

## Impacto do Registro de Consumo de Energia pela Média de Consumo na Detecção de Furto de Energia Utilizando Redes Neurais Artificiais

Sérgio L. M. Lacerda\*. Alexandre C. Oliveira\*\*. Armando J. G. A. Ferreira\*\*\*

\* Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brasil (Tel: 83 2101 1146; e-mail: sergio.lacerda@ee.ufcg.edu.br).

\*\* Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brasil (Tel: 83 2101 1146; e-mail: aco@dee.ufcg.edu.br).

\*\*\* Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brasil (Tel: 83 2101 1146; e-mail: armando.ferreira@ee.ufcg.edu.br).

---

**Abstract:** The energy theft generates significant losses for the electricity distribution concessionaires. Research over the past decade has focused on machine learning as a method of detecting energy theft by analyzing consumption logs. The use of convolutional neural networks is one of these methods. Operating procedures can affect the quality of consumption data, impacting the performance of machine learning methods. In this paper is evaluate how the procedure of concessionaries of use the consumption average affect the performance of the energy theft detection method using a convolutional neural network.

**Resumo:** O furto de energia gera prejuízos significativos para as concessionárias de distribuição de energia elétrica. Pesquisas na última década se concentram no aprendizado de máquina como um método para detectar o roubo de energia a partir da análise dos registros de consumo. O uso de redes neurais convolucionais é um desses métodos. Os procedimentos operacionais das concessionárias podem afetar a qualidade dos dados de consumo, impactando no desempenho dos métodos de aprendizado de máquina. Neste artigo é avaliado como o procedimento das concessionárias de uso da média de consumo afeta o desempenho do método de detecção de roubo de energia utilizando uma rede neural convolucional.

**Keywords:** machine learning; convolutional neural network; artificial neural network; eletricity theft; MSI.

**Palavras-chaves:** aprendizado de máquina; rede neural convolucional; rede neural artificial; furto de energia; ISM.

---

### 1. INTRODUÇÃO

No Brasil, as perdas não técnicas (PNT's) – perdas comerciais – são responsáveis pelo prejuízo anual de R\$ 8 bilhões desde o ano de 2008, segundo Maia, S. (2019), valor equivalente a aproximadamente 10% da energia comercializada, embora varie por estado: de 0,5% a 25%, conforme versão 01/2019 do Relatório ANEEL - Perdas de energia elétrica na distribuição. No mundo, as perdas anuais alcançam o valor de \$ 96 bilhões de dólares, segundo dados divulgados em 2017 no artigo intitulado “\$96 billion is lost every year to electricity theft”, da L. Northeast Group.

A identificação de consumidores com alguma irregularidade no sistema de medição (ISM) via análise de dados, utilizando modelos de classificação com aprendizado de máquina, é uma estratégia que tem merecido atenção, segundo Zhang et al (2018), havendo diversos modelos de classificação baseados em aprendizado supervisionado e não

supervisionado. Os modelos baseados em aprendizado supervisionado fazem bastante uso das redes neurais, como demonstrado nos trabalhos de Zheng, Z. et al (2018), Muniz et al (2009), Huang et al (2018), Zheng, K. et al (2019); e de estratégias baseadas em máquinas de vetor suporte (SVM), como apresentado nos trabalhos de Depuru et al (2011), Jokar et al (2016), Jindal et al (2016). No aprendizado não supervisionado um modelo muito utilizado é o K-Means, como utilizado em Monteiro et al (2018), Al-Wakeel et al (2016), Al-Wakeel et al (2017).

Zheng et al (2018) propôs um modelo de classificação que utiliza rede neural artificial convolucional (RNAC) bidimensional rasa e profunda que se aplica bem ao problema das perdas comerciais por furto de energia, que se enquadra no contexto de irregularidades no sistema de medição.

Para treinar, validar e testar uma RNA convolucional é necessário um conjunto de dados de unidades consumidoras (UC's) regulares e irregulares. Esses dados, antes de serem

utilizados, precisam passar por uma etapa de limpeza (*data cleansing*) para eliminar dados repetidos, incorretos, incoerentes e outros ruídos. Um dos tipos de ruído que comumente é inserido pelas concessionárias brasileiras de distribuição de energia elétrica de forma “consciente” no conjunto de dados de consumo de UC’s é a “leitura pela média”, normalmente feita quando ocorre o impedimento de acesso ao medidor ou a impossibilidade de leitura do consumo. Ocorrendo uma dessas situações, é realizada a média aritmética dos valores faturados nos 12 (doze) últimos ciclos de faturamento, conforme enuncia a Resolução nº 414/2010 da ANEEL. Ainda, segundo essa Resolução, esse procedimento só pode ser adotado por até 3 ciclos consecutivos de leitura de consumo, que deve ser realizado em intervalos de aproximadamente 30 dias, observados o mínimo de 27 e o máximo de 33 dias.

Neste trabalho, buscou-se avaliar o impacto no desempenho de uma RNA convolucional, usada para classificar consumidores, no que tange às ISM’s, quando a concessionária adota o registro pela média de consumo. A RNA convolucional usada tem estrutura equivalente a apresentada por Zheng (2018), tendo sido ela adaptada para o cenário brasileiro onde a granularidade de dados é muito menor que a utilizada por Zheng et al (2018). O desempenho da RNA convolucional é comparado ao de uma rede neural artificial simples (RNAs), com número equivalente de “elementos”, de parâmetros, de modo a comprovar a adequação e efetividade da RNA convolucional no processo de classificação dos consumidores.

## 2. DESEMPENHO DE REDES NEURAIS ARTIFICIAIS NA DETECÇÃO DE CONSUMIDORES IRREGULARES UTILIZANDO O REGISTRO DE CONSUMO DE ENERGIA COM BAIXA GRANULARIDADE

### 2.1 Estrutura das Redes Neurais Artificiais Avaliadas

Duas redes neurais foram avaliadas na detecção de ISM’s com dados de baixa granularidade: uma RNA simples com uma componente rasa, que realiza análises em uma dimensão; e uma RNA convolucional com uma componente rasa e uma profunda, que realiza análises em uma e em duas dimensões. Utilizou-se *batch size* no treinamento  $BS_T = 32$ , *batch size* na validação  $BS_V = 8$ , otimizador Adam e taxa de aprendizado  $LR = 1e^{-5}$ . Essa configuração foi utilizada para o melhor desempenho da RNA convolucional. A RNA simples apresentou seu melhor desempenho para os seguintes ajustes:  $BS_T = 4$ ,  $BS_V = 1$  e  $LR = 1e^{-4}$ . O resultado do desempenho da RNA simples com as duas configurações será apresentado.

A RNA convolucional é estruturada da seguinte forma: componente rasa (unidimensional) e componente profunda (bidimensional). Componente rasa: camadas = 01, neurônios = 12, *dropout* = 0,1 e função de ativação “ReLU”. Componente profunda: camadas = 05, neurônios = 24, filtros por camada = 05, tamanho dos filtros: [7, 7], [5, 5], [5, 5], [3, 3], [3, 3], função *pooling* = MaxPool2D, tamanho do *pooling* = (2,2), *dropout* = 0,1 e função de ativação “TanH” em cada camada. A função de ativação da camada densa (totalmente conectada) ao final das camadas convolucionais é a “ReLU”.

A função de ativação da combinação das componentes rasa e profunda é a “Sigmoid”. Essa configuração de hiperparâmetros faz a rede neural ter uma capacidade de 6.158.616 de elementos que auxiliam o modelo de classificação na avaliação dos dados a serem analisados (alto grau de complexidade), possibilitando a classificação das UC’s em regular ou irregular.

A RNA simples é estruturada da seguinte forma: Componente rasa: camadas = 01, neurônios = 36, *dropout* = 0,1 e função de ativação “ReLU”. Utilizaram-se 36 neurônios para a rede apresentar a mesma quantidade de neurônios utilizados na RNA convolucional (somados os neurônios das duas camadas da RNA convolucional). Com essa configuração de hiperparâmetros a RNA simples tem capacidade de 2.233 elementos que auxiliam o modelo de classificação (baixo grau de complexidade). Para deixá-la com o mesmo grau de complexidade da RNA convolucional, realizou-se outro teste considerando 01 camada com 99.333 neurônios, deixando a RNA simples com 6.158.647 elementos.

### 2.2 Metodologia de Testes

O conjunto de dados usado para treinamento e para teste possui dados de consumo mensal de 60 meses de UC’s regulares e de UC’s irregulares, totalizando 1.477 UC’s regulares e 1.439 UC’s irregulares. A identificação de a UC ser regular ou irregular é fornecida pela concessionária, junto com a informação de consumo mensal. Assim, a UC é apresentada para os modelos de classificação supervisionados que avaliam seus dados de consumo associado a uma “etiqueta” de regular/irregular. Para o conjunto de UC’s regulares foram consideradas apenas aquelas cujo desvio padrão aritmético dos dados de consumo mensal foi menor ou igual a 25% da média de consumo mensal da respectiva UC regular avaliada, considerando os 60 meses de registro informados. A seleção resultou na identificação de 433 UC’s no conjunto de 1.477 UC’s regulares informadas pela concessionária. Para o conjunto de UC’s irregulares foram consideradas apenas as UC’s cujo desvio padrão aritmético dos dados de consumo mensal foi maior ou igual a 75% da média dos dados de consumo da respectiva UC irregular avaliada, considerando os 60 meses de registro informados. A seleção resultou na identificação de 413 UC’s no conjunto de 1.439 UC’s irregulares informadas pela concessionária.

O conjunto de dados das 433 UC’s regulares selecionadas foi segmentado em 3 partições aproximadamente iguais (P1, P2 e P3), bem como o conjunto das 413 UC’s irregulares. As três partições de cada conjunto de UC’s (regulares e irregulares) tiveram suas UC’s classificadas 10 vezes por uma RNA convolucional cuja métrica de avaliação *AUC* foi igual a  $AUC = 80\%$  na etapa de validação, para selecionar as melhores UC’s de cada partição, sendo selecionadas as UC’s que foram classificadas 10/10 corretamente com probabilidade superior a 80% de pertencer à classe correta. Na partição P1 foram obtidas 19 de 144 UC’s regulares e 21 de 137 UC’s irregulares, formando uma partição denominada de P1x; na partição P2 foram obtidas 20 de 144 UC’s regulares e 18 de 138 UC’s irregulares, formando uma

partição denominada P2x; e, na partição, P3 foram obtidas 20 UC's de 145 UC's regulares e 14 de 138 UC's irregulares, formando uma partição P3x. Essas UC's foram adotadas como modelo de UC regular padrão e de UC irregular padrão e sobre os dados destas UC's foram aplicadas as alterações de consumo que simulam a corrupção dos dados pela concessionária, quando a mesma adota o registro de consumo pela média.

Duas das três partições de cada conjunto (regulares, 433 unidades e irregulares, 413 unidades) foram utilizadas na etapa de treinamento e de validação. Na etapa de teste foram utilizadas apenas as UC's regulares das partições P1x, P2x e P3x, sem que elas tenham sido utilizadas na etapa de treinamento ou validação, evitando viciar o resultado.

O modelo de classificação foi treinado 10 vezes e para cada treinamento foi realizado o teste, individual, de cada UC regular das partições P1x, P2x e P3x, sendo as mesmas classificadas 10 vezes como UC regular ou irregular, registrando-se a probabilidade de a UC ser da classe a qual foi classificada, regular ou irregular. No caso da UC ser classificada como irregular, isso denota uma falha do método de classificação.

Com as probabilidades dos 10 resultados foi gerada uma média das 10 probabilidades encontradas em cada classificação para cada UC avaliada (valor da probabilidade entre 0% e 100%). O valor da probabilidade determina se a UC é regular ou irregular, pois dentro desses 10 resultados a UC pode ter sido classificada tanto como regular quanto irregular. Para exemplificar, considere que uma UC foi classificada como regular e atingiu uma probabilidade de 85%; como a probabilidade é superior a 50%, então a mesma é considerada de fato uma UC regular. Foi estabelecido que para uma UC testada ser declarada como regular, a média obtida após a mesma ser testada 10 vezes por um classificador deve ser superior a 50%; caso seja inferior, ela é declarada como irregular. Cada UC foi testada individualmente.

### 2.3 Resultados

Ao realizar a classificação dos modelos de UC's regulares e dos modelos de UC's irregulares, das partições P1x, P2x e P3x, utilizando a RNA simples de baixa complexidade (BC) e a RNA simples de alta complexidade (AC) com  $BS_T = 32$ ,  $BS_V = 8$  e  $LR = 1e^{-5}$  foram obtidos os resultados apresentados nas Tabelas 1 e 2. As Tabelas 3 e 4 apresentam os resultados da RNA simples de baixa complexidade (BC) e a RNA simples de alta complexidade (AC) com  $BS_T = 4$ ,  $BS_V = 1$  e  $LR = 1e^{-4}$ . Os resultados da RNA convolucional (alta complexidade) são apresentados na Tabela 5. A complexidade das Redes Neurais Artificiais foi associada ao número de "ELEMENTOS" que constituem a estrutura de cada RNA.

Os testes registraram quantas UC's de cada partição (P1x, P2x e P3x) foram classificadas corretamente e quantas foram classificadas incorretamente, bem como foi registrado o percentual médio (%med) das UC's classificadas

corretamente (e das classificadas incorretamente) de pertencerem à classe a qual foram classificadas. Onde não houve UC's classificadas como corretas ou como incorretas não existe percentual médio (%med) de classificação, representado na tabela por "-".

**Tabela 1. RNAs - BC ( $BS = 32$  e  $LR = 1e^{-5}$ )**

RNAs - BC	UC's Regulares			UC's Irregulares			UC's Regulares			UC's Irregulares			
	Partição	P1x	P2x	P3x	P1x	P2x	P3x	%med (P1x)	%med (P2x)	%med (P3x)	%med (P1x)	%med (P2x)	%med (P3x)
Classificação													
Correta	0/19	0/20	20/20	0/21	5/18	0/14	-	-	59,5%	-	50,6%	-	-
Incorreta	19/19	20/20	0/20	21/21	13/18	14/14	52,1%	60,3%	-	56%	51,8%	56,1%	-

\*RNAs: RNA simples

**Tabela 2. RNAs - AC ( $BS = 32$  e  $LR = 1e^{-5}$ )**

RNAs - AC	UC's Regulares			UC's Irregulares			UC's Regulares			UC's Irregulares			
	Partição	P1x	P2x	P3x	P1x	P2x	P3x	%med (P1x)	%med (P2x)	%med (P3x)	%med (P1x)	%med (P2x)	%med (P3x)
Classificação													
Correta	19/19	20/20	20/20	6/21	17/18	3/14	68,5%	73,1%	67,3%	51,9%	57,8%	51,2%	-
Incorreta	0/19	0/20	0/20	15/21	1/18	11/14	-	-	-	59,2%	51,3%	55,5%	-

\*RNAs: RNA simples

**Tabela 3. RNAs - BC ( $BS = 1$  e  $LR = 1e^{-4}$ )**

RNAs - BC	UC's Regulares			UC's Irregulares			UC's Regulares			UC's Irregulares			
	Partição	P1x	P2x	P3x	P1x	P2x	P3x	%med (P1x)	%med (P2x)	%med (P3x)	%med (P1x)	%med (P2x)	%med (P3x)
Classificação													
Correta	20/20	19/20	20/20	21/21	18/18	13/14	85%	95,1%	88,4%	90%	93%	92,3%	-
Incorreta	0/20	1/20	0/20	0/21	0/18	1/14	-	83%	-	-	-	-	61%

\*RNAs: RNA simples

**Tabela 4. RNAs - AC ( $BS = 1$  e  $LR = 1e^{-4}$ )**

RNAs - AC	UC's Regulares			UC's Irregulares			UC's Regulares			UC's Irregulares			
	Partição	P1x	P2x	P3x	P1x	P2x	P3x	%med (P1x)	%med (P2x)	%med (P3x)	%med (P1x)	%med (P2x)	%med (P3x)
Classificação													
Correta	18/19	20/20	20/20	21/21	18/18	12/14	99%	97,6%	93,8%	95,1%	97,4%	95,2%	-
Incorreta	1/19	0/20	0/20	0/21	0/18	2/14	65%	-	-	-	-	-	75%

\*RNAs: RNA simples

**Tabela 5. RNAc - AC**

RNAc - AC	UC's Regulares			UC's Irregulares			UC's Regulares			UC's Irregulares			
	Partição	P1x	P2x	P3x	P1x	P2x	P3x	%med (P1x)	%med (P2x)	%med (P3x)	%med (P1x)	%med (P2x)	%med (P3x)
Classificação													
Correta	19/19	20/20	20/20	21/21	18/18	14/14	99,4%	99,3%	99,5%	99,3%	99,5%	99,2%	-
Incorreta	0/19	0/20	0/20	0/21	0/18	0/14	-	-	-	-	-	-	-

\*RNAs: RNA convolucional

A RNA simples apresentou significativa melhoria na classificação das UC's com o  $batch\ size = 1$  e  $LR = 1e^{-4}$ . Contudo, essa melhoria foi acompanhada de um aumento do tempo de treinamento: o tempo de treinamento da RNA simples-BC e da RNA simples-AC com  $batch\ size = 32$  e  $LR = 1e^{-5}$  durou 190ms/época e 270ms/época, respectivamente, enquanto que o treinamento da RNA simples-BC e da RNA simples-AC com  $batch\ size = 1$  e  $LR = 1e^{-4}$  durou 3s/época e 4s/época, respectivamente, representando um aumento aproximado de 15 vezes no tempo de treinamento.

Já a RNA convolucional teve tempo de treinamento de 370ms/época e desempenho melhor do que a RNA simples de baixa ou de alta complexidade com  $BS = 1$  e  $LR = 1e^{-4}$ , além de maior estabilidade na probabilidade de uma UC pertencer à classe a qual foi classificada. Estes resultados confirmam a efetividade e estabilidade da RNA convolucional em classificar corretamente as UC's dos dois tipos, regulares e irregulares, apresentando desempenho superior a RNA simples-BC e RNA simples-AC.



Duas das três partições (P1, P2 e P3) de cada conjunto (regulares e irregulares) foram utilizadas na etapa de treinamento e validação. Uma partição entre as partições P1x, P2x e P3x foi usada na etapa de teste.

O modelo de classificação foi treinado cinco vezes para se obter cinco treinamentos distintos para a RNA convolucional. Cada UC regular das partições P1x, P2x e P3x, foi classificada individualmente na etapa de teste. Com os cinco resultados obtidos foi calculada uma média que representa o resultado do teste de cada UC.

### 3.3 Resultados

Os resultados dos casos A, B e C, indicados na Tabela 6, serão apresentados na Tabela 7. Os resultados dos casos D, E e F, indicados na Tabela 6, serão apresentados na Tabela 8. O símbolo “%C” representa o percentual de classificação da UC como UC regular, isto é, a probabilidade de a UC pertencer à classe de UC’s regulares. A coluna das partições informa quantas UC’s do conjunto foram classificadas de acordo com o %C abordado.

**Tabela 7. Classificação das UC’s para os casos A, B e C**

Caso \ %C	%C ≥ 98%			90% ≤ %C < 98%		
	Partição			Partição		
	P1x	P2x	P3x	P1x	P2x	P3x
Caso A	18/18	17/18	5/20	0	0	8/20
Caso B	14/18	5/18	4/20	0	6/18	4/20
Caso C	9/18	6/18	0	0	2/18	1/20
Caso \ %C	50% < %C < 90%			%C ≤ 50%		
	Partição			Partição		
	P1x	P2x	P3x	P1x	P2x	P3x
Caso A	0	1/18	7/20	0	0	0
Caso B	3/18	6/18	10/20	1/18	1/18	2/20
Caso C	6/18	6/18	9/20	3/18	4/18	10/20

**Tabela 8. Classificação das UC’s para os casos D, E e F**

Caso \ %C	%C ≥ 98%			90% ≤ %C < 98%		
	Partição			Partição		
	P1x	P2x	P3x	P1x	P2x	P3x
Caso D	3/19	3/20	2/20	4/19	4/20	2/20
Caso E	10/19	5/20	5/20	2/19	7/20	2/20
Caso F	12/19	14/20	9/20	3/19	2/20	4/20
Caso \ %C	50% < %C < 90%			%C ≤ 50%		
	Partição			Partição		
	P1x	P2x	P3x	P1x	P2x	P3x
Caso D	9/19	8/20	10/20	3/19	5/20	6/20
Caso E	5/19	4/20	9/20	2/19	4/20	4/20
Caso F	1/19	2/20	5/20	3/19	2/20	2/20

Observando a Tabela 7, percebe-se que à medida que se aumenta a quantidade de meses consecutivos em que o registro do consumo de energia é feito pela média de consumo (caso A=1, caso B=2, caso C=3) as UC’s, mesmo regulares, estão reduzindo seu percentual de classificação de pertencerem à classe de UC’s regulares. A Fig. 1 ilustra esse quadro, na qual se vê 40 de 56 UC’s do caso A (somando as três partições), representando 71,43%, com percentual de classificação igual ou superior a 98%, ou seja, tem 98% de

probabilidade de serem de fato UC’s regulares. Para o percentual de classificação igual ou inferior a 50% no caso A não há nenhuma UC. Já para o caso C, tem-se 15 de 56 UC’s (26,79%) com percentual de classificação igual ou superior a 98% para pertencerem à classe de UC’s regulares, enquanto 17 de 56 UC’s (30,36%) apresentam o percentual de classificação igual ou inferior a 50% - classificando essas 17 UC’s como irregulares.

Quadro semelhante é observado na Fig. 2 – representativo da Tabela 8 (casos D, E e F, na qual o caso D apresenta a maior concentração de leituras pela média de consumo). No caso D, por exemplo, 8 de 59 UC’s (13,56%) apresentaram percentual de classificação igual ou superior a 98% para pertencerem à classe de UC’s regulares, enquanto 14 de 59 UC’s (23,73%) apresentam o percentual de classificação igual ou inferior a 50% - classificando essas 14 UC’s como irregulares.

Ainda é possível observar que os casos C e D apresentam resultados bastante similares, pois, apesar de o caso C ter bem mais incidências da aplicação da média de consumo no registro do consumo mensal (C=24 meses x D=15 meses), ele possui uma periodicidade repetida nas ocorrências da aplicação da média de consumo, enquanto que o caso D é aperiódico nessa aplicação da média de consumo.

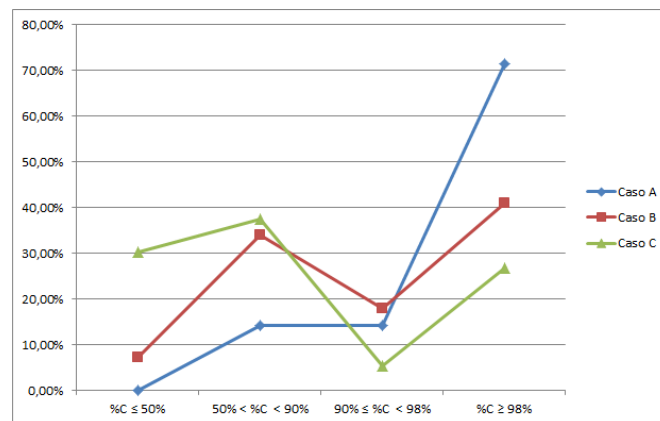


Fig. 1 Classificação das UC’s para os casos A, B e C.

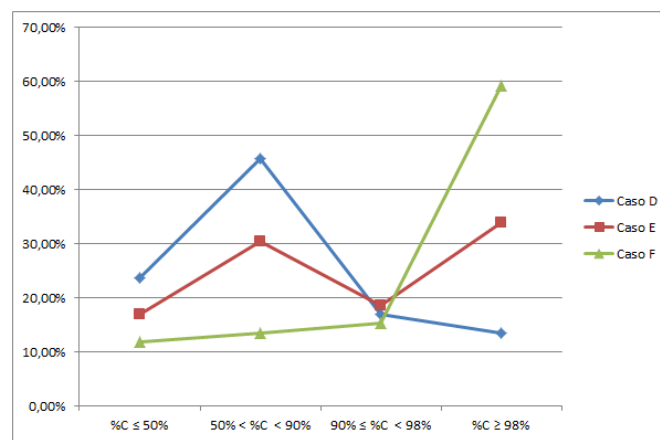


Fig. 2 Classificação das UC’s para os casos D, E e F.

#### 4. CONCLUSÕES E TRABALHOS FUTUROS

Os três tipos de redes neurais artificiais testadas apresentaram comportamentos distintos quando usadas para classificação de unidades consumidoras no que diz respeito à probabilidade de haver, ou não, fraude no medidor ou desvio de energia. A RNA simples-BC mostrou uma alta taxa de falsos positivos para UC's regulares das partições P1x, P2x e P3x e um baixo número de detecção de UC's irregulares das mesmas partições. A RNA simples-AC teve uma baixa taxa de falsos positivos para UC's regulares das partições P1x, P2x e P3x, embora um baixo número de detecção de UC's irregulares das mesmas partições. Entretanto, ao alterar o hiperparâmetro *batch size* de  $BSr = 32$  e  $BSv = 8$  para  $BSr = 4$  e  $BSv = 1$ , e ajustar a taxa de aprendizado de  $LR = 1e^{-5}$  para  $LR = 1e^{-4}$  para acomodação do aprendizado, ambas apresentaram significativa melhoria, apesar de elevar significativamente seu tempo de aprendizado. A RNA convolucional apresentou-se como melhor modelo de classificação, pois classificou corretamente todos os modelos de UC's regulares e irregulares das partições P1x, P2x e P3x que foram testadas (zero falsos positivos), além de uma probabilidade alta de classificação na classe correta e um tempo de aprendizado por época relativamente baixo.

A aplicação da técnica de leitura pela média de consumo, quando feito apenas em um mês, pouco influencia na classificação de uma UC regular. Ao realizar duas leituras seguidas usando o registro de consumo pela média, seja em período aleatório ou em período repetitivo, já influencia no comportamento do classificador, percebendo-se um aumento de classificações incorretas, embora ainda em quantidade baixa. Ao realizar três leituras consecutivas pela média em um período repetitivo, aproximadamente 1/3 das UC's são classificadas incorretamente e 1/3 tem impacto considerável na sua classificação como UC's regulares (redução acentuada do percentual de classificação como UC regular). A realização de uma quantidade menor de três leituras consecutivas pela média, de forma aperiódica, também impacta significativamente na classificação das UC's, fazendo com que aproximadamente 25% das UC's sejam classificadas de forma errada e quase 50% delas tenha sofrido impacto considerável na sua classificação como UC's regulares (redução acentuada do percentual de classificação como UC regular).

Como continuidade do tema, numa evolução do presente trabalho, pretende-se definir uma faixa de desvio padrão na média dos dados de consumo para classificar uma UC do conjunto de UC's como *outlier*, avaliar o desempenho da RNA convolucional ao classificar UC's através da análise de registros de consumo em que foram simulados comportamentos de desvio total e desvio parcial de energia, além também de avaliar o impacto do tamanho da janela de consumo (60, 50 e 40 meses) no processo de classificação das UC's.

#### AGRADECIMENTOS

Os agradecimentos são direcionados à concessionária que cedeu o conjunto de dados para o presente trabalho e para a

Universidade Federal de Campina Grande, que proporcionou os meios e instrumentos necessários à sua realização.

#### REFERÊNCIAS

- Al-Wakeel, A. and Wu, J. K-means based cluster analysis of residential smart meter measurements. *Applied Energy Symposium and Summit 2015: Low carbon cities and urban energy systems*, vol. 88, pp. 754-760, 2016.
- Al-Wakeel, A. Wu, J. and Jenkins, N.. K-means based load estimation of domestic smart meter measurements. *Applied Energy*, vol. 194, pp. 333-342, 2017.
- ANEEL - Agência Nacional de Energia Elétrica. Resolução Normativa ANEEL nº 1.000/2021 – Estabelece as Regras de Prestação do Serviço Público de Distribuição de Energia Elétrica..
- Depuru, S. S. S. R. Wang, L. Devabhaktuni, V. Support vector machine based data classification for detection of electricity theft. in *Proc. IEEE/PES Power Syst. Conf. Expo. (PSCE)*, 2011, pp. 1–8.
- Gusmão, L. Helito, H. Anarelli, T. Conceição, J. R. Ji, T. Barros, G.. “A Customer Journey Mapping Approach to Improve CPFL Energia Fraud Detection Predictive Models”. In *2020 IEEE PES Transmission & Distribution Conference and Exhibition - Latin America*.
- Huang, H. Liu, S. Davis, K. Energy Theft Detection Via Artificial Neural Networks. *IEEE PES Innovative Smart Grid Technologies Conference, ISGT 2018*.
- Jindal, A. Dua, A. Kaur, K. Singh, M. Kumar, N. Mishra, S. Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid. *IEEE Transactions on industrial informatics*, vol. 12, no. 3, June 2016.
- Jokar, P. Arianpoo, N. and Leung, V. C. M. Electricity theft detection in AMI using customers' consumption patterns. *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 216–226, Jan. 2016.
- L. Northeast Group, “\$96 billion is lost every year to electricity theft,” 2017. Disponível em: <<https://electricenergyonline.com/article/organization/24442/636817/-96-Billion-Is-Lost-Every-Year-To-Electricity-Theft.htm>>. Acesso em Dezembro, 2020.
- Maia, Sérgio. (2019). *Perícia em Furto de Energia Elétrica – Da Inspeção Administrativa ao Exame Pericial*. 1ª Ed. Campinas/SP: Millennium Editora.
- Messinis, G. M. Hatzigargyriou, N. D. Unsupervised Classification for Non-Technical Loss Detection. *Power Systems Computation Conference (PSCC)*, 2018.
- Monteiro, M. D. Maciel, R. S. Detection of Commercial Losses in Electric Power Distribution Systems Using Data Mining Techniques. *Simposio Brasileiro de Sistemas Eletricos (SBSE)*, 2018.
- Muniz, C. Vellasco, M. M. B. R. Tanscheit, R. and Figueiredo, K. A neuro-fuzzy system for fraud detection in electricity distribution. In *Proc. Joint Int. Fuzzy Syst. Assoc. World Congr.; Eur. Soc. Fuzzy Logic Technol. Conf.*, Lisbon, Portugal, Jul. 2009, pp. 1096–1101.

- T. Zhang, R. Gao, S. Sun. Theories, Applications and Trends of Non-technical Losses in power utilities using Machine Learning. IEEE/IMCEC 2018.
- Zheng, K. Chen, Q. Wang, Y. Kang, C. Xia, Q.. A Novel Combined Data-Driven Approach for Electricity Theft Detection. IEEE Transactions on Industrial Informatics, vol. 15, n. 3, march 2019, pp 1809 - 1819.
- Zheng, Z. Yang, Y. Niu, X. Dai, H-N. Zhou, Y. Wide & Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids. DOI 10.1109/TII.2017.2785963, IEEE Transactions on Industrial Informatics, vol. 14, n. 4, April 2018, pp 1606 - 1615.