

Classificação de Pacientes Diagnosticados com ELA através de Parâmetros Espetrais

Comparação entre classificadores automáticos e métodos visuais

Classification of Patients Diagnosed with ALS Through Spectral Parameters

Comparison between automatic classifiers and visual methods

Sofia Cotter, Hugo Cordeiro, Gonçalo Marques

Departamento de Engenharia Eletrónica, Telecomunicações e de Computadores

Instituto Superior de Engenharia de Lisboa

Instituto Politécnico de Lisboa

Lisboa, Portugal

a45107@alunos.isel.pt, hugo.cordeiro@isel.pt, goncalo.marques@isel.pt

Resumo — Este trabalho propõe métodos de discriminação automática entre pacientes diagnosticados com Esclerose Lateral Amiotrófica (ELA) bulbar e oradores saudáveis, a partir da extração de parâmetros do espectro de sinais de voz. Os testes com validação cruzada obtiveram uma taxa de acertos máxima num conjunto de teste de 80% e 82.4%, com os classificadores *Linear Discriminant Analysis* e *Support Vector Classifier*, respetivamente. Os resultados obtidos estão em linha com os propostos no estudo prévio quando considerado o declive espectral nas altas frequências.

Palavras Chave – *Análise espectral; Vozes patológicas; Esclerose lateral amiotrófica; Classificação; Avaliação;*

Abstract — This work presents methods of discrimination between patients diagnosed with bulbar Amyotrophic Lateral Sclerosis (ALS) and healthy speakers, based on the extraction of parameters from the voice signal spectrum. Testes using k-cross-validation achieved a maximum accuracy in the test fold of 80% and 82.4% with a *Linear Discriminant Analysis* classifier and a *Support Vector Classifier*, respectively. The results obtained are in line with those proposed in the previous study when considering the spectral slope at high frequencies.

Keywords – *Spectral analysis; Pathological Voices; Amyotrophic lateral sclerosis; Classification; Evaluation;*

I. INTRODUÇÃO

A Esclerose Lateral Amiotrófica, também conhecida como ELA, é uma doença degenerativa que provoca a destruição nos neurónios responsáveis pelo movimento dos músculos voluntários, conduzindo a uma paralisia progressiva que acaba por restringir ou mesmo impedir a realização de tarefas simples como andar, vestir-se, mastigar ou falar [1].

Na Esclerose Lateral Amiotrófica do tipo “bulbar”, os nervos que controlam os músculos da mastigação, deglutição e fala são afetados, tornando estas funções cada vez mais difíceis. A voz é, consequentemente, afetada e pode ganhar um tom nasalado [2].

Não existe cura para a doença, contudo a sua deteção prematura pode ajudar a melhorar a qualidade de vida dos pacientes. Há medicamentos capazes de atrasar a evolução da doença [1], logo seu rastreio é extremamente importante. O uso de métodos não invasivos, como a análise da voz ou do discurso, é muito mais cómodo para os pacientes e, portanto, preferível.

Em [3] propõem-se alguns parâmetros acústicas que podem ser usadas para discriminar oradores com ALS Bulbar, como por exemplo, o comprimento de onda, energia e o *pitch*. Nem todos os parâmetros tiveram o mesmo desempenho. Para os homens, uma única característica acústica pode separar os oradores saudáveis e não saudáveis com 79% de precisão. Para as mulheres, os 15 parâmetros principais conseguem dividir os oradores com 78% de precisão. Em [4], são extraídos parâmetros das vogais /a/ e /i/, consideradas as mais adequadas ao estudo porque a sua pronúncia requer um movimento considerável do músculo da língua (um dos sintomas de ELA Bulbar é atrofio da língua). Mostrou-se que um classificador *Linear Discriminant Analysis* atinge acertos de 88%. Em [5], para a discriminação de oradores com ELA, foram usados classificadores baseados numa rede neuronal convolucional e primeiras e segundas derivadas de um banco de características de filtragem Mel. O melhor resultado foi obtido com redes neurais convolucionais no domínio do tempo, com 90% de precisão.

Em [6], demonstrou-se que a partir da análise espectral do discurso de pacientes, com extração de 3 parâmetros principais, é possível diferenciar entre pacientes com ELA Bulbar e falantes saudáveis, sem necessidade de recorrer a classificadores altamente complexos. Todos os estudos anteriores analisaram fala, ou vogais extraídas da fala contínua.

Este trabalho continua a linha de investigação de parâmetros de discriminação derivados do declive do espectro do sinal de fala apresentados em [6]. O principal objetivo é explorar diferentes classificadores automáticos para distinção de oradores saudáveis e não saudáveis para um conjunto reduzido de oradores e comparar os resultados obtidos com os esperados através da análise gráfica em [6].

A próxima secção deste artigo apresenta o trabalho relacionado, a Secção 3 apresenta uma descrição da base de dados, a Secção 4 descreve o modelo proposto e a Secção 5 expõe os resultados e discussão dos mesmos. O artigo termina com as conclusões.

II. TRABALHO RELACIONADO

Em [6], consta-se que para a caracterização de vozes de pacientes diagnosticados com ELA Bulbar, os espectros de sinais de fala podem ser decompostos em 3 intervalos principais, e os pontos máximos do espectro atingidos em cada intervalo serão cruciais no cálculo dos tais parâmetros, que irão caracterizar cada orador.

Neste trabalho, os espectros do sinal de fala são estimados a partir de uma FFT de 4096 pontos, sendo a duração das tramas de 30ms e o andamento 10ms. Estes espectros são analisados em três bandas de energia, onde são caracterizados os picos de energia e bandas de energia.

O *LBST* (*Low Band Spectral Tilt*) é estimado através do declive entre duas bandas *FBME* (*First Band Maximum Energy*) e *SBME* (*Second Band Maximum Energy*), sendo que a primeira banda contém as duas primeiras harmónicas e a segunda banda é caracterizada pelo intervalo entre a terceira e a décima harmónica, onde se situa o típico primeiro formante da vogal /a/. O parâmetro *HBST* (*High Band Spectral Tilt*) é estimado entre o *SBME* e o *TBME* (*Third Band Maximum Energy*). As três bandas *FBME*, *SBME* e *TBME* são caracterizadas entre [0;420] Hz, [420;1260] Hz e [2100;5880] Hz, respetivamente.

Na análise espectral dos sinais de fala, verifica-se que tipicamente o *LBST* é positivo para oradores saudáveis e negativo para não saudáveis, independentemente do grau da doença. No geral, o *HBST* deverá apresentar menor declive na presença de vozes com maior degradação vocal, devido à maior influência de ruído. Estes declives podem ser observados nas Figuras 1 e 2, que representam uma trama do sinal de fala para um orador saudável e um não saudável, respetivamente.

O parâmetro *BSTA* (*Band Spectral Tilt Angle*) [8] é estimado através dos declives espectrais e representa o ângulo entre os mesmos, sendo tendencialmente superior a 180° no caso dos oradores saudáveis.

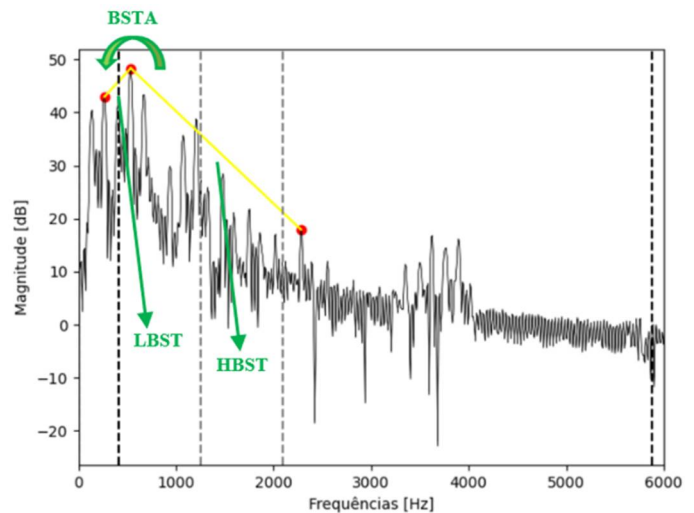


Figura 1 - Espectro de uma trama do sinal de fala de um orador masculino saudável

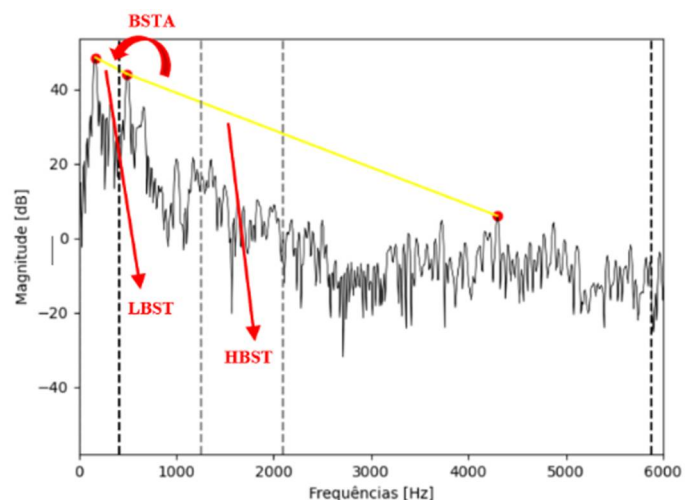


Figura 2 - Espectro de uma trama do sinal de fala de um orador masculino diagnosticado com ELA Bulbar

III. BASE DE DADOS

A base de dados [7] utilizada neste trabalho é constituída por 15 oradores diagnósticos com ELA Bulbar e 39 oradores saudáveis, totalizando 54 oradores. Porém, neste estudo, foram considerados apenas os oradores com idade a partir dos 40 anos, de forma a eliminar "outliers", pois tipicamente, ELA Bulbar é diagnosticada numa faixa etária acima dos 40 anos, portanto descartaram-se alguns oradores saudáveis, sendo que desta forma fica a existir um melhor balanceamento entre oradores saudáveis e não saudáveis. Os oradores diagnosticados com ELA Bulbar estão repartidos entre 6 oradores masculinos e 9 femininos, com uma média de idades de 57,7 anos. Os oradores saudáveis estão repartidos por 8 oradores masculinos e 10 femininos, com uma média de idades de 56.4 anos. Em suma, consideraram-se 15 oradores não saudáveis e 18 saudáveis.

O sinal analisado é a vogal /a/ sustentada, adquirido com vozeamento confortável. Os sinais foram adquiridos com um *smartphone* e *headset*, a uma frequência de amostragem de 44,1 kHz e 16 bit por amostra. A duração média dos sinais é de 4,1 segundos.

IV. MODELO PROPOSTO

A. Extração de Parâmetros

O sinal de fala é analisado com tramas de 30ms e um andamento de 10ms, obtendo para cada trama os parâmetros *LBST*, *HBST* e *BSTA*. A obtenção dos valores médios e desvios-padrão fez-se numa janela de 50 tramas, com andamento de 50 tramas, portanto sem sobreposição de janelas. Dependendo da duração do sinal áudio, os vários oradores terão diferentes quantidades de tramas. Tipicamente, os oradores não saudáveis apresentam menos tramas.

A classificação será então baseada nestes 3 parâmetros, sendo que cada um deles terá valores médios e valores de desvio-padrão, sendo cada vetor de 6 dimensões: [média *LBST*, média *HBST*, média *BSTA*, desvio-padrão *LBST*, desvio-padrão *HBST*, desvio-padrão *BSTA*].

B. Classificadores

Na classificação, houve um cuidado acrescido para garantir que os vetores de informação de um orador se iriam inserir em apenas um conjunto, ou treino ou teste. Uma vez que cada sinal de fala retorna diversas tramas com informação, ao efetuar a divisão dos conjuntos de treino e teste, garantiu-se que as tramas de um utilizador iriam ficar todas no mesmo conjunto. Deste modo, é garantido que não existem tramas do mesmo orador no conjunto de treino e no conjunto de teste.

Foram treinados 6 classificadores, recorrendo à biblioteca Sklearn do Python [9], nomeadamente, o *SGD* (*Stochastic Gradient Descent*), *GNB* (*Gaussian Naive Bayes*), *RF* (*Random Forest*), o método de Regressão Logística, *LR* (*Logistic Regression*), *Linear SVC* (*Support Vector Classifier*), e um simples classificador linear, *LDA* (*Linear Discriminant Analysis*).

SGD é uma abordagem simples, mas muito eficiente, para ajustar classificadores lineares e regressores sob funções de perda convexa, como máquinas de vetor de suporte (linear) e regressão logística. Neste trabalho, usou-se numa máquina de suporte vetorial.

O algoritmo *GNB* é um tipo de algoritmo de *Naive Bayes*, especialmente usado para dados que tenham valores contínuos e assume-se que os parâmetros sigam uma distribuição normal (ou gaussiana), sem covariância (as dimensões são independentes umas das outras).

O *RF* contém conjuntos de árvores de decisão para vários sub-conjuntos do *dataset* e usa uma média para melhorar a precisão preditiva e controlar o *over-fitting*.

O algoritmo *LR* é usado para prever a probabilidade de uma variável dependente binária, 1 (sucesso), 0 (insucesso, falha). É um método adequado se não houver muita correlação entre os dados e presença de “outliers”.

O *LDA* expressa uma variável dependente como sendo a combinação linear dos parâmetros. Isto é, a fronteira entre as classes é linear, não sendo, portanto, adequado para conjuntos de dados bastante complexos.

Por fim, o objetivo de um classificador *Linear SVC* é ajustar-se aos dados, de forma a retornar o melhor hiperplano que

conseguir que faça a separação dos dados nas suas classes. O *kernel* adequado é o *Radial Basis Function Kernel (RBF)*, pois permite desenhar a separação entre dados de diferentes classes com maior flexibilidade do que em apenas uma linha reta (*Linear Kernel*).

Os hiperparâmetros de um modelo são definidos antes do treino ocorrer e influenciam o desempenho do modelo. Para cada classificador e para cada conjunto de *features* consideradas, foram ajustados os hiperparâmetros com métodos de validação cruzada, dividindo o conjunto de treino em treino e validação, e sempre garantindo que as tramas de um mesmo orador iriam ficar em apenas um dos conjuntos.

C. Avaliação

Os resultados obtidos consistem na repetição de vários testes usando validação cruzada. Esta técnica é especialmente útil quando o conjunto de dados é reduzido. Na validação cruzada, o conjunto total de dados é repartido em subconjuntos mutuamente exclusivos e de igual dimensão e, a partir daí, um subconjunto é utilizado para teste e os restantes são utilizados para estimação dos parâmetros, fazendo-se o cálculo da precisão do modelo. Este processo é realizado *k* vezes, alternando de forma circular o subconjunto de teste, de forma que todos os dados tenham participado no teste.

A avaliação dos classificadores foi baseada na taxa de acertos ou *Accuracy (ACC)*, seguindo a Equação 1, e observando ainda o *True Positive Rate (TPR)* ou Sensitividade e o *True Negative Rate (TNR)* ou Especificidade, cujo cálculo é realizado a partir das Equações 2 e 3, respetivamente.

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$TNR = \frac{TN}{TN + FP} \quad (3)$$

V. RESULTADOS E DISCUSSÃO

Nos testes dos classificadores, a validação cruzada foi executada 1000 vezes e estimou-se a média de acertos por conjunto de teste (ou *fold*), de cada repetição, totalizando então 1000 taxas de acerto. No fim, estimou-se a média dos 1000 valores.

A. Resultados dos classificadores

A Tabela 1 contém os resultados para validação cruzada com 5 *folds*, e a Tabela 2 para validação cruzada com 4 *folds*, para todos os classificadores. A validação com 4 *folds* implica usar 25% dos dados para teste, enquanto que 5 *folds* geram conjuntos de teste com 20% da dimensionalidade total dos dados. Pela análise destas tabelas, verifica-se que dos parâmetros espectrais testados, o *HBST* é o que melhor discrimina os oradores na maioria dos classificadores, tanto usando 4 *folds* como 5 *folds*, na validação cruzada. O classificador *RF* foi o único a

demonstrar melhor performance com todos os parâmetros, contudo o resultado obtido não foi, no geral, o melhor. Conclui-se que não há grande vantagem em utilizar todos os parâmetros nos classificadores automáticos.

Tabela 1 - Taxa de acertos média (%) para 6 classificadores e 5 combinações de parâmetros espectrais, em 1000 repetições, com validação cruzada de 5 *folds*.

Param.	Classificadores					
	NBG	SGD	RF	LR	SVC	LDA
LBST, HBST, BSTA	61.7	67.6	71.1	68.0	70.2	69.9
LBST, HBST	63.5	69.5	68.7	68.8	70.7	70.6
LBST	51.0	57.2	62.0	61.0	64.5	63.1
HBST	71.8	65.9	67.8	73.7	72.1	75.0
BSTA	57.9	61.1	64.3	63.4	65.9	65.0

Tabela 2 - Taxa de acertos média (%) para 6 classificadores e 5 combinações de parâmetros espectrais, em 1000 repetições, com validação cruzada de 4 *folds*.

Param.	Classificadores					
	NBG	SGD	RF	LR	SVC	LDA
LBST, HBST, BSTA	61.1	67.9	68.5	69.1	69.0	69.6
LBST, HBST	63.5	65.8	67.9	68.7	68.9	70.3
LBST	50.8	61.5	61.1	61.5	63.3	62.1
HBST	70.6	70.6	67.4	64.9	72.8	74.4
BSTA	56.4	62.2	62.3	63.3	64.8	64.6

Considerando a análise das Tabelas 1 e 2, ao nível dos classificadores, verifica-se que os classificadores *LDA* e *SVC* obtêm os melhores resultados. Usando unicamente o parâmetro espectral *HBST*, e validação de 4 *folds*, o classificador linear *LDA*, atingiu uma média de 74.4% de acertos no conjunto de teste, enquanto o *SVC* atingiu 72.8%. Os mesmos classificadores obtiveram, respetivamente, 75% e 72.1% quando usados 5 *folds*.

No geral, os valores nas Tabelas 1 e 2 não divergem muito, pelo que se pode concluir que os resultados são consistentes, tanto usando 20% ou 25% dos dados para teste, tanto ao nível dos parâmetros avaliados como ao nível dos classificadores.

Tendo em conta os classificadores *LDA* e *SVC*, observe-se os histogramas que representam a distribuição das taxas de acerto para todas as repetições, Figuras 3 e 4.

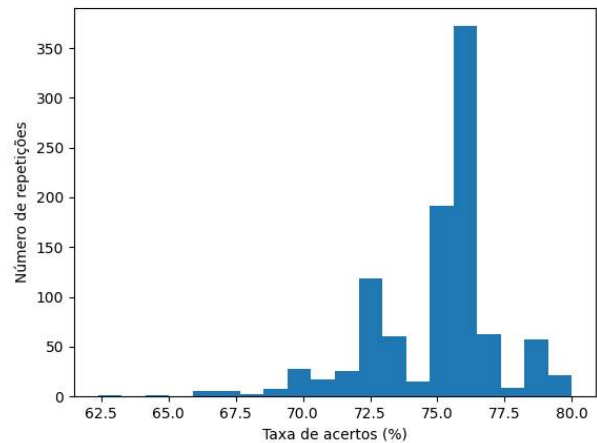


Figura 3 - Histograma de distribuição das taxas de acerto médias para 1000 repetições de validação cruzada com 5 *folds*, usando *LDA* e o parâmetro *HBST*

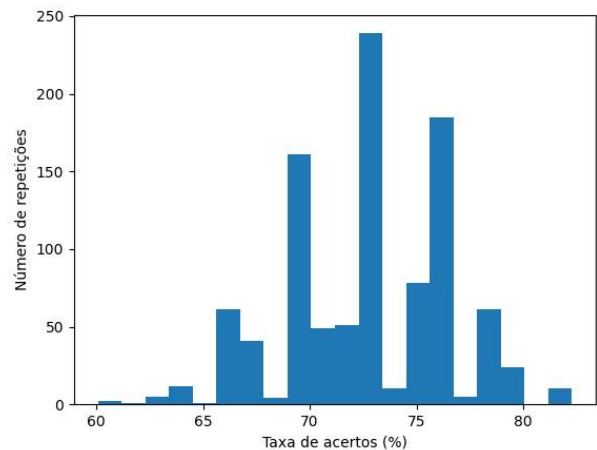


Figura 4 - Histograma de distribuição das taxas de acerto médias para 1000 repetições de validação cruzada com 4 *folds*, usando *SVC* e o parâmetro *HBST*.

Os histogramas revelam que os acertos se concentram acima dos 70% e, no caso do *LDA*, a maioria recai acima dos 75%. Não há acertos abaixo dos 60%, para o classificador *SVC*, e aqui verifica-se uma maior dispersão da classificação, nomeadamente, com valores mais baixos com cerca 64%, mas também com valores mais altos com cerca de 82%, Figura 4. No entanto, conclui-se, que devido à menor amplitude de resultados e tendo resultados de 76% em mais de 35% dos testes, que o discriminador linear é mais robusto.

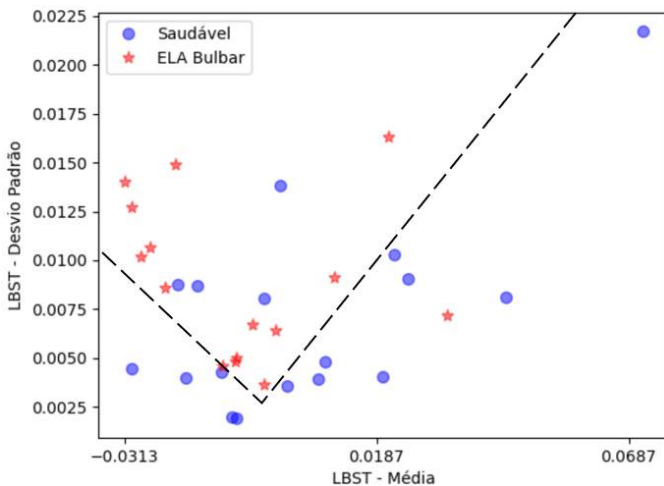
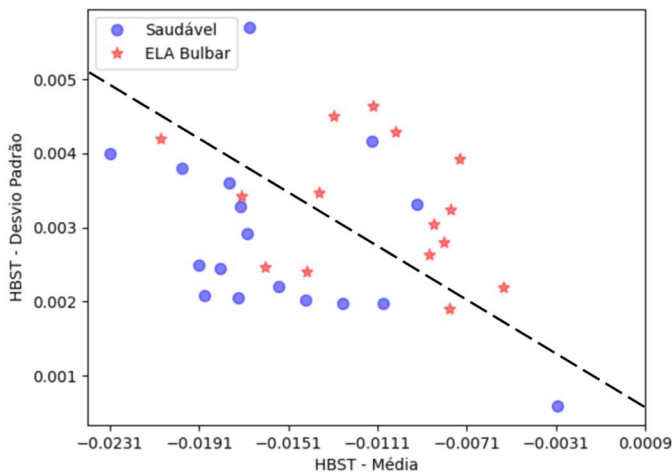
A Tabela 3 apresenta os valores de Sensitividade (*TPR*) e Especificidade (*TNR*), sendo que os mesmos estão acima de 0.5, indicando que os classificadores não são aleatórios. Nomeadamente, o *LDA* consegue ser ligeiramente mais preciso a classificar os verdadeiros positivos e negativos do que o *SVC*.

Tabela 3 - Valores de *TPR* e *TNR* para os classificadores *LDA* e *SVC*

	<i>LDA</i>	<i>SVC</i>
<i>TPR</i>	0.73	0.74
<i>TNR</i>	0.77	0.71

B. Comparação com o trabalho anterior

Como proposto em [6], a partir da observação das médias e desvios-padrão do *LBST* graficamente, consegue-se atingir uma taxa de acertos de 85%, traçando um limiar, Figura 5. Com o *HBST*, esta taxa é de 79%, Figura 6.

Figura 5 – Médias e desvios-padrão do *LBST*, retirado de [6]Figura 6 – Médias e desvios-padrão do *HBST*, retirado de [6]

Em [6], a classificação visual foi realizada usando todos os oradores como conjunto de treino, portanto, não houve divisão em treino e teste. Como tal, a comparação entre os resultados obtidos graficamente no estudo prévio foi feita com os valores de acerto máximo obtidos, nas várias repetições (os tais casos com o “melhor” conjunto de treino), e não os acertos médios, das Tabelas 1 e 2.

Com o parâmetro *LBST*, a classificação automática atingiu um acerto máximo de 76.7%, usando o classificador *SVC*, e 75.7%, usando o classificador *LDA*. Com o *HBST*, a classificação automática atingiu um acerto máximo de 82.4%, usando o *SVC*, e 80%, usando o *LDA*.

Contrariamente ao que se previa no estudo anterior, o *HBST* demonstrou resultados superiores ao *LBST*. A taxa de acertos obtida com o *LBST* em [6] também é superior ao acerto máximo atingido a partir de um classificador automático, em cerca de 10%. Contudo é importante notar que na análise visual não existiu a formulação de conjuntos de teste e treino, sendo que todos os dados foram separados por um limiar ótimo. A aplicação de conjuntos de treino e teste, na avaliação dos classificadores automáticos, e a aplicação da validação cruzada, implica necessariamente a rotação dos oradores. Note-se que estas variações, no caso do parâmetro *LBST*, não permitiram atingir o valor ótimo em nenhum dos classificadores. Por outro lado, a discriminação usando o parâmetro *HBST* conseguiu obter resultados semelhantes à inspeção visual.

VI. CONCLUSÕES

Este trabalho apresenta uma comparação de sistemas de classificação para oradores diagnosticados com ELA Bulbar. Foram considerados 6 classificadores. Após uma pesquisa pelos hiperparâmetros adequados a cada um, realizaram-se 1000 repetições de validação cruzada e estimou-se a taxa de acertos média, para melhor comparação entre os desempenhos dos classificadores.

Obtiveram-se acertos médios de 75% e 73% na discriminação de oradores, em 1000 repetições de validação cruzada, sendo que o uso de apenas 1 parâmetro espectral nos modelos, o *HBST*, conduziu aos melhores resultados. Conclui-se que os classificadores *LDA* e *SVC* são os mais ajustados à discriminação dos dados.

A comparação das taxas de acertos máximas obtidas para estes classificadores e a análise gráfica proposta pelo estudo prévio demonstra que o *HBST* cumpre as previsões de precisão propostas em [6]. Por outro lado, usando classificadores automáticos, a taxa de acerto máxima obtida para o *LBST* foi cerca de 10% menor que o resultado gráfico sugerido no estudo prévio.

Como trabalho futuro, é proposta a continuação do estudo do espectro do sinal de voz, e a procura de novos parâmetros capazes de discriminar indivíduos diagnosticados com ELA Bulbar e indivíduos saudáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Dr. Gonzalo Ramirez, Clínico Geral e Psicólogo, Esclerose Lateral Amiotrófica (ELA): o que é, sintomas e tratamento, novembro 2021, <https://www.tuasaude.com/esclerose-lateral-amiotrofica/>
- [2] Michael Rublin, Esclerose lateral amiotrófica (ELA) e outras doenças do neurônio motor (DNMs), MDCM, New York Presbyterian Hospital-Cornell Medical Center, Última modificação do conteúdo em dezembro 2020, <https://www.msmanuals.com/pt-pt/casa/dist%C3%BArbios-cerebrais-da-medula-espinhal-e-dos-nervos/nervo-perif%C3%A9rico-e-dist%C3%BArbios-relacionados/esclerose-lateral-amiot%C3%B3fica-ela-e-outras-doen%C3%A7as-do-neur%C3%B4nio-motor-dnms>
- [3] R. Norel, M. Pietrowicz, C. Agurto, S. Rishoni, G. Cecchi, “Detection of amyotrophic lateral sclerosis (ALS) via acoustic analysis”, *Proc. Annu.*

Conf. Int. Speech Commun. Assoc. INTERSPEECH. 2018-Septe (2018) pp. 377–381.

- [4] M. Vashkevich, E. Azarov, A. Petrovsky, Y. Rushkevich, “Features extraction for the automatic detection of ALS disease from acoustic speech signals”, in *Signal Processing - Algorithms, Architectures, Arrangements, and Applications Conference Proceedings, SPA*, 2018, vol. 2018-Septe, pp. 321–326.
- [5] K.H. An, M. Kim, K. Teplansky, J.R. Green, T.F. Campbell, Y. Yunusova, D. Heitzman, J. Wang, “Automatic early detection of amyotrophic lateral sclerosis from intelligible speech using convolutional neural networks”, *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*. (2018). pp. 1913–1917.
- [6] H. Cordeiro and C. Meneses, “Voice spectrum energy band and tilt analysis for Bulbar ALS screening,” *Procedia Comput. Sci.*, vol. 181, no. January, pp. 206–213, 2021.
- [7] M. Vashkevich, A. Petrovsky, and Y. Rushkevich, “Bulbar ALS Detection Based on Analysis of Voice Perturbation and Vibrato,” in *Signal Processing - Algorithms, Architectures, Arrangements, and Applications Conference Proceedings, SPA, 2019*, vol. 2019-Septe, pp. 267–272.
- [8] H. Cordeiro and C. Meneses, “Espectro de voz de pacientes diagnosticados com ELA bulbar,” in *16th Iberian Conference on Information Systems and Technologies (CISTI)*, 2021, June, pp. 1-6.
- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, et al. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, 12, 2825-2830.