

## PREVISÃO DE ÓBITOS POR COVID-19 NO BRASIL UTILIZANDO REDES NEURAS ARTIFICIAS COM ABORDAGEM SLIDING WINDOW

**Bruno de Oliveira Rodrigues**<sup>1</sup>, **Danielle Rabelo Costa**<sup>2</sup>

<sup>1</sup>Unicatólica de Quixadá, (rodrigues\_bruno7@hotmail.com)

<sup>2</sup>Unicatólica de Quixadá, (daniellerabelo@unicatolicaquixada.edu.br)

### Resumo

**Objetivo:** A letalidade da doença covid-19 evidenciou a importância de ferramentas no combate a pandemia de maneira eficiente. Por sua vez o trabalho é estruturado em se antecipar a fenômenos como esse para a aplicação de públicas baseadas em dados. Logo o presente estudo se objetivou em elaborar uma rede neural artificial dos dados por óbitos diários no Brasil caudados pelo novo coronavírus usando metodologia específica, podendo auxiliar as medidas tomadas para evitar a propagação do vírus. **Método:** Baseado em estudo e caso de abordagem quantitativa e natureza aplica, o percurso metodológico partiu de quatro etapas: i) obtenção, manipulação e estruturação dos dados; ii) pré-processamento e divisão entre treinamento e teste dos mesmos; iii) seleção da sliding window, construção e escolha da arquitetura do modelo de RNA; iv) acurácia e apresentação dos resultados através das métricas de erro *Mean Square Error* (MSE) e Theil's U. **Resultados e Conclusões:** No que se refere à previsão, o modelo teve resultados aceitáveis, tendo em vista o nível de complexidade da RNA, com acurácia de erro aceitável para MSE e para o índice de Theil's U mostrou-se mais eficiente que um passeio aleatório. Por fim, como proposta de trabalhos futuros, aconselha-se o uso modelos epidemiológicos, modelos combinados, redes neurais otimizadas e a aplicação desta metodologia para outros dados da saúde.

**Palavras-chave:** Covid-19; Previsão; Séries temporais.

**Área Temática:** Inovações e Tecnologias no Enfrentamento à COVID-19.

**Modalidade:** Trabalho completo

## 1 INTRODUÇÃO

A pandemia global causada pelo novo coronavírus ou Covid-19 (SARSCoV-2) mostrou-se uma ameaça ao desenvolvimento da humanidade. Segundo a Organização Mundial da Saúde (OMS) a doença se classifica como uma emergência de saúde pública de interesse internacional (ESPII). Responsável pelas mortes de 3.322 milhões de pessoas no mundo e 428 mil apenas no Brasil até a data de 13 de maio de 2021. Com seu primeiro caso sendo notificado na grade cidade de Wuhan das províncias de Hubei, China, ainda no ano de 2019. (GUAN et al., 2020)

Ainda que a letalidade da doença seja mais baixa em comparativo com de outros vírus da mesma família, sua alta facilidade de propagação ocasionou um número expressivo de mortes (AQUINO et al., 2020). Por se tratar de um vírus novo, os órgãos governamentais ainda estão em processo de formulação para políticas e práticas para o combate a pandemia. Várias medidas de controle foram adotadas durante a pandemia com o intuito e reduzir a propagação do vírus (ANZAI et al., 2020).

Para Rocha et al. (2020) a urgência de ações direcionadas por dados é cada vez mais imprescindível, estimar informações quantitativamente para nortear políticas e reduzir os impactos da doença. Prever o desenvolvimento de uma pandemia é um trabalho bastante difícil, tendo em vista vários fatores como a mudança das séries temporais com a integração de novos aspectos (FIRMINO et al., 2020). Redes Neurais Artificiais (RNA) para previsão de dados ao longo do tempo podem ser utilizadas com sucesso na previsão de informações e padrões futuros (OANCEA e CRISTIAN, 2013).

Dessa forma faz-se necessário a criação de ferramentas tecnológicas que possam auxiliar o enfrentamento da pandemia do novo coronavírus, assim como se antecipar futuros picos que ocasionem mortes em decorrência da doença no país, para que medidas de combate possam ser tomadas de maneiras mais eficientes. O presente estudo tem como objetivo desenvolver um modelo de RNA para previsão da série temporal referente ao número de óbitos diários por Sars-CoV-2 no Brasil, do mesmo modo que verificar a capacidade de predição do modelo em questão.

### 1.1 Sliding window

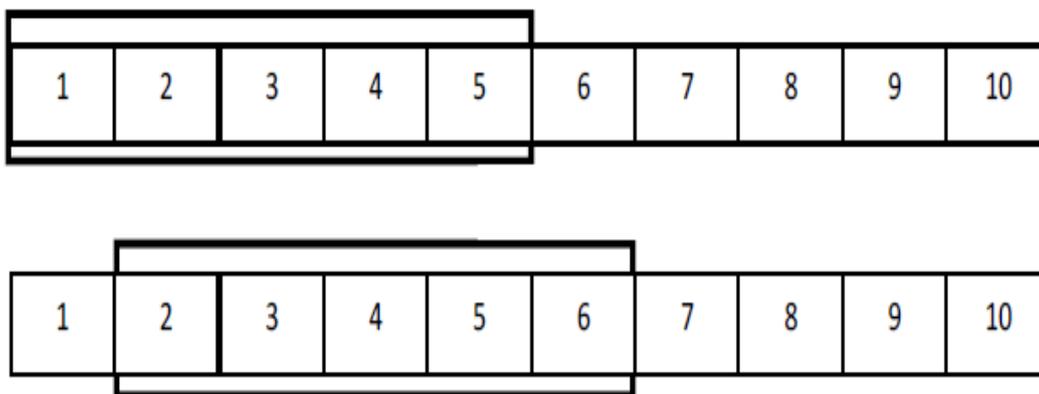
Uma *sliding window* é um valor dado pela segmentação de uma série temporal com o valor real, dessa forma o processo de divisão dos dados para construção da mesma pode

estabelecer parâmetros, como tamanho do bloco segmentado, a quantidade de janelas utilizadas em determinada modelagem, e outros parâmetros de otimização. Em alguns casos uma única série temporal pode ser convertida em um bloco de séries temporais com a utilização dessa metodologia (BENYAHMED et al., 2015).

Na figura 1 mostra o processo de uma *sliding window* em evolução, no qual o tamanho da janela de observações é determinado como 5, na segunda parte da imagem mostra uma *sliding window* com defasagem 1, ou seja, um ponto à frente da série de observações original, esse processo pode ser chamado de “Lag” no caso em exemplo podemos chamar de lag1. Sendo assim a figura mostra a criação de uma *sliding window* de tamanho 5 e lag1 (YU et al., 2014).

Ao utilizar a segmentação de uma janela deslizante, cria-se o objetivo da criação de dados de entrada para a RNA, compostos por partes da série original. Para encontrar um conjunto de pontos de corte para particionar o intervalo em um pequeno número de intervalos. Somente alguns respectivamente. Inicialmente, a janela cobriu de 1 a 5, o que representa aquele

**Figura 1.** A figura mostra o processo de *Sliding window*, em um primeiro momento a inicial e no segundo momento a mesmo fazendo o “deslize” para segmentar o conjunto de dados.



Fonte: Autores, 2021.

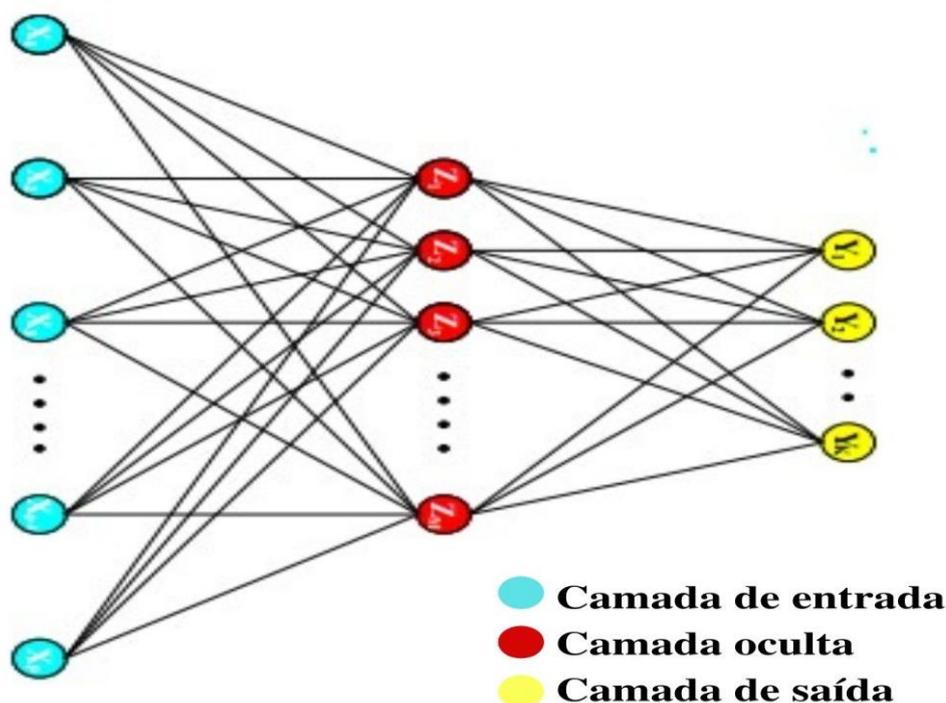
## 1.2 Redes neurais artificiais

As Redes Neurais Artificiais (RNA) são baseadas em tecnologia computacional que simulam o funcionamento dos neurônios biológicos humanos, com tudo utilizam de parâmetros para aprendizado e identificação de padrões. Segundo Sales (2019) são amplamente utilizadas para solução de problemas não lineares, como no caso de séries temporais.

Pode-se classificar uma RNA como um sistema constituído por unidades de processamento, que se chamam neurônios, no qual estão distribuídos em camadas interligadas para processamento de informação (PANIGRAHI; BEHERA, 2017). Cada camada da RNA tem neurônios que estão interligados a pesos para realização do processo de aprendizagem. Dessa forma existe várias arquiteturas para constituição de uma rede.

Existe uma grande variedade de arquiteturas que constituem a estrutura de uma RNA, uma delas é a Perceptron Multicamadas uma das mais comuns (Multilayer Perceptron - MLP). A figura 2 mostra o exemplo de uma rede neural arquitetada, com camadas de entrada, ocultas e de saída. O pressuposto inicial é de que os modelos podem captar padrões provenientes das séries temporais (sazonalidade e tendência, por exemplo) no processo de treinamento e fazer inferência a partir destes para novos conjuntos de dados (teste).

**Figura 2.** Exemplo da estrutura de uma rede neural.



Fonte: Adaptado de Hastie et al., 2009.

### 1.3 Medidas de acurácia

De forma geral, as métricas de erro se classificam como a diferença entre o valor estimado ( $\hat{Z}_t$ ) e o valor real ( $Z_t$ ), o nível de eficiência do modelos de previsão para uma determinada série temporal é medida através de indicadores criados por elas. O presente

trabalho conta com duas medidas de erro, sendo elas *Mean Absolute error* (MAE) e *Mean Square Error* (MSE).

A métrica denominada MSE permite verificar além da acurácia, também a variabilidade, sendo assim uma das mais utilizadas devido sua simplicidade matemática (CASELLA; BERGER, 2002). Seu valor total refere a soma de todos erros elevados ao quadrado.

A medida Theil's U faz um comparativo entre os valores gerados pelo modelo e um passeio aleatório (*Random Walk Model*) da série em questão. Sendo assim quando o valor for superior a 1, sua eficiência é inferior no comparativo, entretanto com o valor inferior a 1, o modelo é melhor que um passeio aleatório.

**Equação 1.** Mean Square Error (MSE)

$$MSE = \frac{\sum_{t=1}^n (z_t - \hat{z}_t)^2}{n}$$

(Eq. 1)

**Equação 2.** U de Theil (THEIL'S U)

$$Theil's U = \frac{\sum_{t=1}^n (z_t - \hat{z}_t)^2}{\sum_{t=1}^n (z_t - z_{t-1})^2}$$

(Eq. 2)

## 2 MÉTODO

O presente estudo tem uma abordagem quantitativa e quanto a natureza se caracteriza como aplicada, o trabalho foi baseado em estudo de caso. Os procedimentos metodológicos utilizados seguiram quatro etapas em ordem cronológica: (i) obtenção, manipulação e estruturação dos dados adquiridos; (ii) pré-processamento e divisão entre treinamento e teste dos mesmos; (iii) construção e escolha da arquitetura do modelo de RNA; (iv) acurácia e apresentação dos resultados através das métricas de erro *Mean Absolute error* (MAE) e *Mean Square Error* (MSE).

i) As observações são referentes ao total de óbitos diários em decorrência da doença Covid-19 no Brasil, foram obtidos através do Coronavírus Brasil, DATASUS do ministério da saúde (Coronavírus Brasil, 2021). Correspondentes a 421 observações, começando em 17 de março

de 2020 a 11 de maio de 2021. Os mesmos foram organizados, tabelados e divididos por meio do ambiente estatístico (RStudio).

ii) A aplicação da metodologia sliding window respeitou os critérios de 6 janelas de tamanho igual ao da série original e com lag's de 1 até 6, respectivamente. A divisão de dados seguiu os seguintes parâmetros, 80% foram utilizados para a fase de treinamento do modelo de RNA e 20% para a realização do teste de eficiência. Ressalta-se que a priori os mesmos foram normalizados seguindo o intervalo de  $[0, 1]$  (salienta-se que esta etapa é crucial para a elaboração do modelo de rede neural).

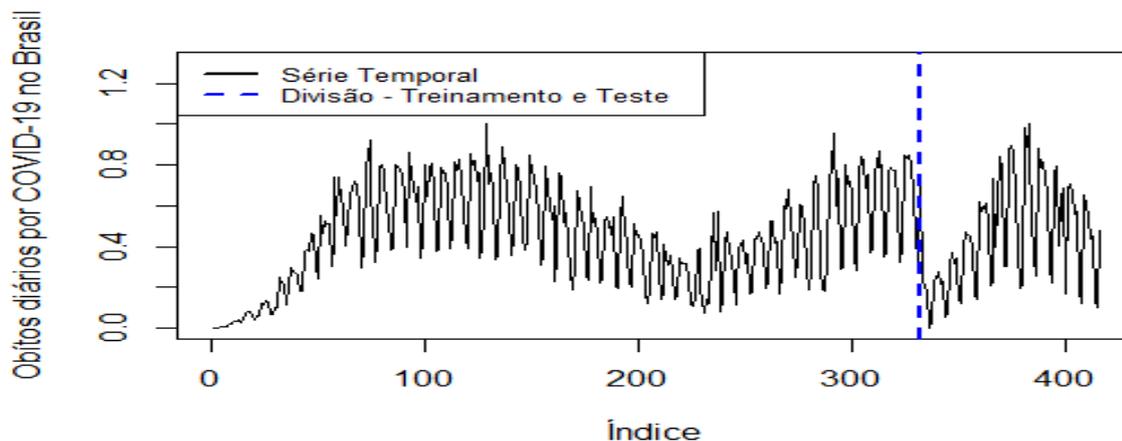
iii) Para desenvolvimento da sliding window foi utilizado o pacote de funções quantmod, já para a elaboração do modelo de RNA e aplicação no conjunto de dados utilizou-se o Pacote Neuralnet, ambos disponíveis no Software RStudio. Por fim foi realizado a previsão (one-step ahead) com o conjunto de teste, vale ressaltar que nessa etapa o algoritmo repete a previsão de um “passo à frente” até a última operação do processo.

iv) Em conclusão as métricas MSE e Theil's U (Eq. 1-2) foram computadas para o ambiente de trabalho estatístico e a posteriori aplicadas no conjunto de teste do modelo para acurácia dos resultados e eficiência da RNA elaborada.

### 3 RESULTADOS E DISCUSSÃO

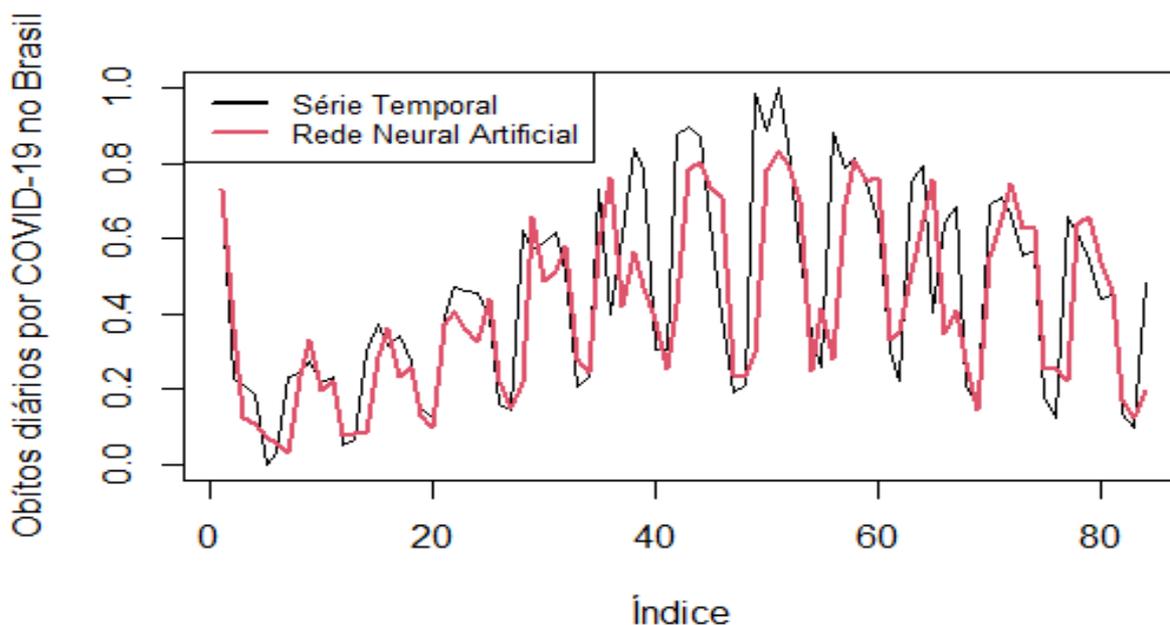
A base de dados apresenta valor mínimo equivalente a 1 e máximo de 4249, com desvio padrão igual a 816.9 e uma média diária com 842 óbitos. Apresentando picos com altos índices de mortes, sendo possível de observar graficamente. A figura 3 mostra a série original normalizada e dividida em fases de treinamento e teste.

**Figura 3.** Série temporal: valores normalizados entre treinamento e teste, separados por linha tracejada em azul.



A Figura 4 apresenta a comparação da RNA com metodologia sliding window e os valores observados apenas para a fase de teste. Nota-se que o modelo acompanha o comportamento da série original. Na etapa de teste o desempenho da rede foi considerável, tendo em vista que o processo de predição alcançou valores aproximados dos reais.

**Figura 4.** Modelo de previsão com valores normalizados para dados de teste.



Fonte: Autores, 2021.

Em relação ao MSE e Theil's U, os valores para as RNA foram 0.0337 e 0.5572, respectivamente. Os valores na métrica MSE ficam em torno dos 3,4% de ineficácia, já a o U de Theil aponta que o modelo se saiu melhor que um passeio aleatório, tendo em vista os valores abaixo de 1, reafirmando o desempenho da RNA. Por fim, vale considerar outras formas de erros residuais tendo em vista que os valores não são suficientes para consolidar o sucesso do modelo (Tabela 1).

**Tabela 1** – Métricas (valores absolutos) para o modelo RNA, calculadas a partir dos valores normalizados.

<b>Valores absolutos*</b>	
<b>Métricas</b>	<b>RNA</b>
<b>MSE</b>	<b>0.0337</b>
<b>Theil's U</b>	<b>0.5572</b>

Fonte: Autores, 2021.

## 4 CONCLUSÃO

Durante todo o trabalho foi elaborado um modelo de RNA para previsão da série temporal do total de óbitos diários no Brasil causados pela pandemia da Covid-19, sendo utilizado diversos parâmetros para construção do mesmo. Os resultados obtidos por pelas métricas utilizadas foram satisfatórios, levando em consideração nível de complexidade da estrutura adotada e a capacidade de acompanhar o comportamento da série original sendo superior a um passeio aleatório.

Para pesquisas futuras, indica-se o uso de modelos epidemiológicos, modelos combinados, redes neurais otimizadas e a aplicação desta metodologia para outros dados da saúde, como nas chances de o paciente desenvolver o estágio grave da doença ou até mesmo poder se antecipar a surtos epidemiológicos em determinadas regiões, assim podendo gerir políticas de controle mais eficientes e antecipadas, evitando sobrecarga no sistema de saúde. Bem como a utilização de métricas mais robustas para acurácia dos erros residuais.

## 5 REFERÊNCIAS

AQUINO, E. M. L. et al. Medidas de distanciamento social no controle da pandemia de COVID-19: potenciais impactos e desafios no Brasil. **Ciência & Saúde Coletiva**, 2020.

ANZAI, A. et al. Assessing the Impact of Reduced Travel on Exportation Dynamics of Novel Coronavirus Infection (COVID-19). **Journal of clinical medicine**, 2020.

BENYAHMED, Y. et al. Adaptive sliding window algorithm for weather data segmentation. **Journal of Theoretical and Applied Information Technology**, Islamabad, v. 80, n. 2, p. 12, Out 2015.

CASELLA, G.; BERGER, R. L. **Statistical inference**. Duxbury Pacific Grove, vol. 83. 2002.

FIRMINO, P. R. A. et al. A non-central beta model to forecast and evaluate pandemics time series. **Chaos Solitons Fractals**, 2020.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2<sup>a</sup>. ed. California: Springer, v. IX, 2009.

LI Q., GUAN X., WU P., WANG X., ZHOU L., TONG Y. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. **The new england journal of medicine**, 2020.

MINISTÉRIO DA SAÚDE. **Coronavírus Brasil**, 2021. Disponível em: <<https://covid.saude.gov.br/>>. Acesso em: 12 mai 2021.

**doity.com.br/conais2021**

ORGANIZAÇÃO MUNDIAL DA SAÚDE. OMS, 12 mai 2021. Disponível em: <<https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019/advice-for-public/qa-coronaviruses>>. Acesso em: 12 mai 2021.

OANCEA, B., e CRISTIAN. C. Ş. Time series forecasting using neural networks. **Challenges of the Knowledge Society. IT in Social Sciences**, 2013.

PANIGRAHI, S.; BEHERA, H. S. A hybrid ETS–ANN model for time series forecasting. **Engineering Applications of Artificial Intelligence**, vol. 66, pp. 49–59, 2017.

ROCHA, J. E. C. et al. Redes neurais artificiais na previsão de contágio e óbitos por COVID-19: um estudo no Estado do Pará. **International Journal of Development Research**, 2020.

SALES, J. P. **Modelling and forecasting sustainable development time series via single and combined approaches**. 104f. Dissertação. Programa de Pós-Graduação em Desenvolvimento Regional Sustentável, Universidade Federal do Cariri, Crato-CE, 2019.

YU, Y. et al. Time Series Outlier Detection Based on Sliding Window Prediction. **Hindawi Open Access publishing for the scientific community**, London, v. 2014, p. 14, Out 2014.