

Avaliação de Métodos Redutores de Dimensionalidade em Sinais de ECG para a Classificação de Anomalias Cardíacas

Ericksson Lucas S. Roque, Diego A. Gomes, Adam Dreytom F. dos Santos

Resumo— Segundo a Organização Mundial de Saúde, os problemas cardíacos são um dos maiores causadores de mortes. Uma das formas de se fazer o diagnóstico desses problemas cardíacos é através da análise de sinais de ECG. Este trabalho apresenta um estudo do efeito combinado da taxa de amostragem e métodos redutores de dimensionalidade sobre o desempenho de classificadores de anomalias cardíacas. Os resultados mostram que a abordagem proposta alcança desempenhos superiores a trabalhos anteriores, chegando a acurácia de 99,94%.

Palavras-Chave— Eletrocardiograma, Redutores de Dimensionalidade, Classificação.

Abstract— According to the World Health Organization, heart problems are a major cause of death. One of the ways to diagnose these heart problems is through ECG signals. This work presents a study of the combined effect of the sampling rate and dimensionality-reducing methods on ECG signals, on the classifier performance. The results show that the proposed approach achieves greater performance than previous works, reaching 99.94% accuracy.

Keywords— Electrocardiogram, Dimensionality Reducers, Classification.

I. INTRODUÇÃO

O desenvolvimento da microeletrônica permitiu a miniaturização de equipamentos de aquisição de sinais fisiológicos, como o Eletrocardiograma (ECG), o qual pode ser utilizado no diagnóstico de doenças cardíacas, como arritmias cardíacas, acidente vascular cerebral isquêmico e hemorrágico [1]. Segundo a Organização Mundial da Saúde (OMS), morrem cerca de 17 milhões pessoas anualmente vítimas de doenças cardíacas, em contra partida, o ECG é capaz de diagnosticar mais de 90% de patologias cardíacas [2]. Em geral, o diagnóstico de tais patologias cardíacas, com o suporte do ECG, é feita por inspeção visual. Todavia, há a possibilidade de interação de novas tecnologias com o serviço da saúde, a fim de auxiliar os profissionais de saúde no diagnóstico dessas enfermidades, como por exemplo os algoritmos de inteligência artificial.

O estudo apresentado em [3] apresenta uma abordagem baseada em *Transformada Wavelet Discreta*, associada com a Análise de Componentes Principais (PCA), aplicados a sinais de ECG para a identificação de *Fibrilação Atrial*, utilizando *Redes Neurais Levenberg-Marquardt*. Os resultados mostram acurácia de 99,7%, mas as simulações foram feitas com apenas uma anomalia cardíaca. Já no trabalho em [4] é proposta a

utilização da Densidade Espectral de Potência (PSD, do inglês Power Spectral Density) para extração de características do ECG. O trabalho utilizou uma base de dados com 17 classes de arritmias cardíacas, aplicada aos algoritmos k-Vizinho Mais Próximos (kNN, do inglês k-Nearest Neighbor), Máquina de Vetores de Suporte (SVM, do inglês Support Vector Machine) e Perceptron Multicamada (MLP, do inglês Multilayer Perceptron) para a classificação, dentre os classificadores o melhor resultado encontrado foi uma acurácia de 98.86% utilizando o SVM.

Como visto acima, esses algoritmos podem alcançar desempenhos satisfatórios, mas exigem um alto custo computacional de sistemas embarcados, que pode ser um problema para tais dispositivos, por conta desses possuírem limitações em termos de processamentos, armazenamento e transmissão de dados. Em paralelo, vem se desenvolvendo a *telemedicina*, e uns dos principais serviços oferecidos é o diagnóstico remoto de anormalidades cardíaca. Futuramente, vários dispositivos serão conectados em rede para o diagnóstico remoto, entretanto, sem a compressão da dimensão dos dados será impossível realizar a transmissão e processamento de dados, além de causar congestionando nas redes, prejudicando outras aplicações [5]. Portanto, é necessário que se tomem precauções em relação a dimensão dos dados, afim de permitir manutenção de níveis de qualidade de serviço. Uma forma de se tomar esta medida é aplicar métodos de compressão de dados, fazendo assim com que a quantidade de dados enviada seja reduzida. Esse paradigma reduz a quantidade total de dados armazenada por aplicação, mantendo as características essenciais desses sinais, e ainda permitindo que mais aplicações possam utilizar este processamento.

Alguns trabalhos propuseram-se sanar este problema. Em [6], o autor utiliza o método *in-group Principal Component Analysis* (igPCA) para extração de características e redução de dimensões de sinais ECG para 16 classes diferentes. O método foi capaz de reduzir o sinal original de 192 para 10 amostras, e alcançou uma acurácia de 79%. Em [7], é proposto um estudo em redução de dimensões em sinais de ECG, utilizando *Independent Component Analysis* (ICA) e PCA como métodos de compressão, que rendeu uma acurácia de 98.67%, mas foram utilizadas apenas 4 classes para a classificação.

Neste trabalho, apresentamos um estudo do efeito combinado da taxa de amostragem e métodos redutores de dimensionalidade em sinais de ECG, sobre o desempenho de classificadores, para um total de 17 classes de arritmias cardíacas.

Ericksson Lucas, e-mail: ericksson.lucas@unifesspa.edu.br; Diego Gomes, e-mail: diagomes@unifesspa.edu.br; Adam Santos, e-mail: adamdreyton@unifesspa.edu.br. Faculdade de Computação e Engenharia Elétrica, Universidade Federal do Sul e Sudeste do Pará, Marabá-PA. Este trabalho foi financiado por pela FAPESPA e pela PROPIT-UNIFESSPA.

O ajuste na taxa de amostragem (reamostragem) foi aplicado como sendo uma primeira medida para redução da dimensão dos sinais, e em seguida aplicamos os métodos Análise da Componentes Principais (PCA do inglês, *Principal Component Analysis*) e Análise de Discriminantes Lineares (LDA, do inglês *Linear Discriminant Analysis*). Como classificadores, utilizamos o kNN, SVM, e o MLP.

II. METODOLOGIA

Para realizar o trabalho, toda a modelagem foi feita dentro da linguagem de programação *Python* [8], que possui a vantagens de dispor de forma gratuita uma grande variedade de bibliotecas para Ciência de Dados e Aprendizado de Máquina, além de ser bastante versátil.

A. Base de Dados

O banco de dados usado nesse estudo foi obtido de *MIT-BIH Arrhythmia Database* [9]. Esse banco de dados é composto por 48 registros de ECG de dois canais, 30 min de duração, obtido de 47 pessoas no Laboratório de Arritmia *BIH (Hospital Beth Israel)*, entre os anos de 1975 e 1979. A aquisição foi realizada com frequência de amostragem de 360 amostras/s por canal, com resolução de 11 bits por amostra. Neste trabalho, foram avaliadas 17 classes de patologias cardíacas diferentes com as quantidades de sinais para cada classe, Fibrilação atrial (135), Flutter atrial (20), Ritmo Idioventricular (10), Ritmo Sinusal (283), Ritmo Acelerado (195), Pré-Excitação Ventricular (45), Taquicardia supraventricular (13), Trigemínia Ventricular (13), Flutter ventricular (10), Batimento ventricular prematuro (66), Síndrome de Wolff-Parkinson-White (21), Contração Ventricular Prematura (133), Bigemínia Ventricular (55), Taquicardia ventricular (10), Fusion Beats (11), Bloqueio do ramo esquerdo (103) e Succinato desidrogenase (10).

B. Pré-Processamento e Extração de Características

A metodologia utilizada para o pré-processamento dos dados de entrada e extração de características foi baseado conforme [4], resumido na Fig. 1.

O processo inicia com a tomada dos dados a partir da base adotada, os quais se apresentam no domínio do tempo, com duração de 10 segundos, totalizando 3600 amostras por sinal. Em seguida, esses sinais foram filtrados, a fim de reduzir os efeitos de *aliasing* após o procedimento de reamostragem [10]. A etapa de reamostragem foi aplicada com o objetivo de verificarmos qual o impacto na escolha da frequência de amostragem para a classificação. Neste trabalho os sinais foram reamostrados para as frequências de amostragem 180, 90 e 45 amostras/s, mas também usamos os sinais na frequência de amostragem original. Em seguida, normalizamos os sinais utilizando a técnica de normalização *z-score*, que transformam os dados para uma distribuição com média 0 e variância igual a 1 [11].

A próxima etapa foi o mapeamento dos sinais para o domínio da frequência, através da Densidade Espectral de Potência (PSD, do inglês *Power Spectral Density*) [12], o qual foi implementado através do método de Welch [13]. A vantagem

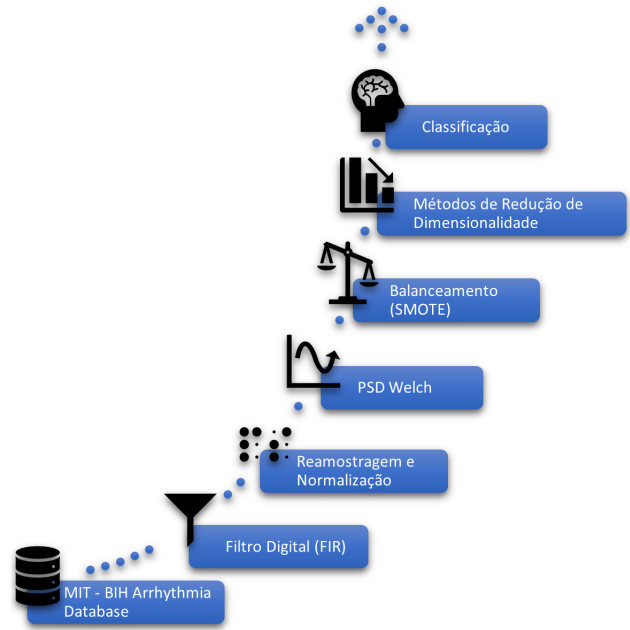


Fig. 1. Processos para extração de características.

de se aplicar esta modelagem, é que a partir da tomada das médias dos vários periodogramas, o sinal obtido ao final do processo é representado por um conjunto de valores menor que o original. Além da extração da PSD, utilizamos o método *Synthetic Minority Oversampling Technique* (SMOTE) para balancearmos a quantidade de amostras entre as classes. Esse passo é importante quando se trabalha com classes com grande disparidade na quantidade de amostras, em que a criação de amostras sintéticas tem sido proposta como um bom meio de aumentar a sensibilidade de um classificador para classes minoritárias, e evitar o *overfitting* das classes majoritárias [14]. Depois dessa etapa, a quantidade total de sinais foi de 1000 para 4811, pois este método atua aumentando a quantidade de sinais nas classes minoritárias.

C. Métodos para Redução de Dimensionalidade

Após o processo de extração de características, os dados são levados aos processos de redução de dimensão, que serão detalhados nesta seção.

1) *Análise da Componentes Principais*: O objetivo do PCA, é reduzir a dimensionalidade de um conjunto de dados, mantendo a variabilidade. Tal característica é alcançada ao obtermos um novo conjunto de dados, as Componentes Principais (PCs), os quais são decorrelacionados entre si, e são ordenados de tal forma que poucos desses conseguem reter a maior parte da variabilidade dos dados originais [15].

Resumidamente, o processo consiste em obter os autovetores λ da matriz de covariância $\Sigma = E[\mathbf{x}\mathbf{x}^T]$, onde \mathbf{x} representa o conjunto de sinais a ser processados, os quais tem dimensão $N \times 1$. Em seguida são selecionados K autovetores que retêm a porcentagem da variância desejada, para formarem então uma matriz $\Lambda = [\lambda_1, \dots, \lambda_K]$, a qual pode então ser

aplicada na transformação dos dados originais como em

$$\hat{\mathbf{x}} = \mathbf{\Lambda}^T \mathbf{x}. \quad (1)$$

A redução de dimensões nesse método ocorre pelo fato que após a transformação acima, os vetores $\hat{\mathbf{x}}$ tem dimensão $K \times 1$, em que $K \ll N$ [15].

2) *Análise de Discriminantes Lineares*: Diferentemente do PCA, a Análise de Discriminantes Lineares (LDA, do inglês *Linear Discriminant Analysis*), tenta reduzir as dimensões do conjunto de dados, mantendo as informações que diferenciam as classes de saída [16]. O processo de execução do LDA é supervisionado, e as novas dimensões são selecionadas com base em sua capacidade de maximizar a distância entre as classes e minimizar a distância entre os pontos de dados em uma classe. Essas novas dimensões formam os discriminantes lineares do conjunto de características [17].

Os passos para a execução do LDA, são os seguintes:

- Dentro das classes, cálculo da matriz de dispersão, S_w :

$$S_w = \sum_{j=1}^C \sum_{i=1}^{L_j} (\mathbf{x}_i^j - \boldsymbol{\mu}_j)(\mathbf{x}_i^j - \boldsymbol{\mu}_j)^T, \quad (2)$$

onde $\boldsymbol{\mu}_j$ é a média da classe j , L_j é quantidade de observações na classe j , C é o número de classes, T representa a operação de transposição.

- Cálculo da matriz de dispersão entre as classes, S_b :

$$S_b = \sum_{j=1}^C (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^T, \quad (3)$$

onde, $\boldsymbol{\mu}$ representa a média de todas as classes;

- Cálculo dos autovetores da matriz de projeção:

$$\mathbf{W} = \text{eig}(\mathbf{S}_w^{-1} \mathbf{S}_b) \quad (4)$$

- Seleção dos D autovetores associados aos maiores autovalores, para então formarmos a matriz $\widehat{\mathbf{W}} = [w_1, \dots, w_D]$;

- Redução da dimensionalidade dos dados de forma semelhante ao que é feito em PCA:

$$\hat{\mathbf{x}} = \widehat{\mathbf{W}}^T \mathbf{x}. \quad (5)$$

III. CLASSIFICADORES

Nesta seção são apresentados os algoritmos utilizados para classificar as anomalias cardíacas.

A. k - Vizinhos Mais Próximos - kNN

O k -Vizinho Mais Próximos (kNN , do inglês *k-Nearest Neighbor*), é um dos classificadores mais simples de ser implementado. A ideia principal do kNN é determinar o rótulo de classificação de um conjunto de dados baseado nas amostras vizinhas advindas de um conjunto de treinamento. Então, são encontrados os k vizinhos mais próximos do padrão de treinamento [18].

B. Máquina de Vetores de Suporte - SVM

A Máquina de Vetores de Suporte (SVM, do inglês *Support Vector Machine*), é um modelo de aprendizado de máquina muito poderoso e versátil. O SVM possui diversos tipos de classificação como linear, polinomial, gaussiana e *sigmoid* [19].

O SVM funciona com suposição de que as classes nos dados de entrada são linearmente separáveis, e então há uma função do tipo hiperplano que realiza esta separação. A escolha do hiperplano ótimo, é feita identificando-se as realizações mais próximas entre classes diferentes, as quais são chamadas de *vetores de suporte*. Em termos conceituais, os vetores de suporte são aqueles pontos de dados que se encontram o mais próximos da superfície de decisão e são, portanto, os mais difíceis de classificar [20]. Então, otimizando-se a separação dos vetores de suporte, consequentemente se otimiza a separação das outras ocorrências das classes.

C. Perceptron Multicamadas - MLP

O Perceptron Multicamada (MLP, do inglês *Multilayer Perceptron*) é caracterizado por possuir em sua arquitetura uma ou mais camadas entre as camadas de entrada e saída [21]. Esta característica permite ao MLP ser utilizado na classificação de padrões que não são linearmente separáveis, e também para a aproximação de funções [22].

IV. RESULTADOS

Nesta seção são apresentados os resultados e discussões, tanto da etapa de pré-processamento como os da classificação. Para a realização das simulações utilizamos os dados descritos na Seção II-A, e um notebook Python, implementado na plataforma *Colaboratory* (Colab) [23], com *unidade de processamento gráfico* (GPU) habilitada.

Em relação a configuração dos métodos redutores de dimensionalidade, para o PCA foram utilizados três valores de variância acumulada, 99%, 95% e 90%. Justifica-se a escolha arbitrária de variâncias acumuladas diferentes, devido ao fato de não haver um consenso para o número de componentes principais de acordo com a variância selecionada [24]. Já para o LDA, a quantidade máxima de componentes válidas é de $(m - 1)$ onde m é a quantidade de classes [25], que nesse caso rende 16 componentes.

A Fig. 2, exibe a comparação entre o LDA e o PCA, em relação a proporção de variância acumulada, de acordo com o número de componentes e taxa de amostragem. Nesta figura as linhas tracejadas denotam o LDA e as contínuas o PCA. Nela podemos notar que o LDA é mais eficaz em reter variância com menos componentes, pois quando comparamos as curvas do LDA com as respectivas representantes no PCA (com a mesma taxa de amostragem), sempre observamos que as primeiras encontram-se mais a esquerda e mais acima, indicando que o LDA consegue acumular mais variabilidade com menos componentes. Por exemplo, nos casos com 360 amostras/s, com apenas 8 componentes o LDA é capaz de representar 90% da variância dos dados, enquanto que o PCA precisa de 12 componentes para manter esta variabilidade.

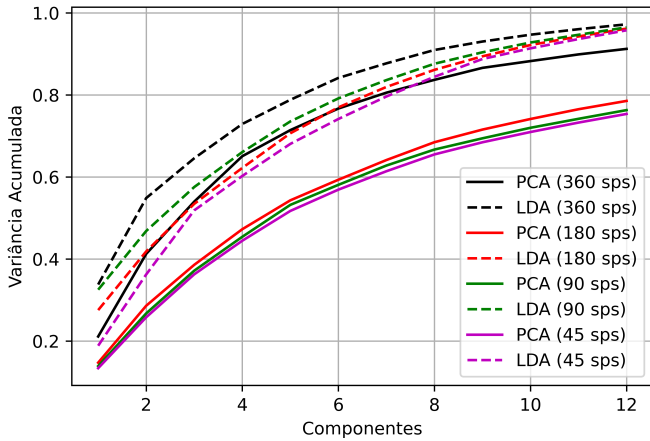


Fig. 2. Proporção de variância acumulada por número de componentes, para o LDA e o PCA, em 4 casos de taxa de amostragem. Nesta figura foram exibidos apenas poucos componentes, a fim de melhorar visualização das curvas. Adicionalmente, com o intuito de reduzir a área ocupada pelas legendas, substituímos “amostras/s” por “sps”.

Na Tabela I, é exibido o comparativo de capacidade de redução de dimensionalidade entre o LDA e PCA, de acordo com a taxa de amostragem (f_a). Nessa tabela percebe-se que o LDA sempre reduz a dimensão dos sinais para 16 amostras, como explicado anteriormente, independente da frequência de amostragem. Quando se trata do PCA, nota-se que quanto mais se diminui a frequência de amostragem, menor é a capacidade do algoritmo em reduzir dimensões. Ainda não temos uma explicação sólida para esta observação, e a trataremos em um trabalho futuro.

TABELA I

QUANTIDADE DE COMPONENTES MANTIDA DA PSD, POR MÉTODO DE REDUÇÃO E TAXA DE AMOSTRAGEM.

f_a (sps)	360	180	90	45
LDA	16	16	16	16
PCA (99%)	31	31	86	91
PCA (95%)	17	35	41	42
PCA (90%)	12	23	26	26

Depois que os dados foram processados, a próxima etapa consiste em lançá-los aos classificadores. A Tabela II exibe as configurações de cada algoritmo. Para o KNN utilizamos uma abordagem com valor de k dinâmico. Para o SVM foi escolhido o *kernel* Gaussiano. Já para o MLP foram utilizadas 2 camadas, e testamos configurações de 10 a 40 neurônios na camada escondida, e treinamento de 500 a 2000 épocas. A melhor configuração para o MLP foi utilizando 40 neurônios e 500 épocas. Cada um dos classificadores foi executado 30 vezes para cada combinação de método redutor de dimensão e taxa de amostragem, e então registramos o acurácia média, a fim de reduzirmos o impacto da inicialização aleatória dos coeficientes. Vale ressaltar também, que estes três classificadores foram implementados utilizando-se a biblioteca *scikit-learn*, do Python.

A Tabela III mostra os resultados de acurácia de teste para os três classificadores citados acima. Nessa tabela é possível verificar que o LDA rendeu os melhores resultados,

TABELA II
CONFIGURAÇÕES DOS CLASSIFICADORES.

Classificador	Parâmetros	Valores
KNN	k	Dinâmico
	Distância	Euclidiana
SVM	Kernel	Gaussiano
	Camadas	2
	Neurônios	40
MLP	Num. Épocas	500
	Solver	<i>rbf</i>
	Função de Ativação	<i>softmax</i>

em todos os classificadores avaliados, em que especificamente a situação com 360 amostras/s rendeu as melhor acurácia em todos os testes. Este resultado pode ser explicado explorando-se a característica do LDA de otimizar a separação entre as classes, que facilita a tarefa dos classificadores. Por outro lado, o PCA, apesar de também ter apresentado bons valores de acurácia, pode ter ficado aquém do LDA, porque o primeiro é baseado em um processo não-supervisionado, o que faz com que este nem sempre preserve a associação entre os dados e as classes [16].

Em relação ao efeito da taxa de amostragem, observa-se que para o LDA, a acurácia tende a reduzir quando a taxa de amostragem diminui. Isto pode estar associado ao fato de que o LDA não permite o ajuste da quantidade de componentes mantida, o que faz com que a perda de informação devido a redução na taxa de amostragem, renda uma piora no desempenho dos classificadores. Já para o PCA, o caso com 360 amostras/s rendeu o melhor desempenho, mas quando se observa a acurácia nas taxas de amostragem menores, não é possível determinar assintoticamente que houve redução de desempenho, pois os valores são bastante próximos. Para o PCA pode-se justificar este comportamento devido o acréscimo de componentes mantida, que tende a guardar mais informação essencial, e assim permitir aos classificadores não sofrerem tanto com a redução de taxa de amostragem, assim como observado na Tabela I.

Quando comparamos o desempenho dos classificadores, verificamos que em relação ao melhor caso (LDA com taxa de amostragem de 360 amostras/s), a diferença na acurácia entre eles é pequena: KNN alcançou 99,94%, o SVM 99,91%, e o MLP 99,88%. Se levarmos em conta apenas esses valores, concluímos que o KNN apresentou a melhor acurácia, mas devido proximidade entre os valores, seria preciso realizar alguns testes estatísticos para confirmarmos esta superioridade, o que será feito em trabalhos futuros. É importante ressaltar que a abordagem proposta neste trabalho rendeu um desempenho superior ao observado em estudos anteriores, como em [4], em que maior acurácia alcançada foi de 98,86%. Este ganho de desempenho pode estar relacionado a forma como o LDA condensa a informação em menos dimensões, e então reduz o espaço de busca e simplifica a separação das classes pelos classificadores.

Algo que merece destaque na análise destas tabelas, é que o LDA, mesmo no caso com taxa de amostragem de 90 amostras/s, apresentou resultado superior a todas as outras situações em que o PCA foi empregado. Este fato indica que é possível empregar uma abordagem de uso de ECG com

taxa de amostragem reduzida, de apenas 90 amostras/s, e com tamanho reduzido dos dados de entrada dos classificadores. Isto implicaria em um barateamento do equipamento devido o uso de um hardware mais simples; redução no tráfego de dados, devido a redução de características do sinal (no caso de um aplicação com inteligência artificial centralizada); e, redução no tempo tanto de treinamento como de classificação, devido a entrada com menor dimensão dos classificadores.

TABELA III

ACURÁCIA DE TESTE PARA O CLASSIFICADORES KNN, SVM E MLP.

f_a (sps)	360	180	90	45
KNN				
LDA	99,94%	99,8%	99,01%	97,27%
PCA (99%)	98,63%	97,18%	97,6%	97,07%
PCA (95%)	98,44%	97,44%	97,19%	97,17%
PCA (90%)	98,17%	97%	97,18%	97,21%
SVM				
LDA	99,91%	99,65%	99,01%	96,90%
PCA (99%)	97,13%	97,25%	97,19%	97,33%
PCA (95%)	95,07%	95,75%	95,17%	95,03%
PCA (90%)	93,48%	91,69%	93,38%	94,62%
MLP				
LDA	99,88%	99,74%	99,05%	96,89%
PCA (99%)	97,82%	96,07%	97,06%	97,30%
PCA (95%)	95,68%	96,14%	96,68%	96,07%
PCA (90%)	97,29%	96,04%	96,67%	96,90%

Finalmente, a fim de considerarmos a complexidade computacional dos métodos redutores de dimensionalidade, computamos o tempo de execução desses métodos, para o qual obtivemos: LDA - 1,35 s; PCA (99%) - 0,93 s; PCA (95%) - 0,92 s; e, PCA (90%) - 0,92 s. É importante ressaltar que estes valores foram obtidos com a base expandida, após a aplicação do SMOTE. Nota-se que nesta comparação o LDA demorou 45% a mais que o PCA. Por este parâmetro o PCA torna-se vantajoso, mas deve-se levar em questão o gargalo de processamento em todo o sistema, para se definir então o melhor algoritmo.

V. CONCLUSÕES

Nesse trabalho apresentamos um estudo sobre o impacto da aplicação dos redutores de dimensionalidade LDA e PCA, na classificação de anomalias cardíacas a partir de sinais de ECG, utilizando classificadores como KNN, SVM e MLP. Os resultados mostram que o LDA é bastante eficiente na compactação da variabilidade com menos componentes, quando comparado com o PCA. Além disso, as acurácia das classificações com os dados reduzidos foram melhores que os encontrados em outros trabalhos da literatura. Nestas simulações chegamos a acurácia de 99,94% com o LDA mantendo 16 componentes, e 98,63% com o PCA mantendo 31 componentes, ambos com taxa de amostragem de 360 amostras/s. Adicionalmente, observou-se que mesmo em taxas de amostragem menores, como 90 amostras/s, o uso do LDA rendeu melhores acurácias que melhor caso com o PCA. Em relação ao desempenho dos classificadores, os melhores resultados com o LDA ficaram próximos a $\approx 99,9\%$.

AGRADECIMENTOS

O autores agradecem a Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA) e a Pró-Reitoria de

Pós-Graduação, Pesquisa e Inovação Tecnológica (PROPIT-UNIFESSPA), pelo financiamento da pesquisa e por promover iniciação científica nas nossas universidades aos redores do Norte do Brasil, tal como a Faculdade de Computação e Engenharia Elétrica (FACEEL) e a Universidade Federal do Sul e Sudeste do Pará (UNIFESSPA).

REFERÊNCIAS

- [1] F. G. Teodoro, D. M. da Costa, S. M. Peres, and C. A. Lima, "Ecg biometric recognition based on fiducial features using support vector machines," in *X Workshop de Visão Computacional-WVC*, 2014.
- [2] O. M. da Saúde, "Doenças cardiovasculares."
- [3] A. Annavarapu, S. Borra, and P. Kora, *ECG Signal Dimensionality Reduction-Based Atrial Fibrillation Detection*, pp. 383–406. Cham: Springer International Publishing, 2018.
- [4] N. Pinho, D. Gomes, and A. dos Santos, "Classifying cardiac rhythms by means of digital signal processing and machine learning," *Journal of Communication and Information Systems*, vol. 35, pp. 25–33, Feb. 2020.
- [5] R. d. F. Dalvi, G. T. Zago, and R. V. Andreão, "Heartbeat classification system based on neural networks and dimensionality reduction," *Research on Biomedical Engineering*, vol. 32, no. 4, pp. 318–326, 2016.
- [6] A. Wosiak, "Principal component analysis based on data characteristics for dimensionality reduction of ecg recordings in arrhythmia classification," *Open Physics*, vol. 17, no. 1, pp. 489–496, 2019.
- [7] M. R. Islam, R. A. Bhuiyan, N. Ahmed, and M. R. Islam, "Pca and ica based hybrid dimension reduction model for cardiac arrhythmia disease diagnosis," in *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, pp. 1–7, IEEE, 2018.
- [8] G. van Rossum, "Python reference manual (3.8)," <https://docs.python.org/3.8/>, Acessado em 7 de Abril de 2020.
- [9] PhysioNet, *PTB Diagnostic ECG Database*, 2003 (Acessado em 3 de Abril de 2020).
- [10] P. S. Diniz, E. A. da Silva, and S. L. Netto, *Processamento Digital de Sinais: Projeto e Análise de Sistemas*. Bookman Editora, 2014.
- [11] A. R. Backes and J. J. d. M. S. Junior, *Introdução à visão computacional usando Matlab*. Alta Books Editora, 2019.
- [12] S. Haykin and B. Van Veen, *Signals and systems*. John Wiley & Sons, 2007.
- [13] C. L. Phillips, J. M. Parr, and E. A. Riskin, *Signals, systems, and transforms*. Prentice Hall Upper Saddle River, 2003.
- [14] D. M. Hawkins, "The problem of overfitting," *Journal of chemical information and computer sciences*, vol. 44, no. 1, pp. 1–12, 2004.
- [15] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.
- [16] H. R. Arabnia, K. Daimi, R. S. C. Soviany, L. Heilig, and K. Brussau, *Principles of Data Science*. Springer Nature Switzerland, 2020.
- [17] Ö. Toygar and A. Adnan, "Face recognition using pca, lda and ica approaches on colored images," *Istanbul University-Journal of Electrical & Electronics Engineering*, vol. 3, no. 1, pp. 735–743, 2003.
- [18] K. Fukunaga and P. M. Narendra, "A branch and bound algorithm for computing k-nearest neighbors," *IEEE transactions on computers*, vol. 100, no. 7, pp. 750–753, 1975.
- [19] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019.
- [20] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [21] T. W. Rauber, "Redes neurais artificiais," *Universidade Federal do Espírito Santo*, 2005.
- [22] K.-L. Du and M. Swamy, "Multilayer perceptrons: Architecture and error backpropagation," in *Neural Networks and Statistical Learning*, pp. 97–141, Springer, 2019.
- [23] Google, "Colaboratory."
- [24] W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng, "Discriminant analysis of principal components for face recognition," in *Face Recognition*, pp. 73–85, Springer, 1998.
- [25] J. M. Giron-Sierra, *Digital Signal Processing with Matlab Examples, Volume 2: Decomposition, Recovery, Data-Based Actions*. Springer, 2016.