maiores valores de especificidade, precisão e F1-*score*.

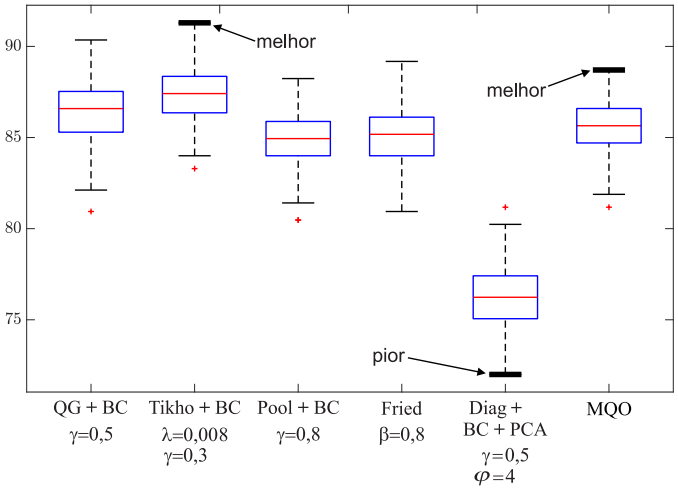


Figura 4: Melhores configurações.

Tabela III: Méritos dos classificadores da Figura 5.

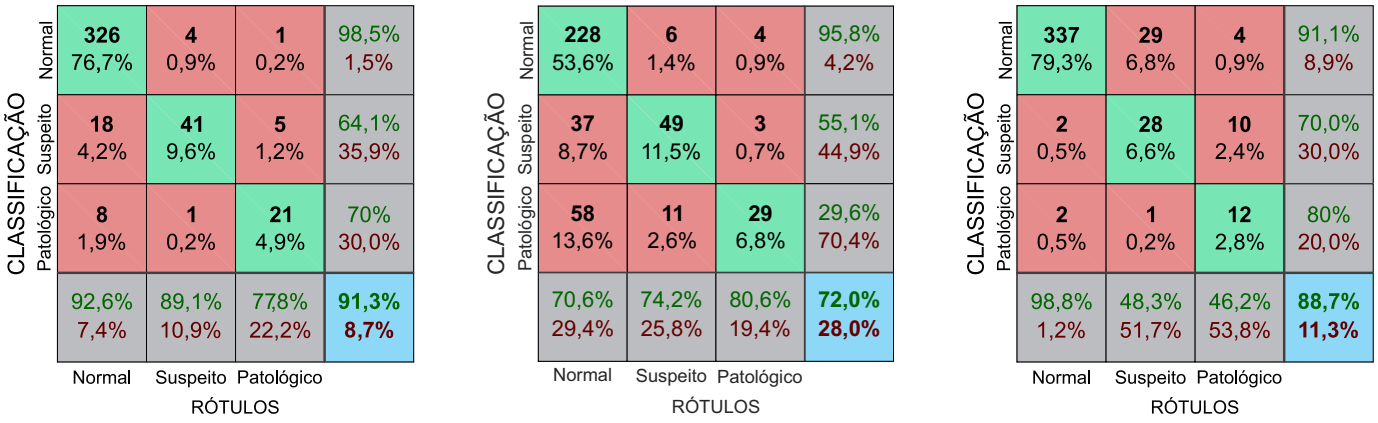
classificador	**mérito**	**valor (%)**
Tikho + BC $\gamma = 0{,}3$ $\lambda = 0{,}008$	sensibilidade	72,41
	especificidade	99,69
	precisão	95,45
	F1-*score*	82,35
Diag + BC + PCA $\gamma = 0{,}5$ $\varphi = 4$	sensibilidade	33,33
	especificidade	98,27
	precisão	87,88
	F1-*score*	48,33
MQO	sensibilidade	85,71
	especificidade	98,82
	precisão	75,00
	F1-*score*	80,00

VII. Conclusões

Este trabalho investigou de forma abrangente o desempenho de classificadores baseados em distância de Mahalanobis sobre dados de cardiotocograma. São utilizadas cinco variantes deste classificador: o classificador quadrático, com regularização de Tikhonov, com regularização de Friedman, com matriz de covariância agregada e classificador QG com matriz de covariância diagonalizada. Técnicas de transformação de Box-Cox e PCA também são utilizadas. Os resultados são comparados com o classificador clássico de mínimos quadrados linear.

A aplicação da transformação de Box-Cox está em quatro dos cinco melhores classificadores, evidenciando sua importância quando utilizada em conjunto com classificadores baseados na distância de Mahalanobis. O classificador quadrático com regularização de Tikhonov ($\lambda = 0{,}008$) e transformação de Box-Cox ($\gamma = 0{,}3$) destacou-se ao alcançar mediana de 87,41% na TA, com valor máximo de 91,3%, especificidade de 99,7%, precisão de 95,4% e posto completo na matriz de covariância. Estes resultados superam diversos classificadores utilizados em [3], como *linear* SVM (81,10%), *subspace discriminant* (81,10%), *logistic regression kernel* (83,30%) e algumas configurações de redes neurais (87,20%).

Embora [3] apresente classificadores com acurácia superior a 90%, como *boosted tree*, *bagged trees* e *cubic* SVM, estes



(a) Tikho + BC ($\gamma = 0{,}3$ e $\lambda = 0{,}008$). (b) Diag + BC + PCA ($\gamma = 0{,}5$ e $\varphi = 4$). (c) MQO.

Figura 5: Matrizes de confusão dos melhores e do pior classificador da Fig. 4.

classificadores podem apresentar certas limitações quando empregados em plataformas embarcadas (dispositivos portáteis, *e.g.*, [5]) com recursos computacionais limitados, especialmente no que se refere à capacidade de processamento e ao uso de memória. Nesse sentido, o classificador quadrático com regularização de Tikhonov e transformação de Box-Cox apresenta-se como uma excelente alterativa não linear, com boa acurácia e baixo custo computacional. É importante destacar que não é necessário o cálculo da inversa da matriz de covariância no dispositivo portátil, uma vez que a matriz já invertida previamente é um parâmetro do modelo.

É importante mencionar que a aplicação da PCA apresenta resultados relevantes, visto que com apenas 9 autovetores, é possível representar $99{,}78\%$ da variância dos dados. Essa redução permitiu que as novas matrizes de covariância fossem de posto completo e melhorou consideravelmente os números de condicionamento recíproco dessas matrizes.

Para tentar aumentar as taxas de acerto, pode-se aplicar a transformação de Box-Cox com um valor de γ distinto para cada um dos 21 atributos. Contudo, essa abordagem é computacionalmente onerosa, pois envolve busca em grade de 21 dimensões. Outra estratégia com potencial para aprimorar acurácia e robustez dos classificadores é a introdução de uma opção de rejeição [18], que evita classificar amostras com alto grau de incerteza. Por fim, recomenda-se a abordagem via classificação unária, uma vez que as classes são bastante desbalanceadas, sendo a classe *normal* a mais abundante.

Referências

[1] M. H. Tran, L. E. Rogalski, and A. Ahamed, "Enhancing fetal health diagnosis: A comprehensive evaluation of machine learning techniques on cardiotocogram data," in *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)*, p. 00599–00604, IEEE, 2025.

[2] I. R. de Vries, R. Melaet, I. A. Huijben, J. O. van Laar, R. D. Kok, S. G. Oei, R. J. van Sloun, and R. Vullings, "Conditional contrastive predictive coding for assessment of fetal health from the cardiotocogram," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 5, p. 3377–3386, 2025.

[3] L. Manavibool, P. Meepadung, N. Supmool, N. Vechpanich, P. Sapphaphab, O. Rinthon, and S. Pechprasarn, "Identifying critical factors in fetal cancer-related deaths using machine learning models and principal components analysis," in *2023 15th Biomedical Engineering International Conference (BMEiCON)*, p. 1–5, IEEE, 2023.

[4] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The mahalanobis distance," *Chemometrics and intelligent laboratory systems*, vol. 50, no. 1, pp. 1–18, 2000.

[5] BIC Diagnóstico, "Doppler fetal portátil S6 - BIC." https://cbemed.com.br/wp-content/uploads/2021/01/Doppler-Fetal-Portatil-S6-DO0101-2.pdf, 2021. Acesso em: 20 maio 2025.

[6] G. E. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 26, no. 2, pp. 211–243, 1964.

[7] Q. Wang, M. Xie, and F. Yang, "Early battery lifetime prediction based on statistical health features and Box-Cox transformation," *Journal of Energy Storage*, vol. 96, p. 112594, 2024.

[8] K. Pearson, "On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, p. 559–572, 1901.

[9] O. Dorabiala, A. Y. Aravkin, and J. N. Kutz, "Ensemble principal component analysis," *IEEE Access*, vol. 12, p. 6663–6671, 2024.

[10] D. Calvetti and E. Somersalo, "Distributed Tikhonov regularization for ill-posed inverse problems from a bayesian perspective," *Computational Optimization and Applications*, Mar. 2025.

[11] K. O. Olorede and W. B. Yahya, "Efficient multiclass prediction using sliced inverse regression with covariance hybridisation: an application to degrees classification problem," *International Journal of Data Science and Analytics*, 2025.

[12] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American statistical association*, vol. 84, no. 405, pp. 165–175, 1989.

[13] M. Mahadi, T. Ballal, M. Moinuddin, T. Y. Al-Naffouri, and U. M. Al-Saggaf, "Regularized linear discriminant analysis using a nonlinear covariance matrix estimator," *IEEE Transactions on Signal Processing*, vol. 72, p. 1049–1064, 2024.

[14] S. Naiem, A. E. Khedr, A. M. Idrees, and M. I. Marie, "Enhancing the efficiency of gaussian naïve Bayes machine learning classifier in the detection of DDOS in cloud computing," *IEEE Access*, vol. 11, p. 124597–124608, 2023.

[15] I. Mala, V. Sladek, and D. Bilkova, "Power comparisons of normality tests based on L-moments and classical tests," *Mathematics and Statistics*, vol. 9, no. 6, p. 994–1003, 2021.

[16] J. Marques de Sá, J. Bernardes, and D. Ayres de Campos, "Cardiotocography data set - machine learning repository." https://archive.ics.uci.edu/ml/datasets/Cardiotocography, 2010.

[17] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, p. 55–63, 1968.

[18] R. Gamelas Sousa, A. R. Rocha Neto, J. S. Cardoso, and G. A. Barreto, "Robust classification with reject option using the self-organizing map," *Neural Computing and Applications*, vol. 26, no. 7, p. 1603–1619, 2015.