

Reconhecimento de Dígitos de Medidores de Energia por meio da Voz no Contexto de um Aplicativo de Autoleitura**Digit Recognition of Energy Meters through Voice in the Context of an Authentication Application**

DOI:10.34115/basrv4n3-098

Recebimento dos originais: 20/04/2020

Aceitação para publicação: 21/06/2020

Paulo Renato Conceição Mendes

Graduado em Ciência da Computação pela Universidade Federal do Maranhão

Instituição: Universidade Federal do Maranhão

Endereço: Av. dos Portugueses, 1966, Bacanga, São Luís – MA, Brasil

E-mail: paulomendes@nca.ufma.br

Carlos Vinícios Martins Rocha

Graduando em Ciência da Computação pela Universidade Federal do Maranhão

Instituição: Universidade Federal do Maranhão

Endereço: Av. dos Portugueses, 1966, Bacanga, São Luís – MA, Brasil

E-mail: carlos.martins@nca.ufma.br

Anderson Matheus Passos Paiva

Mestre em Ciência da Computação pela Universidade Federal do Maranhão

Instituição: Universidade Federal do Maranhão

Endereço: Av. dos Portugueses, 1966, Bacanga, São Luís – MA, Brasil

E-mail: andersonpaiva@nca.ufma.br

José Eduardo de Souza Oliveira

Graduando em Ciência da Computação pela Universidade Federal do Maranhão

Instituição: Universidade Federal do Maranhão

Endereço: Av. dos Portugueses, 1966, Bacanga, São Luís – MA, Brasil

E-mail: joseduardo@nca.ufma.br

Italo Francyles Santos da Silva

Mestre em Ciência da Computação pela Universidade Federal do Maranhão

Instituição: Universidade Federal do Maranhão

Endereço: Av. dos Portugueses, 1966, Bacanga, São Luís – MA, Brasil

E-mail: francyles@nca.ufma.br

Anselmo Cardoso de Paiva

Doutor em Informática pela Pontifícia Universidade Católica do Rio de Janeiro

Instituição: Universidade Federal do Maranhão

Endereço: Av. dos Portugueses, 1966, Bacanga, São Luís – MA, Brasil

E-mail: paiva@nca.ufma.br

Simara Vieira da Rocha

Doutora em Engenharia de Eletricidade pela Universidade Federal do Maranhão
Instituição: Universidade Federal do Maranhão
Endereço: Av. dos Portugueses, 1966, Bacanga, São Luís – MA, Brasil
E-mail: simara.rocha@nca.ufma.br

Aristófanés Correa Silva

Doutor em Informática pela Pontifícia Universidade Católica do Rio de Janeiro
Instituição: Universidade Federal do Maranhão
Endereço: Av. dos Portugueses, 1966, Bacanga, São Luís – MA, Brasil
E-mail: ari@nca.ufma.br

Eliana Márcia G. Monteiro

Instituição: Equatorial Energia - MA
Endereço: Alameda A, 100, Quitandinha, São Luís - MA, Brasil
E-mail: eliana.monteiro@cemar-ma.com.br

Italo Fernandes S. da Silva

Instituição: Equatorial Energia - MA
Endereço: Alameda A, 100, Quitandinha, São Luís - MA, Brasil
E-mail: italo.silva@cemar-ma.com.br

Márcia Izabel A. da Silva

Instituição: Equatorial Energia - MA
Endereço: Alameda A, 100, Quitandinha, São Luís - MA, Brasil
E-mail: marcia.silva@cemar-ma.com.br

Jose Messias dos Santos

Instituição: Equatorial Energia - MA
Endereço: Alameda A, 100, Quitandinha, São Luís - MA, Brasil
E-mail: jose.messias@cemar-ma.com.br

RESUMO

A Agência Nacional de Energia Elétrica (ANEEL) destaca que perdas não-técnicas estão relacionadas a entraves no processo de leitura de consumo. Para a redução dessas falhas, uma alternativa factível e de menor custo seria a leitura realizada pelo próprio consumidor, denominada de autoleitura. Este processo leva em consideração o uso de plataformas digitais, através das quais o consumidor registraria e enviaria as informações de consumo. Uma etapa importante desse processo é o reconhecimento automático de dígitos de medidores por meio da voz. Este trabalho, portanto, propõe um método para a realização dessa tarefa, que utiliza processamento de áudio e inteligência computacional. Para a extração de características de áudio, utiliza-se *Mel-frequency Cepstral Coefficients* (MFCC) e *MelSpectrogram* de forma combinada. O método apresenta *Recall* de 94,74%; *Precision* de 94,91%; F1 score de 94,72% e 0,9419 de índice Kappa utilizando-se o classificador *Support Vector Machine* (SVM).

Palavras-chave: Consumo de Energia, Autoleitura, Reconhecimento, Processamento de Áudio, Aprendizado de Máquina, MFCC, *MelSpectrogram*, SVM

ABSTRACT

According to Brazilian Electricity Regulatory Agency (ANEEL), non-technical losses are related to consumption reading hindrances. In order to reduce these failures, a feasible and low cost alternative

would be the reading done by customers which is called self-reading. This process includes the use of digital platforms and the customer would register and send consumption information. An important step of this process is the digits automatic recognition in energy meters through voice. This work proposes a method to perform this task using audio processing and machine learning. Mel-frequency Cepstral Coefficients (MFCC) and MelSpectrogram are used for feature extraction. The proposed method obtains 94.74% Recall, 94.91% Precision, 94,72% F1 score and 0,9419 Kappa Index using Support Vector Machine (SVM) classifier.

Keywords: Energy Consumption, Self-reading, Recognition, Audio Processing, Machine Learning, MFCC, MelSpectrogram, SVM

1 INTRODUÇÃO

Estudos realizados pela Empresa de Pesquisa Energética (EPE) apontam que, em 2018, o consumo nacional de energia elétrica atingiu cerca de 472.200 GWh (*gigawatts*), sendo que 28% desse valor corresponde somente ao consumo residencial (ANEEL, 2019). Uma justificativa é dada pelo aumento no número de domicílios que, no Brasil, passará de 75 milhões em 2020 (Vidinich and Nery, 2009).

Segundo a Agência Nacional de Energia Elétrica (ANEEL), podem ocorrer no processo de distribuição de energia as chamadas perdas técnicas e não-técnicas (ANEEL, 2019). As perdas técnicas relacionam-se à quantidade de energia elétrica dissipada entre os suprimentos de energia da distribuidora e os pontos de entrega (unidades consumidoras ou distribuidoras supridas). As perdas chamadas de não-técnicas não estão associadas com as situações descritas anteriormente. Apresentam-se, por exemplo, pelos furtos de energia e o impedimento de acesso às unidades consumidoras. Erros de medição e de processamento de faturamento também são exemplos desse tipo de perda. Geralmente, esse tipo de perda é descoberto durante os processos de leitura e fiscalização do consumo de energia.

Para coletar as informações de consumo e imprimir a fatura dos clientes, a Companhia Energética do Maranhão (CEMAR) e as Centrais Elétricas do Pará S.A (CELPA) utilizam um dispositivo móvel e uma impressora portátil. Nesse processo, o leiturista precisa informar manualmente o consumo tal qual é mostrado no registro de energia, o que faz esse procedimento suscetível a erros, acarretando, portanto, em inconsistências no faturamento mensal.

Uma alternativa para evitar inconsistências seria a utilização de medidores inteligentes. No entanto, esta seria uma solução de custo elevado e de longo prazo, dado que envolve a substituição de todos os equipamentos de medição em vigência. Outra alternativa seria a leitura realizada pelo próprio consumidor. Este processo, chamado de autoleitura, apresenta-se como viável e de menor custo, e compreende o uso de plataformas digitais, como sites ou aplicativo desenvolvido para dispositivo móvel, como um meio por onde o consumidor registraria e enviaria suas informações de consumo.

Há muitas vantagens no processo de autoleitura. Ela promove ao consumidor um maior controle sobre o seu consumo de energia. Uma vez obtido o valor do consumo, a companhia pode emitir a fatura

ou fazer uma simulação da mesma, garantindo agilidade no processo de pagamento da fatura. A autoleitura também proporciona um estreitamento na relação de confiança entre as companhias e os consumidores, como também a diminuição dos custos das companhias e a redução dos erros de leituras, principalmente em áreas rurais ou de difícil acesso.

Nesse cenário, métodos computacionais capazes de realizar, de forma automática, a leitura do consumo de energia das unidades consumidoras podem auxiliar na aferição do consumo, tornando-o menos passível de erros. No entanto, para garantir facilidade e agilidade a esse processo, tais métodos precisam fornecer opções de leitura aos usuários, de forma que estes tenham suas preferências atendidas e não encontrem dificuldades para realizar a leitura. Entre essas opções, inclui-se, por exemplo, a autoleitura por meio da voz.

Um método computacional de autoleitura por meio da voz se dá, primeiramente, com a captura do áudio emitido pelo usuário, no qual é falada a sequência de dígitos apresentada pelo display do medidor de energia. Em seguida, cada dígito é segmentado e, por fim, identificado para fins de geração de fatura. Com base nessa proposta, este trabalho apresenta um método para o reconhecimento de dígitos por meio da voz, almejando contribuir com mais eficiência ao processo de leitura e também com a diminuição de perdas relacionadas ao faturamento.

O método proposto utiliza técnicas de processamento de áudio e inteligência computacional para a extração de características e o reconhecimento dos dígitos, sendo parte integrante de uma aplicação, em desenvolvimento, para a implantação do processo de autoleitura. Sendo assim, a funcionalidade de informar os dígitos referentes ao consumo usando a voz surge como uma opção a ser escolhida pelo usuário, atendendo a sua preferência.

Este trabalho está estruturado da seguinte forma: a Seção 2 apresenta trabalhos relacionados a este; a Seção 3 mostra o conjunto de áudios utilizado e como foi sua elaboração; a Seção 4 explica o método proposto; a Seção 5 discute os resultados obtidos por este estudo; e, a Seção 6 apresenta as conclusões.

2 TRABALHOS RELACIONADOS

Existem, na literatura, alguns trabalhos com propostas para o reconhecimento de dígitos em áudio, entretanto, aplicadas em outros contextos. Tabarabae et al. (1994) utilizam uma rede neural híbrida composta por *Perceptrons* e mapas de Kohonen aplicada no reconhecimento de dígitos em língua inglesa, alcançando 93,82% de acurácia geral. Abushariah et al. (2010) também desenvolveram um método para o reconhecimento de dígitos em inglês, no entanto, baseado em Hidden Markov Models (HMM) com a utilização de *Mel-frequency Cepstral Coefficients* (MFCC) para a extração de características. O método alcançou 88% de acurácia quando testado em ambiente ruidoso e 99% no teste em ambiente isolado.

O reconhecimento de dígitos em língua portuguesa é abordado no trabalho de Andrade Bresolin et al. (2008). Esses autores utilizam uma base proprietária, composta por vozes masculinas. Nos experimentos, é comparado o desempenho de vários modelos de transformada de Wavelet. O método alcança taxa de acerto máxima de 99% com o classificador *Support Vector Machine* (SVM). Silva et al. (2012) também abordam o reconhecimento em língua portuguesa, extraindo características com a técnica *Line Spectral Frequencies* (LSF).

Na literatura, existem também propostas para a realização automática da leitura do consumo de unidades consumidoras de energia elétrica, porém, baseadas em imagens (Quintanilha et al., 2017; Zhang et al., 2016). Nossa proposta, no entanto, aborda o reconhecimento de dígitos por áudio no contexto da autoleitura.

No contexto de dispositivos móveis, existem ferramentas disponibilizadas pela Google 1 e Apple 2 para o reconhecimento de voz, que podem ser integradas em aplicações móveis. Uma desvantagem é que, em versões mais antigas dos principais sistemas operacionais móveis em vigência, é necessário conectar-se à internet para utilizar esses recursos. Porém, em versões atuais, já existe o suporte *offline* essas ferramentas. Entretanto, são cobradas taxas para o uso dessas funcionalidades em aplicações desenvolvidas por terceiros. No contexto de um aplicativo de autoleitura, é desejável que a funcionalidade de reconhecimento por voz possa ser utilizada em uma quantidade abrangente de dispositivos. Além disso, não é garantido que o usuário sempre tenha acesso internet no momento em que realiza a leitura, como, por exemplo, aqueles que estão em áreas rurais ou em locais com baixa qualidade de sinal. Por causa desse conjunto de fatores, não seria possível utilizar as opções disponibilizadas pela Google e Apple. Logo, deu-se preferência para o desenvolvimento de um método de reconhecimento de dígitos por voz.

3 BASE DE ÁUDIOS

Foram pesquisadas bases de áudios que possuíssem gravações dos dígitos de 0 a 9 para que fossem utilizadas para o treinamento do módulo de reconhecimento de voz. Nessa busca, foram encontradas as bases *Spoltech Brazilian Portuguese* (CSLU) e *SID Dataset*, porém suas características não atenderam à proposta deste trabalho. A CSLU (Schramm et al., 2006) não possui informação a respeito de dígitos, além de ser uma base privada. E a SID (dos Santos, 2000) não contém áudios separados por dígito. Sendo assim, para o desenvolvimento deste trabalho, optou-se pela elaboração de um conjunto próprio de áudios. Para isso, foi criado um aplicativo com interface bem simples, como mostra a Figura 1.

Figura 1. Tela principal do aplicativo criado para a aquisição dos áudios.



Neste aplicativo, utilizou-se um layout em Grade para dispor os botões de gravação de cada dígito, um seletor de sexo, um botão para reproduzir o áudio antes de salvar, e um botão de salvar para armazenar o áudio no dispositivo.

O aplicativo foi distribuído para 53 pessoas. Sugeriu-se que cada uma gravasse em média dez áudios por dígito, a fim de que obtivéssemos uma quantidade significativa de exemplares. A base foi elaborada, portanto, contendo 11 classes: os dígitos de 0 a 9 e "meia". Esta última foi incluída por causa do hábito brasileiro de usar essa palavra para referir-se ao número 6.

Uma vez finalizado o período de aquisição, todos os áudios gravados foram analisados separadamente com o objetivo de descartar áudios incorretos. Em seguida, notou-se que nem todas as classes possuíam a mesma quantidade de áudios o que poderia acarretar em resultados tendenciosos. Então, foi feito um balanceamento para que todas as classes contivessem a mesma quantidade de áudios. A quantidade de áudios por classes pode ser vista na Tabela 1.

Tabela 1. Quantidade de áudios por classes.

Classe	Feminino	Masculino	Total
0	95	280	375
1	92	283	375
2	94	281	375
3	96	279	375
4	93	282	375
5	91	284	375
6	91	284	375
7	94	281	375
8	98	277	375
9	96	279	375
meia	92	283	375
Total	1032	3093	4125

4 MÉTODO PROPOSTO

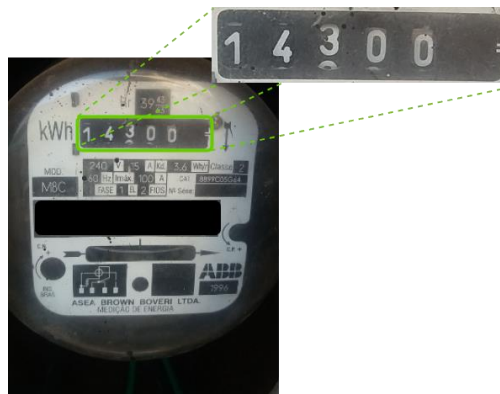
O método proposto para o reconhecimento de dígitos por áudio é composto pela sequência de etapas ilustradas pela Figura 2.

Figura 2. Etapas do Método Proposto



Primeiramente, captura-se o áudio no qual é falada a sequência de dígitos correspondentes à leitura do consumo, que é mostrada no display do medidor como ilustra a Figura 3.

Figura 3. Sequência de dígitos informada pelo medidor



Em seguida, este áudio sofre melhorias na etapa de pré-processamento e passa para a fase de extração de características, em que se aplicam as técnicas *Mel-frequency Cepstral Coefficients* (MFCC) e *MelSpectrogram*. Depois, realiza-se uma seleção de características para, enfim, iniciar a etapa de reconhecimento com o classificador *Support Vector Machine* (SVM).

4.1 PRÉ-PROCESSAMENTO

Na etapa de pré-processamento, utiliza-se a Normalização de Amplitude para deixar o volume mais uniforme em todo o áudio. De acordo com Pessoa et al. (2018), essa uniformização tende a melhorar tarefas de classificação de sons. Esse processo pode ser expresso pela Equação 1

$$Norm(A_i) = \frac{A_i}{\max(abs(A))} \quad (1 \leq i \leq N) \quad (1)$$

Onde A é o vetor com as amostras do sinal de áudio, A_i é a i -ésima amostra do sinal, $\max()$ é a função que retorna o maior valor do vetor, e $abs(A)$ retorna os valores absolutos de A .

4.2 EXTRAÇÃO DE CARACTERÍSTICAS

Para a extração de características de áudio, utiliza-se *Mel-frequency Cepstral Coefficients* (MFCC) (Davis and Mermelstein, 1980) e *MelSpectrogram* (S. Stevens et al., 1937).

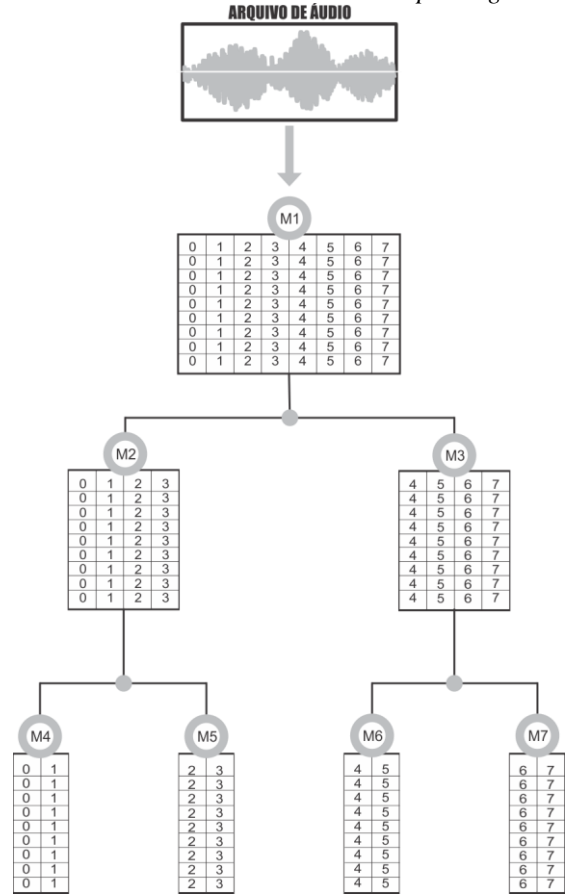
MFCCs são coeficientes calculados a partir do espectro de frequências de pequenas janelas do sinal de áudio, obtido por meio da *Fast Fourier Transform* (FFT) do sinal (da Silveira et al., 2017). Submete-se o espectro a um conjunto de filtros triangulares e igualmente espaçados na Escala Mel. Por fim, os coeficientes são obtidos com a aplicação da transformada discreta de cosseno sobre a saída gerada pelos filtros. Essa técnica gera matrizes com o número de linhas correspondente à quantidade de filtros, e o número de colunas de acordo com a duração do áudio.

O *MelSpectrogram* computa um espectrograma do áudio na Escala de Mel. Assim como o MFCC, o extrator *MelSpectrogram* retorna uma matriz, onde cada linha representa uma característica do espectrograma no decorrer do tempo. Portanto, para um dado áudio, são extraídas duas matrizes (MFCC e *MelSpectrogram*) contendo cada uma, 20 linhas.

Cada uma das linhas dessas matrizes é uma série temporal que varia de acordo com o tamanho do áudio, produzindo, dessa forma, um número variável de colunas. Para corrigir essa variação, é realizado o preenchimento com zero até que o número de colunas seja igual para cada linha. Por fim, para uniformizar a quantidade de características extraídas dos áudios, cada uma das matrizes (MFCC e *MelSpectrogram*) é dividida em seções como ilustrado pela Figura 4.

Uma matriz é dividida ao meio (em relação às colunas) e também em quatro partes, obtendo-se as seguintes seções: matriz completa (M1); primeira e segunda metade (M2 e M3); primeiro, segundo terceiro e quarto quartos (M4 a M7). Para cada seção e também para cada uma das 20 linhas, são calculadas as seguintes métricas: valor mínimo, valor máximo, média, mediana, variância e *Root Mean Square* (RMS).

Figura 4. Divisão das matrizes MFCC e MelSpectrogram em seções.



4.3 SELEÇÃO DE CARACTERÍSTICAS

As representações obtidas no passo anterior são submetidas a uma seleção de características com o objetivo de selecionar um conjunto de atributos menor, porém, bastante representativo. Isso contribui para a geração de modelos menores, minimizando impactos quanto ao armazenamento no contexto de um aplicativo móvel. Para determinar esse conjunto, utilizou-se o algoritmo *Binary Particle Swarm Optimization* (BPSO) (Kennedy and Eberhart, 1997).

O BPSO aborda o conceito de enxame de partículas. Para esse algoritmo, uma partícula consiste em uma máscara binária representando a presença ou a ausência de cada característica. Para cada uma das partículas é calculada uma função de custo. Neste trabalho, utiliza-se uma função de custo baseada no estudo feito por Vieira et al. (2013), a qual pode ser expressa pela Equação 2

$$f_i(u) = \alpha(1 - P) + (1 - \alpha)\left(1 - \frac{N_f}{N_t}\right) \quad (2)$$

Onde u é uma partícula; P é uma medida que avalia o desempenho obtido pelas características selecionadas; N_f é o número de características selecionadas; N_t é quantidade total de características; α é o peso atribuído sobre a medida de desempenho.

4.4 RECONHECIMENTO E MÉTRICAS DE AVALIAÇÃO

Após as etapas de extração e seleção de características, as instâncias são submetidas ao processo de reconhecimento, na qual é utilizado o classificador *Support Vector Machine* (SVM) (Cortes and Vapnik, 1995). Para avaliar o desempenho do classificador, são utilizadas neste trabalho as seguintes métricas: Recall (R), *Precision* (P), F1 score e o índice Kappa. *Recall*, *Precision* e F1 score são definidos, respectivamente, como mostram as equações 3, 4 e 5

$$R = \frac{VP}{VP + FN} \quad (3)$$

$$P = \frac{VP}{VP + FP} \quad (4)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (5)$$

Onde VP e FP são os verdadeiros e falsos positivos; e FN os falsos negativos.

Por fim, o índice Kappa mede o relacionamento entre concordância e causalidade, bem como o desacordo esperado, indicando o quão legítimas são as interpretações dos resultados. A Tabela 2 mostra como os valores de índice Kappa são categorizados, conforme estabelecido por Landis and Koch (1977).

Tabela 2. Classificação dos valores de Kappa.

Índice Kappa (K)	Qualidade
$K < 0,2$	Ruim
$0,2 \leq K < 0,4$	Razoável
$0,4 \leq K < 0,6$	Bom
$0,6 \leq K < 0,8$	Muito Bom
$K \geq 0,8$	Excelente

5 EXPERIMENTOS

No experimento para o reconhecimento de dígitos por meio da voz, foram feitos testes com e sem seleção de características para realizar uma análise comparativa e verificar a relevância desse processo para a obtenção de resultados promissores.

Como relatado na Seção 4.2, as matrizes *MelSpectrogram* e MFCC são divididas em setores. São calculadas 6 métricas para cada setor e para cada uma das 20 linhas das matrizes. Esse processo representa cada áudio em instâncias com 1680 características (6 métricas x 7 setores x 20 linhas x 2 matrizes). Ressalta-se que o tamanho da instância independe da duração do áudio. Para a execução do BPSO, é preciso definir um classificador e uma medida de desempenho, os quais, nestes experimentos, foram o *Logistic Regression* (Pedregosa et al., 2011), devido sua simplicidade e velocidade, e o Recall respectivamente. Também se utilizou $\alpha = 0,88$. Ao fim desta etapa, foram selecionadas 948 características.

Em ambos os testes, as instâncias foram submetidas ao classificador SVM em implementação disponível pela ferramenta EMGU 3. Os parâmetros do SVM foram estimados com o algoritmo Grid Search. A base foi dividida da seguinte forma: 90% para treinamento (3707 áudios, 337 por classe), e 10% para teste (418 áudios, 38 por classe).

Os resultados do teste sem seleção de características são vistos na Tabela 3. O classificador SVM apresentou considerável taxa de acerto para cada classe, com Recall médio de 93,54%. O valor obtido de *Precision* (93,67%) indica uma baixa quantidade de falsos positivos entre as instâncias classificadas.

Tabela 3. Resultados para o teste sem seleção.

Classe	Recall	Precision	F1	Kappa
0	92,11%	94,59%	93,33%	0,9267
1	92,11%	100,00%	95,89%	0,9549
2	100,00%	88,37%	93,83%	0,9316
3	89,47%	87,18%	88,31%	0,8712
4	97,37%	94,87%	96,10%	0,9570
5	97,37%	97,37%	97,37%	0,9710
6	78,95%	90,91%	84,51%	0,8307
7	89,47%	91,89%	90,67%	0,8974
8	94,74%	100,00%	97,30%	0,9703
9	97,37%	92,50%	94,87%	0,9434
Meia	100,00%	92,68%	96,20%	0,9580
Média	93,54%	93,67%	93,49%	0,9284

A Tabela 4 apresenta os resultados do teste com seleção de características. É possível observar a partir deles que o conjunto de atributos selecionados é bastante representativo, visto que os valores obtidos pelas medidas de desempenho são superiores aos do teste sem seleção.

Tabela 4. Resultados para o teste com seleção.

Classe	Recall	Precision	F1	Kappa
0	94,74%	94,74%	94,74%	0,9421
1	92,11%	97,22%	94,59%	0,9407
2	94,74%	100,00%	97,30%	0,9703
3	94,74%	92,31%	93,51%	0,9284
4	97,37%	90,24%	93,67%	0,9301
5	100%	95,00%	97,44%	0,9717
6	84,21%	96,97%	90,14%	0,8923
7	92,11%	94,59%	93,33%	0,9267
8	94,74%	100,00%	97,30%	0,9703
9	97,37%	90,24%	93,67%	0,9301
Meia	100,00%	92,68%	96,20%	0,9580
Média	94,74%	94,91%	94,72%	0,9419

Comparando-se os testes, observa-se que, em ambos, o reconhecimento do dígito 6 apresenta Recall inferior aos demais dígitos. Acredita-se que a questão fonética influencia diretamente nesse resultado, visto que esse dígito tem terminação similar ao número 3, implicando também em falsos positivos para esse último. No entanto, avaliando-se o teste com seleção de características é possível identificar que o reconhecimento melhora para esses casos em comparação ao teste sem seleção.

Nos dois testes, identifica-se, também, que os valores de F1 *score* são similares aos de *Recall* e *Precision*. Isso indica que há um equilíbrio entre essas medidas, e mostra a capacidade do classificador em reconhecer os dígitos de forma não tendenciosa. Ressalta-se também que o índice Kappa obtido em ambos os testes classifica os resultados na categoria "excelente".

Por fim, ressalta-se que o desempenho do classificador no teste com seleção de característica se mostra bastante promissor, indicando que essa etapa foi de grande relevância para a obtenção de melhores resultados. No contexto de um aplicativo móvel de autoleitura, dar-se-ia preferência por esse modelo, pois, como visto na Tabela 4, é capaz de gerar bons resultados e consumiria uma parcela menor do armazenamento interno do dispositivo em comparação com o modelo sem seleção devido a redução de 1680 para 948 características.

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresentou um método para o reconhecimento de dígitos por meio da voz utilizando processamento de áudio e inteligência computacional. O trabalho também relata a elaboração da base de áudios utilizada para o desenvolvimento do teste, no qual as características dos áudios são representadas através de métricas obtidas a partir das matrizes geradas pelas técnicas MFCC e *MelSpectrogram*. Essas características passam pela etapa de seleção na qual se utiliza o algoritmo BPSO.

O método apresenta *Recall* de 94,74%; *Precision* de 94,91%; F1 score de 94,72% e 0,9419 de índice Kappa. Esses resultados foram obtidos com o classificador SVM e são considerados bastante promissores para a aplicação ao qual este trabalho será integrado.

Como trabalho futuro, serão feitos mais testes com o método proposto utilizando outras proporções de treino e teste a fim de obter uma avaliação de desempenho mais abrangente. Pretende-se, também, desenvolver a segmentação automática de cada dígito a partir do áudio da leitura completa do consumo que é observada no medidor. Além disso, busca-se melhorar os resultados obtidos investigando-se outras abordagens para a extração de características, como, por exemplo, Redes Neurais Convolucionais (CNNs). E, por fim, a incorporação do método proposto neste trabalho em uma aplicação de autoleitura para dispositivos móveis.

REFERÊNCIAS

Abusaria, A.A.M., Gunawan, T.S., Khalifa, O.O., and Abushariah, M.A.M. (2010). English digits speech re- cognition system based on hidden markov models. In International Conference on Computer and Communication Engineering (ICCCE'10), 1–5. doi:10.1109/ ICCCE.2010.5556819.

ANEEL (2019). Energia no brasil e no mundo. URL http://www2.aneel.gov.br/arquivos/pdf/atlas_par1_cap2.pdf.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273–297.

da Silveira, F.M., Schueler, C.F., and Cataldo, E. (2017). Utilização da técnica mfcc em conjunto com os parâmetros extraídos do sinal glotal para melhorar o desempenho de um sistema de verificação de locutor.

Davis, S. and Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 357–366. doi:10.1109/TASSP.1980.1163420.

de Andrade Bresolin, A., Neto, A.D.D., and Alsina, P.J. (2008). Digit recognition using wavelet and svm in brazilian portuguese. In 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, 1545–1548. doi:10.1109/ICASSP.2008.4517917.

dos Santos, S.e.A.A. (2000). Reduced sets of subword units for continuous speech recognition of portuguese. *Electronics Letters*, 36(6), 586–588. doi:10.1049/el:20000446.

Kennedy, J. and Eberhart, R.C. (1997). A discrete binary version of the particle swarm algorithm. In 1997 IEEE International Conference on Systems, Man, and Cybernetics. *Computational Cybernetics and Simulation*, volume 5, 4104–4108 vol.5. doi:10.1109/ICSMC.1997.

637339.

Landis, J.R. and Koch, G.G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duches- nay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Pessoa, M.C.P. et al. (2018). Remoção de ruídos aditivos e segmentação de palavras-chave em áudios.

Quintanilha, D.B.P., Costa, R.W.S., Diniz, J.O.B., de Almeida, J.D.S., Braz, G., Silva, A.C., de Paiva, A.C., Monteiro, E.M., Froz, B.R., Pinheiro, L.P.A., and Melho,

W. (2017). Automatic consumption reading on electromechanical meters using hog and svm. In 7th Latin American Conference on Networked and Electronic Media (LACNEM 2017), 57–61. doi:10.1049/ic.2017.0036.

S. Stevens, S., Volkman, J., and B. Newman, E. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, 8, 185–190. doi:10.1121/1.1915893.

Schramm, M., Freitas, L., Zanuz, A., and Barone, D. (2006). Spoltech brazilian portuguese version 1.0 idc2006s16. Philadelphia: Linguistic Data Consortium. Silva, D.F., de Souza, V.M.A., Batista, G.E.A.P.A., and Giusti, R. (2012). Spoken digit recognition in portuguese using line spectral frequencies. In J. Pav' on, N.D. Duque- Méndez, and R. Fuentes-Fernández (eds.), *Advances in Artificial Intelligence – IBERAMIA 2012*, 241–250.

Springer Berlin Heidelberg, Berlin, Heidelberg. Tabarabae, V., Azimisadjadi, B., Zahirazami, S.B., and

Lucas, C. (1994). Isolated word recognition using a hybrid neural network. In *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume ii, II/649–II/652 vol.2. doi:10.1109/ICASSP.1994.389572.

Vidinich, R. and Nery, G. (2009). Pesquisa e desenvolvimento contra o furto de energia. *Revista Pesquisa e Desenvolvimento da ANEEL–P&D*, 15.

Vieira, S.M., Mendonça, L.F., Farinha, G.J., and Sousa,

J.M. (2013). Modified binary pso for feature selection using svm applied to mortality prediction of septic patients. *Applied Soft Computing*, 13(8), 3494–3504.

Zhang, Y., Yang, S., Su, X., Shi, E., and Zhang, H. (2016). Automatic reading of domestic electric meter: an intelligent device based on image processing and zigbee/ethernet communication. *Journal of Real-Time Image Processing*, 12(1), 133–143.